International Series in
Operations Research & Management Science

Berend Wierenga
Ralf van der Lans *Editors*

Handbook of Marketing Decision Models

Second Edition

Operations Research Management Science

Springer

International Series in Operations Research & Management Science

Volume 254

More information about this series at http://www.springer.com/series/6161

Berend Wierenga · Ralf van der Lans
Editors

Handbook of Marketing Decision Models

Second Edition

Springer

Editors
Berend Wierenga
Rotterdam School of Management
Erasmus University
Rotterdam, South Holland
The Netherlands

Ralf van der Lans
Department of Marketing
Hong Kong University of Science and Technology
Kowloon
Hong Kong

ISSN 0884-8289 ISSN 2214-7934 (electronic)
International Series in Operations Research & Management Science
ISBN 978-3-319-56939-0 ISBN 978-3-319-56941-3 (eBook)
DOI 10.1007/978-3-319-56941-3

Library of Congress Control Number: 2017938315

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

We are very pleased to present this second edition of the Handbook of Marketing Decision Models. The field of marketing decision models is in a permanent state of development and growth. Since the publication of the first edition of this Handbook in 2008, new marketing phenomena have come under scrutiny and other areas have been developed more in-depth. This Handbook contains an introductory chapter, followed by seventeen chapters on marketing decision models in different domains. Thirteen of these are entirely new. Four chapters are by the same authors as in the first edition, but represent complete updates and extensions of previous texts. Information technology remains the main driver of developments in marketing decision models. Not surprisingly, this new edition of the Handbook has chapters on models for customer relationship management, customer loyalty management, Web site design, Internet advertising, social media, and social networks. In addition to this, there are new chapters on many other topics. The introductory chapter offers a short description of each chapter. Biographies of the contributing authors can be found at the end of the book.

The Handbook presents state-of-the-art marketing decision models and is highly relevant for several subsets of readers including builders of marketing models, users of marketing models, and academics in marketing departments of business schools, in related departments such as decision sciences and strategy, marketing scientists working outside academia, Ph.D. students, marketing researchers, and consultants. The book is also designed to cover the substantive content in marketing models courses at the graduate level. The Handbook is available in hard copy and in electronic form (also as individual chapters) and will be part of Springer's eBook package for universities.

We would like to thank all our colleagues in the field who have helped us to write this Handbook. Most of all, we thank the authors of the chapters of this book. They are all world-renowned specialists in their fields, people with busy schedules, and they have taken the time and effort to write and revise their chapters. By doing this, they offer the opportunity to others to share their expertise. This is a great service to the field.

We would also like to thank the reviewers. Each chapter was reviewed by two expert colleagues, and the authors have benefited from their comments and recommendations. The names of the reviewers can be found as an Appendix to this preface.

We would like to express our gratitude to our departments, the Marketing Department at the Rotterdam School of Management and the Marketing Department at the Hong Kong University of Science and Technology, for their support during our work on the book.

Finally, we would like to mention the excellent cooperation we received from Fred Hillier, the Editor of the Springer International Series on Operations Research and Management Science, and his successor Camille Price. Furthermore, it was a pleasure to work with Matthew Amboy and all those at Springer who were involved in the preparation, production, and marketing of the book.

The field of marketing decision models started about sixty years ago and has thrived ever since. We hope this Handbook will be a useful guide for the current stage of its life cycle and will inspire many scholars to take the field to its next level.

Berend Wierenga — Rotterdam, The Netherlands
Ralf van der Lans — Kowloon, Hong Kong

Appendix: Reviewers of the chapters for the Handbook of Marketing Decision Models (Second Edition)

Peter Boatwright, Carnegie Mellon University
Michael Braun, Cox Business School, Southern Methodist University
Bart Bronnenberg, Tilburg University
Martijn De Jong, Erasmus School of Economics, Erasmus University
Isaac Dinner, Kenan-Flagner Business School, University of North Carolina
Ceren Kolsarici, Queen's University
Bas Donkers, Erasmus School of Economics, Erasmus University
Daria Dzyabura, Stern School of Business, New York University
Els Gijsbrechts, Tilburg University
Jacob Goldenberg, IDC Herzliya
Sachin Gupta, Cornell University
Raghuram Iyengar, The Wharton School, University of Pennsylvania
Donald R. Lehmann, Columbia University
Song Lin, Hong Kong University of Science and Technology
Andres Musalem, Universidad de Chile and Duke University
Scott Neslin, Amos Tuck School, Dartmouth College
Peren Özturan, VU University Amsterdam
John Roberts, UNSW Australia Business School
Javier Sese, University of Zaragoza
Chung Tuck Siong, Nanyang Technological University, Singapore

Bernd Skiera, Goethe University, Frankfurt
Inseong Song, Seoul National University
Andrew Stephen, Saïd Business School, Oxford University
Stefan Stremersch, Erasmus School of Economics, Erasmus University
Maciej Szymanowski, Rotterdam School of Management, Erasmus University
Seshadri Tirunillai, Bauer College of Business, University of Houston
Olivier Toubia, Columbia University
Harald Van Heerde, Massey University, Auckland (NZ)
Jaap Wieringa, University of Groningen
Kenneth Wilbur, University of California, San Diego
Jie Zhang, University of Maryland

Contents

About the Editors

Berend Wierenga is Professor of Marketing at the Rotterdam School of Management, Erasmus University (RSM). His main research areas are marketing models, managerial decision making in marketing, and marketing management support systems. He has widely published on these topics in journals such as *International Journal of Research in Marketing*, *Journal of Marketing*, *Journal of Marketing Research*, *Management Science, Marketing Science*, and many others. His books include *Marketing Management Support Systems: Principles, Tools and Implementation* (2000) and *Marketing Decision Making and Decision Support* (2010) (both with Gerrit van Bruggen). Berend Wierenga is the Founding Editor of the *International Journal of Research in Marketing* (journal of the European Marketing Academy-EMAC) and is the Editor of the first edition of the *Handbook of Marketing Decision Models* (2008). Before joining Erasmus University, he was a faculty member at Wageningen University where he also obtained his Ph.D. He has held visiting positions at Stanford University, INSEAD, and the Wharton School of the University of Pennsylvania. At Rotterdam School of Management, Berend Wierenga has been Chairman of the Marketing Department, Scientific Director of ERIM (Erasmus Research Institute of Management), and Dean of the School. He is the recipient of the inaugural EMAC Distinguished Marketing Scholar Award, the Ad Fontes Penning of the Erasmus University, and he was appointed Officer in the Order of Oranje-Nassau.

Ralf van der Lans is Associate Professor of Marketing at Hong Kong University of Science and Technology, Hong Kong. He has a M.Sc. in econometrics from Erasmus University Rotterdam and a Ph.D. (both cum laude) from Tilburg University, the Netherlands. His main research interests are in eye-tracking, social networks and choice for which he develops statistical and econometric models to improve marketing decisions. His research appeared in journals such as *Marketing Science*, *Journal of Marketing Research*, *International Journal of Research in Marketing,* and *Journal of the American Statistical Association*. Van der Lans is a member of the editorial board of the *International Journal of Research in Marketing* and has won several research and teaching awards.

Chapter 1
Marketing Decision Models: Progress and Perspectives

Introduction to the Second Edition of the Handbook of Marketing Decision Models

Berend Wierenga and Ralf van der Lans

1.1 Dimensions of Marketing Decision Models

Marketing models are central to modern marketing decision making. The modeling of marketing phenomena with the purpose of supporting and improving marketing decisions started in the 1950s. The first edition of the Handbook of Marketing Decision Models offered a discussion of "The Past, the Present and the Future of Marketing Decision Models" (Wierenga 2008). We do not repeat this here, but take it as the point of departure for the introduction of this New Edition of the Handbook. Other authors have also documented the rich story of marketing decision models, a field often referred to as "Marketing Science". Examples are "The History of Marketing Science", edited by Winer and Neslin (2014) and the Special Section in Marketing Science ("2001: A Marketing Odyssey"), edited by Steckel and Brody (2001).

This new edition of the Handbook shows that since 2008 the field has made impressive progress. Ever since its incubation more than 60 years ago, the field of marketing decision models is in a permanent state of development and growth. New marketing phenomena are coming under scrutiny and new research methodologies are being introduced. Observing the situation at one particular point in time (in this case in 2017), we find several areas of marketing decision models in different stages of development. We classify the current work according to two dimensions. The first dimension is the topical domain, which can be: *existing* (meaning that in the particular domain a substantial amount of modeling work has already been done)

B. Wierenga (✉)
Rotterdam School of Management, Erasmus University (RSM),
Rotterdam, The Netherlands
e-mail: bwierenga@rsm.nl

R. van der Lans
Hong Kong University of Science and Technology, Hong Kong, China
e-mail: rlans@ust.hk

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_1

Table 1.1 Classification of modeling approaches

		Topical domain	
		Existing	*New*
Research methodology	*Existing*	**Accretion** (Chaps. 2–5)	**Exploration** (Chaps. 12–13)
	New	**Sophistication** (Chaps. 6–11)	**Excavation** (Chaps. 14–18)

or *new* (meaning that modeling in the domain has started only recently). The second dimension is research methodology, which can also be: *existing* (meaning that the methodology has been used earlier) or *new* (meaning that new mathematical and statistical methods, and/or new types of data collection and measurement are being used).

Combining the two dimensions produces four quadrants (see Table 1.1). The upper left quadrant refers to modeling efforts directed at existing domains of marketing using existing research methodology. This type of work increases our insights in a particular area. Building on what already exists, new studies produce additional layers of knowledge about the phenomena under study. We use the term **accretion** for this type of work. Research in the lower left quadrant focuses on new research methodologies applied to existing phenomena. New methods allow us to challenge assumptions, to draw more precise conclusions, and take more focused actions (e.g., aimed at individual customers). We call this **sophistication**. Research in the upper right quadrant refers to the modeling of new phenomena using existing research methodologies. This produces a first picture of a new area and we call this **exploration**. Finally, the lower right quadrant refers to the modeling of new phenomena using new research methodology. We use the term **excavation** for this type of research, which involves uncovering and probing new phenomena and at the same time forging the tools needed for this.

Using this framework, we classified the seventeen topical chapters of this Handbook. In the following, we briefly introduce the contributions to the Handbook following this classification.

1.2 The Chapters in This Handbook

1.2.1 Accretion

The Handbook has four chapters in this category.

- Chapter 2 "*Sales Promotion Models*" (Van Heerde and Neslin) presents the most recent insights about the modeling of sales promotions. In terms of share of the marketing budget, sales promotion is the most prominent marketing promotion instrument. Sales promotion has been the object of marketing modeling for several decades, with the area really taking off in the 1980s, stimulated by the advent of scanner data. Online marketing and purchasing have given new impulses to the study of sales promotions. The chapter provides a systematic discussion of a very rich set of sales promotion models, dealing with issues such as measurement at the consumer and at the store level, immediate versus long-term impact, endogeneity, forward buying, pass-through, and descriptive and normative models. It offers a comprehensive picture of the current state-of-the art in the modeling of sales promotions.

- Chapter 3 "*Innovation and New Products Research*" (Fan, Golder, and Lehmann) deals with another long-standing research domain in marketing. From the start with the Bass model in the late 1960s, innovation and new product development have attracted the attention of model builders in marketing. This chapter presents reviews of research in the four stages of new product development, opportunity identification, product design and development, sales forecasting, and commercialization. For each of these stages, the chapter also discusses decision models and provides the most pressing research question Innovation is the lifeblood of economic development, which defines the importance of the research on this topic.
- Chapter 4 "*Models for the Financial-Performance Effects of Marketing*" (Hanssens and Dekimpe) also refers to a classical topic in marketing: how does marketing contribute to the overall financial results of the company? The chapter discusses cash flow effects of marketing, performance metrics, and presents models for the process perspective as well as for the investor perspective. The authors review the many recent papers that focus on the financial performance effects of marketing.
- Chapter 5 "*Loyalty Programs: Current Insights, Research Challenges, and Emerging Trends*" (Bijmolt and Verhoef) discusses recent work on loyalty programs. Customer loyalty programs became popular in the 1990s and received an additional boost from the CRM (Customer Relationship Management) movement. The chapter discusses pre-rewarding and post-rewarding effects, and customer personalization. The authors argue that modeling the effects of loyalty programs may involve various dependent variables, such as purchase incidence, purchase amount, redemption incidence and redemption fraction, and that customer redemption decisions are key to understanding the effect mechanism of loyalty programs. The chapter ends with emerging trends on loyalty programs: digitalization, alignment of loyalty programs with customer experience, and reviving existing loyalty programs.

1.2.2 Sophistication

The Handbook has five chapters in this category.

- Chapter 6 "*Structural Models in Marketing: Consumer Demand and Search*" (Chintagunta) discusses the nature and use of structural models in marketing.

Structural models differ from statistical models in that the relationships between explanatory and output variables are based on the underlying decision processes of consumers and/or firms. Structural models thus aim to go beyond relationships between observed variables. An example is to develop models based on the economic theory that consumers allocate resources that maximize utility. The advantage of structural models is that their predictions may hold "outside the data (extrapolation)" in situations where the market changes (e.g., when one of the competitors leaves the market, when new products are introduced, or when a firm decides to sell through its online channel). This chapter provides a state-of-the-art overview of structural models with a focus on consumer search and demand.

- Chapter 7 "*Economic Models of Choice*" (Allenby, Kim, and Rossi) starts from a rational consumer decision model, where consumers make choices by maximizing utility, taking into account a budget constraint. Such a model allows us to predict the effects of interventions by the supplier (e.g., changes in product quality or price) and the effects of relaxation of the budget constraint. Bayesian statistical methods make it possible to incorporate consumer heterogeneity in choice models, which is very useful in the context of conjoint analysis and market segmentation. The authors discuss the strong points of direct utility models, and develop extensions to consumers buying more than one offering, different error specifications, and multiple constraints.
- Chapter 8 "*Empirical Models of Learning Dynamics: A Survey of Recent Developments*" (Ching, Erdem, and Keane) discusses models for dynamics in choice behavior. Prior history of a consumer or a market may affect a consumer's current utility evaluations and hence influence choices. Starting with models focusing on learning from own purchases, this field has gradually adopted a broader view of dynamics and learning and has developed models for learning from others (e.g., friends or other social network members), from experience with related products (correlated learning), and from examination of expert opinion (search). The chapter also discusses work on learning in the context of strategic interaction where consumers and (competing) firms learn from each other's behavior. Learning models can explain dynamics in consumer behavior for a wide range of products and activities, including drugs, movies, books, restaurants, sexual behaviors, and health insurance plans.
- Chapter 9 "*Measurement Models for Marketing Constructs*" (Baumgartner and Weijters) predominantly has its roots in the behavioral sciences, in contrast with the orientation towards economics of the previous three chapters. For empirical research, the reliable and valid measurement of the constructs under study (e.g., brand attitude, customer orientation, product quality, or customer expertise) is a condition sine qua non. Starting from standard structural equations modeling, this chapter presents a systematic overview of measurement models, congeneric as well as formative, and single-group as well as multiple group models. The chapter also discusses recently published methods for relaxing model assumptions such as zero loadings, uncorrelated factors, normally distributed measures, and methods that deal with response bias in cross-cultural research.

- Chapter 10 "*Marketing Models for the Customer-Centric Firm*" (Ascarza, Fader, and Hardie) focuses on the three drivers of organic growth (and profitability) of customer-centric firms: customer acquisition, customer retention, and customer development. Since the advent of electronic customer databases, this an area of marketing models has developed rapidly. The chapter reviews the key data-based tools and models that are already available and those that are in the making. Models for customer acquisition deal with the issue of which customers to target. Models for managing acquired customers are about computing customer value and handling customers in contractual and non-contractual settings, and include the models of the famous Pareto/NBD family. Furthermore, models have been developed to predict churning (targeting and retaining the most vulnerable customers). Models have been developed for contact customization (adapting the offer based on the customer's characteristics and history), for cross-selling (predicting the next product to buy), and for the allocation of resources across acquisition versus retention. The authors not only review the work on these models in the marketing literature, but also discuss the many contributions in operations research, statistics, and computer science.
- Chapter 11 "*Eye Movement During Search and Choice*" (Van der Lans and Wedel) discusses models for the analysis of eye-tracking data. Eye tracking research (e.g., the study of how consumers watch advertisements) has a long history, but eye-trackers that collect eye-movement data in natural settings in an unobtrusive way have only become available recently. Eye-tracking data contain much more information than the "heat-maps" that are popular in practice. Modeling eye-tracking data helps to predict product search and choice. Predictive analysis can be used to optimize marketing campaigns. This chapter discusses the basics of eye movements and its recording, presents an overview of the recent integrated eye-tracing models for search and choice, and offers a framework for setting up eye-tracking projects.

1.2.3 Exploration

The Handbook has two chapters in this category.

- Chapter 12 "*Business-Cycle Research in Marketing*" (Deleersnyder and Dekimpe) is about a new stream of work. Whereas business cycles have been the object of study in economics for a long time, researchers have only recently started to look at the implications of business cycles for marketing. The chapter starts by summarizing emerging results from three streams of research on the effects of business-cycle fluctuations: their effects on (brand, firm, or industry performance), on marketing conduct, and on marketing effectiveness. Next, the chapter deals with two important methodological issues: how to infer business cycles from time series data and how to link business cycles to marketing variables. The authors discuss the advantages and limitations of several

approaches. This will help further research in this area, aimed at a much-needed better managerial understanding of the effects of business cycles on marketing.
- Chapter 13 "*Marketing Models for the Life Sciences Industry*" (Avagyan, Landsman, and Stremersch) deals with marketing issues in an industry that is specific for a number of reasons. The health industry has a direct impact on people's well-being, it is strictly regulated, and decision making is dispersed over a large number of parties in the healthcare value chain. The chapter consists of two main parts. The first presents typical models, including physician choice models, physician learning models, models for key opinion leaders, and diffusion models. The second part reviews select findings on the role of marketing (including promotion, advertising, pricing) that have emerged from empirical work so far. The area of health marketing is developing fast and the authors discuss future research directions (networks, crowd sourcing, opinion leaders, generic use, etc.). Such future work can build on this chapter in the Handbook.

1.2.4 Excavation

The Handbook has six chapters in this category.

- Chapter 14 "*Marketing Models for Internet Advertising*" (Bucklin and Hoban) deals with the quickly growing area of Internet advertising. The emerging "digital advertising ecosystem" , encompassing consumers, advertisers, websites, advertising platforms, and ad networks is complex and poses major challenges to modelers. The chapter discusses models for the two main formats of internet advertising: paid search sponsored advertising (SSA) and display advertising. Models for Internet display advertising include models for click-through, purchasing, browsing behavior, carryover effects, and temporal spacing of display ads. In SSA, the goal of the advertiser is to obtain a high position in the sponsored listings of the search engine results page (SERP) . For this purpose, the advertiser selects keywords and presents bids on these keywords. The height of these bids affects the position (rank) of the advertiser on the lists, which in turns affects the click-through rate. One of the main problems is that the position of the advertiser on the SERP is endogenous and dependent on the (not transparent) auction mechanism that the search engine uses. The chapter discusses recently developed models that deal with this problem, and reviews recent work on experiments in paid search advertising. Internet advertising is currently one of the hottest areas in marketing model building. The Bucklin and Hoban chapter is a rich account of the state-of-the-art at this point in time and can serve as the departure point for future works.
- Chapter 15 "*Advertising Effectiveness and Media Exposure*" (Danaher) also deals with the quickly-changing media landscape due to the advent of digital media, such as Internet, social media, and smartphones. Most advertising has become multimedia, including (web) display advertising, paid search

advertising, social media, and mobile advertising. The chapter discusses media synergy (including the synergy between traditional and new media), models for multimedia advertising (both for individual and time series data), models for attributing purchases to media, and advertising media selection models for new media (e.g., optimizing for multiple websites). Again, this is an area with many research questions waiting to be answered.

- Chapter 16 "*Social Media Analytics*" (Moe, Netzer, and Schweidel) reviews the work in a new area of marketing science, the analysis of user-generated-content (UCG) in digital environments such as websites, product or brand forums, and other social media. The chapter discusses the drivers and dynamics behind the generation of content by consumers, including the aspect of self-selection of the consumers who post. It examines text-mining, a booming area in computer science that provides methods to convert text data (e.g., from social media) into quantifiable measurements such as attribute scores and sentiments. Current applications of text mining in marketing include studies that describe and monitor markets (e.g., monitoring brand health) and to make predictions (e.g., predicting the effect of negative UGC on stock prices). The authors discuss studies on the effect of product scores from UCG on sales and examine recent work on the effects of social media in combination with other media such as TV and paid online advertising. Research into social media is continuing to grow, driven by the progress in machine learning and related fields. As this chapter demonstrates, this work, coupled with new techniques developed by marketing academics, will greatly improve our ability to characterize social media and incorporate it in subsequent analyses of markets.
- Chapter 17 "*Integrating Social Networks into Marketing Decision Models*" (Chen, Van der Lans, and Trusov) is about how the knowledge of social networks can improve marketing decision making. The emergence of massive social network datasets, such as Facebook, Twitter, and Weibo, has opened a new area of empirical work and offers opportunities for strategies such as viral marketing, crowd sourcing, and crowdfunding. The authors discuss how social networks can be incorporated into marketing decision models. One possibility is to use network measures (e.g., density, connectedness or centrality) directly in models that explain a particular marketing phenomenon (e.g., the diffusion of a new product). Another possibility is to model the social network as a stochastic process and use Markov models to study how consumers' product choices are influenced by the choices of their peers. It is also possible to apply agent-based models. The authors convincingly show the value of network knowledge for marketing purposes. The authors also discuss challenging questions, such as how to sample from large networks, how to attach weights to relationships, and how to solve the endogeneity problem: is the observed behavior explained by the network connection as such, or by the similarity of the individuals that caused the network connection?
- Chapter 18 "*Morphing Theory and Applications*" (Liberali, Hauser, and Urban) is about optimizing a firm's digital marketing strategy. Firms need to customize their marketing efforts to the wishes of their individual consumers,

thereby increasing click-through rates (CTR) and conversion (sales). The authors present a more efficient alternative to conventional A/B testing. In this approach, different "morphs" are presented to consumers. The responses provide the information to identify the cognitive style segment(s) to which a particular consumer belongs. During this process, Bayesian updating effectuates a rapid assignment of consumers to segments. Morphing trades off learning about consumer response (learn) with using that knowledge to display the best banner for the consumer (earn).

1.3 Perspectives for Marketing Decision Models

1.3.1 Information Technology as the Main Driver

As observed earlier, the field of marketing decision models is in a constant state of development. Of the seventeen topical chapters in this Handbook, thirteen (76%) refer to new topical areas, new types of research methodology or both (Table 1.1). Most of the new topical areas (right hand side of Table 1.1), refer to phenomena that did not even exist a few decades ago, such as online marketing, websites, social media, mobile media, user-generated content, digital networks, and (huge) electronic customer databases. Wierenga (2008) pictured marketing decision models in its upstream and downstream context. In terms of upstream factors, it is clear that in the last years, progress in information technology has been the strongest driver behind the developments in marketing decision models. We expect this to continue in the years to come. Besides this, developments in other upstream fields such as econometrics and statistics will continue to help marketing further sharpen its tools, contributing to the ongoing sophistication of its analytical work. For instance, in the last several decades, structural models have emerged in the field of marketing, and have contributed to the development of marketing decision models that also hold in uncertain environments with policy changes (Chap. 6). In combination with Bayesian methodologies that allow for heterogeneity, more advanced models have been built to investigate consumer choices (Chap. 7) and dynamics (Chap. 8). However, developments in computer science, artificial intelligence, data science, machine learning, text-processing and audio-processing may have an even greater impact (Chintaguna et al. 2016). Chapters in this Handbook, for example on social media (Chap. 16) and on the customer-centric firm (Chap. 10) demonstrate the growing importance of techniques such as text mining, data mining, and predictive analytics. The combination of large databases ("Big Data") with machine learning offers a potentially powerful research direction in many fields (Athey 2015; Imbens and Rubin 2015). This is definitely the case for marketing which thrives on ever-larger databases.

1.3.2 Shifts in Marketing Models

Comparing the current work on marketing decision models with the start of the field, for example Kotler's marketing models book of 1971, three shifts in the nature and purpose of marketing models become clear. All three shifts, which we discuss below, are in the direction from more general to more specific. This is a sign of progress and sophistication of the field. The parallel with the natural sciences is clear. For example, physics started out by studying matter at the outside, but as its instruments became sharper, they drilled down to the levels of molecules, particles, and nanostructures. Zooming in with our ever more sophisticated marketing science tools, we are accumulating ever deeper knowledge about marketing phenomena and marketing processes.

1.3.2.1 From Marketing Mix Instruments to Very Specific Marketing Actions

In Kotler (1971), the decision variables are defined at the level of the total marketing budget, and the marketing efforts are defined at the level of the marketing mix instruments of product, price, promotion, and place (distribution). The goal of a marketing program is to find the levels of the marketing expenditures for the different marketing instruments that maximize the profit of the firm. The famous Dorfman-Steiner (1954) conditions are used to check if the allocation of the marketing expenditures over the marketing instruments is optimal. Interestingly, models at the level of marketing instruments were called "micromarketing decision models" (Kotler 1971, p. 285). Most of the current marketing decision models are much more micro. For example, they do not deal with the question of total expenditures on promotion, but they zoom in on issues such as the profitability of a specific sales promotion (Chap. 2), the effectiveness of banners on a specific company website, or how much to bid for particular keyword in search sponsored advertising (Chap. 14). Moreover, these decisions are fine-tuned at different stages in the business cycle (Chap. 12) or at different stages in the development and diffusion of the product life cycle (Chap. 3). This focus on specific marketing actions has made marketing models much more operational and (downstream) more useful for marketing practice and more effective for contributing to firm value (Chap. 4). We expect this development to continue.

1.3.2.2 From Markets to Individual Customers

In traditional marketing, the market to be addressed was the total of all (potential) customers for a specific product or service. Statistics about market turnover and company sales were at the level of the total market. This was followed by the concept of market segmentation (Frank et al. 1972), where a market is segmented into submarkets, groups of customers with specific demand profiles and purchasing behavior. To be able to monitor the purchasing behavior of these customer segments, consumer panels were set up where individual consumers (with known socio-economic profiles) recorded their purchases. The insights obtained in this way were important for marketing practice, and companies started to define their offerings at the level of market segments, although it was not always easy to aim precisely at the pre-defined segments in the market. This changed completely when information technology made it possible to deal with individual customers. In the era of the customer-centric approach (Wierenga 2008; Chap. 10), the individual customer has become the unit of analysis. Companies can monitor the purchasing behavior of individual customers, they can make them tailor-made offers and send them promotional messages to increase loyalty (Chap. 5), in a format morphed in accordance with their specific cognitive style (Chap. 18). This process of individualization will continue, also as a result of an increase in electronic ordering and delivery.

1.3.2.3 From Purchase to Customer Journey

The focal event in marketing has always been a customer making a purchase. Whether or not a customer makes a purchase (and how many) is the most used dependent variable in marketing models. Hence, recording purchases (sales) is the first priority of any marketing database. However, there is more than just sales. Consumers go through decision processes that include a pre-purchase stage with problem recognition and search, a purchase stage with choice, ordering and payment, and a post-purchase stage with usage, consumption and post-purchase engagement such as word of mouth (Lemon and Verhoef 2016). The consumer decision process has been much discussed at a conceptual level, starting with the seminal work of Howard and Sheth (1969). Often, these processes were measured through surveys, in which marketing has observed great developments (Chap. 9). Recently, there have been more electronic means, such as eye-tracking (Chap. 11) or clickstream data (Chap. 14), to actually trace the processes of individual consumers. The "customer journey" as it is often called now (Lemon and Verhoef 2016) has many "touch points", in multiple channels and media. Many of these touch points are company-owned, for example, websites, ordering and payment systems (online and offline), service platforms, sales force, loyalty programs, and brand forums. Registration of a consumer at these different touch points makes it possible to "reconstruct" the customer journey. This can provide important insights in the factors that drive consumer purchasing behavior and their interactions. Related to this is the issue of attribution. If, during the customer journey, the consumer is exposed to various media and channels (e.g., e-mail, search engines, social media, display advertising, print, TV, etc.), how much credit should each of these touch points receive for the eventual purchase? This problem is just starting to get attention (Kannan et al. 2016, Chap. 15). Due to the increased possibilities to monitor customers through a multiple of touch points, we expect that marketing models for the analysis of customer journeys and the attribution of purchase to individual touch points will be booming in the years to come.

Concluding, there are great opportunities for academically interesting and managerially useful work in marketing decision models in the years to come. If, say ten years from now, a third edition of this Handbook would be published, there will be no scarcity of interesting content.

References

Athey, S. 2015. Machine learning and causal inference for policy evaluation. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 5–6.

Chintagunta, P., D.M. Hanssens, and J.R. Hauser. 2016. Marketing science and Big Data. *Marketing Science* 35 (3): 341–342.

Dorfman, R., and P.O. Steiner. 1954. Optimal advertising and optimal quality. *American Economic Review* 44 (5): 826–836.

Frank, R., W. Massy, and Y. Wind. 1972. *Market segmentation*. Englewood Cliffs NJ: Prentice Hall.

Howard, J.A., and J.N. Sheth. 1969. *The theory of buyer behavior*. New York: Wiley.

Imbens, G., and D. Rubin. 2015. *Causal inference for statistics, social, and biomedical sciences: An introduction*. Cambridge UK: Cambridge University Press.

Kannan, P.K., W. Reinartz, and P.C. Verhoef. 2016. The path to purchase and attribution modeling: Introduction to special section on attribution. *International Journal of Research in Marketing* 33 (3): 449–456.

Kotler, P. 1971. *Marketing decision making: A model building approach*. New York: Holt, Rinehart and Winston.

Lemon, K.N., and P.C. Verhoef. 2016. Understanding customer experience throughout the customer journey. *Journal of Marketing*.

Steckel, J.H., and E. Brody. 2001. Special section: 2001: A marketing Odyssey. *Marketing Science* 20 (4): 331–336.

Wierenga, B. 2008. The Past, the Present, and the Future of marketing decision models. In B. Wierenga (Ed.), *Handbook of marketing decision models* (1st ed.), pp. 3–20. New York: Springer Science + Business Media.

Winer, R.S., and S.A. Neslin. 2014. *The history of marketing science*. Hanover, MA: Now Publishers.

Chapter 2
Sales Promotion Models

Harald J. van Heerde and Scott A. Neslin

Firms spend a significant part of their marketing budgets on sales promotions. Retail (2012) indicates that during 1997–2011, promotion accounted for roughly 75% of marketing expenditures for US packaged goods manufacturers; the other 25% was for advertising. In 2011, 58% of the budget was spent on promotion to the trade (i.e., from manufacturers to retailers), and 15% on manufacturer promotions to consumers. Since the impact of promotions on sales is usually immediate and strong (Blattberg et al. 1995), promotions are attractive to results-oriented managers seeking to increase sales in the short term (Neslin 2002). In a meta-analysis, Bijmolt et al. (2005) report that the average short-term sales promotion elasticity is –3.63, which implies that a 20% temporary price cut leads to a 73% rise in sales.[1] There are few, if any, other marketing instruments that are equally effective. Because of this, coupled with the availability of scanner data, marketing researchers have been very active in developing models for analyzing sales promotions. Most applications analyze promotions for consumer packaged goods, and this chapter reflects this practice. Nevertheless, many of the models could be applied to other settings as well.

This chapter discusses models for measuring sales promotion effects. Part I (Sects. 2.1–2.10) focuses on descriptive models, i.e., models that describe and explain sales promotion phenomena. We start by discussing promotions to consumers. Sections 2.1 through 2.5 focus on analyzing the direct impact of promotions on sales and decomposing that impact into a variety of sources. Section 2.6

[1]This figure holds for temporary price cuts without feature or display support. A feature or display may increase the sales effect up to a factor 9 (Narasimhan et al. 1996).

H.J. van Heerde (✉)
Massey University, Auckland 0630, New Zealand
e-mail: heerde@massey.ac.nz

S.A. Neslin
Tuck School of Business, Dartmouth College, Hanover, USA

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_2

examines what happens after the sales promotion bump and describes models for measuring feedback effects, reference price effects, learning effects, permanent effects, and competitive reactions. Section 2.7 discusses endogeneity in sales promotion effect measurement and solutions to address it. Next we turn to descriptive models for promotions aimed at retailers ("trade promotions") in Sect. 2.8, and discuss key issues and models concerning forward buying (Sect. 2.9) and pass-through (Sect. 2.10).

In Part II we discuss normative models, i.e. models that prescribe the best (profit maximizing) sales promotions. Section 2.11 covers models for planning promotions to consumers, Sect. 2.12 provides decision models on trade promotions for manufacturers, and Sect. 2.13 describes normative retailer models for optimal forward buying and pass-through. Part III concludes with a summary (Sect. 2.14), practical model guidelines (Sect. 2.15), and directions for future research (Sect. 2.16).

Part I: Descriptive Models

2.1 Promotions to the Consumer—Introduction

Promotions to the consumer include coupons, rebates, in-store temporary price cuts, feature advertising, and in-store displays. In analyzing promotion effects, we distinguish between *immediate effects* (the impact in the week t the promotion is implemented), *medium-term effects* (the weeks surrounding week t) and *long-term effects* (that take place after the medium-term effects). Much of the work on sales promotion has focused on the sales promotion bump. The goal in modeling this bump is to allocate the increase in sales that occurs in period t to one or more of the sources listed in the second column of Table 2.1. The timing of the sources contributing the bump may be in week t itself (immediate effect) or in the surrounding weeks (medium-term effect). We discuss the immediate and medium-term effects in

Table 2.1 The impact of promotions to the consumer

The sales promotion bump (Sects. 2.1–2.5)		Long-term effects—beyond the sales promotion bump (Sect. 2.6)
Category growth	Increased consumption rate	Purchase-event feedback
	Outside industries	Reference prices
Within-category immediate effects	Cannibalization	Consumer learning
	Brand switching	Long-term effects
	Category switching	Competitive response
	Store switching	
Within-category medium-term effects	Acceleration	
	Deceleration	

Sects. 2.1–2.5. The effects of sales promotions on consumer behavior beyond the sales promotion bump are classified as long-term effects. These effects are list in Table 2.1 as well and discussed in more detail in Sect. 2.6.

We now discuss the decomposition of the sales promotion bump. We assume that the analysis is conducted at the level of an SKU (Stock Keeping Unit), i.e., a particular variety of a brand. The sales promotion sources contributing to the sales promotion bump for an SKU can be classified in three areas. The first is "Category Growth," which means the increase in sales for the promoted SKU does not come at the expense of other products or stores within the category. Promotion-induced purchases may cause households to carry extra inventory, which is consumed at a higher rate. For example, a promotion on potato chips may increase household inventory and cause the household to consume more chips.

The second area is within-category immediate effects, i.e., the purchase draws from within the category in the same time period. This effect consists of:

- Cannibalization: the consumer switches from SKU j' of brand b to SKU j of brand b;
- Brand switching: the consumer switches from brand b' to brand b;
- Category switching: the consumer switches from a product from another category to brand b;
- Store switching: the consumer switches from another store s' to store s to buy brand b;

The third area consists of within-category medium-term effects, i.e., substitution from the period before the promotion (deceleration) or after (acceleration). Deceleration means the consumer postpones the category purchase to week t, because the consumer expects a promotion in week t. Deceleration implies purchase displacement from the past (before t) to now. Acceleration refers to timing acceleration—the consumer buys the category in week t rather than later, or quantity acceleration—the consumer buys more of the category than usual. Both timing and quantity acceleration lead to higher household inventories, and both lead to a purchase displacement from the future ($>t$) to now (t). Note in defining acceleration we assume that the *postpromotion consumption rate does not increase*. If it does, we are in the case of Category Growth, and there is no purchase displacement.

Table 2.2 lists all possible combinations of effects derivable from Table 2.1. We first list Category Growth. Then, within-category substitution can come from four types of products (the item itself in another time period, other items within the same brand, other brands, and other categories), from two places (within the same store, from other stores) and from three time periods (before, during, and after the promotion). Hence in total there are $4 \times 2 \times 3 = 24$ combinations, of which one drops out since a promoted product cannot substitute its own sales in the same period and store. These 23 combinations all imply some form of substitution, as we show in Table 2.2 (listings 2–24). For example, a brand- and store switch (source 7) implies that brand b' in store s' loses sales. The combination of store switching and acceleration (source 17) is sometimes referred to as indirect store switching

Table 2.2 The 24 decomposition effects from manufacturer and retailer perspectives

#	Effect	Promotion of focal SKU of a brand in focal store in focal time period draws sales from…	Increased category consumption	Net unit sales effect for manufacturer	Net unit sales effect for retailer
Category growth					
1	Faster consumption	Household budget	Yes	+	+
Within-category immediate					
2	Cannibalization	Other SKUs from the same brand, in same store and same period	No	0	0
3	Brand switching	Other brands in same store in same period	No	+	0
4	Category switching	Other categories in same store in same period	Yes	+	0
5	Store switching	Same SKU in other stores in same period	No	0	+
6	Cannibalization and store switching	Other SKUs from the same brand, in other stores in same period	No	0	+
7	Brand switching and store switching	Other brands in other stores in same period	No	+	+
8	Category switching and store switching	Other categories in other stores in same period	Yes	+ or 0[a]	+
Within-category medium term					
9	Acceleration	Same SKU in same store in future periods	No	0	0
10	Deceleration	Same SKU in same store in earlier periods	No	0	0
11	Cannibalization and acceleration	Other SKUs from the same brand, in same store and future periods	No	0	0
12	Cannibalization and deceleration	Other SKUs from the same brand in same store in earlier periods	No	0	0
13	Brand switching and acceleration	Other brands in same store in future periods	No	+	0
14	Brand switching and deceleration	Other brands in same store and earlier periods	No	+	0
15	Category switching and acceleration	Other categories in same store in future periods	Yes	+ or 0[a]	0

(continued)

Table 2.2 (continued)

#	Effect	Promotion of focal SKU of a brand in focal store in focal time period draws sales from…	Increased category consumption	Net unit sales effect for manufacturer	Net unit sales effect for retailer
16	Category switching and deceleration	Other categories in same store and earlier periods	Yes	+ or 0[a]	0
17	Store switching and acceleration	Same SKU in other stores in future periods	No	0	+
18	Store switching and deceleration	Same SKU in other stores in earlier periods	No	0	+
19	Cannibalization, store switching and acceleration	Other SKUs from same brand in other stores in future periods	No	0	+
20	Cannibalization, store switching and deceleration	Other SKUs from same brand in other stores in earlier periods	No	0	+
21	Brand switching, store switching and acceleration	Other brands in other stores in future periods	No	+	+
22	Brand switching, store switching and deceleration	Other brands in other stores in earlier periods	No	+	+
23	Category switching, store switching and acceleration	Other categories in other stores in future periods	Yes	+ or 0[a]	+
24	Category switching, store switching and deceleration	Other categories in other stores in earlier periods	Yes	+ or 0[a]	+

[a]If the manufacturer produces the product in the other category that is being substituted, s/he does not benefit from the category switch (0); otherwise s/he does (+)

(Bucklin and Lattin 1992): the consumer visits both stores, but, because of the promotion in store s, she buys the promoted product in store s whereas otherwise she would have bought a product in store s'. Hence a current purchase in store s pre-empts a future purchase in store s'.

Table 2.2 shows where each of the 24 combinations draws from and also whether category consumption increases. This is certainly the case for Category Growth. Category consumption also increases when the contributing source involves category switching (#4, 8, 15, 16, 23, 24).

The effects listed in Table 2.2 are important because each has distinctive managerial implication in terms of unit sales. For example, brand switching benefits manufacturers but not retailers; store switching benefits retailers but not manufacturers; category growth benefits both manufacturers and retailers; category switching is neutral for retailers and may or may not be neutral for manufacturers, depending on whether the manufacturer has products in both categories. The rightmost two columns of Table 2.2 show that some combinations create a "+" for both manufacturers and retailers. So there is potential for conflict as well as "win-win" between manufacturers and retailers (Van Heerde and Gupta 2006). Further complicating matters is that both retailer and manufacturer profit margins may differ. A brand switch might actually benefit the retailer if the switched-to brand has higher margin. This motivates the retailer to demand a trade deal discount. The retailer can "pass through" some of the discount to promoting the brand, while at the same time increasing the margin on that brand. We discuss this in Sect. 2.9.

In sum, it is crucial to measure how the immediate impact of promotion is decomposed into the components shown in Table 2.2. Consequently, the decomposition of the sales promotion bump has gained considerable attention in the literature, and we summarize empirical generalizations in Sect. 2.5. First, however, we discuss the models necessary for measuring the decomposition. We start with individual-level incidence, choice, and quantity models in Sect. 2.2. In Sect. 2.3 we discuss individual-level models for store switching, category switching, cannibalization, and deceleration. In Sect. 2.4 we present aggregate (store-level) models.

2.2 Customer-Level Models—Incidence, Choice, and Quantity

Promotions can influence category incidence, brand choice, and purchase quantity, and historically, these decisions have received the most attention. The unit of analysis in this literature is typically the brand, not the SKU. The models for these household-level decisions are based largely on household panel scanner data, as are the models discussed in Sect. 2.3. However, these models have also utilized conjoint analysis (e.g., Eckert et al. 2012; Ailawadi et al. 2014). The models in Sect. 2.4 are based on aggregate (store- or higher) data, e.g., weekly store data.

The probability that household h buys q_{bt}^{h} units of brand b at shopping trip t is the product of three probabilities:

$$P(Q_{bt}^{h}=q_{bt}^{h})=P(I_{t}^{h}=1)\times P(C_{t}^{h}=b\,|\,I_{t}^{h}=1)\times P(Q_{bt}^{h}\,|\,I_{t}^{h}=1,C_{t}^{h}=b) \qquad (2.1)$$

where

$P(I_t^h = 1)$	is the probability that household h buys the category at trip t (incidence),
$P(C_t^h = b \mid I_t^h = 1)$	is the probability that, conditional on incidence at t, household h buys brand b, and
$P(Q_{bt}^h = q_{bt}^h \mid I_t^h = 1, C_t^h = b)$	is the probability that, conditional on a choice to buy brand b at trip t, the household buys q_{bt}^h units

2.2.1 Category Incidence Models

Category incidence is often modeled as a binary logit (e.g., Bucklin et al. 1998):

$$P(I_t^h = 1) = \frac{1}{1 + e^{-(\gamma_0 + \gamma_1 CV_t^h + \gamma_2 I_{t-1}^h + \gamma_3 \overline{C}^h + \gamma_5 INV_t^h)}} \tag{2.2}$$

where CV_t^h is the "inclusive value", which in a nested logit framework is the expected maximum utility available to household h from buying a brand in the category at time t. It is given by the log of the denominator of the brand choice probability: $CV_t^h = \ln\left(\sum_{b'=1}^{B} \exp\left(u_{b'} + \boldsymbol{\beta} X_{b't}^h\right)\right)$ (see Sect. 2.2.2). I_{t-1}^h is a lagged purchase incidence dummy (Ailawadi and Neslin 1998), and $\overline{C}^h$ is the household's average daily consumption computed from an initialization sample. INV_t^h is the inventory carried by household h at the beginning of shopping trip t. The standard approach to operationalize INV_t^h is:

$$INV_t^h = INV_{t-1}^h + PurQty_{t-1}^h - \overline{C}^h,$$

where $PurQty_{t-1}^h$ is the quantity (in ounces) purchased by household h during trip t-1. Inventory should be mean-centered over time for a given household to remove household differences. The term $\overline{C}^h$ assumes a constant purchase rate for household h. However, Ailawadi and Neslin (1998) propose that the consumption rate for household h at time t $(Consumpt_t^h)$ flexibly varies over time as a function of inventory:

$$Consumpt_t^h = INV_t^h \left[\frac{\overline{C}^h}{\overline{C}^h + (INV_t^h)^f}\right], \tag{2.3}$$

where f is a parameter. A flexible consumption rate (smaller f) means that promotion-induced stockpiling can increase category consumption (Effect #1 in Table 2.2; see also Sun 2005).

2.2.2 Brand Choice Model

The probability that household h buys brand b at time t, conditional on purchasing the category, is often given by a multinomial logit model[2] (Guadagni and Little 1983):

$$P(C_t^h = b \mid I_t^h = 1) = \frac{\exp\left(u_b + \boldsymbol{\beta} X_{bt}^h\right)}{\sum_{b'=1}^{B} \exp\left(u_{b'} + \boldsymbol{\beta} X_{b't}^h\right)}, \tag{2.4}$$

where B is the number of brands and V_{bt}^h is the "deterministic component" of the utility of household h for brand b at time t (Guadagni and Little 1983). A typical formulation would be:

$$V_{bt}^h = u_b + \boldsymbol{\beta} X_{bt}^h = u_b + \beta_1 PRICE_{bt} + \beta_2 FEAT_{bt} + \beta_3 DISP_{bt} + \beta_4 BL_b^h + \beta_5 Last_{bt}^h, \tag{2.5}$$

where u_b is a brand-specific intercept, X_{bt}^h is a vector of marketing and household-specific covariates; and $\boldsymbol{\beta}$ is a vector of response coefficients. The components of X_{bt}^h might include $PRICE_{bt}$, the net price of brand b at time t, $FEAT_{bt}$ and $DISP_{bt}$ as feature and display indicators for brand b, and BL_b^h is the intrinsic loyalty or preference for brand b, calculated as the within-household h market share of brand b in an initialization period and assumed constant over time (Bucklin et al. 1998). The BL term can be eliminated if differences in customer preference (μ_b) are modeled as unobserved heterogeneity (see Sect. 2.2.5). The term $Last_{bt}^h$ is a dummy that is 1 in case brand b was bought last time by household h, and zero else. It captures purchase-event feedback or "state dependence" (see Sect. 2.6.1).

2.2.3 Purchase Quantity Model

Given purchase incidence and choice of brand b, the probability that household h buys $q_{bt}^h = 1, 2, \ldots, n$ units at time t is captured by a Poisson model with a truncation at the zero outcome (Bucklin et al. 1998). This can be written as:

$$P(Q_{bt}^h = q_{bt}^h \mid I_t^h = 1, C_t^h = b) = \frac{\exp\left(-\lambda_{bt}^h\right)\left(\lambda_{bt}^h\right)^{q_{bt}^h}}{\left[1 - \exp\left(-\lambda_{bt}^h\right)\right] q_{bt}^h!}, \tag{2.6}$$

---

[2] Multinomial probit is an alternative to the logit (e.g., Jedidi et al. 1999). The advantage of the probit model is that it avoids the independence of irrelevant alternatives (IIA) assumption of logit models (see Guadagni and Little 1983). However, it does not produce a closed form for the probability of consumer choice.

where λ_{bt}^h is the purchase rate of household h for brand b at time t. This parameter is a function of (mean-centered) inventory, the average number of units purchased by the household, and the size, price, and promotion status of the selected brand:

$$\lambda_{bt}^h = \exp\left(\theta_0 + \theta_1(Inv_t^h - \overline{Inv}_\cdot^h) + \theta_2\overline{Q}^h + \theta_3 SIZE_b + \theta_4 PRICE_{bt} + \theta_5 FEAT_{bt} + \theta_6 DISP_{bt}\right)$$

2.2.4 Estimation

The likelihood function for incidence, choice, and quantity is given by:

$$L = \prod_{h=1}^{H}\prod_{t=1}^{T}\prod_{b=1}^{B}\left(P(I_t^h = 1)^{Y_t^h}(1 - P(I_t^h = 1))^{1-Y_t^h} P(C_t^h = b \mid I_t^h = 1)^{Z_{bt}^h} P\left(Q_{bt}^h = q_{bt}^h \mid I_t^h = 1, C_t^h = b\right)^{Z_{bt}^h}\right)$$

where

Y_t^h Category purchase indicator, equals 1 if household h purchased the category on shopping trip t; 0 otherwise

Z_{bt}^h Brand purchase indicator, equals 1 if household h purchased brand b on shopping trip t; 0 otherwise

Methods used to estimate the model include maximum likelihood, simulated maximum likelihood (Train 2003), and Bayesian "MCMC" estimation (Rossi et al. 2005). Once the incidence, choice, and quantity models have been estimated, we can calculate the incidence elasticity, $\eta_I = \frac{\partial P(I)}{\partial PRICE}\frac{PRICE}{P(I)}$; the choice elasticity, $\eta_{C|I} = \frac{\partial P(C|I)}{\partial PRICE}\frac{PRICE}{P(C|I)}$; and the quantity elasticity, $\eta_{Q|I,C} = \frac{\partial E(Q)}{\partial PRICE}\frac{PRICE}{E(Q)}$, where $E(Q)$ is the expected purchase quantity (see Gupta 1988 for details). Gupta (1988) decomposes the total sales elasticity η_S into that due to purchase incidence (η_I), brand switching $(\eta_{C|I})$ and purchase quantity $(\eta_{Q|I,C})$, so that $\eta_S = \eta_I + \eta_{C|I} + \eta_{Q|I,C}$. For example, a sales elasticity of −3 with respect to promotion might be decomposed as $\underset{(100\%)}{-3} = -\underset{(15\%)}{0.45} \underset{(75\%)}{-2.25} - \underset{(10\%)}{0.3}$, i.e., the brand switching elasticity comprises 75% of the total elasticity, whereas the incidence elasticity is 15% and the quantity elasticity is 10%. We refer to Sect. 2.5 for a more in-depth discussion on how (and how not) to interpret this result.

2.2.5 Heterogeneity

Consumers are naturally heterogeneous in their brand preferences, responsiveness to marketing actions, and how much they learn from the product usage experience. As a result, the parameters of the choice, incidence, and quantity models (Eqs. 2.2–2.6) should be modeled to vary across consumers. For example, Eq. (2.5) could be written as:

$$V_{bt}^{h} = u_{b}^{h} + \boldsymbol{\beta}^{h} X_{bt}^{h} = u_{b}^{h} + \beta_{1}^{h} PRICE_{bt} + \beta_{2}^{h} FEAT_{bt} + \beta_{3}^{h} DISP_{bt} + \beta_{4}^{h} BL_{b}^{h} + \beta_{5}^{h} LAST_{bt}^{h} \tag{2.7}$$

The brand-specific intercept from Eq. (2.5) is now household-specific (u_{b}^{h}) meaning that households can differ in their preferences for various brands. The response coefficients for variable k (β_{k}^{h}) also differ across households. This means that households can differ in their sensitivity to price, feature and display and the other variables in the utility model.

Modeling heterogeneity adds much complexity to choice, incidence, and quantity models. It is worthwhile to make clear why modeling heterogeneity is important:

- *Spurious State Dependence*: In a homogeneous model, the state dependence parameter (β_{5}^{h}) would be over-stated because it soaks up the variation due to heterogeneous preference as well as dynamically changing preference (see Keane 1997a, also Abramson et al. 2000).
- *Segmentation*: Marketing is about segmentation. By learning about heterogeneity, we make our models more useful because we can segment the market based on preference or response.
- *Avoid Independence of Irrelevant Alternatives (IIA)*: Logit models are open to the IIA criticism (see Guadagni and Little 1983). Modeling heterogeneity eliminates IIA problems at the aggregate level, although it still exists at the level of each household. Steenburgh (2008), Rooderkerk et al. (2011), and Liu et al. (2015) develop choice models to overcome IIA.
- *Better Prediction*: Incorporating heterogeneity means that our models incorporate more information; hence they should be expected to predict better.

Researchers face a myriad of decisions in how to model heterogeneity:

- *Distribution of the Individual Parameters*: The distribution of individual-level parameters can be considered to be continuous (Chintagunta et al. 1991), discrete (Kamakura and Russell 1989), or finite mixture (e.g., Varki and Chintagunta 2004).
- *Parameters to be Considered Heterogeneous*: The parameters to model heterogeneously can include preference, most coefficients, or all coefficients.
- *Joint Distribution of the Parameters*: The distribution of the heterogeneous parameters can be considered to be uncorrelated, correlated, or no assumption made.
- *Incorporation of Observed Heterogeneity*: Heterogeneity can be thought of as "observed" versus "unobserved." Observed heterogeneity means that the heterogeneity in any parameter can be captured by measurable variables such as demographics or preference (BL_{b}^{h}). Observable heterogeneity is easy to

incorporate simply by including the observed variables in the model. The concern is that these measures do not capture all the heterogeneity, and so researchers often model unobserved heterogeneity in addition to observed heterogeneity.
- *Choice Set Heterogeneity*: Eq. (2.4) assumes that each household considers the same brands when making a choice. Researchers (e.g., Siddarth et al. 1995) have questioned this assumption and model heterogeneity in choice sets.
- *Estimation*: Maximum Likelihood (ML), Simulated Maximum Likelihood (SML), and Bayesian are possible ways to estimate the model.

The above choices give rise to $3 \times 3 \times 3 \times 2 \times 2 \times 3 = 324$ possible ways to handle heterogeneity. No one route has emerged as most popular. Table 2.3 gives a summary of how a few papers have handled heterogeneity.

While no one single method has been shown to be superior, and sometimes the differences across various approaches are not crucial (e.g., Andrews et al. 2002a, b), the general conclusion is that it is crucial to include some form of unobserved heterogeneity in the model, at a minimum, in the brand-specific intercept. The reasons are (1) Heterogeneity improves fit and prediction (e.g., Chintagunta et al. 1991), (2) Heterogeneity changes the coefficients of other variables (e.g., Chintagunta et al. 1991), although Ailawadi et al. (1999) note that aggregate price elasticities may not change, (3) State dependence can be over-stated when preference heterogeneity is not included (e.g., Keane 1997a; see also Horsky et al. 2006).[3]

So far we have focused on parameter heterogeneity *across consumers*. Another form of heterogeneity is *over-time parameter variation*, even within the same consumer. For example, consumers may display cyclical purchase behavior, where periods of high consumption are alternated with periods of low consumption. Park and Gupta (2011) show this happens in the yogurt category. They use a hidden Markov Model (HMM) that distinguishes between high and low category purchase states. Transition between states is modelled probabilistically. Park and Gupta (2011) demonstrate that the model fits better than benchmark models and show that companies can benefit by targeting households that are in the high category purchase state.

2.2.6 Integrated Incidence, Choice, Quantity Models

Models have been developed that expressly integrate two or more consumer decisions. The integration takes place through correlations between error terms, the formulation of the utility function, or defining the set of decision options. For example, Krishnamurthi and Raj (1988) integrate brand choice and quantity decisions by correlating the error terms of the choice and quantity equations. Nested

---

[3]It is noteworthy that there is some evidence (Abramson et al. 2000; Chiang et al. 1999) that not including choice set heterogeneity significantly distorts parameters.

Table 2.3 Ways to handle heterogeneity in household-level models (selected papers)

	Distribution of parameters	Which parameters heterogeneous	Joint distribution of parameters	Choice set heterogeneity	Including observed heterogeneity?	Estimation
Kamakura and Russel (1989)	Discrete	All	N/A	No	No	ML
Ailawadi et al. (2007b)	Continuous	Most	No stipulation	No	No	SML
Ansari et al. (2006)	Continuous	Most	Correlated	No	No	Bayesian
Gupta and Chintagunta (1994)	Discrete	All	N/A	No	Yes	ML
Seetharaman et al. (1999)	Continuous	All	Correlated	No	Yes	Bayesian
Chiang et al. (1999)	Continuous	All	Correlated	Yes	No	Bayesian

logit posits a utility function that integrates choice and incidence decisions (see Ben-Akiva and Lerman 1991). The incidence portion of the nested logit is what we describe in Sect. 2.2.1. Another set of models integrates choice and incidence by adding a "no-purchase" option, i.e., the consumer is assumed to choose among a set of alternatives, $J - 1$ of which are brands; the Jth is the no-purchase option. See Chib et al. (2004) and the papers to which they refer for examples. Chiang (1991) and Chintagunta (1993) develop integrated models of incidence, choice, and quantity. Bell and Hilber (2006) investigate the relationship between incidence and quantity. They find that consumers with greater storage constraints shop more often and purchase smaller quantities per visit. While modeling the decisions in an integrated way is attractive from econometric and behavioral perspectives, the overall decomposition results do not seem to differ much from separate models. For instance, for the coffee category Gupta (1988) uses non-integrated models whereas Chiang (1991) uses integrated models, but their results on the elasticity decomposition into brand switching, incidence and quantity are almost identical (see Table 2.5 in Sect. 2.5).

2.2.7 Dynamic Structural Models

Another approach to modeling consumer decisions is dynamic structural models. These models begin with the household utility function and include dynamic phenomena such as "forward-looking" consumer behavior, where consumers take into account future utility in making current-period decisions, and consumer learning of brand quality, often in the form of Bayesian learning (See Sect. 2.6.3). Erdem and Keane (1996) develop a dynamic structural model of brand choice that includes forward-looking and learning behavior. Gönül and Srinivasan (1996) develop a dynamic structural model of purchase incidence. Sun et al. (2003) develop a dynamic structural model of incidence and choice. Erdem et al. (2003), Sun (2005), and Chan et al. (2008) develop dynamic structural models of incidence, choice, and quantity. Using structural models versus nonstructural models seems to affect the elasticity decomposition. For instance, Sun et al. (2003) report a brand switching percentage of 56% for a dynamic structural model that accounts for forward-looking customers, whereas the percentage for the non-structural integrated model (nested logit) is 72%.

2.3 Customer-Level Models—Extensions

Researchers have developed models that extend the classical incidence-brand choice-purchase quantity set-up. The key extensions involve store switching (Sect. 2.3.1), cross-category effects (Sect. 2.3.2), SKU-level models (Sect. 2.3.3), and purchase deceleration (Sect. 2.3.4).

2.3.1 Extension 1: Store Switching

Bucklin and Lattin (1992) propose a model that can be used to capture store switching effects. Their store choice model is given by a multinomial logit model that includes store loyalty and store features as explanatory variables.

One challenge in measuring the effects of promotions on store choice is that consumers make store choice decisions based on a host of factors (e.g., location, produce quality, waiting lines) that have little to do with an individual brand's price and promotion. At the same time, price and promotion for brands and categories will affect store choice. Lourenço et al. (2015) offer a new approach to address the question of determining how individual promotions affect the store price image that consumers have.

Another challenge is that the standard logit model for store choice assumes that a consumer knows each store's prices and promotions, which seems unlikely. It seems more likely that there is indirect store switching: a promotional purchase in one store preempts a regular purchase in another store, as we discussed in Sect. 2.1. To tackle these complicating factors, increasingly sophisticated models of store choice are available (Bell and Lattin 1998; Bell et al. 1998; Rhee and Bell 2002; Bodapati and Srinivasan 2006; Singh et al. 2006; Guyt and Gijsbrechts 2014; Van Lin and Gijsbrechts 2014, 2015; Vroegrijk et al. 2013).

2.3.2 Extension 2: Cross-Category Effects

Cross-category effects means that a sales promotion for a brand in a category may either steal sales from brands in other categories (substitution effect) or it may enhance brand sales across categories (complementary effects). This is what retailers hope occurs (Lam et al. 2001). Ailawadi et al. (2007a) call these positive cross-category effects halo effects, and find empirical evidence for them in a major drugstore chain.

To capture cross-category effects, Manchanda et al. (1999) specify a model for shopping basket composition, i.e., what set of categories is bought on a specific shopping trip. They model this as a multivariate probit. Multivariate probits (or logits) differ from multinomial probits (or logits) in that more than one alternative can be chosen on the current purchase occasion. This is the case when we are modeling the *set* of categories purchased. Manchanda et al.'s model includes "complementarity" effects, whereas the price of one category influences sales in another category, and "coincidence" effects, where certain products are bought together.

Mehta and Ma (2012) propose a more elaborate model to capture multi-category brand choice decisions. While Manchanda, Ansari and Gupta only model purchase incidence decisions, Mehta and Ma (2012)'s model captures brand choice, incidence and quantity decisions. Their model allows for cross-category effects of promotions both in the incidence and purchase quantity decisions. The model helps retailers in understanding how they should allocate promotional expenditures across brands within categories and coordinate timing of promotions of brands across categories.

2.3.3 Extension 3: SKU-Level Models

Most choice models are at the brand level. However, consumers buy specific SKUs. The advantage of an SKU-level model is specificity and allowing for cross-SKU elasticities, especially cross-SKU effects between SKUs for the same brand. E.g., does promoting Yoplait's 6-oz. vanilla yogurt take away from Yoplait's 6-oz. blueberry yogurt? The disadvantage of SKU-level models is there can be so many SKUs in a given category that the modeling becomes infeasible.

Fader and Hardie (1996) propose a parsimonious model for SKU choice that addresses this problem. Suppose there are N attributes and let L_n be the number of levels associated with the nth attribute. Define the set $\{l_1, l_2, \ldots, l_N\}$ as the unique set of attribute levels for brand b, SKU j. Fader and Hardie (1996) model the SKU-specific intercept in the logit model (Eq. 2.4) as: $u_{bj} = \sum_{n=1}^{N} m_{bjn}\alpha_n$, where m_{bjn} is an elementary row vector, the l_nth element of which equals 1, and α_n is the vector of preferences over the L_n levels of attribute n. A similar approach was followed by Ho and Chong (2003) and Chintagunta and Dubé (2005). For the covariates X_{bjt}^{h} (Eq. 2.4) we may use SKU-level versions of the variables in the brand choice model.

2.3.4 Extension 4: Deceleration

Deceleration means that consumers anticipate promotions, and consequently they may postpone purchases until a promotion is offered. To capture deceleration, we need a model component for the effect of an expected future promotion on current purchase behavior. Van Heerde et al. (2000) and Macé and Neslin (2004) use actual future prices in a model of brand sales to capture deceleration in a model of brand sales (more about this in Sect. 2.4.2). For household data, Sun et al. (2003) present a structural model for the promotion expectation process. They assume that consumers expect future promotions according to a first-order Markov model. The authors find that the estimated expectations conform rather well with the actual promotion schedule (see also Erdem et al. 2003).

Sun et al. (2003) propose measuring deceleration by adding a variable $PromTime_t^h$ to the category incidence model (Eq. 2.2). This represents the time since the last promotion, and is meant to capture that consumers may hold out until the next promotion. To obtain $PromTime_t^h$ they calculate the average time between promotions in the category. If the time since the last promotion in the category seen by the consumer is greater than this average, $PromTime_t^h$ equals 1; otherwise it equals 0. If $PromTime_t^h$ is 1, a consumer may expect a promotion soon, and defers the current purchase. As a result, they expect (and find) the estimated coefficient for $PromTime_t^h$ to be negative.

2.3.5 Discussion

While the literature provides models for each of the key consumer responses to sales promotion, there are no papers yet that combine all possible responses that are listed in Table 2.2. Van Heerde and Gupta (2006) come close by combining store switching, category incidence, brand and SKU choice, purchase quantity, increased consumption effects and deceleration effects. This allows them to identify 18 of the 24 possible sources of the sales promotion bump from Table 2.2. Since Van Heerde and Gupta (2006) do not model category choice (Sect. 2.3.2), they do not measure effects related to category switching.

To estimate all 24 decomposition sources, one would have to estimate a model for all pertinent consumer decisions. This would require specifying the probability for the decision to choose store s, category k, brand b, SKU j and quantity q_{skbt}^h as:

$$P\left(Q_{skbt}^h = q_{skbt}^h\right) = P\left(S_t^h = s\right) \times P\left(Cat_{skt}^h = 1 \mid S_t^h = s\right) \times P\left(C_{skt}^h = \{b,j\} \mid Cat_{skt}^h = 1, S_t^h = s\right) \\ \times P\left(Q_{bt}^h = q_{skbt}^h \mid S_t^h = s, Cat_{skt}^h = 1, C_{skt}^h = \{b,j\}\right) \tag{2.8}$$

The first three components at the right hand side can be modeled using store, category, and SKU models reviewed in this section. The quantity decision could be modeled using Eq. (2.6). However, it isn't clear whether one could derive an analytical formula to derive the 24-state decomposition in Table 2.2, or whether one would have to use simulation. For example, it would be difficult to distinguish between a category switch and a new category purchase that represents additional expenditures from the household's total budget.

2.4 Store-Level Models of Sales Promotion Effects

A large body of research has developed store-level models for sales promotion effects. These models draw on weekly store-level scanner data that are more readily available, more representative, and easier to process than household-level scanner data (Van Heerde et al. 2004). Table 2.4 shows that the phenomena that are studied with store-level vs household-level data are similar, but the terminology can differ.

Table 2.4 Related terms in models of household- and store data

Household data	Store data
Purchase (Sect. 2.1)	Sales (Sect. 2.4.1)
Category incidence (Sect. 2.2.1)	Number of buyers in category
Brand switching (Sect. 2.2.2)	Cross-brand effects (Sects. 2.4.1 and 2.4.3)
Acceleration (Sect. 2.2.3)	Postpromotion dips (Sect. 2.4.2)
Deceleration (Sect. 2.3.4)	Prepromotion dips (Sect. 2.4.2)

2.4.1 Scan*Pro Model

Perhaps the most well-known store-level model for sales promotion effects is Scan*Pro (Wittink et al. 1988).[4] It is a multiplicative regression model for brand sales, explained by own-brand and cross-brand prices and promotions:

$$S_{bst} = \lambda_{bs}\mu_{bt} \prod_{b'=1}^{B} \left\{ PI_{b'st}^{\beta_{b'b}} \cdot \gamma_{1b'b}^{FEATONLY_{b'st}} \cdot \gamma_{2b'b}^{DISPONLY_{b'st}} \cdot \gamma_{3b'b}^{FEAT\&DISP_{b'st}} \right\} e^{u_{bst}}, \tag{2.9}$$

where

S_{bst}	sales (in units) of brand b in store s in week t
PI_{bst}	price index (ratio of current to regular price) of brand b in store s in week t
$FEATONLY_{bst}$	indicator for feature only: 1 if there is a feature without a display for brand b in store s in week t, 0 else
$DISPONLY_{bst}$	indicator for display only: 1 if there is a display without feature for brand b in store s in week t, 0 else
$FEAT\ \&\ DISP_{bst}$	indicator for feature and display: 1 if there is a feature and display for brand b in store s in week t, 0 else

The model includes a brand-store specific intercept λ_{bs}, a brand-week specific intercept μ_{bt}, own- (β_{bb}) and cross-price ($\beta_{b'b}, b' \neq b$) elasticities, and multipliers for own- (γ_{1bb}, γ_{2bb}, γ_{3bb}) and cross-brand ($\gamma_{1b'b}$, $\gamma_{2b'b}$, $\gamma_{3b'b}$, $b' \neq b$) effects for feature, display, and feature & display. Note by including these latter three variables, one can investigate a possible interaction between Feature and Display. The model is linearized by taking logs and estimated by ordinary least squares.[5]

Van Heerde et al. (2001) propose a semiparametric version of Scan*Pro, estimated by nonparametric techniques. Their results show that the response of sales to the percentage price discount is S-shaped, i.e., there are threshold and saturation effects.

[4]This working paper was reprinted as Chap. 12 of Wieringa et al. (2011).

[5]Another important method is "PromotionScan" (Abraham and Lodish 1993). PromotionScan is a time series approach to estimating the short-term promotion bump. It is based on the "Promoter" method (See Sect. 2.8).

A key independent variable in the Scan*Pro model is the price index (the ratio of actual to regular price). In store data, both actual and regular prices are typically available whereas household data tend to include price paid only. The price index captures the effects of promotional price changes, which may be quite different from regular price effects (Mulhern and Leone 1991; Bijmolt et al. 2005). If there is sufficient variation in regular price, it can be included as a separate predictor variable as in Mulhern and Leone (1991).

Note that Eq. (2.9) allows for asymmetric cross-effects between brands, i.e., the impact of promoting Brand A on sales of Brand B is not the same as the impact of promoting Brand B on sales of Brand A. This is an important feature because asymmetric brand switching has been consistently found in the promotions literature, although its causes are not yet completely explained (Neslin 2002). A downside of modeling all cross effects is that for N brands one needs N^2 parameters per marketing mix instrument. As a result, parameter estimates can be unreliable and lack face validity. Rooderkerk et al. (2013) use a parsimonious specification where the cross effects between SKUs are a function of the similarity of these SKUs.

Aggregate logit models (Sect. 2.4.4) overcome this problem by estimating only one parameter per instrument, from which the N^2 own- and cross effects are derived. In case a modeler wants to consider not only cross-brand effects within categories but also across category effects, the number of cross-effects may become really problematic. To handle this case, Kamakura and Kang (2007) present a parsimonious solution based on a principal-component representation of response parameters.

While the Scan*Pro Model specifies a model for brand sales, an alternative approach is to model both category sales and market share, as brand sales is the product of the two components. Breugelmans and Campo (2011) use this approach to study the effectiveness of in-store displays in a virtual store environment. They use an attraction specification for the market share model, which ensures that market shares add up to one across brands and are bound between zero and one. A difference between this approach and Scan*Pro is that market share models tend to impose a structure on the own and cross effects (e.g., cross effects are a mathematical function of own effects) whereas the Scan*Pro model estimates the cross effects freely. Freely estimated cross effects allow the data to “speak for themselves,” but they do not impose any constraint on the estimates, and hence cross effects with counterintuitive signs are possible in the Scan*Pro model. Incorrect signs for cross effects are less likely in market share models.

2.4.2 Models for Pre- and Postpromotion Dips

Van Heerde et al. (2000) and Macé and Neslin (2004) have used store-level data to measure the aggregate effects of acceleration (i.e., postpromotion dips) and deceleration effects (i.e., prepromotion dips). The (simplified) model is of the form:

$$\ln S_t = \alpha_0 + \alpha_1 \ln PI_t + \sum_{u=1}^{S} \beta_u \ln PI_{t-u} + \sum_{v=1}^{S'} \gamma_v \ln PI_{t+v} + \varepsilon_t, \tag{2.10}$$

where

$\ln S_t$ natural log of sales in week t

$\ln PI_t$ natural log of price index (ratio of current to regular price) in week t

β_u u-week lagged effect of promotion, corresponding to post-promotion dips (acceleration). A positive β_u indicates that a price decrease is followed by a sales dip u weeks later

γ_v v-week lead effect of promotion, corresponding to pre-promotion dips (deceleration). A positive γ_v indicates that a price decrease in v weeks from now is preceded by a decrease in sales now

Macé and Neslin (2004) estimate pre- and postpromotion dips on data spanning 83 stores, 399 weeks, and 30,861 SKUs from 10 product categories. They find that 22% of the sales promotion bump is attributable to postpromotion dips, and 11% to prepromotion dips. Hence, pre- and postpromotion together are one-third of the sales promotion bump, which is remarkably close to the 32% cross-period effect reported by Van Heerde et al. (2004). Macé and Neslin (2004) find that SKU, category, and store-trading area customer characteristics explain significant variation in pre- and postpromotion elasticities.

Note that models for household- and store-level data deal differently with purchase acceleration. Since typical household-level models do not incorporate a store choice model, acceleration effects manifest both *within* the same store (source 9 in Table 2.2) and *across* stores (source 17 in Table 2.2). In store-level models such as (2.10), the aggregate outcome of acceleration (postpromotion dips) is only captured *within* the same store, which is source 9 in Table 2.2. As a result, one may expect larger acceleration effects in household-level than in store-level models.

2.4.3 Store-Level Decomposition Model

Van Heerde et al. (2004) propose a regression-based method for decomposing own-brand effects into cross-brand (brand switching), cross-period (acceleration & deceleration), and category expansion effects. The method uses the identity that total category sales (TCS) during periods $t - S'$ through $t + S$ equals sales of the target brand in period t ("own-brand sales" or OBS) plus sales of other brands in period t ("cross-brand sales" or CBS) plus total category sales in period $t - S'$ through $t + S$, excluding period t ("pre- and post-period category sales" or PPCS). Therefore, TCS = OBS + CBS + PPCS, or –OBS = CBS + PPBC − TCS. The method regresses these four variables on the same set of regressors:

–OBS (own-brand sales) $-S_{bt} = \alpha^{ob} + \beta^{ob} PI_{bt} + \sum_{k=1}^{K} \gamma_k^{ob} X_{kt} + \varepsilon_{bt}^{ob}$

CBS (cross-brand sales) $\sum_{\substack{b'=1 \\ b' \neq b}}^{B} S_{b't} = \alpha^{cb} + \beta^{cb} PI_{bt} + \sum_{k=1}^{K} \gamma_k^{cb} X_{kt} + \varepsilon_{bt}^{cb}$

PPCS (cross-period sales) $\sum_{\substack{u=-S' \\ u \neq 0}}^{S} \sum_{\substack{b'=1 \\ b \neq b}}^{B} S_{b't+u} = \alpha^{cp} + \beta^{cp} PI_{bt} + \sum_{k=1}^{K} \gamma_k^{cp} X_{kt} + \varepsilon_{bt}^{cp}$

–TCS (total category sales) $-\sum_{u=-S'}^{S} \sum_{b=1}^{B} S_{bt+u} = \alpha^{ce} + \beta^{ce} PI_{bt} + \sum_{k=1}^{K} \gamma_k^{ce} X_{kt} + \varepsilon_{bt}^{ce}$

where $\sum_{k=1}^{K} \gamma_K X_{kt}$ captures the effects of covariates such as cross-brand instruments, store dummies, and week dummies. Since –OBS = CBS + PPBC – TCS, the parameters for the price indices (*PI*) add up in the following way:

$$\beta^{ob} = \beta^{cb} + \beta^{cp} + \beta^{ce}, \tag{2.11}$$

Equation (2.11) decomposes the own-brand promotion effect (β^{ob}) into cross-brand switching (β^{cb}), acceleration and decoration (β^{cp}), and category expansion (β^{ce}). Equation 2.11 can be divided through by β^{ob} to provide a percentage decomposition.

All parameters in Eq. (2.11) are expected to be positive. If there is a promotional price discount for brand b, PI_{bt} decreases, own brand sales increases (presumably), and hence minus own brand sales decreases. Consequently, the regression coefficient β^{ob} will be positive. Similarly, a price discount for brand b decreases cross-brand sales (presumably), which implies $\beta^{cb} > 0$. Furthermore, if a decrease in PI_{bt} leads to a decrease in cross-period sales (i.e., pre- and postpromotion dips) $\beta^{cp} > 0$. Finally, if the price discount for brand b manages to increase category sales, then total category sales increase and the negative decreases, and $\beta^{ce} > 0$.

Van Heerde et al. (2004) obtain the decomposition for four types of promotional price discounts: without feature- or display support, with feature-only support, with display-only support, and with feature- and display support. To accomplish this, they use a specific set of independent variables, discussed in the appendix to this chapter.

Van Heerde et al. (2004) provide two extensions of (2.11). One of them defines the model at the SKU level and splits the "cross-brand effect" (or cross-item effect) into within-brand cannibalization (β^{cbw}) and between-brand switching (β^{cbb}), i.e., $\beta^{cb} = \beta^{cbw} + \beta^{cbb}$. The other extension splits the category expansion effect into a cross-store effect (β^{cs}) and a market-expansion effect (β^{me}) (the category-growth effect in Tables 2.1 and 2.2), i.e., $\beta^{ce} = \beta^{cs} + \beta^{me}$. This allows the model to quantify within-brand SKU switching as well as store switching.

Leeflang et al. (2008) extend the model in yet another direction by accounting for cross-category effects. The model allows for positive (complementary) and negative (substitution) cross-category effects. The method uses pairs of categories between which complementary or substitution effects can be expected (e.g., canned beer and bottled beer). Leeflang and Parreña Selva (2012) also study cross-category effects but focus on the moderating factors of these effects, such as the physical distance between categories in a store.

2.4.4 Heterogeneity Across Stores

The original Scan*Pro paper (Wittink et al. 1988) reports strong differences in promotional responses across US regions. Brand managers may exploit these differences by tailoring promotions at the regional level. Several studies have since examined how promotional effectiveness varies across stores. Hoch et al. (1995) and Montgomery (1997) use hierarchical Bayesian methods to allow price response parameters to differ across stores. Hoch et al. (1994) and Montgomery (1997) use store and trading characteristics to explain variation in price sensitivity across stores. Montgomery (1997) shows how the model can be used to adjust prices at the store level for enhanced profitability.

Haans and Gijsbrechts (2011) study how the effectiveness of promotions on category sales varies with the size of the store. They use a hierarchical linear model where log category sales is explained by log price, discount depth, feature and display and a quantity discount variable. The response parameters are store-specific, and they are explained by the size of the store in a second layer. Haans and Gijsbrechts (2011) find that while the percentage lift due to promotions is smaller in larger stores, the absolute effect are larger.

2.4.5 Aggregate Logit Model

A frequent criticism of regression models of promotion is they are not rooted in economic theory. The aggregate logit model overcomes this issue, which is one reason it is increasingly popular in marketing science. Another reason is that it accommodates own- and cross effects with an economy of parameters, which is something that does not hold for the aggregate models discussed in Sects. 2.4.1–2.4.3. The aggregate logit model was introduced by Berry et al. (1995). Its logic is that individual consumers maximize utility and choose brands according to a multinomial logit model.

Estimation of aggregate logit models can take into account price endogeneity (Besanko et al. 1998; Villas-Boas and Winer 1999). For instance in times of a positive demand shock, managers may increase price. To account for this, researchers have correlated price with the error term (Besanko et al. 1998; Villas-Boas and Winer 1999). If the endogenous nature of price is ignored, its coefficient β_2 may be underestimated quite severely as shown in a meta-analysis of price elasticity (Bijmolt et al. 2005).

To complement the demand model (2.17), Besanko et al. (1998) assume a certain model of competitive conduct, and derive a supply model from that. Next,

Table 2.5 Sales bump decomposition results reported in the literature

Study	Category	% secondary demand (=brand switching effect)	% primary demand (=own sales effect not due to brand switching)
		Elasticity decomposition	
Gupta (1988)	Coffee	84	16
Chiang (1991)	Coffee (feature)	81	19
	Coffee (display)	85	15
Chintagunta (1993)	Yogurt	40	60
Bucklin et al. (1998)	Yogurt	58	42
Bell et al. (1999)	Margarine	94	6
	Soft drinks	86	14
	Sugar	84	16
	Paper towels	83	17
	Bathroom tissue	81	19
	Dryer softeners	79	21
	Yogurt	78	22
	Ice cream	77	23
	Potato chips	72	28
	Bacon	72	28
	Liquid detergents	70	30
	Coffee	53	47
	Butter	49	51
Chib et al. (2004)	Cola (price)	78	22
	Cola (display)	68	32
	Cola (feature)	64	36
Van Heerde et al. (2003)	Sugar	65	35
	Yogurt	58	42
	Tuna	49	51
Nair et al. (2005)	Orange juice	65	35
Average elasticity decomposition		71	29
		Unit-sales decomposition	
Pauwels et al. (2002)	Soup	11	89
	Yogurt	39	61
Van Heerde et al. (2003)	Sugar	45	55
	Yogurt	33	67
	Tuna	22	78

(continued)

Table 2.5 (continued)

Study	Category	% secondary demand (=brand switching effect)	% primary demand (=own sales effect not due to brand switching)
Sun et al. (2003)	Ketchup	56	44
Van Heerde et al. (2004)	Peanut butter	43	57
	Shampoo	31	69
	Tuna	31	69
	Bathroom tissue	21	79
Nair et al. (2005)	Orange juice	8	92
Sun (2005)	Tuna	42	58
	Yogurt	39	61
Chan et al. (2008)	Tuna	28	72
	Paper towels	14	86
Leeflang et al. (2008)	Bottled beer	18	82
	Canned beer	11	89
	Concentrated fabric softeners	13	87
	Non-concentrated fabric softeners	29	71
	Concentrated dish detergents	4	96
	Non-concentrated dish detergents	28	72
	Detergents	39	61
Average unit sales effect decomposition		28	72

the demand and supply models are estimated simultaneously. Whereas the original aggregate logit model assumed homogenous consumers, several papers in marketing have relaxed this assumption (e.g., Chintagunta 2001; Dubé et al. 2002). Nevo (2000) provides guidelines how to estimate a heterogeneous aggregate logit model. Moreover, Nair et al. (2005) propose aggregate models that not only capture underlying brand choice decisions, but also incidence and quantity. Their demand elasticity breakdown shows that brand choice accounts for 65% and incidence and quantity for 35%, which is in the same ballpark as the breakdowns obtained from individual-level data (see Table 2.5 in Sect. 2.5).

Park and Gupta (2009) develop a simulated maximum likelihood estimation method for the random coefficient aggregate logit model. The method is especially suitable when there are relatively small samples of shoppers, leading to measurement error. The estimation accounts for endogeneity and it yields unbiased and efficient estimates of the demand parameters.

One drawback of aggregate logit models is that it is difficult to identify unobserved heterogeneity with aggregate data (Bodapati and Gupta 2004). Another one is that it requires the specification of an outside good to account for non-incidence.

This outside good has to be based on assumptions on population size and category consumption, which may be questionable. Yet another drawback of current aggregate logit models is that they typically ignore dynamic and quantity effects such as acceleration, deceleration, and purchase-event feedback. An issue that also needs further development is that the assumed competitive conduct refers to long-term stable prices instead of to Hi-Lo pricing that is the essence of promotional pricing (Neslin 2002, p. 45).

2.5 Generalizations About the Decomposition

Table 2.5 summarizes the findings of the literature on the decomposition of the sales promotion bump. For each study we indicate the product category and the percentage attributed to secondary demand effects (effects that cause substitution from other brands, i.e., brand switching) and the percentage due to primary demand effects (the part of the sales promotion bump that is not due to brand switching). There are two fundamental approaches to calculating the decomposition, the "elasticity" approach and the "unit sales" approach. The elasticity approach is explained in Sect. 2.2.4, and was originated by Gupta (1988). It is based on the mathematical relationship that the elasticity of the probability of buying brand b at time t with respect to (an assumed continuous measure of) promotion equals the sum of the elasticities of brand choice, purchase incidence, and purchase quantity with respect to promotion. The unit sales composition looks at changes in actual sales of the promoted brand as well as other brands in the category. The upper part of Table 2.5 shows the elasticity decomposition, the lower part shows the unit sales decomposition. There are two important findings in Table 2.5:

1. The unit sales decomposition yields lower secondary demand effects compared to the elasticity decomposition, generally in the 10–40% range.
2. Using the elasticity decomposition, there is a general downward trend in the percentage allocated to secondary demand, starting from about 80–85% and now at 45–70%.

The difference in findings for unit sales versus elasticity decompositions is detailed in Van Heerde et al. (2003). The brand choice component of the elasticity decomposition represents the change in the brand choice probability conditional on making a category purchase. The cross-brand component focuses on how many sales other brands lose when the focal brand is promoted (Van Heerde et al. 2003). The key difference is the way they treat the increase in purchase incidence due to the promotion. This increase is ignored in the elasticity decomposition, but in the unit sales decomposition, it is included in the calculation of the net sales loss for the other brands (Van Heerde et al. 2003). As a result, the actual loss in cross-brand sales is much less than what the elasticity-based secondary demand fraction suggests. This difference holds within the same category, as illustrated by the entries for orange juice (Nair et al. 2005) and sugar, yogurt and tuna (Van Heerde et al. 2003) in the top and lower parts of Table 2.5.

One possible reason for the general downward trend in secondary demand effects in elasticity is that more recent models allow for unobserved household heterogeneity. The argument is this (see also Keane 1997b; Sun et al. 2003). Suppose there are two brands, A and B. A large segment of customers lies in wait for brand A to be on promotion, while the other segment lies in wait for brand B's promotions. When brand A is promoted, it is almost exclusively bought by the first segment. As a result, the conditional choice probability for brand A increases spectacularly when it is promoted, while the conditional choice probability for brand B approaches zero. When brand B is promoted, the reverse occurs. Such a phenomenon leads to a very strong conditional brand choice elasticity. However, actually there is very little switching between A and B (promotion is influencing incidence, not switching), and hence there is small cross-brand sales loss in the unit sales decomposition. If this explanation holds, then models that allow for unobserved household heterogeneity should show a lower percentage brand switching than models that assume homogeneity. This could be the reason why the more recent elasticity decomposition results, which tend to be derived from heterogeneous models, show less brand switching. This seems a worthwhile direction for further research.

2.6 Long-Term Impact—Beyond the Immediate Sales Bump

Promotions affect consumers beyond the immediate sales bump. Promotions may lead to purchase-event feedback (Sect. 2.6.1). Promotions may also affect reference prices (Sect. 2.6.2). Over time, consumers learn price promotion patterns (Sect. 2.6.3). Promotions may affect long-term consumer behavior (Sect. 2.6.4). Finally, competitors may react (Sect. 2.6.5).

2.6.1 Purchase-Event Feedback

Purchase event feedback is the degree to which current purchases affect future brand preferences. This is known as "state dependence" in the economics literature (see Roy et al. 1996), and is due to consumer learning from the product purchase and usage experience.

Researchers have captured purchase event feedback by including a lagged purchase indicator such as $Last_{bt}^{h}$ in Eq. (2.5). However, Blattberg and Neslin (1990) distinguish between the purchase effect and the promotion usage effect—the purchase-event feedback from a purchase on promotion may be different than the feedback from a regular purchase. For example, self-perception theory suggests that if the consumer concludes he or she bought the brand because of the promotion rather than brand preference, purchase event feedback will be weakened (Dodson et al. 1978). Behavioral learning theory (Rothschild and Gaidis 1981) suggests promotion purchasing could enhance or detract from purchase event feedback. The effect could be positive if the promotion serves as a reward and thus encourages future purchasing, or negative if promotion merely trains consumers to buy on promotion. To investigate this, Gedenk and Neslin (1999) distinguish whether or not the last purchase was made on promotion. They find that price promotions detract from feedback. This finding is the same as originally reported by Guadagni and Little (1983)—a promotion purchase is less reinforcing than a non-promotion purchase, but better than no purchase at all. It is also the same as found by Seetharaman (2004).

Ailawadi et al. (2007b) propose yet another mechanism for feedback effects of promotions. They postulate that acceleration in the form of larger purchase quantity enhances purchase-event feedback because the household consumes more of the brand over a continuous period of time. In an empirical study of yogurt and ketchup, Ailawadi et al. (2007a, b) find that larger purchase quantity is associated with an increase in repeat purchase rates.

The measurement of purchase event feedback is quite challenging because of its potential confound with customer heterogeneity. Failure to account adequately for customer heterogeneity produces spurious state dependence findings (see Sects. 2.2.6 and 2.6.3).

2.6.2 Reference Prices

The reference price is the standard to which consumers compare an observed price in order to assess the latter's attractiveness (Kalyanaram and Winer 1995). Although there are many ways to operationalize reference price (Winer 1986), a significant body of literature supports the notion that individuals make brand choices based on this comparison. Briesch et al. (1997) conclude that a brand-specific exponentially smoothed reference price provides the best fit and prediction:

$$RP^h_{bt} = \alpha RP^h_{bt-1} + (1-\alpha)\text{Price}_{bt-1}, \tag{2.12}$$

where

RP^h_{bt} household h's reference price of brand b at purchase occasion t
α carryover parameter, $0 \leq \alpha \leq 1$

Prospect theory (Kahneman and Tversky 1979) predicts that consumers react more strongly to price increases than to price decreases (Kalyanaram and Winer 1995). To operationalize this, Erdem et al. (2001) define *LOSS* as the difference between the actual price and the reference price, given that the reference price is lower than the actual price. Similarly, *GAIN* is the difference given that the reference price is higher than the actual price:

$$LOSS_{bt}^{h} = \max\left\{\text{Price}_{bt-1} - RP_{bt}^{h}, 0\right\}$$
$$GAIN_{bt}^{h} = \max\left\{RP_{bt}^{h} - \text{Price}_{bt-1}, 0\right\}$$

To capture the direct effect of price as well as the effects of losses and gains, the utility function in the brand choice model (2.5) can be specified as (Briesch et al. 1997):

$$\boldsymbol{\beta} X_{bt}^{h} = \beta_1 PRICE_{bt} + \beta_2 FEAT_{bt} + \beta_3 DISP_{bt} + \beta_4 BL_{b}^{h} + \beta_5 Last_{bt}^{h} + \beta_6 LOSS_{bt}^{h} + \beta_7 GAIN_{bt}^{h}. \tag{2.13}$$

In (2.13), we expect $\beta_1 < 0$, $\beta_6 < 0$, and $\beta_7 > 0$. If losses loom larger than gains, $|\beta_6| > |\beta_7|$.

Recent papers question the findings regarding loss-aversion (Bell and Lattin 2000), and whether the reference price effect itself has been significantly over-estimated (Chang et al. 1999). The Chang et al. argument is that price sensitive consumers time their purchases to promotions, so observations of purchases with low prices over-represent price sensitive consumers and over-estimate both the loss and gain aspects of reference prices. Further work is needed to take into account these points. From a modeling standpoint, these papers illustrate the subtle but important challenges in modeling household-level data.

2.6.3 Consumer Learning

Frequent exposure to sales promotions may affect consumer perceptions of promotional activity (Krishna et al. 1991) and change their response to promotion. Mela et al. (1997) study the long-term effects of promotion and advertising on consumers' brand choice behavior. They use 8 ¼ years of panel data for a frequently packaged good. Their results indicate that consumers become more price and promotion sensitive over time because of reduced advertising and increased promotions. Mela et al. (1998) conclude that increased long-term exposure of households to promotions reduces their category purchase rate. However, when households do decide to buy, they buy more of the product. Such behavior is indicative of an increasing tendency to "lie in wait" for especially good promotions. This study was among the first to provide evidence of purchase deceleration (see Sect. 2.3.4).

Bijmolt et al. (2005) provide a meta-analysis of 1851 price elasticities reported in four decades of academic research in marketing. A salient finding is that in the period 1956–1999, the average (ceteris paribus) elasticity of sales to price went from −1.8 to −3.5. The relative elasticities (i.e., choice and market share) are quite stable (i.e., no significant change). Thus, the primary demand part of the sales elasticity is increasing over time, whereas the secondary demand part is stable. This finding is consistent with "lie-in-wait" behavior reported by Mela et al. (1998), but inconsistent with an increased sensitivity of the brand choice decision to price reported by Mela et al. (1997).

Consumers' learning of product quality has been examined within the purview of Bayesian learning models. See Sect. 2.2.6 for a discussion of dynamic structural models, several of which incorporate Bayesian learning. Erdem et al. (2008) develop a Bayesian learning model particularly relevant to promotions. They investigate the extent to which consumers learn the quality of a brand through experience, price/promotions, advertising frequency, and advertising content. They find experience is the most important source of learning, while price/promotions and the combined impact of advertising are equally important. They find that price promotions induce negative learning, reducing the total sales impact of promotion by 27% (p. 1122). The authors note however that these effects have to be disentangled from stockpiling (p. 1124).

This highlights the difficulty in measuring learning in structural models. Shin et al. (2012) show that failing to account correctly for customer preference heterogeneity can bias the estimate of learning. Their study is particularly interesting because it combines survey data of brand preferences with the usual consumer panel data. They find that without survey data, the extent of customer learning is significantly over-stated.

DelVecchio et al. (2006) shed further light on learning from promotions by conducting a meta-analysis of 42 studies generating 132 inferred correlations between "the use of sales promotion and post-promotion brand preference" (p. 207). They consider studies that measure choice as well as brand perceptions, based on field or laboratory data. They find that on average promotion does not have a statistically significant association with brand preference. However, there are moderators in both negative and positive directions. Promotions have a negative effect when the discount is 20% or more, when it is an unannounced price reduction, and when it is for a durable good. Promotions have a positive effect when the promotion is a coupon or a premium, when it is a packaged good, and when the brand is competing against several brands. These findings suggest that researchers modeling what consumers learn from promotion need to take into account the form of promotion and the steepness of the promotion discount.

2.6.4 Long-Term Effects

Researchers have also started to investigate the long-term effects of promotions. If sales promotions are successful in attracting (new) consumers to the brand or increase their consumption rate permanently, the sales impact of the promotion should be observed beyond the immediate sales promotion bump. For aggregate (sales) data, there are two primary ways of modeling the long-term effects of sales promotion: Vector Autoregressive Models with X-variables (VARX) and Vector Error-Correction Models. A two-brand VARX model for sales promotion effects could be specified as follows (cf. Nijs et al. 2001):

$$\begin{bmatrix} \Delta \ln S_{1t} \\ \Delta \ln S_{2t} \\ \Delta \ln \text{Price}_{1t} \\ \Delta \ln \text{Price}_{2t} \end{bmatrix} = \begin{bmatrix} c_{0,S1} + \sum_{s=2}^{13} c_{s,S1} SD_{st} + \delta_{S1} t \\ c_{0,S1} + \sum_{s=2}^{13} c_{s,S1} SD_{st} + \delta_{S2} t \\ c_{0,P1} + \sum_{s=2}^{13} c_{s,P1} SD_{st} + \delta_{P1} t \\ c_{0,P1} + \sum_{s=2}^{13} c_{s,P1} SD_{st} + \delta_{P2} t \end{bmatrix} + \sum_{i=1}^{8} \begin{bmatrix} \phi_{11}^i & \phi_{12}^i & \phi_{13}^i & \phi_{14}^i \\ \phi_{21}^i & \phi_{22}^i & \phi_{23}^i & \phi_{24}^i \\ \phi_{31}^i & \phi_{32}^i & \phi_{33}^i & \phi_{34}^i \\ \phi_{41}^i & \phi_{42}^i & \phi_{43}^i & \phi_{44}^i \end{bmatrix} \begin{bmatrix} \Delta \ln S_{1t-i} \\ \Delta \ln S_{2t-i} \\ \Delta \ln \text{Price}_{1t-i} \\ \Delta \ln \text{Price}_{2t-i} \end{bmatrix}$$
$$+ \begin{bmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} & \gamma_{14} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} & \gamma_{24} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} & \gamma_{34} \\ \gamma_{41} & \gamma_{42} & \gamma_{43} & \gamma_{44} \end{bmatrix} \begin{bmatrix} \Delta FEAT_{1t} \\ \Delta DISP_{1t} \\ \Delta FEAT_{2t} \\ \Delta DISP_{2t} \end{bmatrix} + \begin{bmatrix} u_{S1t} \\ u_{S2t} \\ u_{P1t} \\ u_{P2t} \end{bmatrix} \quad (2.14)$$

where

$\Delta \ln S_{bt}$	$\ln S_{bt} - \ln S_{bt-1}$, i.e., current minus lagged sales of brand b.
$\Delta \ln \text{Price}_{bt}$	$\ln \text{Price}_{bt} - \ln \text{Price}_{bt-1}$, i.e., current minus lagged price of brand b
SD_{st}	a 4-weekly seasonal dummy variable (1 during 4-week period s, 0 else)
t	deterministic trend variable
$\Delta FEAT_{bt}$	$FEAT_{bt} - FEAT_{bt-1}$, i.e., the current minus lagged feature dummy for brand b
$\Delta DISP_{bt}$	$DISP_{bt} - DISP_{bt-1}$, i.e., the current minus lagged display dummy for brand b

Equation (2.14) is estimated by OLS or SUR (Seemingly Unrelated Regression). Once the parameters have been estimated, researchers calculate Impulse Response Functions (IRF) to track the incremental impact of a one standard deviation price promotion shock on sales in periods t, $t + 1$, $t + 2$, … The permanent effect of a promotion is the asymptotic value of log sales when $t \to \infty$. Figure 2.1 shows a hypothetical Impulse Response Function. The area under the curve in grey in the figures is called the “long-term effect” or “cumulative effect”. In Fig. 2.1, there is a zero permanent effect of a promotion in week 1. Such a pattern corresponds to no unit root in the sales series (Dekimpe and Hanssens 1995). This means that the long-term (=cumulative) effect is finite. When sales have a unit root, there can be a nonzero permanent effect of a one-time promotion as illustrated in Fig. 2.2. In this case, the long-term effect (area under the curve) is infinite.

Fok et al. (2006) propose a Vector-Error Correction model that directly captures the short- and long-term effects of sales promotions on *brand sales*:

$$\Delta \ln S_t = \beta_0 + \sum_{k=1}^{K} A_k^{sr} \Delta X_{kt} + \Pi \left(\ln S_{t-1} - \sum_{k=1}^{K} A_k^{lr} X_{k,t-1} \right) + \nu_t, \ \nu_t \sim N(0, V), \quad (2.15)$$

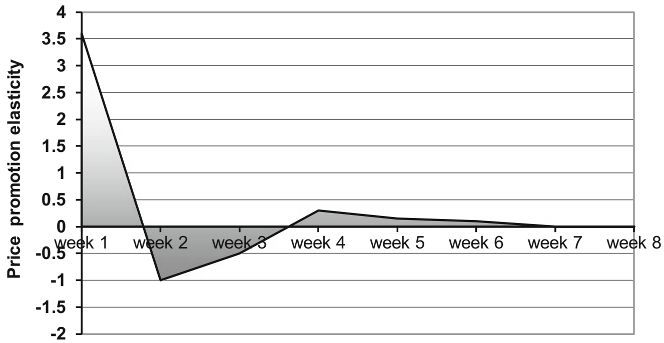


Fig. 2.1 Promotion in week 1 with a zero permanent effect

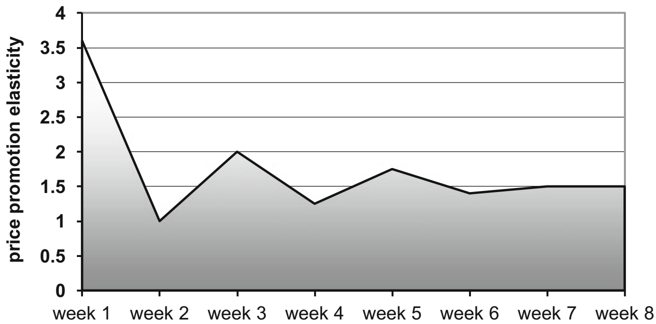


Fig. 2.2 Promotion in week 1 with a positive permanent effect

where

Δ first difference operator: $\Delta X_t = X_t - X_{t-1}$
S_t Vector ($B \times 1$) with sales (in kilo) of brands $b = 1, .., B$ in week t
X_{kt} Vector ($B \times 1$) with marketing-mix variable k ($k = 1, \ldots, K$) of brands $b = 1, \ldots, B$ in week t
β_0 Vector ($B \times 1$) with intercepts of brands $b = 1, \ldots, B$
A_k^{sr} Matrix ($B \times B$) with short-term effects of marketing-mix variable k
A_k^{lr} Matrix ($B \times B$) with long-term effects of marketing-mix variable k
Π Diagonal matrix ($B \times B$) with adjustment effects
ν_t Vector ($B \times 1$) of error terms of brands $b = 1, \ldots, B$ in week t
V Variance-Covariance matrix ($B \times B$) of the error term ν_t

The diagonal elements of A_k^{sr} and A_k^{lr} measure respectively, the short- and long-run effects of the k-th marketing-mix variable of each brand, while the off-diagonal elements capturecross effects. The Π parameters reflect the speed of adjustment towards the underlying long-term equilibrium. We refer to Fok et al.

Table 2.6 Category-demand effects of price promotions across 460 Categories

	Short term effects (%)	Permanent effects (%)
Positive	58	2
Negative	5	0
Zero	37	98

This table is based on Nijs et al. (2001)

Table 2.7 Own-brand sales effects across 442 categories

	Non-significant (%)	Positive own-sales elasticity (%)	Negative own-sales elasticity (%)	Mean own-sales elasticity
Short term effects				
Price promotions	30.96	63.54	5.50	3.989
Advertising	67.00	20.45	12.55	0.014
Permanent effects				
Price promotions	94.99	4.15	0.86	0.046
Advertising	98.23	1.28	0.49	0.000

This table is based on Steenkamp et al. (2005)

(2006) for a formal proof of these properties, and to Van Heerde et al. (2007, 2010, 2013) for applications. Horváth and Franses (2003) provide an in-depth discussion about testing for whether the Error Correction model is appropriate, and the cost of estimating an Error Correction model when it is not appropriate, and vice versa.

Using a VARX model, Nijs et al. (2001) study the effects of consumer price promotions on *category sales* across 460 consumer product categories over a 4-year period. The data describe national sales in Dutch supermarkets and cover virtually the entire marketing mix. The key results are in Table 2.6. In 98% of the cases, there is no permanent effect of promotions on category sales, i.e., the Impulse Response Function resembles Fig. 2.1. This is a sensible result, in that one would not expect permanent category sales effects in the mature categories carried by most supermarkets. An interesting extension of this work would be to look at new categories such as tablet computers or HDTVs.

Steenkamp et al. (2005) use a VARX model to study the permanent effects of promotions and advertising on brand sales for the top three brands from 442 frequently purchased consumer product categories in the Netherlands. Their major results are displayed in Table 2.7. A key conclusion is that in the far majority of cases, there are no permanent effects of promotion and advertising on own-brand sales. In the short term, these effects do exist, and they are more prevalent and stronger for promotions than for advertising.[6]

[6]Note that the final version of this paper (Steenkamp et al. 2005) does not contain these results anymore, since the journal requested the authors to focus on competitive reactions.

Slotegraaf and Pauwels (2008) use persistence (time series) modeling to measure the long-term promotion effects for 100 brands. A key distinguishing feature of the study is that it does not limit itself to just the top three or top five brands per category, but studies 9–25 brands per category, including small brands. Slotegraaf and Pauwels (2008) find that permanent effects are fairly common, as 14% of the brands have a positive sales evolution. On average, price promotions yield a higher permanent elasticity (0.06) than feature (0.003) and display (0.002). Slotegraaf and Pauwels (2008) also document substantial variation in these elasticities across brands. Both permanent and cumulative sales effects from promotions are larger for brands with higher brand equity and more product introductions.

Interestingly, Pauwels et al. (2002) show, based on data from the canned soup and yogurt categories, that permanent promotion effects are virtually absent for brand choice, category incidence, and purchase quantity.

Fok et al. (2006) apply their VEC model (2.20) to seven years of U.S. data on weekly *brand sales* of 100 different brands in 25 product categories. On average, the cumulative promotional price elasticity (−1.91) tends to be smaller in absolute value than the immediate promotional price elasticity (−2.29). Hence, some of the positive effects of a price promotion are compensated in the periods following the promotion by, for example, the effects of acceleration. Actually, the implied size of the post-promotion dip is 17% (100% * ((2.29 − 1.91)/2.29)), which is quite close to 22%, which is what we calculate for the postpromotion dip in Macé and Neslin (2004).[7] It is also interesting to note that the short-term regular price elasticity (−0.54) is much smaller in magnitude than the short-term price promotion elasticity (−2.29).

Ataman et al. (2008) study the long-term sales effects of promotion and other marketing variables. They apply Bayesian time-varying parameter models to analyze the sales for 225 *newly introduced* CPG brands observed across five years. While distribution is the strongest driver of long-term sales, feature and display ranked as the second-most effective instrument. For *mature* CPG brands, Ataman et al. (2010) show that price discounting lead to lower baseline sales in the long run and a to a heightened price sensitivity.

Datta et al. (2015) study the long-term effects of free-trial promotions. Many service providers offer consumers a free use of the service for a limited time in the hopes of retaining these customers. Datta et al. (2015) use a logit model for the monthly decision to keep the service. They show that the acquisition through a free trial affects the baseline retention rate even after the free trial is over. Free-trial customers, on the other hand, are more sensitive to their own usage rates and to marketing activities, creating opportunities for companies to retain these customers.

[7]The 22% cannot directly be calculated from Table 2.4 in Macé and Neslin (2004) because the elasticities reported in that table are point elasticities and therefore do not exactly correspond to the 20% price cut effects calculated in that table. However, calculations using the detailed results summarized in Macé and Neslin's Table 2.4, reveal that on average, 66.2% of the combined pre and post effect is due to post effects. Since the combined effect reported in Macé and Neslin's Table 2.4 is 33.3% of the bump (1 − 0.667 from the last column in the table), the percentage due to postpromotion dips is $0.662 \times 0.333 = 0.2204 = 22.0\%$.

2.6.5 Competitive Reactions[8]

Since promotions affect cross-brand sales and market shares, and they are relatively easy to replicate, competitive reactions are likely. Competitors may either retaliate or accommodate a promotion initiated by a rival brand. Moreover, they may respond in-kind with the same instrument (e.g., price cut followed by price cut) or with another instrument (e.g., price cut followed by volume-plus promotion). Leeflang and Wittink (1992, 1996) specify reaction functions that allow for the measurement of the degree and nature of competitive reactions:

$$\ln\left(\frac{P_{bt}}{P_{b.t-1}}\right) = \alpha_b + \sum_{\substack{b=1 \\ b' \neq b}}^{B} \sum_{t^*=1}^{T^*+1} \beta_{bb't^*} \ln\left(\frac{P_{b',t-t^*+1}}{P_{b',t-t^*}}\right) + \sum_{t^*=2}^{T^*+1} \beta_{bbt^*} \ln\left(\frac{P_{b,t-t^*+1}}{P_{b,t-t^*}}\right) + \sum_{b'=1}^{B} \sum_{t^*=1}^{T^*+1} \sum_{x=1}^{3} \tau_{xbb't^*}\left(wx_{b',t-t^*+1} - wx_{b',t-t^*}\right) + \varepsilon_{bt} \tag{2.16}$$

$\frac{P_{ib}}{P_{b.t-1}}$	ratio of successive prices for brand b in period t
$wx_{b,t} - wx_{b,t-1}$	first difference for the three types of promotions for brand b: $x = 1$ (feature), x = 2 (display), and x = 3 (feature and display)

The parameter $\beta_{bb't^*}$ represents competitive reactions with the same instrument: the price response by brand b to a price change by brand b' that took place t^* periods ago. Parameter $\tau_{xbb't^*}$ captures competitive reactions with different instruments: the price response by brand b to a promotion of type x by brand b' that took place t^* periods ago.

For the grocery category under study, Leeflang and Wittink (1992) find that competitor reactions occur quite frequently, especially using the same marketing instrument as the initiator. By studying competitive reactions based on over 400 consumer product categories over a four-year time span, Steenkamp et al. (2005) test the empirical generalizability of Leeflang and Wittink (1992, 1996). They use VARX models similar to Eq. (2.14). Table 2.8 shows that the predominant reaction to a price promotion attack is no reaction at all. Indeed, for 54% of the brands under price promotion attack, the average short-term promotion reaction is not significantly different from zero. Furthermore, the significant short-term promotion reactions are twice more as likely to be retaliatory than accommodating (30% vs. 16%). Table 2.8 also shows that long-term reactions are very rare. In over 90% of the instances, price promotion attacks do not elicit a persistent or long-term price promotion on the part of the defending brand.

---

[8]Leeflang (2008) provides an in-depth discussion on models for competitive reactions, including structural models.

Table 2.8 Competitive reactions to price promotions

Reaction with price promotion	Short-term effect (%)	Long-term effect (%)
No reaction	54	92
Competitive reaction	30	5
Cooperative reaction	16	3

This table is based on Steenkamp et al. (2005)

Steenkamp et al. (2005) find that absence of reaction corresponds primarily to the absence of harmful cross-sales effects. Only 118 out of 954 brands miss an opportunity in that they could have defended their position, but chose not to. When managers do opt to retaliate, effective retaliation is prevalent (63%). In 56% of these cases the response neutralizes the competitive attack, whereas in 36% of these cases the net effect is positive for the defending brand.

An interesting perspective is provided by Pauwels (2007). He finds that competitive response to promotions plays a relatively minor role in post-promotion effects. The major factor in post-promotion effects is the company's own "inertia" to continue promoting in subsequent weeks. This is a very interesting finding in that it says companies are highly myopic when it comes to formulating promotion policy, basing the frequency of future promotions on the frequency of past promotions, rather than considering the competitive implications.

The competitive reactions studies discussed so far are all in "business as usual" situations. During price wars, which are an extreme form of price competition, competitive reactions may become fiercer, partly fueled by media coverage (Van Heerde et al. 2015).

2.7 Endogeneity

2.7.1 What Is Endogeneity?

Managers may plan sales promotions based on demand factors that are not included in the model. For example, a manager may anticipate a demand increase due to external factors (e.g., favorable weather, events) and decide to cut back on sales promotions or discounts because the product will be in high demand regardless. Alternatively, a manager may capitalize on these demand shocks and offer more or better sales promotion deals than usual to seize the moment. Because these demand shocks are not observed by the researcher (the researcher does not have the data to capture them), they are not explicitly in the model and they become part of the error term. This produces a correlation between the sales promotion variable in the model and the error term. When the model is estimated, the estimate for the effect of the sales promotion variable on utility will be incorrect, because it will partly include the effect of the unobserved demand shock that is correlated with the sales promotion variable. This is the endogeneity problem.

More formally, the endogeneity problem occurs when an independent variable is correlated with the error term. This leads to a biased and inconsistent estimate for the effect of the independent variable on the dependent variable. Endogeneity has become a major theme in the academic marketing literature, especially the endogeneity of pricing. The concern is that managers may set prices strategically and that therefore the observed variation in prices is not random but rather correlated with the error term. Other types of sales promotions (e.g., features, displays) may also be set strategically.

2.7.2 Addressing Endogeneity Through Control Variables

Rather than right away jumping to more technical approaches to address endogeneity, it is essential to have a good understanding of the "easy" fixes (Rossi 2014; Germann et al. 2015). Perhaps the demand factors that managers use to set pricing and sales promotion may be measurable. If so they need to be included in the model. Examples include seasonal dummies (holidays, Christmas, Mother's Day), temperature, and event indicators. Moreover, there are likely to be cross-sectional differences between brands leading to differences in marketing variables. For example, a high-quality, popular brand may be able to charge higher prices than a low-quality, unpopular brand. If the model does not include a brand intercept, the higher utility for the first brand will be attributed to its higher price—which is an incorrect inference. Hence fixed effects for brands (dummies for brands) are required to avoid biases due to parameter inferences based on cross-sectional differences. If this is not feasible, then quality ratings for brands (based on, e.g., consumer reports or online reviews) can be included in the model to make the previously unobservable quality differences observable, which means they are no longer part of the error term or a cause of endogeneity. Including these variables will help the researcher to obtain more correct estimates for sales promotion effects.

The Scan*Pro model includes weekly dummies μ_{bt} to capture market-wide marketing activities that are unobserved by the researcher such as coupon drops or advertising. It also includes store dummies to filter out any cross-store form of endogeneity (e.g., bigger stores charging lower prices). It estimates the model by brand, which means that cross-brand differences (which are potentially endogenous) are not reflected in parameter estimates. Other demand shocks can be captured by including variables for holidays, events and days of the week.

2.7.3 Addressing Endogeneity Through Instrumental Variables

If there is reason to believe that there is still a potential endogeneity problem, a different estimation method than Ordinary Least Squares is required. For aggregate data, the most common approach is Two-Stage Least Squares (2SLS). The idea is to isolate exogenous variation in the endogenous regressor of interest, and use this variable instead of the original endogeneity-plagued variable. 2SLS first predicts the endogenous variable by regressing it on instrumental variables (IVs) and the exogenous variables from the demand model. IVs are variables that are correlated with the endogenous variable but uncorrelated with the error term of the demand model. In the context of sales promotion models, it means that IVs are not correlated with the error term of demand but do explain variation in the endogenous variable of interest (e.g., price, feature, display). For example, ingredient costs are often a good instrument because costs do not directly drive sales yet they are correlated with consumer prices.

In 2SLS, the predicted endogenous variable is included in the focal demand equation instead of the original endogenous variable, and this is the second stage regression. This step ensures that only the exogenous variation in the endogenous variable is considered. The equation can then be estimated using OLS, yielding consistent estimates. The standard errors need to be corrected for the fact that the model includes an estimated regressor; standard packages such as SPSS or STATA do this automatically.

For discrete choice models, the control function approach *should be used rather than 2SLS* (Petrin and Train 2010). The method *also* requir*es instrumental variables (IVs).* The control function approach estimates a first-stage regression for the endogenous variable of interest, e.g., price, explained by the instrumental variable (s) and the exogenous variables in the model. The residual from this regression represents variation in price that cannot be explained by observables, hence represents the endogenous component of price. This residual is added to the model, e.g., in the choice utility function (Eq. 2.3) and is called a control function because it "controls" for the endogeneity of price The original price variable is kept in the model; i.e., it is not replaced by the forecast from the first stage regression, as is done in Two-Stage Least Squares (2SLS). Next, the model that includes the control function is estimated with Maximum Likelihood. Because this step involves the estimated control function rather than the true value, standard errors from standard ML are incorrect. Bootstrapping or asymptotic formulas need to be used to obtain the correct standard errors (Petrin and Train 2010).

2.7.4 Problems in Addressing Endogeneity

The IVs for price or promotion need to be valid (uncorrelated with the error term) and sufficiently strong (have significant explanatory power in the first-stage regression). If the IVs are not valid, the estimator can become even more biased than before the endogeneity correction. If the IVs are not sufficiently strong, the estimator will have large standard errors, leading to insignificance and sometimes counterintuitive signs (Rossi 2014). In practice, these two requirements are hard to meet simultaneously. Strong IVs are often invalid because they may be correlated with current demand. For example, lagged prices will typically explain current prices well, but may also drive current demand due to post-promotion effects, and hence they are invalid IVs. Valid IVs are often weak, because they are too far removed from the endogenous regressor.

Another problem with endogeneity corrections is that they lead to worse model fit (in- and out of sample) of the sales promotion model, and hence predictive performance cannot be used to establish the success of an endogeneity correction (Ebbes et al. 2011). Hence, all in all we caution against a mechanical use of endogeneity correction estimation methods. Much thought needs to be given to whether there is actually a problem once all key control variables have been included.

2.7.5 IV-Free Methods to Address Endogeneity

IVs need to be strong (sufficiently correlated with the endogenous regressor) and valid (uncorrelated with the error term). If proper IVs cannot be found, instrument-free methods can be used. One approach is through Gaussian copulas (Park and Gupta 2012). Copulas are statistical functions that capture the joint distribution between two stochastic variables that each may have different types of marginal distributions. The approach directly models the correlation between the endogenous regressor (X) and the error term (ε), and by doing so, it eliminates the endogeneity problem. Suppose the marginal distribution of the endogenous regressor is H(X) and the marginal distribution of the error term is G(ε). Their joint density is captured through a bivariate Gaussian copula (Park and Gupta 2012, Eq 2.6). This leads to a likelihood function that can be optimized with Maximum Likelihood to obtain consistent estimates of the model parameters.

However, there is also a simpler, equivalent way to estimate the model if the error term has a normal distribution (Park and Gupta 2012, Eq. 2.10). First, the researcher estimates the empirical cumulative distribution of the endogenous variable: H(X). This essentially means sorting the observations from low to high, and then calculating for each observation the proportion of the observations that are less than or equal to the focal observation. Next, the researcher calculates the inverse standard normal CDF for each observation: $\Phi^{-1}(H(X))$. Finally, this term is added to the demand equation of interest, where the original endogenous regressor (X) is kept as is, and this equation is estimated with OLS. Now X will be estimated consistently. Bootstrapping needs to be applied to obtain correct standard errors. A key requirement is that the endogenous variable has a non-normal distribution, which needs to be established first through a normality test.

Gaussian copulas were recently applied by Burmester et al. (2015) and Datta et al. (2015). The copula method by Park and Gupta (2012) can also be applied for discrete choice models (household-level data), for discrete endogenous regressors and for "slope endogeneity," which is an endogeneity problem that arises when the manager decides on marketing actions based on the response parameter heterogeneity.

An alternative instrument-free method is offered by Ebbes et al. (2005). It uses "latent instrumental variables." The idea is that underlying the distribution of the endogenous variables are latent discrete support points, which are the latent IVs. The method then splits the variation of the endogenous regressor into an exogenous part (captured by the discrete support points) and an endogenous part, which is assumed to be distributed bivariate normal together with the error term of the demand equation.

2.7.6 VAR Models and Endogeneity

Finally, we note that VAR models and other vector-based time series models (VARX, VEC) treat multiple variables as endogenous. For example, sales, price and promotion may be part of a three-variate vector that constitutes the dependent (endogenous) variable in these models. These models allow for the measurement of feedback effects of sales on price and promotion and for inertia (e.g., price and promotion depending on own lagged values). However, treating variables as endogenous is not the same as correcting for an endogeneity bias. The VAR model for example does not try to infer a causal effect of one variable on another. In a VAR model, the immediate effect is captured through the covariance matrix of the error term, which means that is the effect is bidirectional (Y1 affecting Y2 as much as Y2 affecting Y1). In other words, there is nothing in a VAR model that tries to correct for endogeneity bias when inferring the effect of one variable on another.

2.8 Promotions to the Trade—Introduction

Manufacturers use promotional discounts to the trade as an incentive for the trade to buy more of the brand, and to sell more to consumers by passing through at least part of the discount. The key two phenomena that determine the effectiveness of trade promotions are forward buying (Sect. 2.8) and pass-through (Sect. 2.9). In

Sect. 2.11 we present decision models for manufacturers who want to optimize their trade promotions, and in Sect. 2.12 we discuss models for retailers who want to optimize pass-through and forward-buying.

2.9 Forward Buying

Trade promotions offered by manufacturers often lead to forward buying by retailers. Forward buying is essentially purchase acceleration by retailers in that retailers purchase during the promotion period to satisfy demand in future periods (Neslin 2002, p. 36). While a retailer may sell part of the extra stock to consumers at a discount, their key incentive to forward buy is to sell the other part at regular price. We show an example in Fig. 2.3a, b.

Suppose a manufacturer offers a trade deal in period $t - 1$ in order to stimulate a retailer to promote in period t. Figure 2.3a shows how a retailer may order higher quantities in period $t - 1$ than usually to benefit from the lower wholesale price offered by the manufacturer in period t.[9] Forward buying implies that the stock bought in period $t - 1$ not only satisfies the extra demand during the consumer promotion in period t (see Fig. 2.3b), but also demand sold at regular price in period $t + 1$ and beyond. Figure 2.3b shows that not only the retailer forward buys but also consumers: the bump in period t is followed by a postpromotion dip in period $t + 1$.

To measure the effectiveness and profitability of trade promotions, Blattberg and Levin (1987) model the interplay between factory shipments, retailer promotions, retailer sales, and retailer inventories:

$$\text{F}\mathit{actoryShipments}_t = f_1(\text{Inventories}_{t-1}, \text{Trade Promotions}_t) \tag{2.17}$$

$$\text{Consumer Promotions}_t = f_2(\text{Trade Promotions}_t, \text{Trade Promotions}_t,\ \text{Inventories}_{t-1}) \tag{2.18}$$

$$\text{Retailer Sales}_t = f_3(\text{Consumer Promotions}_{t-1}) \tag{2.19}$$

$$\text{Inventories}_t = f_4(\text{Inventories}_{t-1}, \text{FactoryShipments}_t,\ \text{Retailer Sales}_{t-1}) \tag{2.20}$$

Equation (2.17) captures the effect of trade promotions and inventories on factory shipments, i.e., how much the retailer orders (and therefore gets shipped). The willingness of retailers to run a consumer promotion depends on the availability of trade promotions and its own inventories (Eq. 2.18). Retail sales are driven by

[9] "Factory Shipments" are shipments from the manufacturer to the retailer and reflect retailer orders or manufacturer sales. Blattberg and Levin (1987) use factory shipments data. Abraham and Lodish (1987) note their method can be applied to factory shipment data as well as other data such as warehouse withdrawals.

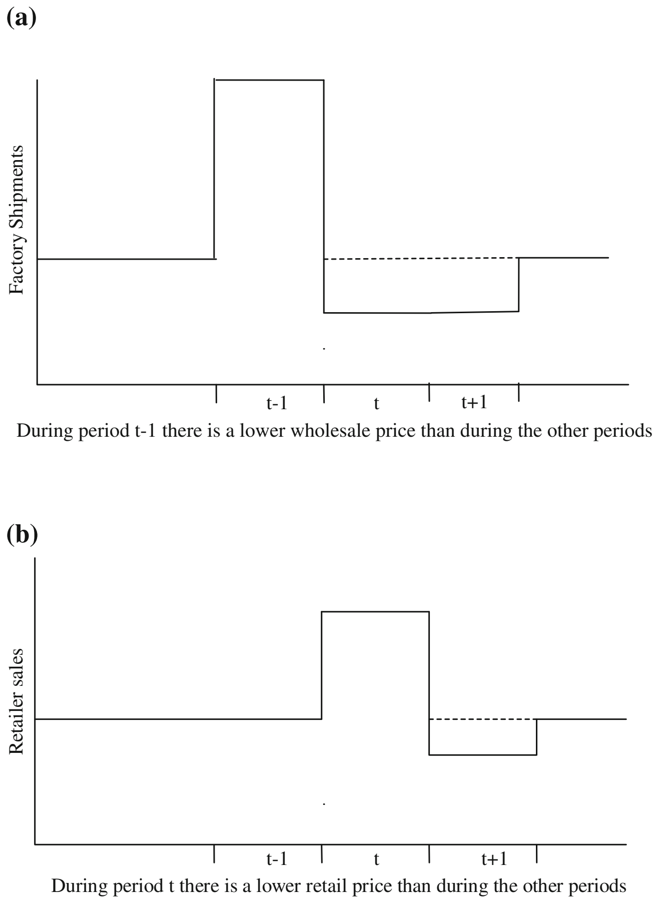


Fig. 2.3 **a** Response of factory shipments to trade promotion, **b** response of retailer sales to consumer promotion

consumer promotions (Eq. 2.19), and Eq. (2.20) shows that retail inventories are a function of its own lag, inflow (Factory shipments) and outflow (Retailer sales).

Another approach to evaluate the effectiveness of the trade promotion is to estimate what factory shipment "sales" (Eq. 2.17) would have been in the absence of the promotion (Abraham and Lodish 1987). Once we estimate baseline sales, the size of the bump can be quantified as the actual factory shipments in the period with the wholesale discount (period $t - 1$ in Fig. 2.3a) minus predicted baseline factory shipments for the same period. To estimate baseline factory shipments, Abraham and Lodish (1987) develop PROMOTER, a time series approach that tries to identify a "base" sales level by extrapolating the sales level during "normal" periods (see Abraham and Lodish 1993 for an application to retail promotions). The PROMOTER model decomposes sales into three components:

$$S_t = B_t + P_t + E_t \tag{2.21}$$

where

S_t Factory shipments at time t
B_t Baseline at time t to be estimated
P_t Promotion effect at time t if any
E_t Noise term

2.10 Pass-Through

The American Marketing Association defines pass-through as: "The number or percentage of sales promotion incentives offered to wholesalers or retailers by manufacturers that are extended to consumers by those channel members" (American Marketing Association 2015). For trade deals, it is the percentage of the discount that is passed on to the consumer in the form of a price reduction. Trade deals constitute close to 60% of manufacturers' marketing budgets (Retail 2012). Manufacturers believe that only 52–66% of their trade spending is passed through to the consumer (Retail 2012). Though published numbers on pass-through range from 0 to 200%, pass-through is often less than 100% (Neslin 2002, p. 34).

Calculating pass-through fundamentally involves the relationship between retailer costs, which decrease when the retailer receives a trade deal, and retailer selling price. Questions have emerged as to how to calculate retailer costs. For example, should the researcher use "average acquisition cost" (AAC) or "transaction cost" (see Nijs et al. 2010)? Transaction cost is the unit cost the retailer paid for product in the current week. AAC can be calculated as a weighted average of the AAC of retailer inventory at the end of the previous week and cost of product the retailer purchased in the current week (Besanko et al. 2005, p. 129). The estimate of pass-through can differ depending on whether one uses AAC or transaction cost.

Table 2.9 shows an example. We assume normal wholesale price (transaction cost for the retailer) is \$4 per case; a trade deal lowers this to \$3 in period 4. The retailer starts with 400 cases in inventory and we assume purchases enough inventory each week exactly to cover demand. The table shows the cases purchased by the retailer, cases purchased by the consumer, and ending inventory. This is used to calculate AAC. For example, in period 3 AAC is (150 × \$4 + (400 − 150) × \$4)/(400) = \$4.[10] However, in period 4, AAC is (250 × \$3 + (400 −

[10]We use the formula provided on page 129 of Besanko et al. (2005): AAC(t) = [(Retailer Purchases in t) × Wholesale price in period t) + [(Inventory at end of t − 1) − (Retail sales in t)] × AAC(t − 1)] × (Inventory at end of t)$^{-1}$.

Table 2.9 Calculating pass-through for hypothetical data

Period	Wholesale price (transaction cost)	Starting inventory	Cases bought by retailer	Cases bought by consumer	Ending inventory	Average acquisition cost (AAC)	Retail price
1	\$4	400	150	150	400	\$4	\$5
2	\$4	400	150	150	400	\$4	\$5
3	\$4	400	150	150	400	\$4	\$5
4	\$3	400	250	250	400	\$3.38	\$4.50
5	\$4	400	150	150	400	\$3.61	\$5
6	\$4	400	150	150	400	\$3.76	\$5
7	\$4	400	150	150	400	\$3.85	\$5
8	\$4	400	150	150	400	\$3.90	\$5

250) × \$4)/(400) = \$3.38. So in period 3, the transaction cost is \$3 while the average AAC is \$3.38. We assume the retailer lowers the retail price from \$5 to \$4.50 in period 4 when the trade deal is offered.

Following BDG, we calculate pass-through by regressing retail price versus AAC (see pages 128–129 of BDG). We obtain:

$$\text{Retail\_}\widehat{\text{Price}}_t = 2.57 + 0.62 \times \text{A}\mathit{cquisition\_Cost}_t$$

According to this linear model, a \$1 decrease in acquisition cost leads to a \$0.62 decrease in retail price. Hence the pass-through rate is 62%. However, from a transaction cost perspective, the \$1 reduction in transaction cost resulted in a \$0.50 pass-through, or 50%. Nijs et al. (2010) assert that using AAC results in inflated estimates of pass-through. The reason is apparent in Table 2.9. AAC is essentially a moving average of previous AAC and current transaction costs. Therefore the full reduction of \$1 in period 4 is cushioned by previous AAC, so period 4 AAC decreases to \$3.38, not \$3. The regression sees a retail price of \$4.50 associated with a cost of \$3.38, so infers that the retail price reduction per dollar cost reduction is larger than if the transaction cost of \$3 were used in period 4. Our purpose is not to adjudicate whether transaction costs or AAC is more appropriate. Our point is that the calculation of relevant costs is crucial for estimating the pass-through rate. The above is a simple example, and does not include carrying costs, forward buying, and other factors that can make cost accounting of inventory a challenge.

The above treats trade deals as discrete incentives applied to a focal product. This may be appropriate for off invoice or "scanback" trade deals (see Gómez et al. 2007 for descriptions). However, much trade promotion spending involves lump sum payments for activities such as advertising programs. The retailer may use lump sum payments in various ways for different SKUs at different times. It thus is difficult to map a particular trade deal payment to a particular retailer promotion action. Ailawadi and Harlam (2009) offer a fresh perspective on pass-through in this environment. They propose to measure pass-through by "adding up the promotion funding, in all its forms, that the retailer receives from a manufacturer in a year and comparing it to the retailer's total spending on price promotions for the manufacturer's products" (p. 783). Using this approach, they find interesting results. For example, pass-through is over 100% in the two years they examined, meaning that the retailer spends more on its promotions than it receives in funding. However, this finding is across all manufacturers, even those who provide no funding yet receive promotions. The pass-through rate among manufacturers that provide funding was 20% (Table 2.3 in Ailawadi and Harlam).

Factors such as price elasticity of the promoted brand, retailer size, retail price, category importance to the retailer, and ability of the promoted brand to take away from others, have been shown to influence pass-through (Besanko et al. 2005; Pauwels 2007; Nijs et al. 2010; Ailawadi and Harlam 2009). One crucial question is the role of market share. Intuition suggests higher share brands should command higher pass-through. However, a simple retailer profit model shows why a high-share brand might not receive high pass-through. The argument is that higher share brands have higher baseline sales, which means the retailer sacrifices more margin by putting it on sale. Let:

B	baseline sales for promoted brand
B_0	baseline sales for rest of category
M	profit margin for promoted brand
M_0	profit margin for rest of category
D	trade deal discount
δ	pass through
η	gain in sales for promoted brand

Then: Retailer Profit with no pass through	$B\,(M + D) + B_0 M_0$
Retailer profit with δ pass through	$B\,(M + D - \delta) + \eta\,(M + D - \delta) + (B_0 - \eta)M_0$
Difference	$-\delta B + \eta\,(M + D - \delta) - \eta M_0$

A high share brand has higher B and this exerts a force for decreasing δ. One could argue that higher share brands have higher η's, but that isn't good if $M_0 > M + D - \delta$, which could be the case if high share brands have lower regular margins for the retailer (they are stronger brands). This leads the retailer to decrease δ to minimize the baseline loss and make $M + D - \delta$ larger than M_0. The manufacturer might argue that promoting its high share brand grows the category through store switching, so that the η incremental units for the brand do not come completely from other brands in the category. In any case, the above analysis shows the challenges the high share brand must overcome to obtain higher pass-through.

Recent studies suggest that despite the pass-through handicap discussed above, higher share brands do command higher pass-through. Although early work by Walters (1989) found market share had no impact on pass-through, more recent studies by BDG, Pauwels (2007), Ailawadi and Harlam (2009), and Nijs et al.

(2010) find that it does. An interesting avenue for future research would be to learn why this occurs despite the baseline handicap discussed above.

Another issue that has received attention is cross-brand pass-through. BDG find that retailers adjust the prices of other brands in the category when passing through a promotion for the focal brand. They find for example that trade deals for large brands are less likely than small brands to generate positive cross-brand pass-through, i.e., large brands do not induce the retailer to reduce the retail price of competing smaller products. In a critique of BDG, McAlister (2007) argues that BDG find so many significant coefficients for cross-brand pass-through because they inadvertently inflate the number of independent observations by a factor of 15 (15 is the number of price zones). When she corrects for this, she finds "the number of stable, significant coefficients for other brands' wholesale prices is lower than one would expect by chance." (p. 876). See Dubé and Gupta (2008) and Duan et al. (2011) for further discourse on this issue. It is also noteworthy that McAlister (2007) argues that the use of average acquisition cost can also distort estimates of own and cross-brand pass-through.

Part II: Normative Models

In Part I we have focused on descriptive models for sales promotions, i.e. models that describe, analyze, and explain sales promotion phenomena. In Part II we discuss normative (decision) models for sales promotions, i.e., models that tell the decision maker what is the best (profit maximizing) decision on sales promotion activities. Section 2.10 focuses on models for promotions to consumers, whereas Sect. 2.11 zooms in on trade promotion models for manufacturers. Section 2.12 takes the perspective of a retailer who tries to optimize forward buying and pass-through in response to trade promotions offered by manufacturers.

2.11 Decision Models for Promotions to the Consumer

2.11.1 Retailer Promotion Optimization

Tellis and Zufryden (1995) formulate a model to maximize cumulative *retailer* category profits over a finite horizon, by optimizing the depth and timing of discounts, and order quantities, for multiple brands. The model is based on an integration of consumer decisions in purchase incidence, brand choice and quantity. The retailer profit objective function is given by:

$$\max_{\{Disc_{bt}, O_{bt}, \delta_{bt}, \xi_{bt}} \left\{ \sum_{b,t} \left(M \cdot S_{bt} \cdot (\text{Price}_{bt} m_{bt} - Disc_{bt}) \right) - \sum_{b,t} \left(\xi_{bt} F_{bt} + h_{bt} \cdot (I_{bt} - I_{bt-1}) / 2 + \delta_{bt} Tag_{bt} \right) \right\}, \tag{2.22}$$

where

$Disc_{bt}$	retailer discount level for brand b, during period $t\,(\geq 0)$
O_{bt}	retailer order quantity for brand b, made at beginning of period $t\,(\geq 0)$
δ_{bt}	integer price-change indicator (=1 if a price change was made for brand b during t relative to $t-1$; 0 otherwise)
ξ_{bt}	integer order time indicator (=1 if an order for brand b is placed during period t; 0 otherwise
M	total household market size
S_{bt}	average sales of brand b per customer during period t, computed as a function of causal variables (including, $Disc_{bt}$), and obtained via models for incidence, brand choice, and quantity
$Price_{bt}$	regular price of brand b during period t,
m_{bt}	regular retailer profit margin (excluding inventory costs) of brand b during period t,
F_{bt}	fixed costs of ordering brand b during period t
h_{bt}	cost per unit of holding inventory of brand b during period t
I_{bt}	retailer inventory for brand b during period t (this depends on orders O_{bt}, sales S_{bt}, and market size M)
Tag_{bt}	cost of retagging shelves if a price change of brand b occurs during period t

The retailer profit function (Eq. 2.22) equals the profit margin before inventory costs less inventory cost for the product category. Inventory costs include the fixed costs of placing an order, the average cost of holding inventory, and the costs for changing retail price such as retagging shelves. This optimization includes constraints that ensure that (1) inventories are updated appropriately and (2) demand is always met (see Tellis and Zufryden 1995 for details).

Natter et al. (2007) present a decision support system for dynamic retail pricing and promotion planning. Their weekly demand model incorporates price, reference price effects, seasonality, article availability information, features and discounts. They quantify demand interdependencies (complements and substitutes) and integrate the net impact of these interdependencies into an optimal pricing model. The methodology was developed and implemented at BauMax, an Austrian Do-It-Yourself (DIY) retailer. Eight pricing rounds with thousands of different Stock Keeping Units (SKUs) served as a testing ground for the approach. The final marketing decision-support system implemented in rounds six through eight increased gross profit on average by 8.1% and sales by 2.1%, relative to predicted baselines.

Divakar et al. (2005) develop a sales forecasting model and decision support system that has been implemented at a major consumer packaged goods company. Managers are able to track forecast versus actual sales in a user-friendly "marketing dashboard" computing environment and drill down to understand the reasons for potential discrepancies. Based on that understanding, managers can adjust price and promotion accordingly. Divakar et al. report the company estimated that the DSS resulted in savings of $11 million on an investment of less than $1 million.

The authors emphasize the importance of organization "buy-in," relevance, and diagnostics for successful real-world adoption of promotion models for decision-making.

An interesting promotion optimization issue is whether the manufacturer should strive for competing stores to alternate promoting its brand, or promote in the same week. The concern is that alternating weeks may allow consumers to switch stores and virtually always buy the brand on deal. Guyt and Gijsbrechts (2014) examine this using a model of store, category, and brand choice. They examine the sales impact of "out-of-phase" (alternating) versus "in-phase" (same-week) promotions. They show how store switching plays a key role in determining whether out-of-phase or in-phase promotion calendars are better for the manufacturer. These conclusions come from simulations, not formal optimization. Formal optimization would be a promising next step. This is a very real issue for manufacturers negotiating the timing of trade deal pass-through with retailers.

2.11.2 Targeting Promotions to Consumers

Since the beginning of the "Age of Addressability" (Blattberg and Deighton 1991), companies have increased their ability to target promotions to individual customers. The promotion can be delivered via email, retailer apps, website customization, or "old fashioned" direct mail. From a modeling perspective, the impetus for promotion targeting comes from parameter heterogeneity. Rossi et al. (1996) were the first to show how individual-level parameters could be used to devise customer-level promotions.

Zhang and Krishnamurthi (2004) provide a method for customizing promotions in online stores. Their approach provides recommendations on when to promote how much to whom. They take the perspective of a manufacturer who wants to optimize its expected gross profit from a household over three shopping trips with respect to a brand's price promotions. The corresponding objective function is:

$$\max_{\{Disc_{bt+s},\, s=0,1,2}\left\{\sum_{s=0}^{2} P(I^h_{t+s}=1, C^h_{t+s}=b)\cdot E(Q^h_{bt+s}|I^h_{t+s}=1, C^h_{t+s}=b)\cdot(m_{bt+s}-Disc_{bt+s})\right\} \tag{2.23}$$

(all symbols have been defined previously; see Eqs. (2.1) and (2.22)). The optimization is subject to the constraint that discounts are nonnegative and that they do not exceed a fixed fraction of the regular price, to prevent brand equity erosion. The authors demonstrate that their approach may lead to much higher profits, especially because it prevents wasting money on price discounts that are too steep.

Zhang and Wedel (2009) followed up this work by studying targeted promotions in online versus offline stores. They estimated response models (incidence, choice, quantity) separately for the online and offline channels of a focal retailer. A key difference between online and offline targeting is the authors assume that online redemption rate is 100% since the delivery would be through customer-level modification of the website. Redemption for offline promotions was assumed to be 15% because delivery would be through coupons made available at checkout. The authors also distinguished between "loyalty" promotions offered to customers who purchased the focal brand last time, and "competitive" promotions offered to customers who purchase a different brand last. The authors find that loyalty promotions were more profitable in online stores, while competitive promotions were more profitable in offline stores. The reason is the difference in customer behavior. Customers in the online store were strongly state dependent. This meant it was difficult to get them to switch brands and easier to get them to stick with the current brand, so loyalty promotions were more profitable. Customers in offline stores were less state dependent hence it was easier to get them to switch brands with a promotion.

Zhang and Wedel's optimization used a three-period "rolling horizon". The advantage of this approach is that decisions are made on a periodic basis using the exact data from the customer's recent purchase history, not their estimated data (see also Neslin et al. 2009). However, the optimization has to be re-run each decision period. Another approach is dynamic programming. Dynamic programming produces a policy function that specifies, given the "state variables" that characterize the customer, whether the firm should or should not offer a promotion at a particular time to that customer. Khan et al. (2009) used finite horizon dynamic programming to determine when to offer which customers free shipping or coupons for an online grocery retailer. Neslin et al. (2013) used an infinite horizon dynamic program to determine when to offer which customers email or direct mail promotions for a meal preparation service.

There are several modeling issues to resolve for optimal promotion targeting. These include the optimization method as well as the particulars of the response model. One implementation issue is what should be the level of customization—optimize the timing but do not consider cross-sectional heterogeneity in response (mass targeting), optimize timing and consider segment-level heterogeneity, and optimize timing and consider customer-level heterogeneity. If model estimates are precise and computation time is not an issue, customer-level targeting is preferred. However, in real world applications, customer-specific parameters may be imprecise, and running optimization at the customer level can be computationally difficult. Zhang and Wedel (2009) as well as Khan et al. (2009) investigated this issue. Khan et al. found that in comparison to baseline, targeting based on timing alone increased profits by 7.8%, targeting based on timing at the segment level increased profits by 10.9%, and targeting based on timing at the individual level increased profits by 13.2%. Zhang and Wedel's results are similar in that mass targeting does quite well, and there are decreasing returns to segment-level and then customer-level customization. Everything however depends on the response function—the nature of heterogeneity and the precision with which individual-level parameters can be measured. Future research is needed to decide what level of optimization is best.

The above research uses estimated response models and formal optimization to target promotions. Another approach is to generate a targeting policy without the benefit of modeling or optimization and to test the policy with a field experiment. The concern of course is the firm may be "leaving money on the table". See Venkatesan and Farris (2012) for an analysis of a supermarket targeted coupon strategy, and Luo et al. (2014) for an analysis of mobile-delivered coupon.

2.12 Manufacturer Decision Models for Trade Promotions

Silva-Risso et al. (1999) present a decision support system that permits to search for a manufacturer's optimal trade promotion calendar. By modeling the purchase incidence (timing), choice and quantity decisions they decompose total sales into incremental and non-incremental. The manufacturer's objective function is given by:

$$\begin{aligned}\max_{\kappa_{bt}, FEAT_{bt}, DISP_{bt}} \Bigg\{ & \sum_{t=1}^{T} \rho^t \cdot M \cdot E(\Delta S_{bt}) \cdot (\mathrm{Price}_{bt} \cdot (1 - \kappa_{bt} \cdot DSTEP) - MCOST_{bt}) \\ & - \sum_{t=1}^{T} \rho^t \cdot M \cdot E(B_{bt}) \cdot (\mathrm{Price}_{bt} \cdot \kappa_{bt} \cdot DSTEP) \\ & - \sum_{t=1}^{T} (\rho^t \cdot \delta_{bt} \cdot Tag_{bt} + FEAT_{bt} \cdot FCOST_{bt} + DISP_{bt} \cdot DCOST_{bt}) \\ & + \sum_{t=T+1}^{T+13} \rho^t \cdot M \cdot E(\Delta S_{bt}|_{\text{no promotion}}) \cdot (\overline{\mathrm{Price}}_b - \overline{MCOST}_b) \Bigg\}\end{aligned} \tag{2.24}$$

where δ_{bt} and Tag_{bt}, have been defined previously (Sect. 2.10), and

κ_{bt}	0, 1, 2, …, 10. This is the discount multiplier for brand-size b in week t. If $\kappa_{bt} = 0$, brand-size b is sold at the base price in week t. When the manufacturer offers a discount, it is computed as a multiple of a discount step level, e.g., 5%
ρ	discount rate
M	average number of category consumers that shop in the store or chain
$E\,(\Delta S_{bt})$	expected number of incremental units of brand-size b in week t due to the promotion
Price_{bt}	wholesale base price of brand-size b in week t
$DSTEP$	base discount step, e.g., 5%
$MCOST_{bt}$	manufacturer's marginal cost of brand-size b in week t
$E\,(B_{bt})$	expected number of baseline plus borrowed units of brand-size b in week t

$FCOST_{bt}$	fixed cost charged by retailer to run a feature ad for brand-size b in week t
$DCOST_{bt}$	fixed cost charged by retailer to set up a display for brand-size b in week t
$E\left(\Delta S_{bt}\|_{\text{no promotion}}\right)$	expected number of incremental units of brand-size b in week $t = T+1, \ldots, T + 13$ due to carry-over effects from promotions in the period $t = 1, \ldots, T$
$\overline{\text{Price}}_b$	average wholesale base price of brand-size b
$\overline{MCOST}_b$	average manufacturer's marginal cost of brand-size b

The objective function has four components: (1) the expected contribution from incremental units, (2) the expected opportunity cost of selling at a discount to consumers who would have bought the brand at the regular price, (3) the fixed costs associated with promotion decisions, and (4) the carry-over effects from consumption and purchase feedback over a 13-week period subsequent to the planning horizon. The objective function is maximized subject to constraints on the minimal and maximal number of promotions. Furthermore, the retailer may insist on a minimum level of category profits.

Neslin et al. (1995) develop a model that optimizes trade promotions and advertising. Their dynamic optimization model considers the actions of manufacturers, retailers, and consumers. The manufacturer attempts to maximize its profits by advertising directly to consumers and offering periodic trade deal discounts in the hope that the retailer will in turn pass through a retailer promotion to the consumer. Neslin et al. (1995) specify a multi-equation model for the retailer order and pass-through decisions, and for the effects of advertising and promotion on aggregate consumer demand. One of their findings is an intrinsic negative relationship between optimal levels of advertising and promotion. Higher advertising begets higher baseline sales, which increases the cost of promotion in the form of lost margin on baseline sales. Higher levels of promotion erode margin, thereby decreasing the incremental contribution of advertising. The result is that forces that tend to increase promotion tend to decrease advertising. This could be offset if advertising enhances consumer response to promotions, but the point is, *absent this interaction*, this research suggests optimal promotion and advertising expenditures are negatively related.

2.13 Retailer Decision Models for Forward Buying and Pass-Through

Retailers may benefit from trade promotions by forward buying. Blattberg and Neslin (1990, p. 460) derive the optimal amount of forward buying:

$$W^* = \frac{52 \cdot (G \cdot P - HC)}{(PC \cdot P \cdot CC + 13 \cdot SC)} \tag{2.25}$$

where

W^* optimal number of weeks supply to forward buy
G increase in profit margin per case from trade deal (trade deal discount in dollars)
P number of cases per pallet
HC handling cost per pallet
PC purchase cost per case (regular wholesale price – G)
CC cost of capital (borrowing costs to finance the forward buy)
SC storage cost per pallet per month

Equation (2.25) shows the rich interplay among variables that determine the amount of forward buying. For example, a larger trade deal discount (G) directly encourages more forward buying. It also decreases the retailer's purchase cost (PC), which decreases the retailer's financing costs (PC · P · CC). This encourages even more forward buying. Equation (2.25) also shows that all else equal, larger costs of capital (CC), storage (SC), and handling (HC) discourage forward buying.

Equation (2.25) is an inventory management model that does not take into account the increase in demand that results from the retailer passing through a portion of the trade deal discount to decrease retail price. There have been a few studies that have derived the optimal level of pass-through for a retailer. While not taking into account all the inventory factors in Eq. (2.25), Tyagi (1999) found that pass-through depends on the following function:

$$\varphi = \frac{S(\text{Price}^*)S''(\text{Price}^*)}{S'(\text{Price}^*)}, \tag{2.26}$$

where S is the demand function at the retail level, Price* is the optimal retail price, and the primes stand for the first or second derivatives of the demand function. Specifically, if $\varphi < 1$, retailer pass-through is less than 100%; if $\varphi = 1$, retailer pass-through is 100%, and if $\varphi > 1$, retailer pass-through is greater than 100%. Tyagi (1999) shows that for the linear and all concave consumer demand functions, optimal pass-through is less than 100%, However, for commonly used demand functions such as the constant elasticity demand function (e.g., the Scan*Pro model from Sect. 2.4.1), a rational retailer engages in greater than 100% pass-through. Moorthy (2005) generalizes Tyagi's formulation in several directions. First, besides whole price changes, Moorthy (2005) considers other cost components as well, such as inventory and labor costs. Second, Moorthy (2005) considers multiple retailers and multiple brands. Moorthy finds for example that cross-brand pass-through may be optimal, i.e., when decreasing the price of Brand A, it may be optimal to increase the price of Brand B.

As noted earlier, off-invoice trade promotions imply that retailers obtain a discount on the wholesale price for every unit *bought* on promotion. However, manufacturers often lose money on these deals as a result of forward-buying by retailers (Drèze and Bell 2003). Current trade promotion practice often shuns off-invoice trade deals in favor of trade deals that compensate retailers based on how much they *sell*, not *buy* (see Gomez et al. 2007) An example is the scanback deal, which gives retailers a discount on units *sold* during the promotion. Drèze and Bell (2003) develop a theory to compare retailer pricing decisions and profitability under scan-back and traditional off-invoice trade deals. They derive that for a given set of deal parameters (regular price, deal size, and deal duration), the retailer always prefers an off-invoice deal (because of the benefits of forward buying), whereas the manufacturer always prefers a scan-back. However, manufacturers can redesign the scanback to replicate the retailer profits generated by the off-invoice deal. The redesign makes the retailer indifferent between the off-invoice and the scan-back and makes the manufacturer strictly better off. The benefit of scan-back deals for retailers is that they economize on excess inventory costs, since scan-back deals do not lead to forward buying. For a redesigned scan-back in which the deal length is the same as off-invoice, but the deal depth is increased, consumer demand increases.

Part III: Conclusions

This part concludes this chapter on sales promotion models. Section 2.13 presents a summary of the key empirical findings on sales promotion effectiveness. Next, Sect. 2.14 offers practical guidelines in model implementation, and Sect. 2.15 elaborates on avenues of further research in the sales promotion realm.

2.14 Summary

This chapter has presented several models for the effects of sales promotions. In order to determine which of these models may be most relevant, we now summarize their key findings:

- Promotions to consumers lead to very strong sales promotion bumps in the short term. Hence it is essential that a model captures short-term effects.
- As a generalized finding across many categories and brands, brand switching effects expressed in unit sales are about 1/3 of the bump (Table 2.5), acceleration and deceleration effects are also 1/3 (Macé and Neslin 2004), and the remaining 1/3 is sometimes labeled “category expansion” (Van Heerde et al. 2004). Therefore it is important that any short-term model distinguishes at least among these main sources. See Table 2.2 for deeper insights of the sources for the short-term promotion bump.

- There are significant purchase-event feedback effects of promotions. In the first couple of weeks buying on promotion, a consumer's purchase behavior is affected by that promotion. Hence it is important to accommodate feedback effects in models (Sect. 2.6.1).
- Permanent effects of promotions on brand and category sales are rare (Sect. 2.6.4). That is, the effect of a promotion typically dies out after a number of weeks. Hence it may not be necessary to model these permanent effects.
- Two key factors that drive the profitability of trade promotions are pass-through and forward buying (Sects. 2.8 and 2.9). Incremental sales at retail are driven by pass-through combined with consumer response to promotions (Sect. 2.11). Any optimization model for trade dealing needs at least to include these phenomena.
- The rise of scanback deals may call for new models for pass-through (Sect. 2.12).

2.15 Practical Modeling Guidelines

In this section we provide a number of practical guidelines for building sales promotion models. Irrespective of whether the aim is to build a descriptive model (Part I) or normative model (Part II), it is important first to evaluate the available data. Do they match the modeling objective? If the goal is to learn about consumer heterogeneity, data at the consumer level are required. If the goal is to understand aggregate promotion effects, data at the store-level or at a higher aggregation level are sufficient. Once the data have been collected, it is important that the most important causal drivers of the performance measure of interest are available. In other words, the independent variables in the dataset should be able to explain a significant proportion of the variation in the dependent variable. A next step is to check descriptive statistics (means, variances, time series plots) and identify (and possibly correct or delete) outliers.

The subsequent step is to specify a descriptive model. A descriptive model may be the end goal in itself or constitute the building block of a normative model. We provide in Table 2.10 a number of descriptive models, with a few (admittedly subjective) pros and cons of each. As for the individual-level models, our view is that the minimum requirements include heterogeneity (see Sect. 2.2.5) and purchase-event feedback (Sect. 2.6.1). In aggregate regression- and time series models, it is important to include dynamic effects (Sect. 2.4.2). While aggregate logit models offer the benefits of (1) consistency with individual-level utility maximization and (2) parsimony in modeling cross-effects, they currently lack dynamic effects.

Software is increasingly available to estimate the models included in Table 2.10. There is no need to program maximum likelihood estimation for logit models, regression models, or Poisson models, as these are readily available in SPSS,

Table 2.10 A selection of models for sales promotion effects

Model	Dependent variable	Model	Advantage	Disadvantage	Key studies
Household-level data					
Individual-level purchase behavior models	Brand choice	Multinomial logit model	• Consistent with utility theory	IIA-assumption (can be avoided by using a probit model)	• Guadagni and Little (1983) • Gupta (1988)
	Purchase incidence	Binomial logit model	• Proper model	–	• Gupta (1988)
	Purchase quantity	Poisson model	• Easy to estimate	Mean-variance equality	Bucklin et al. (1998)
	Consumption	Consumption model	• Allows for flexible consumption		Ailawadi and Neslin (1998)
	Store choice	Multinomial logit model	• Consistent with utility theory	IIA-assumption	Bucklin and Lattin (1992)
	Category choice	Multivariate Probit	• Proper model		Manchanda et al. (1999)
	SKU choice	Multinomial logit model	• Parsimonious	IIA-assumption	Fader and Hardie (1996)
	Incidence, choice, and quantity	Dynamic structural model	• Complete, integrated	Difficult to estimate	Erdem et al. (2003), Sun (2005)
Weekly store-level data					
Baseline model	Brand sales to consumers or to retailers	Time series smoothing procedure to distinguish baseline and promotional sales	• Intuitively appealing • Easy to implement	• Difficult to correct for all confounds • No parameterized model for running policy simulations	• Abraham and Lodish (1987, 1993)

(continued)

Table 2.10 (continued)

Model	Dependent variable	Model	Advantage	Disadvantage	Key studies
Regression model	Brand sales at weekly store level	Scan*Pro model: multiplicative regression model	• Fits data well • Tested in many applications	• No dynamic effects • Many independent variables	Wittink et al. (1988)
		Scan*Pro model with lead and lagged effects	• Allows to measure pre- and postpromotion dips	• Many independent variables	Van Heerde et al. (2000)
		System of linear additive models	• Gives decomposition effects	• Linearity assumption • Many independent variables	Van Heerde et al. (2004)
Time series model	Brand sales at market level, or category sales at market level	Vector autoregressive model	• Allows for distinction between long- and short-term effects	• Many parameters	• Nijs et al. (2001) • Steenkamp et al. (2005)
	Brand sales at market level, or category sales at market level	Vector error correction model	• Separate parameters for long- and short-term effects		• Fok et al. (2006)
Aggregate logit model	Brand sales at store level	Aggregate logit model, derived from aggregating individual logit models	• Consistent with utility theory • Parsimonious	• Hard to implement • No dynamic effects • Identification of heterogeneity • Specification of outside good	• Berry et al. (1995) • Dubé et al. (2002)

Eviews, STATA, SAS, Limdep and other statistical packages. STATA also provides multinomial probit estimation. For time series models, Eviews, Stata and SAS/ETS are good choices. Some models (such as multivariate probit models, consumption models, dynamic structural models, aggregate logit models) are not (yet) available in the major commercial statistical programs. Custom programs are required to estimate these models, which can be accomplished in Gauss, Matlab, Stata, SAS, and the free statistical platform R (see Rossi et al. 2005 for several Bayesian models in R).

2.16 Future Research

We hope this chapter enables researchers to implement models that lead to more effective and profitable promotions. We also hope this chapter stimulates new research in areas that have not yet obtained sufficient attention. One such area is the effects of sales promotions for non-grocery products. While most of the models discussed in this chapter are for grocery products, it is unclear whether they are applicable to promotion effects for other items such as durables or services. Though some headway has been made (e.g., Van Heerde et al. (2005) present a promotion model for clothing stores), there is ample room for additional model development. We expect that deceleration effects are stronger in categories that are more expensive (per item) than grocery products. For example, consumers anticipate last-minute holiday deals, many consumers postpone purchasing clothes until the sales season starts (Van Heerde et al. 2005), and the same may apply to car and furniture purchases.

Promotions involve dynamic effects, whose effects are not yet fully captured by current models. For example, we lack optimal retailer models that take into account consumer learning and expectations. One could take the model of Sun et al. (2003) as a starting point and next optimize profit from the firm perspective. Another gap in the literature is store-level models that disentangle state dependence, reference prices, and purchase timing effects.

The decision models for consumer promotions and trade promotions are typically based on descriptive models of demand responses to promotions. However, these decision models tend to exclude competitor responses. If competitors respond to a player's "optimal promotion plan", the outcome may become suboptimal for this player. It seems worthwhile to develop decision models that explicitly account for competitive reactions.

While the literature provides many insights into the effects of price promotions, features and displays, relatively little is known about other promotion types. For example, it is not clear whether promotions that offer more value for the same price (20% extra volume, buy-one-get-one free) are more or less effective than equivalent price promotions (20% lower price, 50% lower price).

The field of online promotions is virtually untapped. For example: are the effects of emailed coupons different from the effects of traditional paper coupons? What are the impacts of free shipping promotions or price discounts communicated in banner ads or on a retail website? Can we design an optimal contact model for email promotions for frequent shoppers that maximizes both retailer and manufacturer profit?

It seems that a key input that is difficult to collect is promotion activities and customer behavior for the *non-focal* store. This is in contrast to scanner data. Online stores and manufacturers have lots data as they pertain to their store. These can and are being used to tailor promotions, (see Zhang and Krishnamurthi 2004; Zhang and Wedel 2009) but ideally data will become available for non-focal stores. In short, promotion models have focused largely on frequently purchased products and estimated using scanner data. There are several issues that need to be resolved in that industry, and the digitized environment opens up an entirely new opportunity for promotions research.

Appendix: Variable Definition for the Decomposition in Sect. 2.4.3

Van Heerde et al. (2004) obtain decomposition (2.11) for price index variables with four types of support (with/without feature, with/without display). To achieve this, they transform the original four promotion variables (PI, FEATONLY, DISPONLY, FEAT&DISP) from the Scan*Pro model into seven new variables: price index with feature-support (PF), price index with display-only support (PD), price index with feature and display support (PFD), price index without support (PWO), plus FWO (Feature without price cut), DWO (Display without price cut), and FDWO (Feature&Display without price cut). Regular price is indicated by a price index with value “1”. A 20% discount would be indicated by 0.8. The PWO, PF, PD, and PFD variables are defaulted to “1” if there is no price discount, but change depending on whether there is a discount and if so how it is supported. The FWO, DWO, and FDWO variables default to “0” and can equal “1” only if there is a feature, display, or both, without a price cut.

To illustrate the transformation, Table 2.11 contains the four original and seven new variables. In case # 1 there is no promotion, and the original price index (PI) equals 1 while the FEATONLY, DISPONLY, FEAT&DISP are zero, as defined in the Scan*Pro model. Since there is no (supported or unsupported) price discount in case #1, the four new price index variables (PWO, PF, PD, PFD) are all at their nonpromotional value of 1. The FWO, DWO, and FDWO variables are zero since there is no feature or display without a price cut. In case # 2 there is a twenty percent price discount without any support, which shows up in the original variables as a price index of 0.8 while FEATONLY, DISPONLY, FEAT&DISP remain zero. Since this is a price cut without support, among the new price indices only the price index without support (PWO) variable is decreased to 0.8. The other three price indices PF, PD, and PFD stay at their nonpromotional level of 1, while the FWO, DWO, and FDWO variables stay at their default value of 0. Case # 3

Table 2.11 Transforming the four Scan*Pro variables into seven new variables

	Scan*Pro variables					Seven new variables (Van Heerde et al. 2000, 2004)						
	Price index	Feature-only	Display-only	Feature and display		Price index without support	Price index with feature-only support	Price index with display only support	Price index with feature and display support	Feature-only without price cut	Display-only without price cut	Feature and display without price cut
Case #	PI	FEAT	DISP	FEAT & DISP		PWO	PF	PD	PFD	FWO	DWO	FDWO
1	1	0	0	0		1	1	1	1	0	0	0
2	0.8	0	0	0		0.8	1	1	1	0	0	0
3	0.8	1	0	0		1	0.8	1	1	0	0	0
4	0.8	0	1	0		1	1	0.8	1	0	0	0
5	0.8	0	0	1		1	1	1	0.8	0	0	0
6	1	1	0	0		1	1	1	1	1	0	0
7	1	0	1	0		1	1	1	1	0	1	0
8	1	0	0	1		1	1	1	1	0	0	1

represents a twenty percent price cut with a feature-only, and hence price index with feature-only support (PF) is lowered to 0.8 (and again, all other variables are at their nonpromotional levels). Analogously, in cases # 4 and 5 the variables PD and PFD become 0.8, in turn. In case # 6 there is a feature without a price cut, which can be seen from the original variables since FEATONLY becomes 1 while PI remains 1. Consequently, among the new variables, FWO becomes 1, while DWO and FDWO remain 0, and PWO, PF, PD, and PFD stay 1 since there is no price cut. Cases #7 and 8 show how DWO and FDWO are defined.

Since price cuts tend to be communicated with feature and or display, the four Scan*Pro Variables tend to be highly correlated. The seven new variables, in contrast, describe seven mutually exclusive promotion situations, and they tend to be much less correlated. While the seven new variables are larger in number than the four Scan*Pro variables, a few of them typically do not vary and can therefore be excluded from models (especially the FDWO variable). Researchers who are concerned about multicollinearity in their (store-level) model may consider using the new set of seven variables proposed in this appendix.

References

Abraham, Magid M., and Leonard M. Lodish. 1987. Promoter: An automated promotion evaluation system. *Marketing Science* 6 (2): 101–123.

Abraham, Magid M., and Leonard M. Lodish. 1993. An implemented system for improving promotion productivity using store scanner data. *Marketing Science* 12 (3): 248–269.

Abramson, Charles, Rick L. Andrews, Imran S. Currim, and Morgan Jones. 2000. Parameter bias from unobserved effects in the multinomial logit model of consumer choice. *Journal of Marketing Research* 37 (4): 410–426.

Ailawadi, Kusum L., and Bari A. Harlam. 2009. Findings—Retailer promotion pass-through: A measure, its magnitude, and its determinants. *Marketing Science* 28 (4): 782–791.

Ailawadi, Kusum L., and Scott A. Neslin. 1998. The effect of promotion on consumption: Buying more and consuming it faster. *Journal of Marketing Research* 35 (August): 390–398.

Ailawadi, Kusum L., Karen Gedenk, and Scott A. Neslin. 1999. Heterogeneity and purchase event feedback in choice models: An empirical analysis with implications for model building. *International Journal of Research in Marketing* 16 (3): 177–198.

Ailawadi, Kusum L., Jacques Cesar, Bari Harlam, and David Trounce. 2007a. Quantifying and improving promotion profitability at CVS. *Marketing Science* 26 (4): 566–575.

Ailawadi, Kusum L., Karen Gedenk, Christian Lutzky, and Scott A. Neslin. 2007b. Decomposition of the sales impact of promotion-induced stockpiling. *Journal of Marketing Research* 44 (3): 450–467.

Ailawadi, Kusum L., Karen Gedenk, Tobias Langer, Yu Ma, and Scott. A. Neslin. 2014. Consumer response to uncertain promotions: An empirical analysis of conditional rebates. *International Journal of Research in Marketing* 31 (1): 94–106.

American Marketing Association. 2015. https://www.ama.org/resources/Pages/Dictionary.aspx?dLetter=P. Accessed 13 Aug 2015.

Andrews, Rick L., Andrew Ainslie, and Imran S. Currim. 2002a. An empirical comparison of logit choice models with discrete versus continuous representations of heterogeneity. *Journal of Marketing Research* 39 (November): 479–487.

Andrews, Rick L., Asim Ansari, and Imran S. Currim. 2002b. Hierarchical Bayes versus finite mixture conjoint analysis models: A comparison of fit, prediction, and partworth recovery. *Journal of Marketing Research* 39 (1): 87–98.

Ansari, Asim, Carl F. Mela, and Scott A. Neslin. 2006. Customer channel migration. Working paper, Columbia Business School, New York.

Ataman, Berk, Carl. F. Mela, and Harald J. van Heerde. 2008. Building brands. *Marketing Science* 27 (November–December): 1036–1054.

Ataman, Berk, Harald J. van Heerde, and Carl. F. Mela. 2010. The long-term effect of marketing strategy on brand sales. *Journal of Marketing Research* 47 (5): 866–882.

Bell, David R., Jeongweng Chiang, and V. Padmanabhan. 1999. The decomposition of promotional response: An empirical generalization. *Marketing Science* 18 (4): 504–526.

Bell, David R., and Christian A.L. Hilber. 2006. An empirical test of the theory of sales: Do household storage constraints affect consumer and store behavior? *Quantitative Marketing and Economics* 4 (2): 87–117.

Bell, David R., and James M. Lattin. 1998. "Shopping behavior and consumer preference for store price format: Why "large basket" shoppers prefer EDLP." *Marketing Science* 17 (1): 66–88.

Bell, David R., and James M. Lattin. 2000. Looking for loss aversion in scanner panel data: The confounding effect of price response heterogeneity. *Marketing Science* 19 (2): 185–200.

Bell, David R., Teck-Hua Ho, and Christopher S. Tang. 1998. Determining where to shop: Fixed and variable costs of shopping. *Journal of Marketing Research* 35 (August): 352–369.

Bell, David R., and James M. Lattin. 1998. "Shopping Behavior and Consumer Preference for Store Price Format: Why "Large Basket"' Shoppers Prefer EDLP." *Marketing Science* 17 (1): 66–88.

Ben-Akiva, Moshe, and Steven R. Lerman. 1991. *Discrete choice analysis*. Cambridge, MA: MIT Press.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* 63 (4): 841–890.

Besanko, David, Sachin Gupta, and Dipak Jain. 1998. Logit demand estimation under competitive pricing behavior: An equilibrium framework. *Management Science* 44 (11, Part 1 of 2): 1533–1547.

Besanko, David, Jean-Pierre Dubé, and Sachin Gupta. 2005. Own-brand and cross-brand retail pass-through. *Marketing Science* 24 (1): 123–137.

Bijmolt, Tammo H.A., Harald J. Van Heerde, and Rik G.M. Pieters. 2005. New empirical generalizations on the determinants of price elasticity. *Journal of Marketing Research* 42 (May): 141–156.

Blattberg, Robert C., and John Deighton. 1991. Interactive marketing: Exploiting the age of addressability. *Sloan Management Review* 33 (1): 21–32.

Blattberg, Robert C., and Alan Levin. 1987. Modelling the effectiveness and profitability of trade promotions. *Marketing Science* 6 (2): 124–146.

Blattberg, Robert C., and Scott A. Neslin. 1990. *Sales promotion, concepts, methods and strategies*. Englewood Cliffs, New Jersey: Prentice Hall.

Blattberg, Robert C., Richard Briesch, and Edward J. Fox. 1995. How promotions work. *Marketing Science* 14 (3, Part 2 of 2): g122–g132.

Bodapati, Anand V., and Sachin Gupta. 2004. The recoverability of segmentation structure from store-level aggregate data. *Journal of Marketing Research* 41 (August): 351–364.

Bodapati, Anand V., and V. Srinivasan. 2006. The impact of feature advertising on customer store choice. Working paper, Stanford Graduate School of Business.

Breugelmans, Els, and Katia Campo. 2011. Effectiveness of in-store displays in a virtual store environment. *Journal of Retailing* 87 (1): 75–89.

Briesch, Richard A., Lakshman Krishnamurthi, Tridib Mazumdar, and S.P. Raj. 1997. A comparative analysis of reference price models. *Journal of Consumer Research* 24 (September): 202–214.

Bucklin, Randolph E., and James M. Lattin. 1992. A model of product category competition among grocery retailers. *Journal of Retailing* 68 (3): 271–293.

Bucklin, Randolph E., Sunil Gupta, and S. Siddarth. 1998. Determining segmentation in sales response across consumer purchase behaviors. *Journal of Marketing Research* 35 (May): 189–197.

Burmester, Alexa B., Jan U. Becker, Harald J. van Heerde, and Michel Clement. 2015. Pre- and post-launch effects of publicity and advertising on sales. *International Journal of Research in Marketing*.

Chan, Tat, Chakravarthi Narasimhan, and Qin Zhang. 2008. Decomposing promotional effects with a dynamic structural model of flexible consumption. *Journal of Marketing Research* 45 (August): 487–498.

Chang, Kwangpil, S. Siddarth, and Charles B. Weinberg. 1999. The impact of heterogeneity in purchase timing and price responsiveness on estimates of sticker shock effects. *Marketing Science* 18 (2): 178–192.

Chiang, Jeongwen. 1991. A simultaneous approach to the whether, what and how much to buy questions. *Marketing Science* 10 (4): 297–315.

Chiang, Jeongwen, Siddhartha Chib, and Chakravarthi Narasimhan. 1999. Markov Chain Monte Carlo and models of consideration set and parameter heterogeneity. *Journal of Econometrics* 89 (1–2): 223–248.

Chib, Siddhartha, P.B. Seetharaman, and Andrei Strijnev. 2004. Model of brand choice with a no-purchase option calibrated to scanner-panel data. *Journal of Marketing Research* 41 (May): 184–196.

Chintagunta, Pradeep K. 1993. Investigating purchase incidence, brand choice, and purchase quantity decisions of households. *Marketing Science* 12 (2): 184–208.

Chintagunta, Pradeep K. 2001. Endogeneity and heterogeneity in a probit demand model: Estimation using aggregate data. *Marketing Science* 20 (4): 442–456.

Chintagunta, Pradeep K., Dipak C. Jain, and Naufel J. Vilcassim. 1991. "Investigating heterogeneity in brand preferences in logit models for panel data." *Journal of Marketing Research* 28 (4): 417–428.

Chintagunta, Pradeep K., and Jean-Pierre Dubé. 2005. Estimating a stockkeeping-unit-level brand choice model that combines household panel data and store data. *Journal of Marketing Research* 42 (August): 368–379.

Datta, Hannes, Bram Foubert, and Harald J. van Heerde. 2015. The challenge of retaining customers acquired by free trials. *Journal of Marketing Research* 52 (April): 217–234.

Dekimpe, Marnik G., and Dominique M. Hanssens. 1995. The persistence of marketing effects on sales. *Marketing Science* 14 (1): 1–21.

DelVecchio, Devon, David H. Henard, and Traci H. Freling. 2006. the effect of sales promotion on post-promotion brand preference: A meta-analysis. *Journal of Retailing* 82 (3): 203–213.

Divakar, Suresh, Brian T. Ratchford, and Venkatesh Shankar. 2005. CHAN4CAST: A multichannel multiregion forecasting model for consumer packaged goods. *Marketing Science* 24 (3): 333–350.

Dodson, Joe A., Alice M. Tybout, and Brian Sternthal. 1978. Impact of deal and deal retraction on brand switching. *Journal of Marketing Research* 15 (May): 72–81.

Drèze, Xavier, and David R. Bell. 2003. Creating win-win trade promotions: Theory and empirical analysis of scan-back trade deals. *Marketing Science* 22 (1): 16–39.

Duan, Jason A., Leigh McAlister, and Shameek Sinha. 2011. Commentary-reexamining Bayesian model-comparison evidence of cross-brand pass-through. *Marketing Science* 30 (3): 550–561.

Dubé, Jean-Pierre, Pradeep K. Chintagunta, Amil Petrin, Bart J. Bronnenberg, P.B. Ron Goettler, K.Sudhir Seetharaman, Raphael Thomadsen, and Ying Zhao. 2002. Structural applications of the discrete choice model. *Marketing Letters* 13 (3): 207–220.

Dubé, Jean-Pierre, and Sachin Gupta. 2008. Cross-brand pass-through in supermarket pricing. *Marketing Science* 27 (3): 324–333.

Eckert, Christine, Jordan J. Louviere, and Towhidul Islam. 2012. Seeing the forest despite the trees: Brand effects on choice uncertainty. *International Journal of Research in Marketing* 29 (3): 256–264.

Ebbes, Peter, Michel Wedel, Ulf Böckenholt, and Ton Steerneman. 2005. Solving and testing for regressor-error (in)dependence when no instrumental variables are available: With new evidence for the effect of education on income. *Quantitative Marketing & Economics* 3 (4): 365–392.

Ebbes, Peter, Dominik Papies, and Harald J. van Heerde. 2011. The sense and non-sense of holdout sample validation in the presence of endogeneity. *Marketing Science* 30 (6): 1115–1122.

Erdem, Tülin, and Michael P. Keane. 1996. Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science* 15 (1): 1–20.

Erdem, Tülin, Susumu Imai, and Michael P. Keane. 2003. Brand and quantity choice dynamics under price uncertainty. *Quantitative Marketing and Economics* 1 (1): 5–64.

Erdem, Tülin, Glenn Mayhew, and Baohong Sun. 2001. Understanding reference-price shoppers: A within- and cross-category analysis. *Journal of Marketing Research* 38 (November): 445–457.

Erdem, Tülin, Michael P. Keane, and Baohong Sun. 2008. A dynamic model of brand choice when price and advertising signal product quality. *Marketing Science* 27 (6): 1111–1125.

Fader, Peter S., and Bruce G.S. Hardie. 1996. Modeling consumer choice among SKUs. *Journal of Marketing Research* 33 (November): 442–452.

Fok, Dennis, Csilla Horváth, Richard Paap, and Philip Hans Franses. 2006. A hierarchical Bayes error correction model to explain dynamic effects of price changes. *Journal of Marketing Research* 43 (August): 443–461.

Gedenk, Karen, and Scott A. Neslin. 1999. The role of retail promotion in determining future brand loyalty: Its effect on future purchase event feedback. *Journal of Retailing* 75 (4): 433–459.

Germann, Frank, Peter Ebbes, and Rajdeep Grewal. 2015. The chief marketing officer matters! *Journal of Marketing* 79 (3): 1–22.

Gómez, Miguel I., Vithala R. Rao, and Edward W. McLaughlin. 2007. Empirical analysis of budget and allocation of trade promotions in the U.S. supermarket industry. *Journal of Marketing Research* 44 (3): 410–424.

Gönül, Füsun, and Kannan Srinivasan. 1996. Estimating the impact of consumer expectations on purchase behavior: A dynamic structural model. *Marketing Science* 15 (3): 262–279.

Guadagni, Peter M., and John D.C. Little. 1983. A logit model calibrated on scanner data. *Marketing Science* 2 (3): 203–238.

Gupta, Sunil. 1988. Impact of sales promotion on when, what and how much to buy. *Journal of Marketing Research* 25 (November): 342–355.

Gupta, Sachin, and Pradeep K. Chintagunta. 1994. On using demographic variables to determine segment membership in logit mixture models. *Journal of Marketing Research* 31 (1): 128–136.

Guyt, Jonne Y., and Els Gijsbrechts. 2014. Take turns or March in sync? The impact of the national brand promotion calendar on manufacturer and retailer performance. *Journal of Marketing Research* 51 (6): 753–772.

Haans, Hans, and Els Gijsbrechts. 2011. One-deal-fits-all? On category sales promotion effectiveness in smaller versus larger supermarkets. *Journal of Retailing* 87 (4): 420–436.

Ho, Teck-Hua, and Juin-Kuan Chong. 2003. A parsimonious model of stockkeeping-unit choice. *Journal of Marketing Research* 40 (August): 351–365.

Hoch, Stephen J., Byung-Do Kim, Alan L. Montgomery, and Peter E. Rossi. 1995. Determinants of store-level price elasticity. *Journal of Marketing Research* 32 (1): 17–29.

Hoch, Stephen J., Xavier Dreze, and Mary E. Purk. 1994. "EDLP, Hi-Lo, and Margin Arithmetic." *Journal of Marketing* 58 (4): 16–27.

Horsky, Dan, Sanjog Misra, and Paul Nelson. 2006. Observed and unobserved preference heterogeneity in brand-choice models. *Marketing Science* 25 (4): 322–335.

Horváth, Csilla, and Philip Hans B.F. Franses. 2003. Deriving dynamic marketing effectiveness from econometric time series models. In *ERIM report series research in management*. Rotterdam: Erasmus Research Institute of Management.

Jedidi, Kamel, Carl F. Mela, and Sunil Gupta. 1999. Managing advertising and promotion for long-run profitability. *Marketing Science* 18 (1): 1–22.

Kahneman, Daniel, and Amos Tversky. 1979. Prospect theory: An analysis of decision under risk. *Econometrica* 47 (March): 263–291.

Kalyanaram, Gurumurthy, and Russell S. Winer. 1995. Empirical generalizations from reference price research. *Marketing Science* 14 (3): G161–G169.

Kamakura, Wagner A., and Wooseong Kang. 2007. Chain-wide and store-level analysis for cross-category management. *Journal of Retailing* 83 (2): 159–170.

Kamakura, Wagner A., and Gary J. Russell. 1989. A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research* 26 (November): 379–390.

Kantar Retail. 2012. *2012 trade promotion industry benchmarking study*. Wilton, CT: Kantar Retail.

Keane, Michael P. 1997a. Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business and Economic Statistics* 15 (3): 310–327.

Keane, Michael P. 1997b. Current issues in discrete choice modeling. *Marketing Letters* 8 (3): 307–322.

Khan, Romana, Michael Lewis, and Vishal Singh. 2009. customer management and the value of one-to-one marketing. *Marketing Science* 28 (6): 1073–1079.

Krishna, Aradhna, Imran S. Currim, and Robert W. Shoemaker. 1991. Consumer perceptions of promotional activity. *Journal of Marketing* 55: 4–16.

Krishnamurthi, Lakshman, and S.P. Raj. 1988. A model of brand choice and purchase quantity price sensitivities. *Marketing Science* 7 (1): 1–20.

Lam, Shun Yin, Mark Vandenbosch, John Hulland, and Michael Pearce. 2001. Evaluating promotions in shopping environments: Decomposing sales response into attraction, conversion, and spending effects. *Marketing Science* 20 (2): 194–215.

Leeflang, Peter S.H. 2008. Modeling competitive responsiveness. *In Chapter 7 in handbook of marketing decision models*, ed. Berend Wierenga, 211–251. International Series in Operational Research & Management Science. New York (USA): Springer Science and Business Media.

Leeflang, Peter S.H., and Josefa Parreña Selva. 2012. Cross-category demand effects of price promotions. *Journal of the Academy of Marketing Science* 40 (4): 572–586.

Leeflang, Peter S.H., and Dick R. Wittink. 1992. Diagnosing competitive reactions using (aggregated) scanner data. *International Journal of Research in Marketing* 9 (1): 39–57.

Leeflang, Peter S.H., and Dick R. Wittink. 1996. Competitive reaction versus consumer response: Do managers overreact? *International Journal of Research in Marketing* 13 (2): 103–120.

Leeflang, Peter S.H., Josefa Parreña Selva, Albert van Dijk, and Dick R. Wittink. 2008. Decomposing the sales promotion bump accounting for cross-category effects. *International Journal of Research in Marketing* 25 (3): 201–214.

Liu, Qiang, Thomas J. Steenburgh, and Sachin Gupta. 2015. The cross attributes flexible substitution logit: Uncovering category expansion and share impacts of marketing instruments. *Marketing Science* 34 (1): 144–159.

Lourenço, Carlos J.S., Els Gijsbrechts, and Richard Paap. 2015. The impact of category prices on store price image formation: An empirical analysis. *Journal of Marketing Research* 52 (2): 200–216.

Luo, Xueming, Michelle Andrews, Zheng Fang, and Chee Wei Phang. 2014. Mobile targeting. *Management Science* 60 (7): 1738–1756.

Macé, Sandrine, and Scott A. Neslin. 2004. The determinants of pre- and postpromotion dips in sales of frequently purchased goods. *Journal of Marketing Research* 41 (August): 339–350.

Manchanda, Puneet, Asim Ansari, and Sunil Gupta. 1999. The "shopping basket": A model for multicategory purchase incidence decisions. *Marketing Science* 18 (2): 95–114.

McAlister, Leigh. 2007. Comment—cross-brand pass-through: Fact or artifact? *Marketing Science* 26 (6): 876–898.

Mehta, Nitin, and Yu Ma. 2012. A multicategory model of consumers' purchase incidence, quantity, and brand choice decisions: Methodological issues and implications on promotional decisions. *Journal of Marketing Research* 49 (August): 435–451.

Mela, Carl F., Sunil Gupta, and Donald R. Lehmann. 1997. The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing Research* 34 (May): 248–261.

Mela, Carl F., Kamel Jedidi, and Douglas Bowman. 1998. The long-term impact of promotions on consumer stockpiling behavior. *Journal of Marketing Research* 35 (May): 250–262.

Montgomery, Alan L. 1997. Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing Science* 16 (4): 315–337.

Moorthy, Shridhar. 2005. A general theory of pass-through in channels with category management and retail competition. *Marketing Science* 24 (1): 110–122.

Mulhern, Francis J., and Robert P. Leone. 1991. Implicit price bundling of retail products: A multiproduct approach to maximizing store profitability. *Journal of Marketing* 55 (October): 63–76.

Nair, Harikesh, Jean-Pierre Dubé, and Pradeep Chintagunta. 2005. Accounting for primary and secondary demand effects with aggregate data. *Marketing Science* 24 (3): 444–460.

Narasimhan, Chakravarthi, Scott A. Neslin, and Subrata K. Sen. 1996. promotional elasticities and category characteristics. *Journal of Marketing* 60 (April): 17–30.

Natter, Martin, Andreas Mild, Thomas Reutterer, and Alfred Taudes. 2007. An assortment-wide decision-support system for dynamic pricing and promotion planning. *Marketing Science* 26 (4): 576–583.

Neslin, Scott A. 2002. *Sales promotion*. Relevant knowledge series. Cambridge, MA: Marketing Science Institute.

Neslin, Scott A., Stephen G. Powell, and Linda G. Schneider Stone. 1995. The effects of retailer and consumer response on optimal manufacturer advertising and trade promotion strategies. *Management Science* 41 (5): 749–766.

Neslin, Scott A., Thomas P. Novak, Kenneth R. Baker, and Donna L. Hoffman. 2009. An optimal contact model for maximizing online panel response rates. *Management Science* 55 (5): 727–737.

Neslin, Scott A., Gail Ayala Taylor, Kimberly D. Grantham, and Kimberly R. McNeil. 2013. Overcoming the 'recency trap' in customer relationship management. *Journal of the Academy of Marketing Science* 41 (3): 320–337.

Nevo, Aviv. 2000. A practitioner's guide to estimation of random-coefficient logit models of demand. *Journal of Economics & Management Strategy* 9 (4): 513–548.

Nijs, Vincent, Marnik G. Dekimpe, Jan Benedict Steenkamp, and Dominique M. Hanssens. 2001. The category demand effects of price promotions. *Marketing Science* 20 (1): 1–22.

Nijs, Vincent, Kanishka Misra, Eric T. Anderson, Karsten Hansen, and Lakshman Krishnamurthi. 2010. Channel pass-through of trade promotions. *Marketing Science* 29 (2): 250–267.

Park, Sungho, and Sachin Gupta. 2009. Simulated maximum likelihood estimator for the random coefficient logit model using aggregate data. *Journal of Marketing Research* 46 (August): 531–542.

Park, Sungho, and Sachin Gupta. 2011. A regime-switching model of cyclical category buying. *Marketing Science* 30 (3): 469–480.

Park, Sungho, and Sachin Gupta. 2012. Handling endogenous regressors by joint estimation using copulas. *Marketing Science* 31 (4): 567–586.

Pauwels, Koen. 2007. How retailer and competitor decisions drive the long-term effectiveness of manufacturer promotions for fast moving consumer goods. *Journal of Retailing* 83 (3): 297–308.

Pauwels, Koen, Dominique M. Hanssens, and S. Siddarth. 2002. The long-term effects of price promotions on category incidence, brand choice and purchase quantity. *Journal of Marketing Research* 39 (November): 421–439.

Petrin, Amil, and Kenneth Train. 2010. A control function approach to endogeneity in consumer choice models. *Journal of Marketing Research* 47 (1): 3–13.

Rhee, Hongjai, and David R. Bell. 2002. The inter-store mobility of supermarket shoppers. *Journal of Retailing* 78 (4): 225–237.

Rooderkerk, Robert P., Harald J. van Heerde, and Tammo H.A. Bijmolt. 2011. Incorporating context effects into a choice model. *Journal of Marketing Research* 48 (August): 767–780.

Rooderkerk, Robert P., Harald J. van Heerde, and Tammo H.A. Bijmolt. 2013. Optimizing retail assortments. *Marketing Science* 32 (5): 699–715.

Rossi, Peter. 2014. Even the rich can make themselves poor: A critical examination of IV methods in marketing applications. *Marketing Science* 33 (5): 655–672.

Rossi, Peter E., Robert E. McCulloch, and Greg M. Allenby. 1996. The value of purchase history data in target marketing. *Marketing Science* 15 (4): 321–340.

Rossi, Peter E., Greg M. Allenby, and Robert McCulloch. 2005. *Bayesian statistics and marketing*. Chichester, West Sussex, England: Wiley.

Rothschild, Michael L., and William C. Gaidis. 1981. Behavioral learning theory: Its relevance to marketing and promotions. *Journal of Marketing* 45 (2): 70–78.

Roy, Rishin, Pradeep K. Chintagunta, and Sudeep Haldar. 1996. A framework for investigating habits, 'the hand of the past,' and heterogeneity in dynamic brand choice. *Marketing Science* 15 (3): 280–299.

Seetharaman, P.B. 2004. Modeling multiple sources of state dependence in random utility models: A distributed lag approach. *Marketing Science* 23 (2): 263–271.

Seetharaman, P.B., Andrew Ainslie, and Pradeep K. Chintagunta. 1999. Investigating household state dependence effects across categories. *Journal of Marketing Research* 36 (4): 488–500.

Shin, Sangwoo, Sanjog Misra, and Dan Horsky. 2012. Disentangling preferences and learning in brand choice models. *Marketing Science* 31 (1): 115–137.

Siddarth, S., Randolph E. Bucklin, and Donald G. Morrison. 1995. Making the cut: Modeling and analyzing choice set restriction in scanner panel data. *Journal of Marketing Research* 32 (3): 255–266.

Silva-Risso, Jorge M., Randolph E. Bucklin, and Donald G. Morrison. 1999. A decision support system for planning manufacturers' sales promotion calendars. *Marketing Science* 18 (3): 274–300.

Singh, Vishal, Karsten Hansen, and Robert Blattberg. 2006. Market entry and consumer behavior: An investigation of a Wal-Mart supercenter. *Marketing Science* 25 (2): 457–476.

Slotegraaf, Rebecca J., and Koen Pauwels. 2008. The impact of brand equity and innovation on the long-term effectiveness of promotions. *Journal of Marketing Research* 45 (3): 293–306.

Steenburgh, Thomas J. 2008. The invariant proportion of substitution property (IPS) of discrete-choice models. *Marketing Science* 27 (2): 300–307.

Steenkamp, Jan Benedict E.M., Vincent R. Nijs, Dominique M. Hanssens, and Marnik G. Dekimpe. 2005. Competitive reactions to advertising and promotion attacks. *Marketing Science* 24 (1): 35–54.

Sun, Baohong. 2005. Promotion effects on endogenous consumption. *Marketing Science* 24 (3): 430–443.

Sun, Baohong, Scott A. Neslin, and Kannan Srinivasan. 2003. Measuring the impact of promotions on brand switching when consumers are forward looking. *Journal of Marketing Research* 40 (November): 389–405.

Tellis, Gerard J., and Fred S. Zufryden. 1995. Tackling the retailer decision maze: Which brands to discount, how much, when and why? *Marketing Science* 14 (3, Part 1 of 2): 271–299.

Train, Kenneth. 2003. *Discrete choice methods with simulation*. Cambridge, UK: Cambridge University Press. http://elsa.berkeley.edu/books/choice2.html.

Tyagi, Rajeev K. 1999. A characterization of retailer response to manufacturer trade deals. *Journal of Marketing Research* 36 (November): 510–516.

Van Heerde, Harald J, and Tammo HA Bijmolt. 2005. "Decomposing the promotional revenue bump for loyalty program members versus non-members." *Journal of Marketing Research* 42 (November): 443–457.

Van Heerde, Harald J., Sachin Gupta. 2006. The origin of demand: A system to classify the sources of the sales promotion bump. Working paper, University of Waikato.

Van Heerde, Harald J., Peter S.H. Leeflang, and Dick R. Wittink. 2000. The estimation of pre- and postpromotion dips with store-level scanner data. *Journal of Marketing Research* 37 (August); 383–395.

Van Heerde, Harald J., Peter S.H. Leeflang, and Dick R. Wittink. 2001. Semiparametric analysis to estimate the deal effect curve. *Journal of Marketing Research* 38 (May): 197–215.

Van Heerde, Harald J., Sachin Gupta, and Dick R. Wittink. 2003. Is 75% of the sales promotion bump due to brand switching? No, only 33% is. *Journal of Marketing Research* 40 (November): 481–491.

Van Heerde, J. Harald, Peter S.H. Leeflang, and Dick R. Wittink. 2004. Decomposing the sales promotion bump with store data. *Marketing Science* 23 (3): 317–334.

Van Heerde, J. Harald, Kristiaan Helsen, and Marnik Dekimpe. 2007. The impact of a product-harm crisis on marketing effectiveness. *Marketing Science* 26 (2): 230–245.

Van Heerde, J. Harald, Shuba Srinivasan, and Marnik G. Dekimpe. 2010. Estimating cannibalization rates for a pioneering innovation. *Marketing Science* 29 (6): 1024–1039.

Van Heerde, Harald J., Maarten Gijsenberg, Marnik G. Dekimpe, and Jan-Benedict E.M. Steenkamp. 2013. Advertising and price effectiveness over the business cycle. *Journal of Marketing Research* 50 (April): 177–193.

Van Heerde, Harald J., Els Gijsbrechts, and Koen Pauwels. 2015. Fanning the flames? How media coverage of a price war impacts retailers, consumers and investors. *Journal of Marketing Research*.

Van Lin, Arjen, and Els Gijsbrechts. 2014. Shopper loyalty to whom? Chain versus outlet loyalty in the context of store acquisitions. *Journal of Marketing Research* 51 (3): 352–370.

Van Lin, Arjen, and Els Gijsbrechts. 2015. The battle for health and beauty: What drives supermarket and drugstore category-promotion lifts? *International Journal of Research in Marketing*. http://dx.doi.org/10.1016/j.ijresmar.2015.09.002.

Varki, Sajeev, and Pradeep K. Chintagunta. 2004. The augmented latent class model: Incorporating additional heterogeneity in the latent class model for panel data. *Journal of Marketing Research* 41 (2): 226–233.

Venkatesan, Rajkumar, and Paul W. Farris. 2012. Measuring and managing returns from retailer-customized coupon campaigns. *Journal of Marketing* 76 (1): 76–94.

Villas-Boas, J. Miguel, and Russell S. Winer. 1999. "Endogeneity in brand choice models." *Management Science* 45 (10): 1324–1338.

Vroegrijk, Mark, Els Gijsbrechts, and Katia Campo. 2013. Close encounter with the hard discounter: A multiple-store shopping perspective on the impact of local hard-discounter entry. *Journal of Marketing Research* 50 (5): 606–626.

Walters, Rockney G. 1989. An empirical investigation into retailer response to manufacturer trade promotions. *Journal of Retailing* 65 (2): 253–272.

Wieringa, Jaap E., Peter C. Verhoef, and Janny C. Hoekstra. 2011. *Liber Amicorum in honor of Peter S.H. Leeflang*. Rijksuniversiteit Groningen.

Winer, Russell S. 1986. A reference price model of brand choice for frequently purchased products. *Journal of Consumer Research* 13 (September): 250–256.

Wittink, Dick R., Michael J. Addona, William J. Hawkes, and John C. Porter. 1988. The estimation, validation, and use of promotional effects based on scanner data. Working paper, Cornell University.

Zhang, Jie, and Lakshman Krishnamurthi. 2004. Customizing promotions in online stores. *Marketing Science* 24 (4): 561–578.

Zhang, Jie, and Michel Wedel. 2009. The effectiveness of customized promotions in online and offline stores. *Journal of Marketing Research* 48 (2): 190–206.

Chapter 3
Innovation and New Products Research: A State-of-the-Art Review, Models for Managerial Decision Making, and Future Research Directions

Tingting Fan, Peter N. Golder and Donald R. Lehmann

3.1 Introduction

Innovation is the lifeblood of economies. It spawns new firms, revitalizes established organizations, enriches entrepreneurs, builds professional reputations, and raises living standards throughout the world. Over the years, innovation has received considerable attention in the marketing literature, which is not surprising given its importance to both companies and consumers.

Our purpose in writing this chapter is threefold. First, we provide a literature review of major papers in the field of new products research. We organize our review into four tables, one for each of the four stages of the new product development process, and then by topic within each of these stages. Moreover, we provide a short summary of each paper in the tables. These tables and short summaries provide an overview of research and findings in the field, plus direct readers to papers particularly suited to their interests. Second, we highlight specific models within each stage of the new product development process. These models are useful for marketing researchers and managers tackling challenges in the new products domain. Third, after reviewing the literature, we suggest numerous general research directions as well as specific research questions to guide future investigations in this area. We believe this will be particularly useful to those new to the field of new products research, especially those interested in applying quantitative models (e.g., business school Ph.D. students, consultants, practitioners).

T. Fan
The Business School of Chinese University of Hong Kong, Sha Tin, Hong Kong

P.N. Golder (✉)
Tuck School of Business at Dartmouth, Hanover, NH, USA
e-mail: Peter.N.Golder@tuck.dartmouth.edu

D.R. Lehmann
Columbia Business School, New York, NY, USA

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_3

Several researchers have previously documented the innovation subfield of the marketing discipline.[1] In addition to this chapter, we refer interested readers to chapters by Lehmann and Golder (2014) and Rao (2014), prior journal articles by Hauser et al. (2006) and Peres et al. (2010), and the entire *Handbook of Research* on New Product Development (Golder and Mitra 2017). Next, we briefly summarize these writings, and then provide the motivation and positioning of our chapter.

Lehmann and Golder (2014) wrote their chapter on new products research for the *History of Marketing Science*. In it, they organize important new products research beginning in the 1960s going through the 2010s. They find that research is limited on the topics of opportunity identification and concept generation. However, they find that researchers have paid greater attention to design and development (particularly conjoint analysis), sales forecasting, and some aspects of strategy (e.g., competitive response, order of entry).

Rao (2014) provides a focused, comprehensive review of conjoint analysis, in contrast with other authors who discuss conjoint analysis as part of the broader new products literature. Rao's (2014) chapter updates earlier reviews by Wittink and Cattin (1989), Green and Srinivasan (1990), and Wittink et al. (1994).

Hauser et al. (2006) organize research on innovation into five fields of inquiry: (i) consumer response to innovation (e.g., consumer innovativeness, new product growth models, network externalities), (ii) organizational impact (e.g., contextual drivers, adoption of new methods), (iii) market entry strategies (e.g., technological evolution, portfolio management), (iv) prescriptive techniques for product development (e.g., fuzzy front end, design tools), and (v) outcomes from innovation (e.g., market rewards, defending against new entrants, internal rewards for innovation).

Peres et al. (2010) review research on innovation diffusion. Building on prior reviews (e.g., Mahajan et al. 1995), they incorporate more recent research on interpersonal influence, network externalities, and social signals. Their review covers within-market effects, cross-market effects, and cross-country effects for all social influences, not just word of mouth.

The *Handbook of Research on New Product Development* (Golder and Mitra 2017) contains 19 chapters that provide depth on specific aspects of new product development and innovation. These chapters describe the frontiers of new products research as well as offer numerous insights for extending these frontiers of our knowledge base.

Empirical findings across a range of new products research studies can be found in *Empirical Generalizations about Marketing Impact* (Hanssens 2015). Topics in this book of interest to new products researchers include customer satisfaction and product reviews, objective and perceived quality, order of entry, sales diffusion and social influence, product innovation, and competitive reaction.

---

[1] The *Journal of Product Innovation Management* is also a valuable repository of new products research.

Our chapter for this handbook provides a unique perspective on innovation and new products research. In contrast with earlier works, we organize our review by the major stages of the new product development process: (i) opportunity identification, (ii) product design and development, (iii) sales forecasting, and (iv) commercialization. Our review is more comprehensive than Lehmann and Golder (2014) since their purpose was to provide exemplars of research themes illustrating the historical evolution of the field. One aspect of innovation covered in Lehmann and Golder (2014) that we do not cover is how to value innovations. Most importantly, we focus on models useful for managerial decision-making in various innovation and product development contexts. Therefore, we provide details of several models so that managers and researchers can better understand how these models could be used and the insights that can be generated for informing their decision making.

Overall, our chapter has three overriding objectives: (i) to organize the vast marketing literature on innovation and new products according to the four stages of the new product development process, (ii) to provide detailed discussions of decision-making models useful for each stage of the new product development process, and (iii) to identify the most promising opportunities for future research in each of these stages. Next, we discuss research categorized into the four stages of the new product development process. Within each of these stages, we elaborate on key models. We conclude with our agenda for future research.

3.2 Organizing Research on Innovation and New Products

The new product development process can be broken down into four stages: (i) opportunity identification, (ii) product design and development, (iii) sales forecasting, and (iv) commercialization. The first stage subsumes idea generation and idea screening, which are sometimes treated as separate stages. The final stage, commercialization, includes some research on post-commercialization activities but does not attempt to cover the vast literature on the product management of mature products.

3.2.1 Opportunity Identification

Current research on opportunity identification focuses on three areas: (i) opportunities identified by investigating lead users, (ii) opportunities identified by using online platforms, and (iii) opportunities identified through innovation templates (see Table 3.1). Research on lead users began in the 1980s and is largely associated with the work of Eric von Hippel. Lead users are consumers who have needs ahead of the general market and also create make-shift solutions for those needs. When firms can identify lead users, they benefit in two ways. First, they become aware of unmet needs that exist in the marketplace. These needs are important enough that

Table 3.1 Opportunity identification

Topics	Literature	Summary
Opportunity identified from lead users	Von Hippel (1986)	Lead users are "users whose present strong needs will become general in a marketplace months or years in the future". Lead users can help companies forecast emerging needs, provide new product concepts, and expedite new product development.
	Urban and Von Hippel (1988)	Apply the lead user methodology in the development of a new industrial product (i.e., computer-aided systems for the design of printed circuit boards (PC-CAD)). The authors demonstrate that lead users provide useful data and people prefer the new product concepts generated from lead users.
	Lilien et al. (2002)	Test the effect of lead users in a natural experiment conducted in the 3M Company. The authors found that the forecast annual sales of lead user projects are more than eight times higher than those for the traditional projects and that divisions funding lead user projects generated their highest rate of major product lines in the past 50 years.
Opportunity identified from online platforms	Urban and Hauser (2004)	The authors develop a method which "listens in" to conversations between customers and web-based virtual advisers and demonstrates its value to identify valuable opportunities.
	Bayus (2013)	New product ideas can come from crowdsourcing communities where a dispersed crowd of consumers generates ideas. They found that serial ideators are more likely than consumers with a single idea to generate valuable ideas, but they are unlikely to repeat their early success once their ideas are implemented.
	Stephen et al. (2015)	When new product ideas come from consumers through online platforms, higher clustering/interconnectivity among consumers generate less innovative ideas because ideas from clustered consumers are more likely to be similar or redundant.
	Luo and Toubia (2015)	On online idea generation platforms, high-knowledge consumers generate better ideas when they are exposed to abstract cues such as problem decomposition, while low-knowledge consumers generate better ideas when they see concrete cues such as other consumers' ideas.

(continued)

Table 3.1 (continued)

Topics	Literature	Summary
Opportunity identified from innovation templates	Goldenberg et al. (1999)	Identify five “innovation templates” that most successful new products fit into. These templates are based on product attributes: inclusion, exclusion, linking, unlinking, joining, and splitting.
	Goldenberg et al. (2001)	New product success can be predicted by inspecting the new product idea (innovation templates) and the circumstances of its emergence (protocol). In particular, successful products tend to conform to one of the templates and solve customer problems. New products that are developed only by the inventor or mimicking popular trends are likely to fail.
	Goldenberg et al. (2003)	Apply the “innovation templates” on new products (e.g., DVD player—subtraction, Gillette double-bladed razor—multiplication, Caesarea Creation Industries’ rug for children’s rooms—division, defrosting filament in an automobile windshield with enhanced radio reception—task unification, Elgo’s indoor sprinkler kit—attribute dependency change).
	Boyd and Goldenberg (2013)	Successful applications of the new product templates are discussed.
Opportunities identified from other ways	Chandy and Tellis (1998)	How willingness to cannibalize prior investments may be more important than firm size as a driver of radical product innovation.
	Moorman and Miner (1997)	Greater organizational memory dispersion increases new product creativity and performance.
	Ofek and Sarvary (2003)	How leaders and followers will invest in R&D versus advertising with next-generation technology products. They find that firm investments depend on whether current leadership is based on R&D competence or reputation.
	Kornish and Ulrich (2011)	Parallel search is a common approach to innovation. Firms identify a large number of opportunities and then select some to develop, with only a few successful cases. The authors develop a method to extrapolate unique ideas from a lot of redundant ideas.
	Burroughs et al. (2011)	Firms can enhance creativity by leveraging the monetary reward program and a creative training program in combination.

consumers actively generate their own solutions. Second, potential new product ideas are generated through lead users' solutions. While often these rudimentary solutions are not firms' final solutions, they can be very useful in contributing to firms' attempts to develop products for a broad market. Research has identified many successful market applications of the lead user concept.

Another area of research on opportunity identification to emerge was the development and application of innovation templates. This research stream offered a structured approach to what might otherwise be an undefined process with random outcomes. For example, by identifying five templates of innovation that most successful products belong to, Goldenberg et al. (2001) show how future new products can be identified and refined more easily.

A third and more recent area of research on opportunity identification is the use of online platforms such as Dell's Idea Storm (Bayus 2013). This research tends to investigate how the structure and process of online communities contribute to or detract from opportunity identification and idea generation. Unlike the previous two areas of research, we need better documentation of the marketplace success of new products generated through online platforms.

3.2.1.1 Modeling Opportunity Identification

The most common goal in this research area is to predict the success of new product ideas. Here, researchers have usefully employed logit models (e.g., Bayus 2013; Goldenberg et al. 2001). For example, Bayus (2013) used an individual-effects logit model where the dependent variable (y_{it}) is a binary variable that indicates whether ideator i proposes an idea that is eventually implemented (1 for yes and 0 otherwise).

$$\Pr(y_{it} = 1) = \Lambda(\alpha_i + \beta x_{it} + \gamma z_i), \tag{3.1}$$

where Λ is a logit model $\Lambda(w) = e^w/(1 + e^w)$. The independent variables (x_{it} and z_i) include the past success of this ideator in generating implemented ideas, the diversity of this ideator's past comments, and control variables such as the ideator's demographic information.

In addition to predicting the success of an idea, researchers are also interested in the number of ideas generated by an ideator. One model used in this context is the Poisson model. For example, Bayus (2013) uses an individual-effect Poisson model where the dependent variable (y_{it}) is the number of ideas generated by ideator i. This count variable takes on positive integer values, which are assumed to follow the Poisson distribution. The model is specified as follows.

$$\Pr(y_{it}) = \lambda^{y_{it}} e^{-\lambda}/y_{it}! \tag{3.2}$$

where λ is an empirically-estimated parameter.

3.2.2 Product Design and Development

Once firms have identified potential opportunities, generated new product ideas, and screened those ideas for the most promising alternatives, firms need to move these ideas into product design and development. Unlike opportunity identification, product design and development has a rich and reasonably large literature in marketing (see Table 3.2 for an overview organized by topic). Much of this literature is due to one model-based area of research: conjoint analysis. Over several decades, many researchers have made important contributions to this literature, which in turn has had a substantial impact on management practice. Conjoint analysis may be the most impactful method developed in marketing research and it almost certainly is such within the area of innovation and new products research. Rao (2014) provides a focused, comprehensive review of conjoint analysis, which updates earlier reviews by Wittink and Cattin (1989), Green and Srinivasan (1990), Wittink et al. (1994), and Rao (2008).

One of the key challenges in conducting conjoint analysis research is that so many product variants exist in multi-attribute product space. As a result, much of the research in conjoint analysis has dealt with methodological and practical approaches by asking consumers about a subset of alternatives while still being able to estimate their preferences across the multi-attribute space.

Besides conjoint analysis, researchers have proposed other approaches for incorporating customers' preferences into product design and development. These approaches include quality function deployment (QFD) (Hauser and Clausing 1988) and voice of the customer (VOC) (Griffin and Hauser 1993). More recently, the idea of heterogeneous design or product morphing has been proposed as a way to take advantage of flexible production in order to better satisfy consumers' specific preferences.

3.2.2.1 *Models for Product Design*

Homogeneous Product Design. Traditional models of product design often use McFadden's (1974) random utility model of consumer choice where all consumers have the same value for attributes. Specifically, the utility u_{sj} for alternative j in choice set s is

$$u_{sj} = x_{sj}^{'}\beta + \varepsilon_{sj}, \tag{3.3}$$

where x_{sj} is a vector of the attribute levels and β captures the corresponding weights (or part-worths) consumers have for each attribute level. β s are homogeneous across consumers and therefore assume that the product design is the same for all consumers. ε_{sj} is an error term following an extreme value distribution. The maximization of consumer utility generates the probability that alternative j is chosen from choice set s:

Table 3.2 Product design and development

Topics	Literature	Summary
Concept and measurement	Homburg et al. (2015)	Product design is measured along three dimensions: aesthetics, functionality, and symbolism. Product design has positive influence on consumer willingness to pay, purchase intention, and word of mouth.
Technology evolution	Fisher and Pry (1971)	Technological evolution follows an S-shape path.
	Sood and Tellis (2005)	The authors challenge the traditional S-shaped technological evolution and found that technological evolution follows a step function with sharp improvements in performance following long periods of no improvement.
	Golder et al. (2009)	Examine 29 radical innovations from initial concept to mass-market commercialization and report findings on when (duration times), by whom (product development leaders), and how (technology borrowing) radical innovations are developed.
Successful product designs	Landwehr et al. (2011)	Prototypical but complex car designs are more preferred and generate more sales than unusual and complex designs.
	Koukova et al. (2012)	The design of multi-format digital products depends on usage situation. Consumers prefer complementary formats because they can use the multi-format product in different usage situations.
	Landwehr et al. (2013)	The influence of a product design on product liking depends on the different stages of exposure. Typical design is preferred at lower exposure levels, while atypical design is preferred at higher exposure levels. In the long run atypical designs are more likely to succeed.
	Ma et al. (2015)	Consumers are more likely to adopt a really new innovation as a detachable peripheral component than integrate it into the base product.
New product planning	Cooper (1990, 1994)	Stage-gate method which consists of concept development, design, testing, and launch.
	Ding and Eliashberg (2002)	Investigate how to structure the new product development pipeline by selecting the appropriate number of projects to fund at each stage in order to have a successful product emerge in the end.

(continued)

Table 3.2 (continued)

Topics	Literature	Summary
	Griffin (1997)	Ways to increase the speed and improve the chances of success in moving through the "funnel."
	Sethi and Iqbal (2008)	A strictly applied stage-gate process may inhibit the development of really new products.
	Ederer and Manso (2013)	Examine the effect of pay for performance on innovation. They found that the combination of tolerance for early failure and reward for long-term success is effective in motivating innovation.
Organizational factors that influence new product development	Rindfleish and Moorman (2001)	Different impacts of horizontal and vertical alliances on new product development. They find that alliances with higher overlap in firms' knowledge bases and higher quality relationships lead to higher new product creativity and faster new product development.
	Sethi et al. (2001)	Moving beyond functional boundaries to identify with the cross-functional team promotes the innovativeness of new products.
	Ganesan et al. (2005)	Examine geographic proximity of alliance partners and find that strong relational ties may be more important than simple geographic proximity and that e-mail communication, in contrast to face-to-face communication enhances new product creativity and development speed.
	Slotegraaf and Atuahene-Gima (2011)	The degree of stability in a new product development project team has a curvilinear relationship to new product advantage.
	Cui and O'Connor (2012)	Examine the resource diversity of multiple alliance partners and its contribution to firm innovation.
	Sethi et al. (2012)	Examine how to use micropolitical strategies to win approval for development of new-to-the-firm products (with market and technology newness).
	Borah and Tellis (2014)	Empirically study firms' choice of and payoff from making, buying, or allying for innovations.
	Tracey et al. (2014)	New product outcomes (e.g., product novelty and speed to market) are

(continued)

Table 3.2 (continued)

Topics	Literature	Summary
		influenced by regional cluster's macro-level configuration and its micro-level governance processes.
	Wies and Moorman (2015)	The influence of going public on firms' innovation: after going public, firms innovate at higher levels and introduce higher levels of variety with each innovation while also introducing less risky innovation, characterized by fewer breakthrough innovations and fewer innovations into new-to-the-firm categories.
How to design a new product? Conjoint analysis		
Foundational papers	Luce and Turkey (1964)	The authors developed procedures for simultaneously measuring the joint effects of two or more variables from rank-ordered data.
	Green and Rao (1971)	The foundational paper about conjoint analysis in marketing.
	Green and Srinivasan (1978)	A review paper: traces the development of conjoint method and discusses the implementation of the method.
	Green and Srinivasan (1990)	An update and extension of their 1978 review of conjoint analysis.
Commercial applications	Wind et al. (1989)	Marriott used conjoint analysis to design Courtyard by Marriott.
	Wittink and Cattin (1989)	Document 1062 conjoint applications in the United States between 1981 and 1985.
	Wittink et al. (1994)	Document 1000 conjoint applications in Europe between 1986 and 1991.
Simple products	Johnson (1974)	Application of conjoint using two-attribute trade off analysis.
	Wind (1973)	Application of conjoint method on full profile ratings.
Multi-attribute products	Shocker and Srinivasan (1974)	LINMAP (LiNear programming techniques for Multidimensional Analysis of Preferences) is used to determine individual consumer's ideal point and salience weights for product attributes.

(continued)

Table 3.2 (continued)

Topics	Literature	Summary
	Green et al. (1981)	For complex products that have many attributes, the authors treat the attributes separately and use data based on conjoint analysis to determine consumer choice. The method is called POSSE (Product Optimization and Selected Segment Evaluation).
	Johnson (1987)	Adaptive conjoint: individuals give initial estimates, which are then revised based on choices between pairs of options purposely selected by the researcher to provide the most useful additional information. This approach was popularized commercially by Sawtooth Software.
	Toubia et al. (2003)	A polyhedral, choice-based design and estimation algorithm for adaptive conjoint analysis. This method converges quickly on a respondent's part-worth utilities with a limited number of questions.
	Toubia et al. (2004)	A new polyhedral choice-based conjoint analysis question-design method.
	Netzer and Srinivasan (2011)	A web-based adaptive self-explicated approach for conjoint analysis of products with ten or more attributes.
Incorporate customer preferences	Hauser and Clausing (1988)	Introduce a planning matrix from Quality Function Deployment (QFD), House of Quality, to relate customer preferences to a firm's design, manufacturing, and marketing.
	Griffin and Hauser (1993)	Incorporate the "Voice of Customer" in order to identify, structure, and provide priorities for customer needs.
	Hoeffler (2003)	In order to mimic consumer coping mechanism when facing uncertainty, the author incorporates both mental simulation and analogies into a standard preference measurement technique (e.g., conjoint analysis) and demonstrates its superior predictive accuracy.
	Coviello and Joseph (2012)	Major innovations are more likely to succeed in small and young technology firms when customers are involved with new product development process.

(continued)

Table 3.2 (continued)

Topics	Literature	Summary
	Kim et al. (2014)	Incorporate the influence of peers in conjoint analysis and demonstrate a PIE (P, the physical product attributes; I, the individual characteristics of the choice maker; E, characteristics of an external peer group) framework of preference, which has superior predictive performance in a conjoint task.
Incorporate multi-media tools to create simulated products	Urban et al. (1996)	Incorporate a multi-media virtual-buying environment to predict consumer response to really-new products. The new method conditions consumers for future situations, simulates user experience, and encourages consumers to actively search for information on the product.
	Urban et al. (1997)	Demonstrate that the internal validity of multi-media stimuli is high and its external validity is comparable with traditional lab stimuli.
Heterogeneous designs	Sandor and Wedel (2005)	First to propose the use of heterogeneous designs where different customers get different designs.
	Liu and Tang (2015)	More efficient heterogeneous choice designs for large number of subdesigns.
	Hauser et al. (2014)	Improve web morphing by incorporating switching costs, potential website exit, and the impact of all clicks to decide the optimal timing of morphing for each customer.

$$p_{sj} = \frac{\exp(x'_{sj}\beta)}{\sum_{j=1}^{J} \exp(x'_{sj}\beta)} \tag{3.4}$$

Recent studies incorporate consumer heterogeneity where consumers have different values for the attributes. Here, a common model is the mixed logit, where β s contain random effects. For example, Sandor and Wedel (2002, 2005) assume that β is multivariate normally distributed with mean μ_B and variance Σ. If Σ is assumed to be a diagonal matrix, β can be written as $\beta = \mu_B + V\sigma_B$ with vector $\sigma_B = (\sigma_1, \ldots, \sigma_m)'$, and V is an $m * m$ diagonal matrix. Assuming utility maximization, the probability that product i is chosen from choice set s is

$$\varphi_{sj} = \int p_{sj}(v)f(v)dv, \text{ where } p_{sj}(v) = \frac{\exp\{x_{sj}^{'}(\mu_B + V\sigma_B)\}}{\sum_{j=1}^{J} \exp\{x_{sj}^{'}(\mu_B + V\sigma_B)\}}, \quad (3.5)$$

where v is the vector containing the m diagonal elements of matrix V.

Determining the optimal product design requires evaluating the information matrix of this mixed logit model. A widely used measure of the information matrix is the inverse of the determinant of the Fisher information matrix, which is called the D-error (Sandor and Wedel 2002).

$$D-error = \det[I(\mu_B, \sigma_B|X]^{-1/2K}, \quad (3.6)$$

where I is the Fisher information matrix and K is the number of attribute level combinations. Minimizing this D-error generates the optimal product design.

One limitation of the mixed logit model is the computation infeasibility when the number of possible designs is too large. To address this limitation, Liu and Tang (2015) propose a particular conjoint choice context, which can be specified as $\{(C_k, w_k)\}, k = 1, \ldots, K$, where C_k is a choice set and w_k is the continuous weight for the choice set with the constraint such that $0 \le w_k \le 1$ and $\sum_{k=1}^{K} w_k = 1$. Instead of searching for the globally optimal product design, this approach finds a globally optimal continuous design by minimizing the D-errors over the entire space of continuous designs. It generates completely heterogeneous designs for each individual respondent in the choice experiment and more importantly, it is computationally feasible.

Conjoint Analysis. Traditional choice-based conjoint analysis relies on consumers' evaluations of product attributes, resulting in the desirability of attribute levels and the importance of each attribute. However, there are difficulties when consumers need to evaluate a large number of product attributes at one time. Recent research extends the traditional stated preference methods and solves this problem by adaptively choosing a subset of attribute comparisons and interpolating the importance of other attributes (Netzer and Srinivasan 2011).

In particular, Netzer and Srinivasan (2011) ask respondents to allocate 100 points across attributes to reflect the importance of each attribute. To alleviate the problem of too many product attributes, they break down the constant-sum allocation across the full set of attributes into a subset of constant-sum allocations between two attributes at a time. In order to reduce the number of paired comparison questions, they propose an adaptive approach where respondents are asked to compare only a subset of all possible paired comparison questions. Then researchers interpolate the importance of the attributes not included in the subset of comparison questions. Meanwhile, researchers can select the attributes for the next comparison to minimize the interpolation errors. Then the importance ratios between two attributes ($r_{j_1 j_2} = W_{j_1}/W_{j_2}$, where W_{j_k} is the importance of attribute j_k) can be calculated from the paired comparisons. With the set of attribute importance ratios, a log-linear multiple regression is used to estimate relative attribute importance. The essence of their approach is to improve the estimation of individual-level attribute importances (W_j) in a way that avoids respondent overload.

3.2.3 Sales Forecasting

Forecasting can be done from the time an idea has been generated until after it has been commercialized. However, most research has focused on forecasting sales at the end of the product design and development stage.

Research on the sales forecasting of new products has been conducted in marketing for several decades. The most well known sales forecasting model in marketing is the Bass model which spawned a huge literature on diffusion models. Conceptually, these models are based on the sociological literature on diffusions of innovation, largely driven through a process of communications (e.g., Rogers 2003). We elaborate on this model in discussing one of its extensions later in this section.

The forecasting literature for non-durables primarily uses stochastic models applied to both first-time purchases and repeat sales. Another class of forecasting models, flow models, incorporate test market results to project initial sales to full-market sales. Regression models have also been used to predict new product sales. Yet another approach combines elements of flow models, regression models, and stochastic approaches to forecast the sales of new consumer nondurables. The well known ASSESSOR model has improved forecasting accuracy because researchers and managers can project laboratory results to market results. The most recent research on sales forecasting has either incorporated online-enabled approaches or focused on forecasting approaches for online environments (e.g., online word-of-mouth, online reviews, virtual markets, sentiment analysis, and blogs; e.g., Chevalier and Mayzlin 2006) (Table 3.3).

3.2.3.1 Sales Forecasting Models

In this section, we provide details on diffusion model extensions and customer lifetime analysis.

Extensions of Diffusion Models. The standard Bass (1969) diffusion model incorporates a hazard rate whereby consumers who have not yet adopted a new product do so at time t:

$$h(t) = p + qF(t), \tag{3.7}$$

where $F(t)$ is the proportion of consumers who have adopted this new product at time t; p captures the intrinsic tendency to adopt (coefficient of innovation), which can be influenced by consumer characteristics, innovation appeal, etc.; and

Table 3.3 Sales forecasting

Topics	Literature	Summary
Forecast with consumer panel statistics (model free)	Fourt and Woodlock (1960)	Customer purchase frequency information can be used to predict the success of grocery products.
Forecast with stochastic processes	Kuehn (1962)	Uses stochastic processes to describe and model new product adoption.
	Massy (1969)	Stochastic evolutionary adoption model (STEAM) uses purchase incidence data to simulate household future purchase and forecast sales.
	Parfitt and Collins (1968)	Use a stochastic model with purchase data to predict the market share for newly launched brands and the market share of established brands after promotions.
	Ehrenberg (1972)	Uses stochastic models and repeat purchase data to predict future sales.
	Schmittlein et al. (1987)	Develop the Pareto/NBD model based on the number and timing of the customers' previous purchase to predict future purchase.
	Fader and Schmittlein (1992)	The predicted sales by the Dirichlet model are usually lower than the actual sales of high-share brands. It is probably because distinct consumer segments favor large brands.
	Fader et al. (2005)	The beta-geometric/NBD (BG/NBD) model is developed to predict the future purchase with easier implementation than Pareto/NBD model and similar results.
	Jerath et al. (2011)	The POD (periodic death opportunity) model relaxes the assumption about customer attrition and thus has better prediction of future sales.
	Bemmaor and Glady (2012)	The Gamma/Gompertz/NBD model is developed to better predict future sales.
Forecast with test market results (Flow models)	Urban (1970)	The SPRINTER model, which is based on the behavioral process of the diffusion of innovation, uses test-market data to predict sales of new frequently purchased consumer products.

(continued)

Table 3.3 (continued)

Topics	Literature	Summary
	Urban (1975)	The PERCEPTOR model, which tracks consumers through states of awareness, trial, purchase, and repurchase, helps estimate the market share for alternate new brand designs.
	Assmus (1975)	The NewProd model traces the number of potential buyers who are at one of the 11 stages of adoption process and predicts the market share for the first year after the product is introduced into the market.
Forecast with regression-based models	Claycamp and Liddy (1969)	The AYER new product model, where both advertising recall and trial purchase are modeled and advertising recall is one factor in the trial purchase equation, generates good predictions based on several months of test market data.
	Blattberg and Golanty (1978)	The Tracker model incorporates awareness and repeat sales with other factors (e.g., advertising, price) and predicts year-end test market sales with early (3-month) test market results. It also helps managers with new product positioning, redesign, and market planning.
	Blackburn and Clancy (1980)	The LITMUS model combines early test market results with survey data to forecast new product sales and provides diagnostic information on a new product's strengths and weaknesses, and feedback on a product's entire marketing mix.
	Pringle et al. (1982)	The NEWS model predicts consumer awareness, trial, repeat purchase, usage, sales, and market share for a new brand.
Forecast with pre-test-market model	Silk and Urban (1978)	The ASSESSOR model which has two parts—an awareness-trial-repeat model and a preference model—uses constant sum preference data to predict sales of new packaged goods before they are test marketed.
	Urban and Katz (1983)	Demonstrate that the ASSESSOR model can predict sales accurately and is commercially viable.

(continued)

Table 3.3 (continued)

Topics	Literature	Summary
Forecast with other factors (e.g., word-of-mouth, observational learning)	Godes and Mayzlin (2004)	Online word-of-mouth is measured. Its dispersion across user communities can explain new TV show ratings.
	Chevalier and Mayzlin (2006)	An improvement in a book's reviews leads to an increase in book sales. Negative reviews have greater impact on sales than positive reviews.
	Liu (2006)	The volume of online word-of-mouth can help explain movie box office revenue.
	Dahan et al. (2011)	Securities trading of concepts (STOC) can measure aggregate consumer preferences on new product concepts.
	Chen et al. (2011)	While negative word-of-mouth has more influence on sales than positive word-of-mouth, positive observational learning (OL) increases sales but negative OL has no effect.
	Iyengar et al. (2011)	Social contagion influences new product adoption through network ties. This influence is moderated by both the recipients' perception of their opinion leadership and the sources' volume of product usage.
	Sonnier et al. (2011)	Sentiment analysis of online communication shows that positive and neutral word-of-mouth help sales and negative word-of-mouth hurts sales.
	Sood et al. (2012)	Develop a model called SAQ (Step And Wait) for predicting the path of technological innovation.
	Sun (2012)	Higher variance of a book's online ratings improves its sales rank when its average rating is low.
	Gopinath et al. (2013)	Opening day movie box office is influenced by prerelease blog volume and advertising, whereas postrelease movie box office is influenced by postrelease blog valence and advertising.
	Tang et al. (2014)	Neutral user-generated content has non-neutral impact on product sales and its impact differs between mixed-neutral and indifferent-neutral user-generated content.

(continued)

Table 3.3 (continued)

Topics	Literature	Summary
	Risselada et al. (2014)	The effects of social influence and direct marketing on high-technology product adoption change over time. The effect of social influence from cumulative adoptions decreases from the product introduction onward, whereas the influence of recent adoptions remains unchanged. The effect of direct marketing also decreases from the product introduction onward.
	Gopinath et al. (2014)	The valence of online word-of-mouth influences sales and its impact increases over time, whereas its volume has no impact. The effect of attribute-oriented advertising on sales decreases faster than emotion-oriented advertising.
	Aral and Walker (2014)	Random experiments on the adoption of a Facebook application demonstrate that the influence of peers on new product adoption is moderated by the tie strength and structural embeddedness of the social network.
	Toubia et al. (2014)	Develop an approach for using individual-level data on social interactions to improve the aggregate penetration forecasts made by diffusion models.
	Kornish and Ulrich (2014)	It is important to predict the success of new products from raw ideas. Ideas that are one standard deviation better have 50% higher sales.

q measures the effect of social contagion (coefficient of imitation). The proportion of new product adoption at time t can be written as

$$f(t) = \frac{dF(t)}{dt} = h(t)[1 - F(t)] = [p + qF(t)][1 - F(t)]. \tag{3.8}$$

The solution of this equation can be used to predict the cumulative penetration of a new product, in other words, the proportion of consumers who have adopted a new product at time t. Specifically,

$$F(t) = \left[1 - \exp\left(-g - \frac{p+q}{t}\right)\right] \Big/ \left[1 + \left(\frac{q}{p}\right)\exp\left(-g - \frac{p+q}{t}\right)\right], \tag{3.9}$$

where g is a location parameter.

One limitation of the original Bass model is that consumers are homogeneous and influenced by the same factors. One extension is the asymmetric influence model (AIM) where two consumer segments, i.e., influentials and imitators differ in the factors that drive their adoption behavior (Van den Bulte and Joshi 2007). The hazard functions for these two segments are

$$\text{For influentials (denoted as 1)}, h_1(t) = p_1 + q_1 F_1(t); \tag{3.10}$$

$$\text{For imitators (denoted as 2)}, h_2(t) = p_2 + q_2[wF_1(t) + (1-w)F_2(t)], \tag{3.11}$$

where w denotes the relative importance that imitators attach to influentials' versus other imitators' adoption behavior $(0 \le w \le 1)$. Assuming the proportion of influentials is θ and the proportion of imitators is $1-\theta$, the overall cumulative market penetration is:

$$F_m(t) = \theta F_1(t) + (1-\theta)F_2(t). \tag{3.12}$$

And the fraction of population adopting at time t is:

$$f_m(t) = \theta f_1(t) + (1-\theta) f_2(t). \tag{3.13}$$

From Eqs. 3.6 and 3.7, the population hazard function can be derived as

$$h_m(t) = f_m(t)/[1 - F_m(t)] = [\theta f_1(t) + (1-\theta) f_2(t)]/(1 - F_m(t)). \tag{3.14}$$

As individual consumer level social interaction data become available, researchers have extended the Bass model by incorporating data such as social ties and new product recommendations among consumers (Toubia et al. 2014). Assume that consumer i receives r_{it} recommendations in period t. r_{it} is assumed to follow a binomial distribution specified as:

$$\begin{aligned} & r_{it} \sim Bin(ties_i, aF_{t-1}) \\ & \Rightarrow P(r_{it}|ties_i) = \binom{ties_i}{r_{it}} (aF_{t-1})^{r_{it}} (1 - aF_{t-1})^{ties_i - r_{it}} \end{aligned} \tag{3.15}$$

where $ties_i$ is the number of consumer i's social ties, F_{t-1} is the cumulative penetration in period $t-1$, and a is the probability that a given tie would recommend the product to consumer i conditional on the tied consumer having adopted.

The hazard rate is influenced by the social recommendation r_{it} and is specified as

$$h(r_{it}) = 1 - (1-p)(1-q)^{r_{it}} \tag{3.16}$$

where p and q have similar interpretation as in the Bass model.

Toubia et al. (2014) show that the aggregate diffusion process can be derived from Eqs. 3.9 and 3.10. Let $P(ties)$ be the probability mass function of the number of social ties, and let f_t^{ties} and F_t^{ties} be the marginal and cumulative aggregate penetration in period t among consumers with ties. The marginal penetration f_t^{ties} is equal to the proportion of non-adopters, $1 - F_t^{ties}$, multiplied by the expected value of the hazard rate in period t among these consumers:

$$\begin{aligned} f_t^{ties} &= \left(1 - F_{t-1}^{ties}\right) E_{r_t}[h(r_t)|ties] \\ &= \left(1 - F_{t-1}^{ties}\right) \sum_{r_t=0}^{ties} h(r_t) P(r_t|ties) \end{aligned} \tag{3.17}$$

The Bass model can be further extended by incorporating the number of consumer recommendations (Toubia et al. 2014).

Extensions of Customer Lifetime Models. Another common model used to predict sales is the customer lifetime model, where statistical models predict how long a customer will stay with a company and how much he will buy (i.e., purchase rate). For more information about customer lifetime models and related issues, we refer to Chap. 10 of this Handbook: "Marketing Models for the Customer-Centric Firm" by Ascarza, Fader, and Hardie. The most influential model in this stream is the Pareto/NBD model (Schmittlein et al. 1987). Here, the time at which a customer becomes "dead" (i.e., no longer buy from a company) is denoted τ. For any time $T > 0$, if the customer is still alive at T (so $\tau > T$), the number of purchases (x) in $(0, T]$ is assumed to follow the Poisson distribution:

$$P[X = x|\lambda, \tau > T] = e^{-\lambda T} \frac{(\lambda T)^x}{x!}; \quad x = 0, 1, 2, \ldots, \tag{3.18}$$

where the purchase rate λ is assumed to follow a gamma distribution:

$$\mathrm{g}(\lambda|\mathrm{r}, \alpha) = \frac{\alpha^{\mathrm{r}}}{\Gamma(\mathrm{r})} \lambda^{\mathrm{r}-1} \mathrm{e}^{-\alpha\lambda}; \quad \lambda > 0; \quad r, \alpha > 0. \tag{3.19}$$

The time τ until becoming "dead" is assumed to follow an exponential distribution:

$$f(\tau|\mu) = \mu e^{-\mu\tau}; \quad \tau > 0, \tag{3.20}$$

where the death rate μ is also assumed to follow a gamma distribution:

$$h(\mu|s, \beta) = \frac{\beta^s}{\Gamma(s)} \mu^{s-1} e^{-\beta\mu}; \quad \mu > 0; \; s, \beta > 0. \tag{3.21}$$

With these equations, the purchases made while a customer is "alive" follow the NBD model and have the distribution:

$$P[X = x|r, \alpha, \tau > T] = \binom{x + r - 1}{x} \left(\frac{\alpha}{\alpha + T}\right)^r \left(\frac{T}{\alpha + T}\right)^x; \quad x = 0, 1, 2, \ldots \tag{3.22}$$

"Deaths" for a sample of customers follow the Pareto distribution:

$$f(\tau|s,\beta) = \frac{s}{\beta}\left(\frac{\beta}{\beta+\tau}\right)^{s+1}, \quad r > 0 \tag{3.23}$$

Overall, this combined purchase event/duration model is called the Pareto/NBD. One challenge with this model is its complicated likelihood function and numerous evaluations of the Gaussian hypergeometric function. To address this limitation, Fader et al. (2005) develop a beta-geometric/NBD (BG/NBD) model, a variation of the Pareto/NBD, which is easier to implement. Similar to the Pareto/NBD model, the BG/NBD model assumes that the number of purchases follows the Poisson distribution and the purchase rate follows the gamma distribution. The difference between the Pareto/NBD and BG/NBD is how/when customers become inactive. The Pareto/NBD assumes that customers can "die" at any point in time, independent of the occurrence of actual purchases, whereas the BG/NBD assumes that customers "die" immediately after a purchase. Specifically, the BG/NBD assumes that after any purchase, a customer becomes "dead" with a probability p which follows a beta distribution with p.d.f:

$$f(p|a,b) = \frac{p^{a-1}(1-p)^{b-1}}{B(a,b)}, \quad 0 \le p \le 1, \tag{3.24}$$

where $B(a, b)$ is the beta function, which can be written as $B(a,b) = \Gamma(a)\Gamma(b)/\Gamma(\mathrm{a,b})$.

The point at which the customer "dies" is distributed across purchases according to a (shifted) geometric distribution with p.m.f

$$\begin{aligned} &P(\text{inactive immediately after } j\text{th transaction}) \\ &\quad = p(1-p)^{j-1}, \quad j = 1, 2, 3, \ldots \end{aligned} \tag{3.25}$$

Fader et al. (2005) show that the BG/NBD model can be estimated easily in Microsoft Excel and therefore is usable in most business applications.

Researchers also extend the Pareto/NBD model by making it more flexible and more powerful for sales prediction. For example, Bemmaor and Glady (2012) develop the gamma/Gompertz/NBD (G/G/NBD) model. Similar to the Pareto/NBD model, the G/G/NBD model assumes that the number of purchases follows the Poisson distribution and the purchase rate follows the gamma distribution. The difference between the Pareto/NBD and G/G/NBD is that the probability that a customer dies before time τ is assumed to follow a Gompertz distribution:

$$F(\tau|\eta) = 1 - \exp\left(-\eta\left(e^{b\tau} - 1\right)\right), \quad \eta, b > 0, \tau > 0 \tag{3.26}$$

Compared with the Pareto/NBD model, the G/G/NBD model is more flexible because the p.d.f. of the Gompertz distribution can be skewed left or right and it can exhibit a mode at zero or an interior mode. Bemmaor and Glady (2012) also show that the G/G/NBD predicts sales better than the Pareto/NBD model.

3.2.4 Commercialization

The final stage of the new product development process is commercialization. This stage has received much attention in marketing because it lies at the nexus of firm strategy and consumer response to new products (e.g., Boulding and Christen 2003; Golder and Tellis 1993; Min et al. 2006; Robinson and Fornell 1985). Research on commercialization can be categorized into several areas.

First, there is a large and rich literature on entry timing strategies. Initial research in this area found that market pioneers or first movers enjoyed substantial advantages over later entrants. However, this research suffered from several limitations including survivor bias and misclassifying successful firms. Correcting for these limitations resulted in the finding that market pioneers tend to have much higher failure rates, lower market shares, and lower rates of market leadership than previously believed.

Second, another stream of research has attempted to identify the factors associated with new product success. These include differentiation (both meaningful and seemingly meaningless) (Carpenter et al. 1994), introducing innovative product attributes (Shankar et al. 1998), and market characteristics like network effects (Wang et al. 2010). A related stream of research looks specifically at the marketing mix variables associated with new product success (e.g., Bruce et al. 2012; Kopalle and Lehmann 2006; Spann et al. 2015).

Finally, another stream of research examines how incumbents defend against new entrants. Much of this work is analytical (theoretical) in nature (e.g., Hauser and Shugan 1983). Some of these papers make strong prescriptive recommendations of strategies for firms to follow. Surprisingly, some empirical research shows that firms actually do very little to respond to competitors' innovations (Table 3.4).

3.2.4.1 Modeling Commercialization

In this section, we provide details on models of channel acceptance and customer lifetime value.

Channel acceptance of new products. When a new product is introduced to a market, there are two major challenges: how to position it so that consumers will choose it over alternative products and how to convince retailers to accept it. Luo et al. (2007) develop an approach to positioning and pricing a new product that directly incorporates the consumers' preferences and retailer's acceptance criteria.

The authors propose two frameworks of market estimation before and after a new product's entry (see Figs. 3.1 and 3.2). In these two frameworks, consumers' preferences are first estimated with a random utility choice model for a

Table 3.4 Commercialization

Topics	Literature	Summary
Entry timing	Robinson and Fornell (1985)	Market pioneers have higher market shares because of firm-based superiority (better marketing mix and more cost savings) and consumer information advantages.
	Urban et al. (1986)	Market pioneers have higher market shares across 24 categories.
	Robinson (1988)	Market pioneers have higher market shares in industrial goods industries because of their stronger products compared with competitors' products and the characteristics of the industrial goods industries.
	Carpenter and Nakamoto (1990)	Market pioneers have advantages in acculturating consumers' preferences for the pioneer rather than for later entrants.
	Kalyanaram and Urban (1992)	Later entrants suffer a long-term market disadvantage in 8 categories of consumer packaged goods.
	Golder and Tellis (1993)	A historical analysis shows that 47% of market pioneers fail and their market share is overstated in the literature because of a survival bias whereby failed pioneers are not included in the data sets and successful later entrants are misclassified as pioneers.
	Kalyanaram et al. (1995)	Main conclusions: (1) for consumer packaged goods, order of market entry has stronger negative relationship with trial penetration than with repeat purchase; (2) pioneers have broader product lines than late entrants; (3) skill and resource profiles differ across market pioneers, early followers, and late entrants; and (4) order of market entry is not related to long-term survival rates.
	Narasimhan and Zhang (2000)	Firm with a larger pioneering premium may choose to wait, while a firm with a smaller pioneering premium speeds to the market.
	Bohlmann et al. (2002)	Pioneers are better off in product categories where consumers value variety, whereas pioneers are worse off in categories where consumers value quality.
	Boulding and Christen (2003)	Market pioneers may suffer a long term profit disadvantage because of greater average cost.
	Min et al. (2006)	Pioneers have first-mover advantages with incremental innovations, but they are likely to fail with a really new product.

(continued)

Table 3.4 (continued)

Topics	Literature	Summary
	Boulding and Christen (2008)	Market pioneers benefit from two cost advantages—experience curve effects and preemption of input factors, while they suffer from three cost disadvantages—imitation, vintage effects, and demand orientation.
	Wang et al. (2010)	In markets with strong network effects, pioneer survival advantage occurs when their product is cross-generation compatible but within-generation incompatible. In contrast, in markets with weak network effects, pioneer survival advantage is likely to occur when their product is cross-generation incompatible but within-generation compatible.
Strategies to succeed	Carpenter and Nakamoto (1990)	Examine optimal positioning, advertising, and pricing strategies for a late entrant and conclude that a differentiated strategy is optimal and a "me-too" positioning is often sub-optimal.
	Carpenter et al. (1994)	Meaningless differentiation can result in a meaningfully differentiated brand.
	Montoya-Weiss and Calantone (1994)	Meta-analysis of the factors that contribute to the success of new products.
	Nowlis and Simonson (1996)	Introducing new product features adds substantial value and increases brand choice.
	Shankar et al. (1998)	For a later entrant, being innovative can create advantages with higher market potential and a higher repeat purchase rate than either the pioneer or non-innovative late entrant.
	Henard and Syzmanski (2001)	Meta-analysis of the factors that contribute to the success of new products.
	Haenlein and Libai (2013)	Demonstrate that when targeting potential adopters of a new product, firms should target customers with high lifetime value, or "revenue leaders".
Marketing mix in new market	Horsky and Nelson (1992)	An analysis of the optimal positioning and pricing strategy of a new brand using game theory.
	Cooper (2000)	An approach to marketing planning for radically new products.
	Kopalle and Lehmann (2006)	Examine optimal advertised quality, actual quality, and price for a firm entering a market. They found that it is optimal to overstate quality when customers rely relatively less on advertising to form quality expectations and

(continued)

Table 3.4 (continued)

Topics	Literature	Summary
		customers' intrinsic satisfaction with a product is high.
	Hitsch (2006)	Explore optimal entry and exit policy when demand of a new product is uncertain.
	Luo et al. (2007)	Develop an approach to positioning and pricing a new product that incorporate the retailer's acceptance criteria into the development process.
	Narayanan and Manchanda (2009)	Examine the optimal allocation of marketing communication across consumers and over time for the launch of a new product.
	Bruce et al. (2012)	Study the dynamic effects of advertising and word-of-mouth on demand for new products at different stages. They found that increased advertising is more effective at an earlier stage and increased word-of-mouth is more effective at a later stage.
	Spann et al. (2015)	Analyze dynamic pricing strategies in the introduction and early growth phases of new products.
Strategies to defend against a new entrant	Hauser and Shugan (1983)	Analyze how a firm should adjust its marketing expenditures and price to defend its position when attacked by a competitive new product.
	Robinson (1988)	Incumbents' most common response to a new entrant is either no response or only a single reaction with one marketing variable.
	Bowman and Gatignon (1996)	Investigate the influence of order-of-entry on the effectiveness of a firm's marketing mix and found that late entry reduces sensitivity to price, promotion, and quality.
	Gatignon et al. (1997)	Empirically examine the effectiveness of different defensive strategies against new product entry and found that faster reaction is more successful whereas greater breadth of reaction (number of marketing mix variables used) is less successful.
	Roberts et al. (2005)	Develop a model to set an incumbent's defensive marketing strategy prior to a new entrant's launch.
Strategies to shift away from a failing new product	Boulding et al. (1997)	Managers remain committed to a new product launch even when confronted with strong evidence of failure. This commitment is lessened by precommitment to a predetermined stopping rule or introducing a new decision maker.

(continued)

Table 3.4 (continued)

Topics	Literature	Summary
	Biyalogorsky et al. (2006)	Develop and test a conceptual framework that explains why new product managers maintain or escalate their commitment to a failing new product. They argue that it may not be possible to eliminate commitment bias at the individual level, and that organizational processes must be used instead.
The outcomes of innovations	Gielens (2012)	Explore when and to what extent new products change national brands' market position.
	Rubera and Kirca (2012)	A meta-analysis of the effect of firm innovativeness on its value and financial position.
	Dotzel et al. (2013)	Examine the determinants of service innovativeness and its interrelationships with firm-level customer satisfaction, firm value, and firm risk.
	Rubera (2015)	Empirically study the effect of design innovativeness (i.e., the degree of novelty in a product's design) on new product sales' evolution.

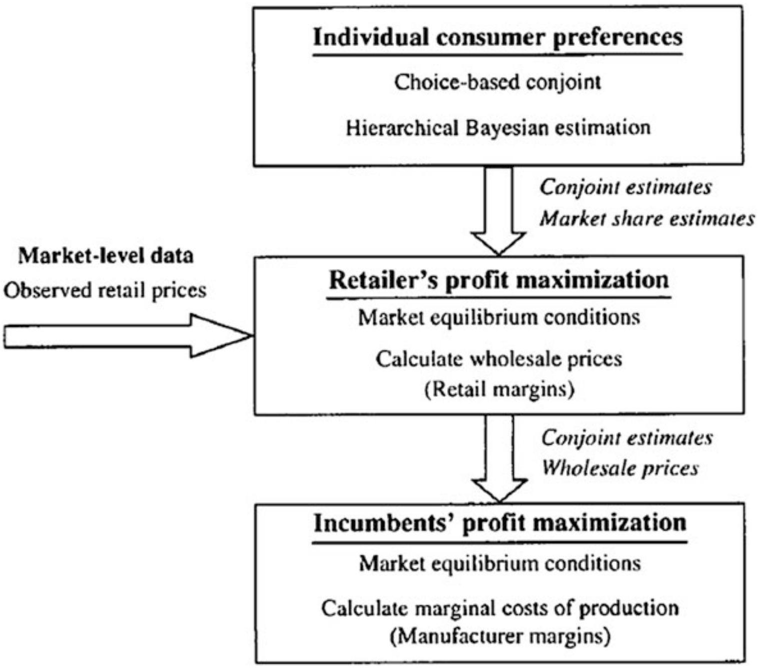


Fig. 3.1 Estimation of market specifics—before entry. *Source* Luo et al. (2007)

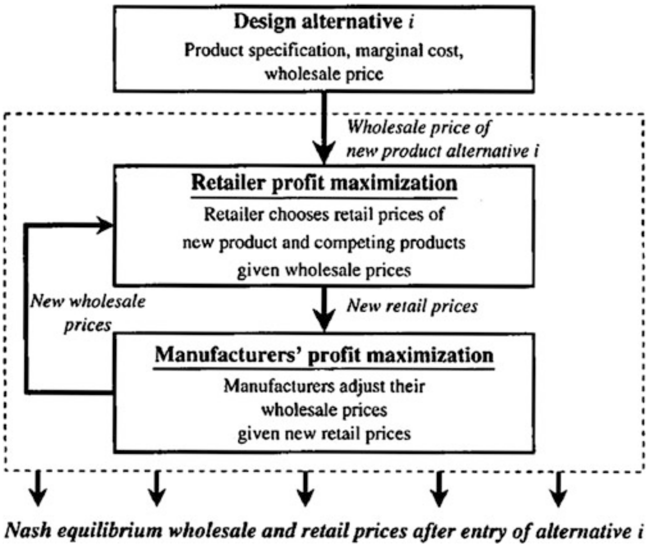


Fig. 3.2 Market scenario development—after entry of design alternative. *Source* Luo et al. (2007)

conjoint choice experiment with N individual consumers evaluating K choice sets with G alternative product designs. The utility of consumer i for product design g in choice k is defined as:

$$U_i(x_{gk}, p_{gk}) = \left(x'_{gk}\beta_{ix} + p_{gk}\beta_{ip}\right) + \varepsilon_{igk}, \quad (3.27)$$

where x_{gk} is a vector of product attributes of design g, p_{gk} is the price, and $\varepsilon_{igk,}$ is the random component of the utility. The probability of consumer i choosing design g can be derived from Eq. (3.1). Specifically, it is expressed using the logit expression.

$$\Pr_{igk} = \frac{\exp(x'_{gk}\beta_{ix} + p_{gk}\beta_{ip})}{\sum_{g'=1}^{G}[\exp(x'_{g'k}\beta_{ix} + p_{g'k}\beta_{ip})] + \exp(a_i)}, \quad (3.28)$$

where a_i is the utility of no-choice option for consumer i.

The next step is to estimate wholesale prices and marginal costs of incumbent products. First, the wholesale prices can be determined by maximizing the retailer's profits. In particular, before a new product is introduced to the market, the retailer's profit maximization is specified as:

$$\max_{p_1, p_2, \ldots, p_J} \pi^r = \left\{\sum_{j=1}^{J}\left[m_j * (p_j - w_j) * S\right]\right\} - sc * J, \quad (3.29)$$

where π^r is the retailer's profit, m_j is product j's market share, w_j is its wholesale price, S is market size, and sc is the marginal shelf cost.

Alternatively, the marginal costs of incumbent products can be estimated by maximizing manufacturer's profits. That is,

$$\max_{w_j} \pi_j^m = \left(w_j - c_j\right) * m_j * S - F_j \quad j - 1, \ldots, J, \tag{3.30}$$

where c_j is product j's marginal cost, and F_j is its fixed cost.

The manufacturer's goal is to select a product design and a wholesale price so that the product will be accepted by retailers, and be more profitable than other designs. As shown in Fig. 3.2, we need to estimate the new market scenario after the entry of the new product by estimating new wholesale and retail prices. The procedure includes solving two optimization problems iteratively: the retail profit maximization problem (second block in Fig. 3.2) and the manufacturer profit maximization problem (third block in Fig. 3.2). The maximization equations are similar to Eqs. 3.29 and 3.30 but with new wholesale and retail prices.

An Application of Customer Lifetime Value (*CLV*). When a new product is introduced to the market, managers can use customer lifetime value (CLV) to identify and target the most profitable customers. Customer lifetime value (CLV) is the present value of all future profits obtained from a customer over his life of relationship with a firm. It is specified as (Gupta et al. 2004; Reinartz and Kumar 2003):

$$CLV = \sum\nolimits_{t=0}^{T} \frac{(p_t - c_t) r_t}{(1+i)^t} - AC \tag{3.31}$$

where p_t is the price paid by the customer at time t, c_t is the direct cost of serving the customer at t, i is the discount rate for the firm, r_t is the probability of a customer being "alive" at time t, AC is the acquisition cost, and T is the time horizon for estimating CLV. If the margin $(p_t - c_t)$ and retention rate are constant over time and the time horizon is assumed to be infinite, CLV can be simplified to (Gupta and Lehmann 2003):

$$CLV = \sum_{t=0}^{\infty} \frac{(p-c) r^t}{(1+i)^t} = m \frac{r}{(1+i-r)} \tag{3.32}$$

Haenlein and Libai (2013) use CLV to identify profitable customers ("revenue leaders"). They argue that targeting "revenue leaders" can accelerate these customers' new product adoption and therefore create an earlier and larger cash flow. More important, these customers can create higher-than-average social value. This effect is due to "network assortativity"—a phenomenon whereby people tend to be connected with others who are like them.

In order to assess the value of "revenue leaders," Haenlein and Libai (2013) use stochastic network-based cellular automata, an ABM (Agent-Based Model) technique, to simulate new product adoption based on local interactions among individual customers. The basic idea of the ABM technique is to start with a social network where no customer has yet adopted the product. Then the utility of buying the product is randomly generated and assigned to each customer and customers whose utility is larger than the product price will adopt the new product. Customers who have adopted the new product influence other customers by word-of-mouth and thus more customers adopt the new product in their social network. Details of the ABM technique are given in Goldenberg et al. (2002). One benefit of using the ABM technique is that it allows researchers to explore the effectiveness of various seeding programs and helps mangers target the most profitable customers. In Haenlein and Libai (2013), the value created by a seeding program is:

Total value = Direct value (i.e., new product adoptions by customers who are seeded) + Social value (i.e., new product adoptions by customers who are connected with the seeded ones) − Cost of the seeding program. Haenlein and Libai (2013) demonstrate that targeting "revenue leaders" is more profitable than targeting "opinion leaders."

3.3 Future Research Opportunities

While prior research has contributed much to our understanding of innovation and new products, many opportunities remain to fulfill the potential in this important area of research. We begin by discussing key areas of emphasis that transcend particular stages of the product development process.

One key topic of future research should be to focus more on metrics than models. Prior research has already done much to develop useful models. However, going forward, developing appropriate metrics for firms to use systematically over time offers great potential benefits. With these metrics, researchers would be able to enhance our understanding of new product development, show firms how they should reduce the inherent inefficiencies, and help them deliver successful innovations on a more regular basis.

A second key topic of future research is business model innovation. Nearly all marketing research use the new product as the level of analysis. However, the product is only one part of the overall offering delivered to customers. Firms can grow through other elements of the marketing mix (i.e., pricing, communication, channels) and other aspects of their business model (e.g., financing, sourcing, partnering), and are influenced by actions that lie outside the firms' control.

A third key topic of future research is to more thoroughly document the process of generating and commercializing the most innovative new products. Currently, we mainly know anecdotes and selected pieces of complete innovation success stories. The first step in repeating these successes is to at least thoroughly understand how they occurred in the past and to see what differentiates them from failures.

Next, we outline some more specific questions for future research. We organize these using the same four stages of new products research that we used to organize prior research: (i) opportunity identification, (ii) product design and development, (iii) sales forecasting, and (iv) commercialization.

3.3.1 Opportunity Identification

The following research questions are most important to address in the area of opportunity identification:

1. How should firms identify the most relevant lead users?
2. Which lead users are predictive of the future preferences of the general consumer market?
3. How can online platforms (e.g., user groups, Facebook, snapchat, and Instagram) be used to identify potential opportunity areas, generate new product ideas, and screen those ideas?
4. What are the best approaches for generating or moderating business-to-consumer communications and consumer-to-consumer communications?
5. When and how are new technologies incorporated into new products? Do these new technologies lead or lag firms' efforts to identify new opportunities?

3.3.2 Product Design and Development

Research questions important to address in the area of product design and development include:

1. How do firms document and learn from failures during the new product development process? How should they do this?
2. How should firms make use of online platforms to design and develop their new products? What are the best ways to involve consumers at various points during design and development? What are the downsides of doing so?
3. What are the similarities and differences in designing and developing new products versus new services versus integrated products and services?
4. When is it appropriate for firms to rely on product champions versus cross-functional teams?
5. How should intrapreneurship be encouraged within organizations? How should firms fund and reward innovators? When and how should firms use think tanks or skunk works?
6. How and when should firms pursue joint development projects with other firms or with potential customers (especially business customers)?

7. When and how should firms abandon new products during design and development?
8. When and how should firms incorporate new technologies into new products?
9. When and what should new products borrow from past products, e.g., to be compatible with behavior or expectations?
10. Are portfolio approaches useful for managing risks in new product development projects?

3.3.3 Sales Forecasting

The following research questions are important to address in the area of sales forecasting:

1. How should firms use online platforms and social media (e.g., Facebook, Instagram, snapchat, etc.) to forecast sales?
2. How should the use of these platforms vary for business-to-business versus business-to-consumer products?
3. How can sales forecasting techniques be more diagnostic by decomposing the overall sales forecasts into the various elements of each new product or service offering?
4. What testing techniques provide better information about ultimate market acceptance earlier in the product design and development process?
5. How can firms generate better estimates of cannibalization across their product lines?
6. How do forecasts themselves impact strategy and success?

3.3.4 Commercialization

While this stage is typically ignored by academics, it is often the most critical one. Important questions to address in this area include:

1. How should firms document, learn from, and apply lessons learned from failed new product launches?
2. How should firms use social media to promote new product launches? How should these efforts differ between business and consumer markets?
3. What is the role of opinion leaders in markets with high social media activity?
4. Who are the opinion leaders in markets with high social media activity and how do they differ from opinion leaders in markets with low social media activity?
5. When and how should firms kill new products after commercialization?

6. What does the concept of relative product advantage really mean? How is it measured? How much does it contribute to a new product's success?
7. How do new categories obtain their names? Should firms be more proactive about promoting new category names?
8. How important are informational cascades in driving new product adoption?
9. What are the contextual factors that determine when being fast to market is more or less important? What are the differences between incremental innovations and radical innovations?
10. What is the appropriate scale of entry for new products? What are the factors that determine when it should be large or small?
11. How should firms manage consumer disadoption and disposal?

3.4 Conclusion

This review has highlighted some key marketing research on innovation and new products. Where we provided less extensive coverage, we refer readers to other useful references. For each of the four stages of the new product development process (opportunity identification, product design and development, sales forecasting, and commercialization), we organize literature by sub-topics within each of these stages. This hopefully will give readers a good sense of the state-of-the-art in each of these research areas. We also provide thoughts on some important research to conduct going forward. Overall, much has been learned already. Nonetheless, given the importance of new product innovation to firms, to individuals, and to societies, we hope that tomorrow's researchers continue to generate newer and richer insights in this vitally important field of investigation.

References

Aral, Sinan, and Dylan Walker. 2014. Tie strength, embeddedness, and social influence: A large-scale networked experiment. *Management Science* 60 (6): 1352–1370.

Assmus, G. 1975. NEWPROD: The design and implementation of a new product model. *Journal of Marketing* 39 (1): 16–23.

Bass, Frank M. 1969. A new product growth model for consumer durables. *Management Science* 15 (5): 215–227.

Bayus, Barry L. 2013. Crowdsourcing new product ideas over time: An analysis of the Dell IdeaStorm community. *Management Science* 59 (1): 226–244.

Bemmaor, Albert C., and Nicolas Glady. 2012. Modeling purchasing behavior with sudden 'Death': A flexible customer lifetime model. *Management Science* 58 (5): 1012–1021.

Blackburn, J.D., and K.J. Clancy. 1980. LITMUS: A new product planning model. In *Proceedings: Market measurement and analysis,* ed. Robert P. Leone, 182–193. Providence, R.I.: The Institute of Management Sciences.

Blattberg, R., and J. Golanty. 1978. Tracker: An early test market forecasting and diagnostic model for new product planning. *Journal of Marketing Research* 15 (2): 192–202.

Biyalogorsky, E., W. Boulding, and R. Staelin. 2006. Stuck in the past: Why managers persist with new product failures. *Journal of Marketing* 10 (2): 108–121.

Bohlmann, Jonathan D, Peter N. Golder, and Debanjan Mitra. 2002. Deconstructing the pioneer's advantage: Examining vintage effects and consumer valuations of quality and variety. *Management Science* 48 (9): 1175–1195.

Borah, Abhishek, and Gerard J. Tellis. 2014. Make, buy, or ally? choice of and payoff from announcements of alternate strategies for innovations. *Marketing Science* 33 (1): 114–133.

Boulding, W., and M. Christen. 2003. Sustainable pioneering advantage? Profit implications of market entry order. *Marketing Science* 22 (3): 371–392.

Boulding, William, and Markus Christen. 2008. Disentangling pioneering cost advantages and disadvantages. *Marketing Science* 27 (4): 699–716.

Boulding, W., R. Morgan, and R. Staelin. 1997. Pulling the plug to stop the new product drain. *Journal of Marketing Research* 34 (1): 164–176.

Bowman, D., and H. Gatignon. 1996. Order of entry as a moderator of the effect of the marketing mix on market share. *Marketing Science* 15 (3): 222–242.

Boyd, D., and J. Goldenberg. 2013. *Inside the box*. New York, NY: Simon and Schuster.

Bruce, Norris I., Natasha Zhang Foutz, and Ceren Kolsarici. 2012. Dynamic effectiveness of advertising and word of mouth in sequential distribution of new products. *Journal of Marketing Research* 49 (4): 469–486.

Burroughs, James E., Darren W. Dahl, C. Page Moreau, Amitava Chattopadhyay, and Gerald J. Gorn. 2011. Facilitating and rewarding creativity during new product development. *Journal of Marketing* 75 (4): 53–67.

Carpenter, G.S., R. Glazer, and K. Nakamoto. 1994. Meaningful brands from meaningless differentiation: The dependence on irrelevant attributes. *Journal of Marketing Research* 31 (3): 339–350.

Carpenter, G.S., and K. Nakamoto. 1990. Competitive strategies for late entry into a market with a dominant brand. *Management Science* 36 (10): 1268–1278. Focused Issue on the State of the Art in Theory and Method in Strategy Research.

Chandy, R.K., and G.J. Tellis. 1998. Organizing for radical product innovation: The overlooked role of willingness to cannibalize. *Journal of Marketing Research* 35 (4): 474–487.

Chen, Yubo, Qi Wang, and Jinhong Xie. 2011. Online social interactions: A natural experiment on word-of-mouth versus observational learning. *Journal of Marketing Research* 48 (2): 238–254.

Chevalier, J.A., and D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* 43 (3): 343–354.

Claycamp, H.J., and L.E. Liddy. 1969. Prediction of new product performance: An analytical approach. *Journal of Marketing Research* 6 (4): 414–420.

Cooper, L.G. 2000. Strategic marketing planning for radically new products. *Journal of Marketing* 64 (1): 1–16.

Cooper, R.G. 1990. Stage-gate systems: A new tool for managing new products. *Business Horizons* 33 (3): 44–54.

Cooper, R.G. 1994. Perspective third-generation new product processes. *Journal of Product Innovation Management* 11 (1): 3–14.

Coviello, Nicole E., and Richard M. Joseph. 2012. Creating major innovations with customers: Insights from small and young technology firms. *Journal of Marketing* 76 (6): 87–104.

Cui, Anna S., and Gina O'Connor. 2012. Alliance portfolio resource diversity and firm innovation. *Journal of Marketing* 76 (4): 24–43.

Dahan, E., A.J. Kim, A.W. Lo, T. Poggio, and N. Chan. 2011. Securities trading of concepts (STOC). *Journal of Marketing Research* 48 (3): 497–517.

Ding, M., and J. Eliashberg. 2002. Structuring the new product development pipeline. *Management Science* 48 (3): 343–363.

Dotzel, Thomas, Venkatesh Shankar, and Leonard L. Berry. 2013. Service innovativeness and firm value. *Journal of Marketing Research* 50 (2): 259–276.

Ederer, Florian, and Gustavo Manso. 2013. Is Pay for Performance Detrimental to Innovation? *Management Science* 59 (7): 1496–1513.

Ehrenberg, A.S.C. 1972. *Repeat-buying: Theory and application*. Amsterdam: North Holland Press.

Fader, P.S., B.G.S. Hardie, and K.L. Lee. 2005. Counting your customers the easy way: An alternative to the Pareto/NBD model. *Marketing Science* 24 (Spring): 275–284.

Fader, P.S., and D.C. Schmittlein. 1992. Excess behavioral loyalty for high-share brands: Deviations from the Dirichlet model for repeat purchasing. *Journal of Marketing Research* 30 (November): 478–493.

Fisher, J.C., and R.H. Pry. 1971. A simple substitution model of technological change. *Technological Forecasting and Social Change* 3: 75–88.

Fourt, L.A., and J.W. Woodlock. 1960. Early prediction of market success for new grocery products. *Journal of Marketing* 25 (2): 31–38.

Ganesan, S., A.J. Malter, and A. Rindfleisch. 2005. Does distance still matter? Geographic proximity and new product development. *Journal of Marketing* 69 (4): 44–60.

Gatignon, H., T.S. Robertson, and A.J. Fein. 1997. Incumbent defense strategies against new product entry. *International Journal of Research in Marketing* 14 (2): 163–176.

Gielens, Katrijn. 2012. New products: The antidote to private label growth? *Journal of Marketing Research* 49 (3): 408–423.

Godes, D., and D. Mayzlin. 2004. Using online conversations to study word-of-mouth communication. *Marketing Science* 23 (4): 545–560.

Goldenberg, Jacob, David Mazursky, and Sorin Solomon. 1999. Toward identifying the templates of new products: A channeled ideation approach. *Journal of Marketing Research* 36 (2): 200–210.

Goldenberg, J., R. Horowitz, A. Levav, and D. Mazursky. 2003. Finding your innovation sweet spot. *Harvard Business Review* 81 (3): 120–130.

Goldenberg, Jacob, Donald R. Lehmann, and David Mazursky. 2001. The idea itself and the circumstances of its emergence as predictors of new product success. *Management Science* 47 (1): 69–84.

Goldenberg, Jacob, Barak Libai, and Eitan Muller. 2002. Riding the saddle: How cross-market communications can create a major slump in sales. *Journal of Marketing* 66 (2): 1–16.

Golder, Peter N., and Debanjan Mitra (editors). 2017. *Handbook of research on new product development*. Edward Elgar.

Golder, Peter N., Rachel Shacham, and Debanjan Mitra. 2009. Innovations' origins: When, by whom, and how are radical innovations developed? *Marketing Science* 28 (1): 166–179.

Golder, Peter N., and Gerard J. Tellis. 1993. Pioneer advantage: Marketing logic or marketing legend? *Journal of Marketing Research* 30 (May): 158–170.

Gopinath, Shyam, Pradeep K. Chintagunta, and Sriram Venkataraman. 2013. Blogs, advertising, and local-market movie box office performance. *Management Science* 59 (12): 2635–2654.

Gopinath, Shyam, Jacquelyn S. Thomas, and Lakshman Krishnamurthi. 2014. Investigating the relationship between the content of online word of mouth, advertising, and brand performance. *Marketing Science* 33 (2): 241–258.

Green, P.E., J.D. Carroll, and S.M. Goldberg. 1981. A general approach to product design optimization via conjoint analysis. *Journal of Marketing* 45 (Summer): 17–37.

Green, P.E., and V.R. Rao. 1971. Conjoint measurement for quantifying judgmental data. *Journal of Marketing Research* 8 (August): 355–363.

Green, P.E., and V. Srinivasan. 1978. Conjoint analysis in consumer research: Issues and outlook. *Journal of Consumer Research* 5 (2): 103–123.

Green, P.E., and V. Srinivasan. 1990. Conjoint analysis in marketing: New developments with implications for research and practice. *Journal of Marketing* 54 (4): 3–19.

Griffin, A. 1997. PDMA research on new product development practices: Updating trends and benchmarking best practices. *Journal of Product Innovation Management* 4 (6): 429–458.

Griffin, A., and J.R. Hauser. 1993. The voice of the customer. *Marketing Science* 12 (1): 1–27.

Gupta, Sunil, and Donald R. Lehmann. 2003. Customers as assets. *Journal of Interactive Marketing* 17 (1): 9–24.

Gupta, Sunil, Donald R. Lehmann, and Jennifer A. Stuart. 2004. Valuing Customers. *Journal of Marketing Research* 41 (1): 7–18.

Haenlein, Michael, and Barak Libai. 2013. Targeting revenue leaders for a new product. *Journal of Marketing* 77 (3): 65–80.

Hanssens, Dominique M. 2015. *Empirical generalizations about marketing impact*, 2nd ed. Cambridge, MA: Marketing Science Institute.

Hauser, J.R., and D. Clausing. 1988. The house of quality. *Harvard Business Review* 66 (3): 63–73.

Hauser, John R., Guilherme Liberali, and Glen L. Urban. 2014. Website morphing 2.0: switching costs, partial exposure, random exit, and when to morph. *Management Science* 60 (6): 1594–1616.

Hauser, John R., and S.M. Shugan. 1983. Defensive marketing strategies. *Marketing Science* 2 (4): 319–360.

Hauser, John R., Gerard J. Tellis, and Abbie Griffin. 2006. Research on innovation: A review and agenda for marketing science. *Marketing science* 25 (6): 687–717.

Henard, D.H., and D.M. Szymanski. 2001. Why some new products are more successful than others. *Journal of Marketing Research* 38 (3): 362–375.

Hitsch, G.A. 2006. An empirical model of optimal dynamic product launch and exit under demand uncertainty. *Marketing Science* 25 (1): 25–50.

Hoeffler, Steve. 2003. Measuring preferences for really new products. *Journal of Marketing Research* 40 (4): 406–420.

Homburg, Christian, Martin Schwemmle, and Christina Kuehnl. 2015. New product design: Concept, measurement, and consequences. *Journal of Marketing* 79 (3): 41–56.

Horsky, D., and P. Nelson. 1992. New brand positioning and pricing in an oligopolistic market. *Marketing Science* 11 (2): 133.

Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente. 2011. Opinion leadership and social contagion in new product diffusion. *Marketing Science* 30 (2): 195–21.

Jerath, Kinshuk, Peter S. Fader, and Bruce G.S. Hardie. 2011. New perspectives on customer 'death' using a generalization of the Pareto/NBD model. *Marketing Science* 30 (5): 866–880.

Johnson, R.M. 1974. Tradeoff analysis of consumer value. *Journal of Marketing Research* 11 (2): 121–127.

Johnson, R.M. 1987. *Adaptive conjoint analysis.* Paper presented at the Sawtooth Software Conference.

Kalyanaram, G., W.T. Robinson, and G.L. Urban. 1995. Order of market entry: Established empirical generalizations, emerging empirical generalizations, and future research. *Marketing Science* 14 (3 Supplement): G212–G221.

Kalyanaram, G., and G.L. Urban. 1992. Dynamic effects of the order of entry on market share, trial penetration, and repeat purchases for frequently purchased consumer goods. *Marketing Science* 11 (3): 235–250.

Kim, Hye-Jin, Young-Hoon Park, Eric T. Bradlow, and Min Ding. 2014. PIE: A holistic preference concept and measurement model. *Journal of Marketing Research* 51 (3): 335–351.

Kopalle, P., and D.R. Lehmann. 2006. Setting quality expectations when entering a market: What should the promise be? *Marketing Science* 25 (1): 8–24.

Kornish, Laura J., and Karl T. Ulrich. 2011. Opportunity spaces in innovation: Empirical analysis of large samples of ideas. *Management Science* 57 (1): 107–128.

Kornish, Laura J., and Karl T. Ulrich. 2014. The importance of the raw idea in innovation: Testing the sow's ear hypothesis. *Journal of Marketing* 51 (1): 14–26.

Koukova, Nevena T., P.K. Kannan, and Amna Kirmani. 2012. Multiformat digital products: How design attributes interact with usage situations to determine choice. *Journal of Marketing Research* 49 (1): 100–114.

Kuehn, A.A. 1962. Consumer brand choice as a learning process. *Journal of Advertising Research* 2: 10–17.

Landwehr, R., Aparna A. Labroo, and Andreas Herrmann. 2011. Gut liking for the ordinary: Incorporating design fluency improves automobile sales forecasts. *Marketing Science* 30 (3): 416–429. http://pubsonline.informs.org/action/doSearch?text1=Landwehr%2C+J+R&field1=ContribJan.

Landwehr, Jan R., Daniel Wentzel, and Andreas Herrmann. 2013. Product design for the long run: Consumer responses to typical and atypical designs at different stages of exposure. *Journal of Marketing* 77 (5): 92–107.

Lehmann, Donald R., and Peter N. Golder. 2014. New products research. In *The history of marketing science*, ed. Russell S. Winer and Scott A. Neslin, now publishers.

Lilien, G.L., P.D. Morrison, K. Searls, M. Sonnack, and E.V. Hippel. 2002. Performance assessment of the lead user idea-generation process for new product development. *Management Science* 48 (8): 1042–1059.

Liu, Y. 2006. Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing* 70 (3): 74–89.

Liu, Qing, and Yihui Tang. 2015. Construction of heterogeneous conjoint choice designs: A new approach. *Marketing Science* 34 (3): 346–366.

Luce, D.R., and J.W. Turkey. 1964. Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology* 1: 1–27.

Luo, L., P.K. Kannan, and B.T. Ratchford. 2007. New product development under channel acceptance. *Marketing Science* 26 (2): 149–163.

Luo, Lan, and Olivier Toubia. 2015. Improving online idea generation platforms and customizing the task structure on the basis of consumers' domain-specific knowledge. *Journal of Marketing* (forthcoming).

Ma, Zhenfeng, Tripat Gill, and Ying Jiang. 2015. Core versus peripheral innovations: The effect of innovation locus on consumer adoption of new products. *Journal of Marketing Research* 52 (3): 309–324.

Mahajan, Vijay, Eitan Muller, and Frank M. Bass. 1995. Diffusion of new products: Empirical generalizations and managerial uses. *Marketing Science* 14 (3): G79–G88.

Massy, W.F. 1969. Forecasting the demand for new convenience products. *Journal of Marketing Research* 6 (4): 405–412.

McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in econometrics*, ed. Zarembka, 105–42. New York: Academic Press.

Min, S., M.U. Kalwani, and W.T. Robinson. 2006. Market pioneer and early follower survival risks: A contingency analysis of really new versus incrementally new product-markets. *Journal of Marketing* 70 (1): 15–33.

Montoya-Weiss, M.M., and R. Calantone. 1994. Determinants of new product performance: A review and meta-analysis. *Journal of Product Innovation Management* 11 (5): 397–417.

Moorman, C., and A.S. Miner. 1997. The impact of organizational memory on new product performance and creativity. *Journal of Marketing Research* 34 (1): 91–106.

Narasimhan, C., and Z.J. Zhang. 2000. Market entry strategy under firm heterogeneity and asymmetric payoffs. *Marketing Science* 19 (4): 313–325.

Narayanan, Sridhar, and Puneet Manchanda. 2009. Heterogeneous learning and the targeting of marketing communication for new products. *Marketing Science* 28 (3): 424–441.

Netzer, O. and V. Srinivasan. 2011. Adaptive, self-explication of multiattribute preferences. *Journal of Marketing Research* 48 (1): 140–156.

Nowlis, S.M., and I. Simonson. 1996. The effect of new product features on brand choice. *Journal of Marketing Research* 33 (1): 36–46.

Ofek, E., and M. Sarvary. 2003. R&D, marketing and the success of next generation products. *Marketing Science* 22 (3): 355–370.

Parfitt, J.H., and B.J.K. Collins. 1968. Use of consumer panels for brand-share prediction. *Journal of Marketing Research* 5 (2): 131–145.

Peres, Renana, Eitan Muller, and Vijay Mahajan. 2010. Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing* 27 (2): 91–106.

Pringle, L.G., R.D. Wilson, and E.I. Brody. 1982. News: A decision-oriented model for new product analysis and forecasting. *Marketing Science* 1 (1): 1–29.

Rao, Vithala R. 2014. Conjoint analysis. In *The history of marketing science*, ed. Russell S. Winer and Scott A. Neslin. World Scientific - Now Publishers Series in Business.

Rao, V.R. 2008. Developments in conjoint analysis. In *Handbook of marketing decision models*, 1st ed, ed. Berend Wierenga, 23–53. New York: Springer Science+Business Media.

Reinartz, Werner J., and Vita Kumar. 2003. The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing* 67 (1): 77–99.

Rindfleisch, A., and C. Moorman. 2001. The acquisition and utilization of information in new product alliances: A strength-of-ties perspective. *Journal of Marketing* 65 (2): 1–18.

Risselada, Hans, Peter C. Verhoef, and Tammo H.A. Bijmolt. 2014. Dynamic effects of social influence and direct marketing on the adoption of high-technology products. *Journal of Marketing* 78 (2): 52–68.

Roberts, John H., Charles J. Nelson, and Pamela D. Morrison. 2005. a prelaunch diffusion model for evaluating market defense strategies. *Marketing Science* 24 (1): 150–164.

Robinson, W.T. 1988. Marketing mix reactions to entry. *Marketing Science* 7 (4): 368–385.

Robinson, W.T., and C. Fornell. 1985. Sources of market pioneer advantages in consumer goods industries. *Journal of Marketing Research* 22 (3): 305–317.

Rogers, Everett M. 2003. *Diffusion of innovation*, 5th edn. Free Press.

Rubera, Gaia. 2015. Design innovativeness and product sales' evolution. *Marketing Science* 34 (1): 98–115.

Rubera, Gais, and Ahmet H. Kirca. 2012. Firm innovativeness and its performance outcomes: A meta-analytic review and theoretical integration. *Journal of Marketing* 76 (3): 130–147.

Sándor, Zsolt, and Michel Wedel. 2002. Profile construction in experimental choice designs for mixed logit models. *Marketing Science* 21 (4): 455–475.

Sándor, Zsolt, and Michel Wedel. 2005. Heterogeneous conjoint choice designs. *Journal of Marketing Research* 42 (2): 210–218.

Schmittlein, David C., Donald G. Morrison, and Richard Colombo. 1987. Counting your customers: Who are they and what will they do next? *Management Science* 33 (1): 1–24.

Sethi, R., and Z. Iqbal. 2008. Stage-gate controls, learning failure, and adverse effect on novel new products. *Journal of Marketing* 72 (1): 118–134.

Sethi, Rajesh, Zafar Iqbal, and Anju Sethi. 2012. Developing new-to-the-firm products: The role of micropolitical strategies. *Journal of Marketing* 76 (2): 99–115.

Sethi, R., D.C. Smith, and C.W. Park. 2001. Cross-functional product development teams, creativity, and the innovativeness of new consumer products. *Journal of Marketing Research* 38 (1): 73–85.

Shanker, V., G.S. Carpenter, and L. Krishnamurthi. 1998. Late mover advantage: How innovative late entrants outsell pioneers. *Journal of Marketing Research* 35 (February): 54–70.

Shocker, A.D., and V. Srinivasan. 1974. A consumer-based methodology for the identification of new product ideas. *Management Science* 20 (6): 921–937.

Silk, A.J., and G.L. Urban. 1978. Pre-test-market forecasting of new packaged goods: A model and measurement methodology. *Journal of Marketing Research* 15 (2): 171–191.

Slotegraaf, Rebecca, and Kwaku Atuahene-Gima. 2011. Product development team stability and new product advantage: The role of decision-making processes. *Journal of Marketing* 75 (1): 96–108.

Sonnier, Garrett P., Leigh McAlister, and Oliver J. Rutz. 2011. A dynamic model of the effect of online communications on firm sales. *Marketing Science* 30 (4): 702–716.

Sood, Ashish, Gareth M. James, Gerard J. Tellis, and Ji Zhu. 2012. Predicting the path of technological innovation: SAW vs. Moore, Bass, Gompertz, and Kryder. *Marketing Science* 31 (6): 964–979.

Sood, A., and G.J. Tellis. 2005. Technological evolution and radical innovation. *Journal of Marketing* 69 (3): 152–168.

Spann, Martin, Marc Fischer, and Gerard J. Tellis. 2015. Skimming or penetration? Strategic dynamic pricing for new products. *Marketing Science* 34 (2): 235–249.

Stephen, Andrew T., Peter Pal Zubcsek, and Jacob Goldenberg. 2015. Lower connectivity is better: The effects of network structure on redundancy of ideas and customer innovativeness in interdependent ideation tasks. *Journal of Marketing* (forthcoming).
Sun, Monic. 2012. How does the variance of product ratings matter? *Management Science* 58 (4): 696–707.
Tang, Tanya, Eric Fang, and Feng Wang. 2014. Is neutral really neutral? The effects of neutral user-generated content on product sales. *Journal of Marketing* 78 (4): 41–58.
Toubia, Olivier, Jacob Goldenberg, and Rosanna Garcia. 2014. improving penetration forecasts using social interactions data. *Management Science* 60 (12): 3049–3066.
Toubia, Olivier, John R. Hauser, and Duncan I. Simester. 2004. Polyhedral methods for adaptive choice-based conjoint analysis. *Journal of Marketing Research* 41 (1): 116–131.
Toubia, O., D.I. Simester, J.R. Hauser, and E. Dahan. 2003. Fast polyhedral adaptive conjoint estimation. *Marketing Science* 22 (3): 273–303.
Tracey, Paul, Jan B. Heide, and Simon J. Bell. 2014. Bringing 'Place' back in: Regional clusters, project governance, and new product outcomes. *Journal of Marketing* 78 (6): 1–16.
Urban, G.L. 1970. SPRINTER MOD III: A model for the analysis of new frequently purchased consumer products. *Operations Research* 18 (5): 805–854.
Urban, G. L. 1975. PERCEPTOR: A model for product positioning, *Management Science*, 21 (8): Application Series, 858–871.
Urban, G.L., T. Carter, S. Gaskin, and Z. Mucha. 1986. Market share rewards to pioneering brands: An empirical analysis and strategic implications. *Management Science* 32 (June): 645–659.
Urban, G.L., and J.R. Hauser. 2004. "Listening In" to find and explore new combinations of customer needs. *Journal of Marketing* 68 (2): 72–87.
Urban, G.L., J.R. Hauser, W.J. Qualls, B.D. Weinberg, J.D. Bohlmann, and R.A. Chicos. 1997. Validation and lessons from the field: Applications of information acceleration. *Journal of Marketing Research* 34 (February): 143–153.
Urban, G.L., and E. Von Hippel. 1988. Lead user analysis for the development of new industrial products. *Management Science* 34 (5): 569–582.
Urban, G.L., and G.M. Katz. 1983. Pre-test-market models: Validation and managerial implications. *Journal of Marketing Research* 20 (3): 221–234.
Urban, G.L., B.D. Weinberg, and J.R. Hauser. 1996. Premarket forecasting of really new products. *Journal of Marketing* 60 (1): 47–60.
Van den Bulte, Christophe, and Yogesh V. Joshi. 2007. New product diffusion with influentials and imitators. *Marketing Science* 26 (3): 400–421.
Von Hippel, E. 1986. Lead users: A source of novel product concepts. *Management Science* 32 (7): 791–805.
Wang, Qi, Yubo Chen, and Jinhong Xie. 2010. Survival in markets with network effects: Product Compatibility and order-of-entry effects. *Journal of Marketing* 74 (4): 1–14.
Wies, Simone, and Christine Moorman. 2015. Going public: How stock market listing changes firm innovation behavior. *Journal of Marketing Research* (forthcoming).
Wind, Y. 1973. A new procedure for concept evaluation. *Journal of Marketing* 37 (4): 2–11.
Wind, J., P.E. Green, D. Shifflet, and M. Scarbrough. 1989. Courtyard by Marriott: Designing a hotel facility with consumer-based marketing models. *Interfaces* 19 (1): 25–47.
Wittink, D.R., and P. Cattin. 1989. Commercial use of conjoint analysis: An update. *Journal of Marketing* 53 (3): 91–96.
Wittink, D.R., M. Vriens, and W. Burhenne. 1994. Commercial use of conjoint analysis in Europe: Results and critical reflections. *International Journal of Research in Marketing* 11: 41–52.

Chapter 4
Models for the Financial-Performance Effects of Marketing

Dominique M. Hanssens and Marnik G. Dekimpe

4.1 Introduction

Several of the others chapters of this Handbook focus on models for different aspects of the marketing mix. From a managerial perspective, such models are important at the functional level, such as the optimal deployment of sales resources, the scheduling of promotional activities, or the configuration of new-product features. The logical "end point" of these models is typically an assessment of the sales or market-share lift that can be attributed to the marketing tactic, followed by a profitability assessment.

The current chapter complements this work by focusing on performance criteria that are relevant to the entire enterprise, not just the marketing function. We use *financial* criteria for that purpose, as they provide metrics that are comparable across the marketing mix (an internal criterion), and also relate well to investors' evaluation of the firm (an external criterion). As such, we treat marketing as an investment in customer value creation and communication that ultimately must create shareholder value as well. The mechanism connecting these two has been referred to as the "chain of marketing productivity" (Rust et al. 2004).

It is well known that investor or shareholder value is created by expectations of future cash flows. These cash flows are transformed into a present value by using a discount factor that reflects the risk or volatility around these expectations. Therefore, we argue that *marketing performance models should ultimately relate to*

D.M. Hanssens (✉)
University of California, Los Angeles, Los Angeles, USA
e-mail: dominique.hanssens@anderson.ucla.edu

M.G. Dekimpe
Tilburg University, Tilburg, The Netherlands

M.G. Dekimpe
KU Leuven, Leuven, Belgium

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_4

the creation of these cash flows. This puts a special condition on the models, i.e. the output variable should be intrinsically linked to financial behavior at the firm level. Compared to the vast array of existing marketing models that explore various aspects of customer and competitor behavior (e.g. choice models, game-theoretic models), financial-performance models tend to be structurally simpler, i.e. they typically have fewer constructs and less behavioral detail. On the other hand, the models must account for temporal patterns such as trends and volatility, and for a substantial forward-looking (expectation) component in the data. Not all marketing models are suitable for that purpose. For example, complex models of brand switching and/or variety seeking may be *cash-flow neutral* if, in any given period, the number of in-switchers versus out-switchers remains approximately the same. Such models are not discussed in this chapter.

4.2 Marketing and Cash Flows

Shareholder value is driven by a flow metric, i.c. current and anticipated net (or "free") cash flows. According to Srivastava et al. (1998), marketing can enhance shareholder value in three different ways:

- by increasing the magnitude[1] of the net cash flows (i.e. higher profitability)
- by accelerating the cash flows (i.e. faster profitability)
- by lowering the volatility of the cash flows (i.e. safer profitability)

These impacts are often indirect (Joshi and Hanssens 2010), as marketing's primary role is in creating and stimulating *demand*, which is typically measured by *sales* or *revenues*. Thus, in order to trace marketing's role in net cash-flow generation, we opt to start with models of sales or revenue generation, which are commonly known as market-response models or marketing-mix models (see e.g. Hanssens et al. 2001 for a detailed coverage). Market-response models should then be combined with the *cost* structure of marketing, which may be fixed (e.g. an advertising campaign), variable (e.g. sales commissions),[2] or a combination of both (e.g. the costs of a sales-promotion campaign). Since current accounting standards enforce that marketing actions are expensed, as opposed to capitalized, the profits and cash flows

[1]Higher cash flows are, for example, obtained when marketing helps to acquire additional customers, or convinces consumer to spend more. However, the role of marketing could also be to prevent a decline in sales or cash flows through appropriate defensive actions (see, for example, Gatignon et al. 1997; Hauser and Shugan 1983; Roberts 2005 or Roberts et al. 2005 for discussions on defensive marketing strategies).

[2]In later models, we will also allow for a direct impact of marketing support on financial metric (the "direct route" in the terminology of Joshi and Hanssens 2010).

derived from marketing are equivalent. Note that we make abstraction of possible delays between the booking of revenue and the receipt of payments.[3]

"Marketing investment spending", such as brand-building advertisements and customer-loyalty-building service enhancements, is not recognized as an investment under current accounting rules. Only the fruits of these marketing efforts are recognized in accounting performance measurers, typically with a lag. These "fruits" may include increased unit sales, higher price premiums and/or a higher revenue base (i.e. the portion of revenue that is realized without marketing effort). Thus the task of quantifying the investment qualities of marketing spending relies on tying financial performance data to these spending levels, which requires the skills of a marketing model builder. The first task in this process is making a careful distinction between *stock* and *flow* performance metrics. This distinction, which originated in the system dynamics literature (e.g. Forrester 1961), is between variables representing accumulations (inventories, or stocks) and changes in these accumulations (flows). A stock in and of itself does not produce cash, but it may enable or enhance future cash flows, and thus plays an important indirect role for financial performance.

The chapter is organized as follows. We begin with a review of financial marketing data and performance metrics, and formulate some criteria for the use of such metrics. Next, we investigate in some detail how key performance metrics are related to marketing activities, using different models as needed. First, we describe how marketing can create cash flows, after which we discuss models that capture how the investment community perceives the firm's marketing actions. In the process, we indicate various areas in need of further research, and discuss managerial implications.

4.3 Criteria for Good Performance Metrics

In the spirit of "what you can measure, you can manage", recent years have seen an emergence of marketing performance metrics that help make marketing financially accountable, and that steer marketing resource allocation in a productive direction (see e.g. Ambler 2003). An overview of commonly-used metrics may be found in Fig. 4.1. The figure illustrates that, despite the strategic importance of these metrics, only a subset is routinely reported to the senior levels in the organization. As Farris et al. (2010) point out, firms still use a financial jargon at senior levels, and it will take some time before customer- or marketing-oriented metrics become commonplace.

[3]By accounting definition, "free" or "net" cash flow is operating profit minus investment. Investment is the net change in the firm's capital. However, "marketing induced capital" such as brand equity or customer equity is currently not recognized on the firm's balance sheet. For example, a $20 million investment in a plant or equipment is recognized as an asset, whereas a $20 million advertising campaign for a brand is not.

	U.S. (n = 224)	Japan (n = 117)	Germany (n = 120)	U.K. (n = 120)	France (n = 116)	Overall
Marketing Metric						
Market share	73	57	97	80	90	79
Perceived product/service quality	77	68	84	71	75	77
Customer loyalty/retention	67	56	69	58	65	64
Customer/segment profitability	73	40	74	65	59	64
Relative price	65	48	84	53	63	63
Actual/potential customer/segment lifetime value	32	35	51	32	58	40
Average	64	51	77	60	68	

Fig. 4.1 Percent of firms reporting various metrics to the board. *Source* Barwise and Farley (2003)

When choosing metrics, we start with the objectives of the measurement process. In marketing there are generally two classes of objectives: evaluation of the impact of past marketing actions, and choice of future marketing actions, i.e. resource allocation (Ambler and Roberts 2008). The former is part of the accounting and control function of the firm, and the latter is part of marketing strategy and planning. In addition, Quelch and McGovern (2006) have formulated desirable properties performance metrics should have from a board-room perspective. We expand on their view by focusing on metrics that are usable in a modeling context as well, and thus are helpful for marketing performance evaluation and resource allocation. We propose the following criteria:

- *Financial relevance*. Firms need to create shareholder value, and therefore any intermediate marketing performance metrics (such as market share, customer satisfaction, etc.) must ultimately be tied to that value. According to the well-known Efficient Markets Hypothesis (EMH) in finance, investors are fully and accurately able to incorporate any new information that has value relevance. Provided marketing drives firm performance, new marketing developments could (should) be value relevant.[4]
- *Actionable*. It must be possible, at reasonable cost, to collect data on the performance metric, and to relate it analytically to marketing investments. This is where a number of empirically-tested models from the marketing-science literature are called for, such as models of trial and repeat purchasing, models of the diffusion of innovations, or models on the creation of brand and/or customer equity.

[4]An extensive literature exists whether it is reasonable to assume that the investor reaction mechanism is always accurate and complete (given that many of them are not marketing experts, or because investors may be influenced by persuasive, but potentially misleading, communications by company executives and/or other mediating factors). An in-depth coverage of this debate is beyond the scope of the current chapter. We refer to Srinivasan and Hanssens (2009) for a more extensive coverage.

- *Stable behavior*. Highly volatile metrics are difficult to interpret and manage, and should be avoided where possible.[5] For example, using sufficiently large samples for attitudinal metrics will avoid unduly large sample variation.
- *Reliable long-term guidance*. This is the "leading indicator" aspect of a metric, i.e. are positive movements in the metric indicative of improving health for the brand or firm?

Using these four criteria as a guide, we now turn to marketing models that support various performance metrics. First, we address the *process* perspective, i.e. we describe how marketing can create financial cash flows, along with other antecedents of performance. If a firm understands these causal connections (i.e., marketing evaluation), it is in a stronger position to make productive marketing resource allocation decisions (i.e., marketing planning). However, that does not necessarily imply that the outside world, in particular the investment community, will immediately recognize this know-how (the last arrow in Fig. 4.1). Thus we must also address how investors *perceive* the firm's marketing actions and their impact on its financial outlook. Finally, we make some observations on the linkages between the process and the perception perspective.

4.4 The Process Perspective

4.4.1 The Core Sales-Response Model

We begin with a core sales response model that explains variations in customer demand for the firm's products and services, and which is therefore at the source of cash flow generation. The basic sales response function is the following multiplicative model[6]

$$S_t = e^c\ M_t^\beta\ X_t^\gamma\ Z_t^\delta\ e_t^u, \tag{4.1}$$

where S_t refers to sales or another performance metric in period t (for example, week t), M_t is marketing support in that week, X_t refers to other firm-controlled variables, Z_t corresponds to uncontrollable (environmental) factors, and u_t is an error term. The core response model may be estimated across time periods t, but could also be specified over cross-sectional units $i = 1, \ldots, I$, or both. We expect

[5]This limited-volatility criterion has received less attention in the marketing literature. A notable exception is Fischer et al. (2016), who consider various drivers of (potentially sub-optimal) revenue and cash-flow volatility.

[6]We opt for the multiplicative, rather than the linear, model as our base model as (i) it allows for diminishing returns to scale, (ii) offers direct elasticity estimates, and (iii) automatically allows for interaction effects.

$0 < \beta < 1$ in estimation, a condition which results in diminishing returns to scale, or concavity of response.

The base model (4.1) implies that infinite marketing support results in infinite sales. In practice, however, there will be a limit or *ceiling* to sales, usually determined by prevailing market conditions. While there are other ways to represent concavity (see e.g. Hanssens et al. 2001, pp. 100–102), the multiplicative function is particularly appealing as it recognizes that marketing-mix effects interact with one another, i.e. the marginal sales effect of an incremental marketing dollar depends on the other elements in the equation. In addition, taking logarithms linearizes the model as follows:

$$\ln(S_t) = c + \beta \ln(M_t) + \gamma \ln(X_t) + \delta \ln(Z_t) + u_t, \tag{4.2}$$

making it easily estimable. Finally, the response parameters are readily interpreted as response elasticities, which are helpful in making comparisons and deriving empirical generalizations of marketing impact.[7]

In some cases, the response is S-shaped, i.e. there is a minimum or threshold-level of marketing spend below which there is little or no impact, followed by a range of spending with rapidly increasing sales response. At even higher spending levels (i.e. past the inflection point), the usual diminishing returns appear. The core model (4.1) can readily be extended to an "odds" model that allows for S-shaped response, as demonstrated by Johansson (1979):

$$(S_t - I)/(K - S_t) = e^c M_t^\beta X_t^\gamma Z_t^\delta e_t^u, \tag{4.3}$$

where *I* is the minimum sales level (e.g. the level at zero marketing spend), and *K* is the ceiling level. For example, if sales is expressed in relative terms (e.g. market share), *I* could be set at 0% and *K* at 100%. For marketing response parameters $0 < \beta < 1$, model (4.3) is still concave, but for $\beta > 1$, the function is S-shaped. Johansson (1979) discusses the formal estimation of (4.3) with maximum-likelihood methods, as well as an easy approximation based on ordinary least squares.

For all practical purposes, concavity and S-shape are sufficient functional forms to capture the essence of marketing response.[8] Naturally, the core response model (4.1) will need to be extended in order to accommodate some specific behavioral marketing phenomena. For example, marketing initiatives often impact demand in time periods after the expenditure has ended. Such lagged effects may be incorporated directly by using a dynamic response function in the lag operator L (i.e. $L^k X_t = X_{t-k}$). The response model then generalizes to

[7] See in this respect Hanssens (2015).

[8] We refer to Hanssens et al. (2001) for a review of other functional specifications that have been regularly used in the literature (pp. 94–115), and for a more in-depth discussion on the use of the lag operator (p. 181).

$$S_t = e^c M_t^{\beta(L)} X_t^{\gamma(L)} Z_t^{\delta(L)} e_t^u, \tag{4.4}$$

with $\beta(L) = \beta_0 + \beta_1 L + \beta_2 L^2 + \cdots$, and similarly for the other dynamic parameters. We will discuss additional extensions to the core response model as needed for incorporating different aspects of cash-flow generation of marketing.

4.4.2 Cash-Flow Generation

How does the core response model generate cash flows? Assuming a constant profit-margin, the net cash flows (CF) in period t—excluding non-marketing costs—may be expressed as

$$CF_t = S_t * \text{margin} - M_t \tag{4.5}$$

The Return on the Investment in Marketing M, sometimes referred to as ROMI, is then defined as

$$\text{ROMI} = [CF(M) - CF(M = 0)]/M \tag{4.6}$$

Note that ROMI is a ratio, which is useful for an ex-post assessment of the return of a specific marketing campaign or investment. However, ROMI should *not* be used to determine optimal levels of marketing spending. Doing so will often result in under-investing on marketing, because ROMI typically declines monotonically with higher spending (see Ambler and Roberts 2008; Farris et al. 2015 for an elaboration). Instead, the optimal marketing spend M* may be derived from maximizing the cash-flow function (4.5) based on the response model (4.4):

$$M^* = \left[e^{c'} * \beta(L) * \text{margin}\right]^{1/[1-\beta(L)]}, \tag{4.7}$$

where we have incorporated the effects of other firm-controlled variables X and environmental conditions Z into the adjusted baseline $e^{c'}$ for ease of exposition.

Importantly, the relationship between marketing spending and cash flow generation depends on (i) the natural size (the baseline) of the business, (ii) the productivity of marketing spending β(L), and (iii) the prevailing profit margin. Taken together, they fully determine optimal short-run marketing-resource allocation. At the same time, these determinants are exogenous; for example, it is assumed that more aggressive marketing spending has no impact on either the baseline or marketing effectiveness itself. Thus, the decision rule in (4.7) may be thought of as a harvesting or reactive view of marketing resource allocation.

However, a prevailing belief among practitioners and academics is that well-placed marketing spending not only stimulates sales, but also builds *future assets* for the firm. In order to represent that capability of marketing, we must extend the core response model to account for endogenously created assets that, in turn, will generate future cash flows, as illustrated in Fig. 4.1. This is done by considering *stock metrics* of market performance in addition to cash flows.

4.4.3 Flow and Stock Metrics

The demand or revenue generation process above is naturally expressed as a flow metric. Similarly, flow metrics are used to express the ongoing cost of marketing. For example, a firm may routinely spend \$2 million a month on marketing communications, which result in an incremental \$3 million in gross profits. The net monthly cash flow due to marketing communication would be \$1 million, and the ROMI would be \$1 million/\$2 million = 50% (using Eq. 4.6).

Ideally, these ongoing marketing expenditures will also create beneficial cumulative effects, which would be assessed as *stock* metrics. For example, the cumulative sales of a new technology durable, or installed base, is a stock variable that is instrumental in convincing other users to adopt the product as well. Such a stock generates future cash flows without additional marketing expenditures, which is financially attractive to the firm. Similarly, many attitudinal measures are stock metrics, e.g. the percent of the target market that is aware of a product, or the overall price image of a retail store. Brand equity and customer equity, too, are stock measures. From a financial performance perspective, our task is to *gauge the cash flows that are drawn from these stocks*, *independent of (or on top of) current marketing expense.*

In what follows, we explore how marketing can create or enhance such stock metrics, and how the core response model may be extended to capture these effects. Analytically, this is the case when the *revenue baseline is allowed to change (grow) over time*, i.e. a higher level of firm revenue is obtained independent of current marketing spending. We identify three sources of such expanded baseline revenue:

- *External forces*. making strategic choices that expand the scope of the business, such as tapping new markets, new segments or distribution channels. Other baseline-driving forces are outside firm control, for example rising disposable incomes in the target market or the entry of a new competitor in the category.
- *Experiential quality to the customer*. When the product or service quality is high, the resulting customer satisfaction may increase repeat-purchase rates and/or word of mouth, even without additional marketing investments. This leads to the development of customer equity, i.e. the long-term value of the customer to the firm has increased.
- *Brand equity building*. Higher equity brands tend to have higher baseline sales, all else (including current marketing expenditures) equal (see e.g. Kamakura and Russell 1993). While the sources of brand equity and customer equity may be

very different, their financial outcomes for the firm are similar, i.e. higher baseline revenue.[9]

"Stock" sources of cash flows are inherently long-run oriented, and strategic in nature. For example, a brand's quality reputation among customers tends to lag objective reality by several years, so it takes time for a brand to reap the financial benefits of investments in product quality (Mitra and Golder 2006). Also, once a stock of hard-core loyal customers has been created, revenues will accrue to the firm for an extended period of time. By contrast, the optimal marketing spending rule in (4.7) only impacts current (or short-run) flows, either through improved marketing effectiveness (e.g. a better media-mix allocation), which lifts β(L), or through more aggressive spending, which lifts M. Improving β(L) is the focus of much of the current interest in marketing accountability, as discussed in detail by Ambler (2003). More aggressive spending is naturally limited by the realities of decreasing returns to marketing and competitive reaction. Thus changes in β(L) or M are typically more tactical in nature.

Extending the core response model to account for the "stock building" function of marketing allows for a more complete short-run and long-run accountability of marketing activity. We first discuss two models that explicitly account for this stock building potential of marketing: (i) time-varying baseline models (Sect. 4.4.3.1), and (ii) generalized diffusion models (Sect. 4.4.3.2). Next, we discuss two stock metrics that have received considerable attention in the recent marketing literature: brand equity (Sect. 4.4.3.3) and customer equity (Sect. 4.4.3.4). Finally, we comment on the usefulness of intermediate performance measures (Sect. 4.4.3.5) in financial-performance models.

4.4.3.1 Time-Varying Baseline Models

Srivastava et al. (2005) make the interesting observation that most market response models assess marketing's influence on sales variations above the baseline, but that the baseline itself does not change. The baseline in revenue is an intuitive measure of brand equity, after adjusting for external determinants such as market size, per capita income and competition. Given sufficiently long time-series data, time-varying parameter models may be used to assess the evolution of baselines, and in particular the evolution that can be attributed to past marketing.[10] The following time-varying market response model for brand i at time t, adapted from

[9]Apart from brand and customer equity, one could also consider a third source of market-based assets: channel equity emerging from collaborative channel relationships (see, for example, Srivastava et al. 2006 for a discussion).

[10]Cross-sectional comparisons of brand equity cannot monitor the *formation* of brand strength, only the equilibrium result of the branding process. By contrast, longitudinal data, possibly across several brands or markets, allow us to infer how marketing spending builds brands over time.

Pauwels and Hanssens (2007), and linearized for ease of exposition, captures this process:

$$S_{i,t} = c_{i,t} + {}_k\Sigma\beta_{ki}(L)M_{ki,t} + \varepsilon_{i,t} \tag{4.8}$$

$$c_{i,t} = c_{i,t-1} + {}_k\Sigma\gamma_{ki}(L)M_{ki,t} + \eta_{i,t} \tag{4.9}$$

where the parameters $\beta_{ki}(L)$ measure the standard sales response effects of marketing instrument *k* of the brand (M_{ki}), and the parameters $\gamma_{ki}(L)$ capture the baseline expansion effects of M_{ki} (assuming positive impact). This representation gives rise to the following combinations of demand-generation and brand-building impact of marketing instrument *k* (see also Leeflang et al. 2009):

	γ_{ki} (L) = 0	γ_{ki} (L) > 0
β_{ki} (L) = 0	Ineffective marketing	Marketing builds the brand
β_{ki} (L) > 0	Marketing generates sales (profits)	Marketing generates sales and builds the brand

In the ideal situation, marketing spending offers short-term returns via demand generation, but also builds the brand. In that case, the brand-building effect is a financial bonus or windfall, as the incremental cash flows from demand-stimulation may already be sufficient to generate a positive ROMI (and not just a positive sales effect).

The more ambiguous scenario is where demand generation is insufficient to generate a positive ROMI, however the sustained[11] marketing spending builds the brand in a "cash-flow invisible" way (behavioral explanations for this scenario exist, but are beyond the scope of our chapter). Such a policy would be a true investment in that short-run losses are incurred for the purpose of increasing long-term benefits. Indeed, as time moves on, an increasing portion of revenue accrues to the firm without marketing spending, and that portion is demonstrably related to previous band-building spending.

From an econometric perspective, the dynamic system of (4.8) and (4.9) may be estimated by state-space methods such as the Kalman filter that provide a time path of brand equity (see Naik 2015 for a detailed review). Sriram and Kalwani (2007) used a logit-model version of this approach to demonstrate that sales promotions for orange juice brands lift sales revenue, while at the same time eroding brand equity. In a similar vein, Ataman et al. (2008) used dynamic linear modeling to show how marketing activities can be instrumental in building new brands and in managing existing brands for sustainable growth.

[11]Dekimpe and Hanssens (1999) also discuss the case where temporary (rather than sustained) spending has persistent brand-building effects, a situation they call hysteresis. Hanssens et al. (2016), in turn, identify windows of opportunities for opportunistic "brand-building" expenditures.

4.4.3.2 Generalized Diffusion Models

The notion that marketing expenditures can contribute to an asset or stock which, in turn, generates future cash flows is also reflected in many diffusion models, which may be viewed as special cases of the time-varying baseline models discussed above. The exponential surge in the first-generation sales of consumer and industrial durables such as the fax machine and the iPod cannot be explained by the growth in population or purchasing power alone, nor by the advertising spending patterns in the later stages of the life cycle. Instead, a process of *internal influence* (imitation) from early adopters of the product accounts to a large extent for the exponential growth, even though this imitation effect subsequently dies out as the market reaches maturity. Such a diffusion process can occur spontaneously, but it can also be *accelerated* by marketing spending, as in the model by Bass et al. (1994):

$$S_t = [\text{market size} - Y_{t-1}] * [p_1 + q_1 Y_{t-1} + p_2 * f(M_t) + q_2 * Y_{t-1} * f(M_t)] \quad (4.10)$$

where

Y_{t-1} installed base at the beginning of period *t*, i.e. $S_0 + S_1 + S_2 + \cdots + S_{t-1}$
p_1 the strength of external innovation in the market
p_2 the impact of marketing on innovation
q_1 the degree of imitation in the market
q_2 the impact of marketing on imitation
$f(M)$ the market response function for innovation and imitation, which could be multiplicative, as in Eq. (4.1).

This model is sometimes referred to as the "generalized Bass model", as it expands the basic diffusion model due to Bass (1969), which is obtained by setting $p_2 = q_2 = 0$. The spontaneous growth in sales and cash flows comes from the installed base, or stock of cumulative sales (Y_{t-1}) and the strength of consumer imitation (q_1). This factor is largely responsible for the spectacular growth in revenues and earnings in the first decade of high-technology companies such as Microsoft, Dell and Google.

The cash-flow *acceleration* function of marketing comes from two sources: creating awareness of the new product among innovative prospects, and encouraging imitation among imitative customers. However, overall market size is not affected, so these marketing actions *shift forward* a fixed ultimate demand for the product, which is consistent with the cash-flow-acceleration function of marketing in Srivastava et al. (1998). A recurring example of this form of marketing is in the motion-picture industry, where aggressive pre-launch advertising campaigns are often used to attract viewers to the theaters on opening weekend (Elberse and Eliashberg 2003). Other marketing investments are aimed more at increasing the long-run market potential of the innovation, for example by proposing and communicating new usage situations for the product.

The cash-flow implications from the diffusion of innovations are complex, not only because of the nonlinearities involved, but also because the marketing impact may differ in different stages of the life cycle. In addition, profit margins may change with the learning curve and increased competition. Consider, for example, the study by Horksy and Simon (1983) on the impact of advertising on the diffusion of a new electronic banking service. The authors found that life-cycle cash flows for the bank were maximized by initial aggressive advertising for the new product, and then gradually reducing advertising support over time. Moreover, the installed base (or stock) for one technology may positively influence the diffusion of later generations of that product (see e.g. Norton and Bass 1987) and of complementary products (Srinivasan et al. 2004).

In recent years, the diffusion of innovation literature has benefited from new data sources coming from the internet, in particular the rise of social media. Such data allow researchers to investigate more qualitative and strategic aspects of diffusion, such as the study of different seeding strategies to accelerate diffusion, which is often referred to as "viral marketing". For example, Hinz et al. (2011) conducted large-scale field experiments in the telecommunications sector and found that targeting "hubs" and "bridges" outperformed random seeding strategies by 39–100%. Onishi and Manchanda (2012), in turn, show in two different settings (movies and cellular phone services) how pre-release advertising stock and pre-release blogging activity behave synergistically to speed up and increase post-release performance.

In conclusion, while the basic diffusion model remains, its managerial relevance is much enhanced by the use of both social-media metrics and concepts developed in the social-network (see, for example, the chapter by Chen, Van der Lans and Trusov in this Handbook) and agency-based modelling (see, for example, Rand and Rust 2011) literature.

4.4.3.3 Brand Equity

Perhaps the most frequently-studied stock metric in the marketing literature is the concept of brand equity. Keller and Lehmann (2001) considered three broad classes of brand-equity measures: customer mindset measures, product-market measures and financial-market based measures. An excellent review is given in Ailawadi et al. (2003), which is not repeated here. These authors propose the *revenue premium* as a financially-relevant measure for the value of a brand in a given industry.

The revenue premium is defined as the difference in revenue realized by branded versus unbranded competitors, i.e.

$$\text{Revenue premium} = \text{volume}_{\text{brand}} * \text{price}_{\text{brand}} - \text{volume}_{\text{non-brand}} * \text{price}_{\text{non-brand}} \tag{4.11}$$

This reflects the idea that brand equity may boost sales volume, allow for a price premium, or both. Put differently, brand-building activities may enhance future cash flows as a result of realizing a higher sales volume, and/or a higher price. The measure is shown to be actionable, stable over time, and to have considerable diagnostic value in terms of the brand's long-run health, thereby conforming to our earlier criteria. Interestingly, Ailawadi et al. (2003) also demonstrate how branded products exhibit asymmetric up- and downward (market-share) price elasticities. Using data from a variety of consumer-packaged products, they derive that low-revenue premium brands have an average down price elasticity of –1.195, and an average up elasticity of –0.921. High-equity brands, in contrast, have an average down share elasticity of –0.747, and an up elasticity of only –0.183. Hence, while brands with a higher revenue premium gain some share when they reduce their prices, they lose comparatively less share when they increase their price. As such, brand equity is a stock metric that enhances future cash flows through three different routes described earlier: higher baseline sales (volume premium), higher profit margins (price premium), and increased marketing effectiveness (differential $\beta(L)$).

Note that some marketing activity may deteriorate brand equity. For example, Mela et al. (1997) used time-varying response models to demonstrate that increasing the frequency of sales promotions may increase customers' price sensitivity to the brand. As a result, either a smaller percent of sales is generated at full price, or the brand's price premium is lowered. Both scenarios result in damage to the brand's equity.

From a model building perspective, the revenue premium that captures brand equity in (4.11) is typically estimated using the sales-response model (4.4) for different brands in a category, and examining differences in the intercept and slope parameters. The time-varying model (4.8), (4.9) may also be used in this context.

The measurement and impact of brand equity continues to be a major research area, and some empirical generalizations about brand impact have begun to appear. Among those, of a particular interest is the recent meta-analysis of nearly 500 elasticities in Edeling and Fischer (2014). They conclude that the brand and customer asset → market value elasticity averages 0.5, which is strong evidence that marketing's contribution to brand building results in substantial long-term financial benefits for the firm.

4.4.3.4 Customer Equity

While brand equity focuses on the supply side, i.e. the offerings of the firm, customer equity (CE) is an asset valued on the demand side, with specific reference to the firm's customer base. Customer lifetime value (CLV) is generally defined as the present value of all future profits obtained from a customer over his/her life of relationship with a firm (Gupta et al. 2004):

$$CLV = \sum_{t=0}^{T} \frac{(p_t - c_t) r_t}{(1+i)^t} - AC \tag{4.12}$$

where

- p_t revenue generated by a consumer at time t,
- c_t direct cost of servicing the customer at time t,
- i discount rate or cost of capital for the firm,
- r_t probability of customer repeat buying or being "alive" at time t,
- AC customer acquisition cost,
- T time horizon for estimating CLV.

Customer equity (CE) is the sum of the firm's customers' lifetime values. CLV and CE measure "net present value" from a customer asset perspective, and thus speak to both shareholder value and customer value.

Marketing spending may impact customer equity in several ways: through acquiring new customers (at a cost AC per customer), through retaining existing customers (at a servicing cost c_t in each period) and through increasing per-customer revenue, which is sometimes referred to as 'share of wallet'. Different models elaborate on different aspects of marketing's role in customer equity building; see, for example, the chapter by Bijmolt van Verhoef in this Handbook.

In order to connect customer equity with financial-performance models, we must aggregate customer-specific records and link them with firm performance. In relationship businesses such as insurance and financial services, this can be done through direct counting of customers and aggregation of their CLVs. In that case, the models reviewed in the chapter by Ascarza, Fader and Hardie in this Handbook are highly relevant.

In most cases, however, the direct-count approach is not feasible or practical, and we should infer marketing's impact on customer equity at a more aggregate level (see e.g. Rust et al. 2004). This may be achieved by examining marketing's role in *purchase reinforcement*, i.e. using an existing sale to create more future sales from that customer. Purchase reinforcement modeling applies mainly in frequently purchased product and service categories, where consumers have reason to expect a similar-quality experience between one purchase occasion and the next. Givon and Horsky (1990) developed a market-share model that contrasts the impact of purchase experience (β) relative to marketing-induced retention (λ) as follows:

$$\text{Share}_t = \alpha(1-\lambda) + (\beta+\lambda)\,\text{Share}_{t-1} - \beta\lambda\,\text{Share}_{t-2} + \gamma M_t + e_t. \tag{4.13}$$

This model is a special case of the dynamic core response function (4.5) with two-period dynamics. Thus it lends itself to calculations of the cash-flow impact (and therefore return) of investments in marketing versus customer service provision. In their empirical investigation of four frequently purchased product categories, the authors reported that $\beta > \lambda$, i.e. the impact of purchase experience exceeds that of marketing spending. As such, even without renewed instantaneous marketing support, a stock effect is at work that results in future sales.

Since then, more complex models have been developed that infer movements in customer equity from sales transactions data and brand-related marketing actions in a variety of sectors. For example, Hanssens et al. (2008) explored the impact of various marketing activities and external factors on the growth in customer equity for a major financial institution. Customer equity has, in various studies, been found to be an actionable and stable metric, which offers reliable guidance and an explicit linkage to financial performance (see e.g. Gupta and Lehmann 2005 for a review). We refer to Kumar and Shah (2015) for a comprehensive overview of research on customer lifetime value and customer equity.

Last, but not least, the two major marketing assets, brand equity and customer equity, have been formally linked in a comprehensive study of the U.S. automobile market (Stahl et al. 2012). They found, among other things, that customer knowledge of a brand has a strong positive relationship with customer acquisition, retention and profitability. Brand differentiation, on the other hand, has a positive connection with customer profitability, but impacts acquisition and retention negatively. Such findings imply that the financial implications of various strategic marketing decisions are not straightforward and may involve important tradeoffs.

4.4.3.5 Intermediate Performance Variables and Marketing Dashboards

Financial performance models have a shareholder-value orientation that may be outside the decision perimeter of most functional marketing decision makers. Marketing dashboards may be used to represent intermediate results that are directly relevant for these functional managers, and to provide the "big picture" of performance evolution for top management. As already illustrated in Fig. 4.1, various intermediate metrics are regularly reported to the board, along with more sales and market-share related measures. A detailed discussion of marketing dashboards may be found in Lehmann and Reibstein (2006). They note that a complete dashboard should integrate the impact of marketing spending on the *interim marketing metrics* and their impact on the financial consequences. In addition, dashboards should show both the short-term as well as the long-term impact of marketing, i.e. they should be not only historical but forward looking as well. Most corporate dashboards, however, have not yet advanced to this stage.

In the present modeling context, we are mainly concerned with the usefulness of intermediate metrics such as brand awareness and customer satisfaction in evaluating marketing's financial performance.[12] From an econometric perspective, an intermediate metric is *redundant* if it does not add predictive power above and

[12]An excellent review on the link between perceptual marketing metrics and financial performance is given in Gupta and Zeithaml (2006; see e.g. their Table 1).

beyond that provided by the core response model (4.1) or its extension. We illustrate this condition with the intermediate variable brand awareness (A). The core response model in implicit form, and omitting time subscripts and error terms for ease of exposition, is:

$$S = f(M), \tag{4.14}$$

and the intermediate response model is

$$A = g(M), \tag{4.15}$$

which could be a standard awareness model discussed in an advertising context in Mahajan et al. (1984). The integrated financial response model

$$S = h(A, M) \tag{4.16}$$

may be compared to the core model (4.14), for example on the basis of its residual mean squared error in a forecast sample. If model (4.16) is superior, then the intermediate metric A should be tracked and included in the dashboard, as it contains financially valuable information above and beyond that already reflected in revenue and marketing spending. This may occur, for example, when advertising-induced-awareness is persistent, thus creating a stock that facilitates demand creation.

If model (4.16) fails the comparison test, the intermediate metric A may still be valuable at the functional level (assuming Eq. 4.15 produces strong results), but it need not be incorporated in the financial valuation of marketing. This may occur when advertising-induced-awareness loses its relevance quickly due to frequent product innovation, for example in high-technology categories.

In conclusion, we propose that the important question of how many intermediate performance metrics to include in a marketing dashboard be addressed using the notion of *incremental predictive capability*, for which good analytical criteria exist. This question has been explored recently in empirical investigations of market response models that combine transactional metrics (such as sales, prices and advertising spending) and attitudinal metrics (such as consumer advertising awareness and brand liking), see e.g. Srinivasan et al. (2010) and Hanssens et al. (2014). The consensus finding is that such combined models improve the sales predictions obtained by either transactional or attitudinal models in isolation. In a holdout test on several brands in four consumer product categories, these combined models reduced the sales forecast errors by up to 50% (Hanssens et al. 2014). In addition, some interesting estimates were obtained of the *conversion rates* of attitude metrics to sales performance. For example, the elasticity of brand sales with respect to brand *liking* scores is around 0.5, and is higher than the elasticities for upper-funnel metrics such as awareness and consideration.

4.5 The Investor Perspective

Thus far, we discussed how marketing can create cash flows for the firm, either directly (through the current and lagged effects in the core response model (4.4)), or by contributing to stock variables that result in future cash flows even when new marketing expenditures are absent. The question remains, however, to what extent marketing's contribution to these cash flows is recognized by an important external audience, the shareholder or investor. More specifically, we consider to what extent this contribution is reflected in (changes in) the firms' market value.

The valuation of public firms is captured in their stock price, or market capitalization (stock price times shares outstanding). The movement of these stock prices produces *stock returns*, which is the conventional profit measure for investors.[13] We use *stock-return response modeling* to assess the degree to which marketing actions and industry conditions improve the outlook on a firm's cash flows and thereby lift its valuation. A separate set of financial models deals with the valuation of brands as intangible assets, specifically the portion of a firm's overall market capitalization that may be attributed to brand equity. These models are outside the scope of our review, and we refer the interested reader to Madden et al. (2006) for a comprehensive discussion. Similarly, the relationship between customer equity and market capitalization is discussed in Gupta and Zeithaml (2006).

Stock-return response models are similar to the internal market response models discussed previously, with one important point of difference: the dependent variable is *future* or *expectations* oriented. Indeed, stock prices may be viewed as *consensus forecasts* that react only to *new* information that is deemed relevant. Thus, the basic value assessed by internal financial performance models may already be contained in the firm's existing stock price. As such, stock-return response modeling establishes whether the information contained in one or more marketing actions is associated with changes in expectations of future cash flows and, hence, stock price and returns (we refer to Mizik and Jacobson 2004 and Srinivasan and Hanssens 2009 for a detailed review). We will first discuss two approaches to stock-return modeling that have been used to date: a single-equation method based on the efficient markets hypothesis, and a system's (vector-autoregressive) approach. Next we will summarize some key findings from the use of these models.

4.5.1 Single-Equation Approach

The stock-market valuation of a firm depicts the consensus expectation of its discounted future cash flows. The efficient market hypothesis (EMH) developed in the

---

[13]Apart from stock returns, marketing flow and stock metrics have also been linked to two other components of shareholder value, systematic risk and idiosyncratic risk (see, among others, Tuli and Bharadwaj 2009 or Osinga et al. 2011).

finance literature implies that stock prices follow random walks: the current price reflects all known information about the firm's future earnings prospects (Fama and French 1992). For instance, investors may expect the firm to maintain its usual level of advertising and price promotions. Developments that positively affect cash flows result in increases in stock price, while those negatively affecting cash flows result in decreases. In our context, regressing stock returns on changes in the marketing mix provides insights into the stock market's expectations of the associated long-term changes in cash flows. In particular, we test for *incremental* information content, that is the degree to which marketing actions explain stock price movements above and beyond the impact of current accounting measures such as revenue and earnings.

Stock-return models are highly specific to the marketing and industry characteristics of each firm. We illustrate the principles in the context of the automobile sector, in particular the role of product innovation, advertising and sales promotions (Srinivasan et al. 2009). However, the models all start with a benchmark return model, based on the Capital Asset Pricing Model (CAPM) developed in the finance and accounting literature. Following Fama and French (1992, 1993), the CAPM model is augmented with firm-specific risk factors that control for the size of the company (assets), its market-to-book ratio, and its momentum (past trends in stock return). Indeed, smaller firms are expected to outperform larger firms, and stocks with lower market-to-book ratios are expected to outperform those with a higher market-to-book ratio. Both of these effects imply that riskier stocks are characterized by higher returns. These factors reflect the a priori investor expectations in stock returns that are based on the past operations of the firm, and thus they are lagged in the model. As such, the benchmark model takes on the following form:

$$\begin{aligned}\mathrm{RET}_{i,t} = \alpha_0 + \alpha_1 \mathrm{ASSETS}_{i,t-1} + \alpha_2 \mathrm{VBR}_{i,t-1} + \alpha_3 \mathrm{MNT}_{i,t} + \alpha_4 \mathrm{EARN}_{i,t} + \alpha_5 \mathrm{SP500}_t \\ + \Sigma \alpha_j \mathrm{SEAS}_{j,t} + \varepsilon_{it}\end{aligned} \tag{4.17}$$

where RET_{it} is the stock return for firm i at time t, ASSETS_{it-1} the firm size at time $t-1$, VBR_{it-1} the market-to-book ratio (in logs) at time $t-1$, MNT_{it} measures the momentum in stock returns, EARN_{it} is the firm income, and ε_{it} is the error term. Additionally, the model may control for macro-economic movements by including covariates such as the S&P 500 Index ($\mathrm{SP500}_t$). Depending on the nature of the business, the model may also control for seasonal and holiday dummy variables (SEAS_{it} in this case).

The financial benchmark model (4.17) is subsequently augmented with marketing variables in order to assess hypotheses on their impact on future cash flows. They are expressed in changes or shocks (denoted in (4.18) through the difference operator Δ), i.e. deviations from past behaviors already incorporated in investors' expectations. Such a model has the following form[14]:

$$\begin{aligned} RET_{i,t} = {} & \alpha_0 + \alpha_1 ASSETS_{i,t-1} + \alpha_2 VBR_{i,t-1} + \alpha_3 MNT_{i,t} + \alpha_4 EARN_{i,t} + \alpha_5 SP500_t \\ & + \Sigma\alpha_j SEAS_{j,t} + \beta_1\,\Delta\,ADVi,t + \beta_2\,\Delta\,PROM_{i,t} + \beta_2\,\Delta\,INNOV_{i,t} + \varepsilon_{it} \end{aligned} \tag{4.18}$$

where the β-parameters allow to test whether changes in firm *i*'s advertising (ADV), promotional support (PROM) or innovation level (INNOV) have additional explanatory power above and beyond the variables already contained in the benchmark model. Equation (4.18) can be extended to control for other industry-relevant characteristics. Tuli et al. (2012), for example, consider the stock-return implications of changes in same-store sales, a key metric in the retailing industry. Likewise, the set of marketing variables can be expanded to reflect specific firm characteristics (see e.g. Srinivasan et al. 2009). Thus, the stock-return model augments traditional financial valuation models with *changes* in marketing strategy. In the case study above, the stock market was found to react positively to product innovation,[15] especially when combined with advertising spending. Investors were also found to react negatively to sales promotion initiatives.

A special case of the stock-return model is the *marketing event study*. Methodologically, event studies are similar in design, however the input variable is one or more isolated interventions, as opposed to ongoing marketing-mix activities. For example, event studies have been used to measure the impact on stock returns of company name changes (Horsky and Swyngedouw 1987), internet channel additions (Geyskens et al. 2002), new-product announcements (Chaney et al. 1991), foreign-market entries (Gielens et al. 2008), outsourcing decisions (Raassens et al. 2012), and opening-weekend box office results of motion pictures (Joshi and Hanssens 2009), among others. An in-depth discussion on the use of marketing event studies is given in Srinivasan and Bharadwaj (2004).

4.5.2 Vector-Autoregressive Approach

The Efficient Markets Hypothesis may not always hold, due to incomplete information available to investors and biases in their interpretation. In particular, researchers have questioned the assumption of *immediate* dissemination of all available information. For example, Fornell et al. (2006) found that

[14] As before, also non-linear specifications could be considered, which may be more appealing from a (subsequent) optimization point of view.

[15] Other studies on the stock-market reaction to firms' innovation activities include, among others, Sorescu et al. (2007) and Sorescu and Spanjol (2008).

publicly-available information about firms' customer satisfaction levels is slow to be reflected in stock prices, leaving a substantial arbitrage opportunity. It is even more difficult to gauge the impact of single marketing actions, and therefore one should not expect that they will be fully incorporated in stock prices either. Instead, investors will *update* their evaluation of these actions over time. Therefore, the short-term investor reaction may be adjusted over time until it stabilizes in the long run, and becomes so predictable that it loses its ability to further adjust stock prices. This behavior motivates the use of long-run or persistence models instead of event windows to study the impact of marketing on firm value.

Vector-autoregressive (VAR) models are well suited to measure the dynamic performance response and interactions between performance and marketing variables (Dekimpe and Hanssens 1999). Both performance variables and marketing actions are endogenous, i.e. they are explained by their own past and the past of the other endogenous variables. Specifically, VAR models not only measure direct (immediate and lagged) response to marketing actions, but also capture the performance implications of complex feedback loops. For instance, a successful new-product introduction will generate higher revenue, which may prompt the manufacturer to reduce sales promotions in subsequent periods. The combination of increased sales and higher margins may improve earnings and stock price, and thereby further enhance the over-time effectiveness of the initial product introduction. Because of such chains of events, the full performance implications of the initial product introduction may extend well beyond the immediate effects. We refer to Dekimpe and Hanssens (2017) for more methodological detail on these models.[16]

We illustrate the use of stock-performance VAR models through a recent example in the automobile sector, described in detail in Pauwels et al. (2004). Following the results of various unit-root tests, a VAR model is specified for each automotive brand j (e.g. Chevrolet, Saturn and Cadillac) from firm i (General Motors) in category k (e.g. the SUV category):

$$\begin{bmatrix} \Delta VBR_{i,t} \\ \Delta INC_{i,t} \\ \Delta REV_{i,t} \\ NPI_{ijk,t} \\ SPR_{ijk,t} \end{bmatrix} = C + \sum_{n=1}^{N} B_n \times \begin{bmatrix} \Delta VBR_{i,t-n} \\ \Delta INC_{i,t-n} \\ \Delta REV_{i,t-n} \\ NPI_{ijk,t-n} \\ SPR_{ijk,t-n} \end{bmatrix} + \Gamma \times \begin{bmatrix} \Delta S\&P500_t \\ \Delta Construct_t \\ \Delta Exchange_t \\ \Delta EPS_{i,t} \end{bmatrix} + \begin{bmatrix} u_{VBRi,t} \\ u_{INCi,t} \\ u_{REVi,t} \\ u_{NPIijk,t} \\ u_{SPRijk,t} \end{bmatrix} \tag{4.19}$$

with B_n, Γ matrices of coefficients, and $[u_{VBRi,t}, u_{INCi,t}, u_{REVi,t}, u_{NPIijk,t}, u_{SPRijk,t}]' \sim N(0, \Sigma_u)$. The B_n matrices contain the autoregressive parameters capturing the dynamic effects among the endogenous variables, while the Γ matrix links the endogenous

[16]Through the feedback loops, VAR models also capture the role that (past) stock-price variations play in managerial decision making. It would be useful to extend these models to also incorporate anticipated stock-price variations (foresight) into dynamic marketing models.

variables to a set of exogenous control variables. In this system, the first equation explains changes in firm value, operationalized as the ratio of the firm's market value to book value (*VBR*). This variable reflects a firm's potential growth opportunities, and is used frequently for assessing a firm's ability to achieve abnormal returns relative to its investment base. The second and third equations explain the changes in, respectively, bottom-line (*INC*) and top-line financial performance (*REV*) of firm *i*. The fourth and fifth equations model firm *i*'s marketing actions, i.e. new-product introductions (*NPI*) and sales promotions (*SPR*) for brand *j* in product category *k*. The model also includes various exogenous factors, seasonal demand variations (such as Labor Day weekend, Memorial Day weekend,and the end of each quarter), fluctuations in the overall economic and investment climate (S&P 500, the Construction Cost index and the dollar-Yen exchange rate), and accounts for the impact of stock-market analyst earnings expectations (*EPS*).

Overall, VAR models require extensive time-series data as they contain many more parameters than stock-return models. As the quality of financial and marketing databases increases, we expect these models to be used more frequently in the future (see Dekimpe and Hanssens 2017 for a more extensive discussion). In this particular application, Pauwels and his co-authors found that new-product introductions increased long-term financial performance and firm value, while promotions did not. Moreover, investor reactions to new product introductions were found to grow over time, and to yield the highest stock market benefits for entries into new markets.

4.5.3 Key Findings on Marketing and Firm Value

Several empirical studies have examined various aspects of the marketing-firm value relationship. The overarching conclusion is that marketing *does* impact investor sentiment and therefore firm value (see, for example, Srinivasan and Hanssens 2009 or Luo et al. 2012 for recent reviews). Furthermore, and importantly, investors react positively to marketing actions that are known to have beneficial effects on long-term business performance, and vice versa. This finding goes against the popularized view that investors care only about short-term earnings.

As a case in point, the impact of product innovation has a positive impact on stock returns. In the short run, new product announcements tend to be associated with positive abnormal returns (Sood and Tellis 2009). In the long run, using a 1-year window, product innovation again has a positive impact, which is stronger for radical as opposed to incremental innovations (Sorescu and Spanjol 2008). These findings contrast with those on the investor effects of sales promotions, which has been found to be negative (Srinivasan et al. 2004). Despite the fact that such promotions tend to have significant and immediate sales effects, investors are concerned that they will erode brand equity and thus eventually harm the brand's financial outlook.

Interestingly, marketing metrics often have a differing impacts on, respectively, the stock-return component of shareholder wealth and its two risk components (systematic and idiosyncratic risk; see, e.g. Bharadwaj et al. 2011). In addition, several of these effects have been found to be context-dependent, calling for appropriate contingency frameworks. A comprehensive overview of the models and findings on the marketing-finance interface may be found in Ganesan (2012).

4.6 Conclusion

Every year, companies spend a sizeable portion of their revenues on a variety of marketing activities. In the U.S., advertising and sales force expenditures alone sum to more than 1.6 trillion dollars, about 10% of the Gross National Product (Dekimpe and Hanssens 2011). These activities should be viewed as investments that ultimately return value to the firm's shareholders (Srinivasan and Hanssens 2009). Thus the assessment of the financial performance of marketing investments is an important task for marketing scientists and marketing managers alike. This assessment involves both flow metrics and stock metrics.

Our chapter has presented a framework for financial performance models from two perspectives: internal—i.e. describing how marketing creates value to the firm's shareholders—and external—i.e. describing how outside investors react to changes in marketing strategy. Starting from the core market response model, we first derived the standard measure of return of investment to marketing. We also isolated the three determinants of marketing spending that drive cash flows to the shareholders, viz. baseline business revenue, marketing effectiveness and profit margin. We then expanded the value creation of marketing to include *stock* metrics, in particular those created by diffusion of innovation, brand equity and customer equity. Marketing's total financial performance contribution is the sum of its impacts on these stock and flow metrics.

The shareholders' valuation of marketing is driven by their expectations on future cash flows and their perceptions on how marketing influences these cash flows. Investor valuation models should therefore focus on the *new* information contained in various marketing strategies and actions. We have discussed, in turn, the single-equation stock return response model, and the vector-autoregressive system's model as viable alternatives to measure marketing's impact on stock returns. Taken together, process models of value creation and investor valuation models provide a comprehensive and powerful resource to gauge marketing's impact on financial performance.

In terms of managerial implications, two important conclusions emerge. First, there *are* formal links between marketing actions and financial outcomes, and thus the marketing executive can and should participate in decisions that impact the financial outlook of the firm. Second, in so doing, the marketing executive should draw a careful distinction between actions that enhance or protect revenue flow and actions that build brand or customer equity. The latter two are not easily visible in the short run, but the metrics and models we have discussed above provide an implementable framework to answer all-important questions about the financial return on marketing and the role of marketing in the modern enterprise.

References

Ailawadi, Kusum L., Donald R. Lehmann, and Scott A. Neslin. 2003. Revenue premium as an outcome measure of brand equity. *Journal of Marketing* 67 (October): 1–17.

Ambler, Tim. 2003. *Marketing and the bottom line*, 2nd ed. London, U.K.: FT Prentice Hall.

Ambler, Tim, and John H. Roberts. 2008. Assessing marketing performance: Don't settle for a silver metric. *Journal of Marketing Management* 24 (7–8): 733–750.

Ataman, Berk, Carl F. Mela, and Harald J. van Heerde. 2008. Building brands. *Marketing Science* 26 (November–December), 1036–1054.

Barwise, Patrick, and John U. Farley. 2003. Which marketing metrics are used and where. *Marketing Reports* 2: 105–107.

Bass, Frank M. 1969. A new product growth model for consumer durables. *Management Science* 15 (January): 215–227.

Bass, Frank M., Trichy V. Krishnan, and Dipak C. Jain. 1994. Why the Bass model fits without decision variables. *Marketing Science* 13 (3): 204–223.

Bharadwaj, Sundar G., Kapil R. Tuli, and Andre Bonfrer. 2011. The impact of brand quality on shareholder wealth. *Journal of Marketing* 75 (September): 88–104.

Chaney, Paul K., Timothy M. Devinney, and Russell S. Winer. 1991. The impact of new product introductions on the market value of firms. *Journal of Business* 64 (4): 573–610.

Dekimpe, Marnik G., and Dominique M. Hanssens. 1999. Sustained spending and persistent response: A new look at long-term marketing profitability. *Journal of Marketing Research* 36 (November): 1–31.

Dekimpe, Marnik G., and Dominique M. Hanssens. 2011. The hidden powers of advertising investments. In *Liber Amicorum in honor of Peter S.H. Leeflang*, ed. Jaap E. Wierenga, Peter C. Verhoef, and Janny C. Hoekstra. Rijksuniversiteit Groningen.

Dekimpe, Marnik G., and Dominique M. Hanssens. 2017. Time-series models in marketing: Insights from persistence modeling. In *Handbook of marketing analytics: Methods and applications in marketing, public policy, and litigation support*, ed. Natalie Mizik and Dominique M. Hanssens. Edward Elgar Publishing Ltd. (in print).

Edeling, Alexander, and Marc Fischer. 2014. *Marketing's impact on firm value: Generalizations from a meta-analysis*, Report No. 14-107. Cambridge, MA: Marketing Science Institute.

Elberse, Anita, and Jehoshua Eliashberg. 2003. Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing Science* 22 (3): 329–354.

Fama, Eugene, and Kenneth French. 1992. The cross-section of expected stock returns. *Journal of Finance* 47 (2): 427–465.

Fama, Eugene, and Kenneth French. 1993. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics* 33 (1): 3–56.

Farris, Paul W., Neil T. Bendle, Phillip E. Pfeifer, and David J. Reibstein. 2010. *Marketing metrics, the definitive guide to measuring marketing performance*, 2nd ed. Wharton School Publishing.

Farris, Paul W., Dominique M. Hanssens, James D. Lenskold, and David J. Reibstein. 2015. Marketing return on investment: Seeking clarity for concept and measurement. *Applied Marketing Analytics* 1 (3): 267–282.

Fischer, Marc, Hyun S. Shin, and Dominique M. Hanssens. 2016. Brand performance volatility from marketing spending. *Management Science* 62 (1): 197–215.

Fornell, Claes, Sunil Mithas, Forrest Morgeson, and M.S. Krishnan. 2006. Customer satisfaction and stock prices: High returns, low risk. *Journal of Marketing* 70 (January): 3–14.

Forrester, Jay W. 1961. *Industrial dynamics*. Cambridge, MA: MIT Press.

Ganesan, Shankar (ed.). 2012. *Handbook of marketing and finance*. Edward Elgar Publishing.

Gatignon, Hubert, Thomas S. Robertson, and Adam J. Fein. 1997. Incumbent defense strategies against new product entry. *International Journal of Research in Marketing* 14 (May): 163–176.

Geyskens, Inge, Katrijn Gielens, and Marnik G. Dekimpe. 2002. The market valuation of internet channel additions. *Journal of Marketing* 66 (April): 102–119.

Gielens, Katrijn, Linda M. Van de Gucht, Jan-Benedict E.M. Steenkamp, and Marnik G. Dekimpe. 2008. Dancing with a giant: The effect of Wal-Mart's entry into the United Kingdom on the stock prices of European retailers. *Journal of Marketing Research* 45 (October): 519–534.

Givon, Moshe, and Dan Horsky. 1990. Untangling the effects of purchase reinforcement and advertising carryover. *Marketing Science* 9 (2): 171–187.

Gupta, Sunil, and Donald R. Lehmann. 2005. *Managing customers as investments*. Upper Saddle River, NJ: Wharton School Publishing/Pearson-Financial Times.

Gupta, Sunil, and Valarie Zeithaml. 2006. Customer metrics and their impact on financial performance. *Marketing Science* 25 (6): 718–739.

Gupta, Sunil, Donald R. Lehmann, and Jennifer Ames Stuart. 2004. Valuing customers. *Journal of Marketing Research* 41 (February): 7–18.

Hanssens, Dominique M. (ed.). 2015. *Empirical generalizations about marketing impact*, 2nd ed. Cambridge, MA: Marketing Science Institute.

Hanssens, Dominique M., Leonard J. Parsons, and Randall L. Schultz. 2001. *Market response models*, 2nd ed. Boston, MA: Kluwer Academic Publishers.

Hanssens, Dominique M., Daniel Thorpe, and Carl Finkbeiner. 2008. Marketing when customer equity matters. *Harvard Business Review* May, 117–123.

Hanssens, Dominique M., Koen H. Pauwels, Shuba Srinivasan, Marc Vanhuele, and Gokhan Yildirim. 2014. Consumer attitude metrics for guiding marketing mix decisions. *Marketing Science* 33 (July–August), 534–550.

Hanssens, Dominique M., Fang Wang, and Xiao-Ping Zhang. 2016. Performance growth and opportunistic marketing spending. *International Journal of Research in Marketing*.

Hardie, Bruce, Pete Fader, and Eva Ascarza. 2016. Customer base analysis, customer analytics, customer relationship. In *Handbook of marketing decision models*, ed. Berend Wierenga and Ralf van der Lans (New edition in print).

Hauser, John R., and Steven M. Shugan. 1983. Defensive marketing strategies. *Marketing Science* 2 (Autumn): 319–360.

Hinz, Oliver, Bernd Skiera, Christian Barrot, and Jan Becker. 2011. An empirical comparison of seeding strategies for viral marketing. *Journal of Marketing* 75 (November): 55–71.

Horsky, Dan, and Leonard S. Simon. 1983. Advertising and the diffusion of new products. *Marketing Science* 2 (Winter): 1–17.

Horsky, Dan, and Patrick Swyngedouw. 1987. Does it pay to change your company's name? A stock market perspective. *Marketing Science* 6 (November): 320–335.

Johansson, Johny K. 1979. Advertising and the S-curve: A new approach. *Journal of Marketing Research* 16 (August): 346–354.

Joshi, Amit M., and Dominique M. Hanssens. 2009. Movie advertising and the stock market valuation of studios: A case of great expectations? *Marketing Science* **28** (March-April), 239–250.

Joshi, Amit M., and Dominique M. Hanssens. 2010. The direct and indirect effects of advertising spending on firm value. *Journal of Marketing* 74 (January): 20–33.

Kamakura, Wagner A., and Gary J. Russell. 1993. Measuring brand value with scanner data. *International Journal of Research in Marketing* 10 (March): 9–22.

Keller, Kevin L., and Donald R. Lehmann. 2001. The brand value chain: Linking strategic and financial performance. Working paper, Tuck School of Business, Dartmouth College.

Kumar, V., and Denish Shah (eds). 2015. *Handbook of research on customer equity in marketing*. Edward-Elgar Publishing.

Leeflang, Peter S.H., Tammo H.A. Bijmolt, Jenny van Doorn, Dominique M. Hanssens, Harald J. van Heerde, Peter C. Verhoef, and Jaap E. Wieringa. 2009. Creating lift versus building the base: Current trends in marketing dynamics. *International Journal of Research in Marketing* 26 (March): 13–20

Lehmann, Donald R., and David J. Reibstein. 2006. *Marketing metrics and financial performance*. Cambridge, MA: Marketing Science Institute, Relevant Knowledge Series.

Luo, Xueming, Koen H. Pauwels, and Dominique Hanssens. 2012. Time-series models of pricing the impact of marketing on firm value. In *Handbook of marketing and finance*, ed. Shankar Ganesan, 43–65. Northampton, MA: Edward Elgar.

Madden, T.J., Frank Fehle, and Susan Fournier. 2006. Brands matter: An empirical demonstration of the creation of shareholder value through branding. *Journal of the Academy of Marketing Science* 34 (2): 224–235.

Mahajan, Vijay, Eitan Muller, and Subhash Sharma. 1984. An empirical comparison of awareness forecasting models of new product introduction. *Marketing Science* 3 (Summer): 179–197.

Mela, Carl F., Sunil Gupta, and Donald R. Lehmann. 1997. The long-term impact of promotion and advertising on consumer brand choice. *Journal of Marketing Research* 34 (2): 248–261.

Mitra, Debanjan, and Peter N. Golder. 2006. How does objective quality affect perceived quality: Short-term effects, long-term effects, and asymmetries. *Marketing Science* 25 (May), 230–247.

Mizik, Natalie, and Robert Jacobson. 2004. Stock return response modeling. In *Assessing marketing strategy performance*, ed. Christine Moorman, and Donald R. Lehmann. Cambridge, MA: Marketing Science Institute.

Naik, Prasad. 2015. *Marketing dynamics: A primer on estimation and control.* Foundations and Trends in Marketing.

Norton, John A., and Frank M. Bass. 1987. A diffusion theory model of adoption and substitution for successive generations of high-technology products. *Management Science* 33 (September): 1069–1086.

Onishi, Hiroshi, and Puneet Manchanda. 2012. Marketing activity, blogging and sales. *International Journal of Research in Marketing* 29 (3): 221–234.

Osinga, Ernst C., Peter S.H. Leeflang, Shuba Srinivasan, and Jaap E. Wieringa. 2011. Why do firms invest in consumer advertising with limited sales response? A shareholder perspective. *Journal of Marketing* 75 (January): 109–124.

Pauwels, Koen, and Dominique M. Hanssens. 2007. Performance regimes and marketing policy shifts. *Marketing Science* 26 (3): 293–311.

Pauwels, Koen, Jorge Silva-Risso, Shuba Srinivasan, and Dominique M. Hanssens. 2004. New products, sales promotions and firm value, with application to the automobile industry. *Journal of Marketing* 68 (October): 142–156.

Quelch, John, and Gail J. McGovern. 2006. Boards must measure marketing effectiveness. *Directors & Boards*, 53–56.

Raassens, Neomie, Stefan H.K. Wuyts, and Inge Geyskens. 2012. The market valuation of outsourcing new product development. *Journal of Marketing Research* **49** (5), 682–695.

Rand, William M., and Roland T. Rust. 2011. Agent- based modeling in marketing: Guidelines for Rigor. *International Journal of Research in Marketing* 28 (June): 181–193.

Roberts, John H. 2005. Defensive marketing: How a strong incumbent can protect its position. *Harvard Business Review* **83** (November), 150–157, 168.

Roberts, John H., Charles J. Nelson, and Pamela D. Morrison. 2005. A prelaunch diffusion model for evaluating market defense strategies. *Marketing Science* 24 (Winter): 150–164.

Rust, Roland T., Tim Ambler, Gregory S. Carpenter, V. Kumar, and Rajendra K. Srivastava. 2004a. Measuring marketing productivity: Current knowledge and future directions. *Journal of Marketing* 68 (October): 76–89.

Rust, Roland T., Katherine N. Lemon, and Valarie A. Zeithaml. 2004b. Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing* 68 (January): 109–127.

Sood, Ashish, and Gerard J. Tellis. 2009. Do innovations really payoff? Total stock market returns to innovation. *Marketing Science* 28 (3): 442–456.

Sorescu, Alina, Venkatesh Shankar, and Tarun Kushwaha. 2007. New product preannouncements and shareholder value: Don't make promises you can't keep. *Journal of Marketing Research* 44 (3): 468–489.

Sorescu, Alina, and Jelena Spanjol. 2008. Innovation's effect on firm value and risk: Insights from consumer packaged goods. *Journal of Marketing* 72 (2): 114–132.

Srinivasan, Raji, and Sundat Bharadwaj. 2004. Event studies in marketing strategy research. In *Assessing marketing strategy performance*, ed. Christine Moorman, and Donald R. Lehmann. Cambridge, MA: Marketing Science Institute.

Srinivasan, Raji, Gary L. Lilien, and Arvind Rangaswamy. 2004. First in, first out? The effects of network externalities on pioneer survival. *Journal of Marketing* 68 (January): 41–58.

Srinivasan, Shuba, and Dominique M. Hanssens. 2009. Marketing and firm value: Metrics, methods, findings and future directions. *Journal of Marketing Research* 46 (June): 293–312.

Srinivasan, Shuba, Koen Pauwels, Jorge Silva-Risso, and Dominique M. Hanssens. 2009. Product innovations, advertising, and stock returns. *Journal of Marketing* 73 (January): 24–43.

Srinivasan, Shuba, Marc Vanhuele, and Koen H. Pauwels. 2010. Mindset metrics in market response models: An integrative approach. *Journal of Marketing Research* (4), 672–684.

Sriram, S., and Manohar U. Kalwani. 2007. Optimal advertising and promotion budgets in dynamic markets with brand equity as a mediating variable. *Management Science* 53 (January): 46–60.

Srivastava, Rajendra, David J. Reibstein, and William S. Woodside. 2005. *Metrics for linking marketing to financial performance*. Cambridge, MA: Marketing Science Institute.

Srivastava, Rajendra, David J. Reibstein, and Yogesh V. Joshi. 2006. Link marketing metrics to financial performance. Technical report, Zyman Institute of Brand Science, Emory University.

Srivastava, Rajendra, Tasadduq A. Shervani, and Liam Fahey. 1998. Market-based assets and shareholder value: A framework for analysis. *Journal of Marketing* 62 (January): 1–18.

Stahl, Florian, Mark Heitmann, Donald R. Lehmann, and Scott A. Neslin. 2012. The impact of brand equity on customer acquisition, retention, and profit margin. *Journal of Marketing* 76 (July): 44–63.

Trusov, Michael, Ralf van der Lans, and Xi Chen. 2016. Social networks. In *Handbook of marketing decision models*, ed. Berend Wierenga and Ralf van der Lans (New edition in print).

Tuli, Kapil R., and Sundar G. Bharadwaj. 2009. Customer satisfaction and stock returns risk. *Journal of Marketing* 73 (November): 184–197.

Tuli, Kapil, Anirban Mukherjee, and Marnik G. Dekimpe. 2012. On the value relevance of retailer advertising spending and same-store sales growth. *Journal of Retailing* 88 (4): 447–461.

Chapter 5
Loyalty Programs: Current Insights, Research Challenges, and Emerging Trends

Tammo H.A. Bijmolt and Peter C. Verhoef

5.1 Introduction

With the rise of relationship marketing in the 1990s (Berry 1995; Sheth and Parvatiyar 1995), interest in creating long-term relationships between customers and firms has increased. The growing attention to customer relationship management (CRM) in the first decade of this century has only emphasized this interest more strongly. For example, Payne and Frow (2005) underscore the importance of developing appropriate long-term relationships with customers in their definition of CRM. Similarly, in their conceptualization of CRM, Reinartz et al. (2004) focus strongly on relationship development.

Creating strong attitude-based relationships is a real challenge, especially in mass markets. Consumers purchase products and services from dozens of companies, and they are bombarded with advertisements on fantastic new offers. For firms, it is difficult to provide consistent service quality throughout all customer touchpoints; yet the need to do so has become more prominent with the increasing number of customer touchpoints both online and offline (Verhoef et al. 2015).

To foster customer relationships, especially in consumer markets, firms have implemented specific relationship-building programs, or so-called loyalty programs (LPs). These programs provide monetary benefits (e.g. through direct discounts or

---

Author order is alphabetical; both authors contributed equally to this chapter. We thank the editors and two anonymous reviewers for their helpful comments.

T.H.A. Bijmolt (✉) · P.C. Verhoef
Faculty of Economics and Business, Department of Marketing,
University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands
e-mail: t.h.a.bijmolt@rug.nl

P.C. Verhoef
e-mail: p.c.verhoef@rug.nl

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_5

rewards) and/or more soft benefits by focusing on creating commitment to the firm among customers through excellent service or giving special treatment to customers. Although specific LPs (e.g. stamps) have been around for decades, these programs became popular in the 1990s with the introduction of plastic loyalty cards that could be linked to a customer database using scanning and chip technology. The American Marketing Association[1] defines LPs as 'continuity incentive programs offered by a retailer to reward customers and encourage repeat business'. Bijmolt et al. (2011) suggest that this definition also includes manufacturers as potential offerees. They also provide five criteria to distinguish LPs from other marketing instruments (Bijmolt et al. 2011, p. 201):

1. LPs foster customer loyalty;
2. LPs are structured in such a way that customers should become a member of a program;
3. LPs have a long-term focus;
4. LPs reward customers who are members of the program, typically based on their purchase behavior; and
5. LPs are supplemented with ongoing marketing efforts (e.g. targeted mailings, personalized offers).

Previous reviews have summarized the available knowledge on LPs (see Bijmolt et al. 2011; Breugelmans et al. 2015). In this chapter, we build on and extend these reviews. First, we begin by discussing LPs from both a consumer and a firm perspective. Next, we discuss three mechanisms underlying the LP effect, the LP design, and models that can be used to analyze LP data. We also provide a discussion on emerging topics in LPs, specifically addressing increasing digitalization, empowered customers, and the prevalence of big data. We conclude with a discussion on some pressing research questions.

5.2 Loyalty Program Phases

In this chapter, we structure the discussion on LPs around three phases: introduction, maintenance, and termination (see Table 5.1). We consider two main decision makers: the firm and its customers. The firm typically initiates an LP for several reasons; the main reason is to increase customer loyalty, by stimulating repurchase behavior and cross-buying, thereby leading to increased share of wallet (Leenheer et al. 2007; Verhoef 2003). Research suggests that firms' internal marketing culture, customer characteristics, and the market environment are drivers of the decision to develop LPs (Leenheer and Bijmolt 2008). First, customer-oriented firms are often likely to introduce LPs (Gable et al. 2008; Smith et al. 2004). Second, LPs are also more frequently used in markets with a high purchase frequency and in which profit

[1]See https://www.ama.org/resources/Pages/Dictionary.aspx.

Table 5.1 Loyalty program phases and respective decisions

	LP phases		
Perspective	LP Introduction	LP Maintenance	LP dissolution
Firm	LP introduction LP type and design	LP design adaptations LP analytics Personalized marketing LP renewal	LP termination
Customer	LP enrollment and usage	Customer loyalty (behavior and attitude) Reward redemption	LP leaving

heterogeneity among customers is high (Leenheer and Bijmolt 2008; Zeithaml et al. 2001). Finally, LPs are frequently introduced in response to a competitor's introduction of such a program (Leenheer and Bijmolt 2008; Liu and Yang 2009; Ziliani and Bellini 2004). When setting up LPs, firms must decide on the design of the program. This decision involves multiple issues, which we discuss in Sect. 5.4 while focusing on three specific LP formats—short-term LPs, multi-vendor LPs, and hierarchical LPs (HLPs).

In response to the introduction of an LP, customers may or may not enroll in the program. Enrollment occurs during the whole LP life cycle, as (new) customers are constantly confronted with the LP and asked to participate. Being members of the LP should have an effect on customers. The assumption is that customers will use the program and, in line with the firm's objectives, become more loyal, in terms of both attitudinal and behavioral loyalty. This assumption has been one of the most debated issues in LP research. Do LPs really foster loyalty (Dowling and Uncles 1997)? Given the importance of this issue, we devote substantial attention to the LP effectiveness question in Sect. 5.3.

From a customer perspective, usage of an LP involves not only purchasing but also reward redemption. That is, rewarding customer behavior is deemed important to create enduring engagement in the LP. Important questions are whether and how firms can influence reward redemption and whether and how this redemption subsequently creates a stronger attitudinal and behavioral response to LPs (Dorotic et al. 2014; Stourm et al. 2015). Thus, firms should consider rewarding an explicit part of the LP. Although this should initially be done when designing the program, firms may also adapt the program's design and reward structure during the program's life cycle. This design change is part of a firm's LP strategy during the LP life cycle, which also involves analytics of the LP database and personalized marketing to LP members (Blattberg et al. 2008). Personalized marketing in LPs can potentially influence customer behavior as well (Dorotic et al. 2014). Beyond that, firms can use the outcomes of LP analytics to adapt the program. Importantly, design changes can also be negative, as rewards may be reduced to save costs, which has, for example, occurred with multiple frequent-flier reward programs. Given the omnipresence of LPs and increasing competitive pressure for firms to attract and retain customers, customers are likely to reduce their usage of an LP or leave the program altogether if the LP rewards are not valuable enough to outweigh their investment in time and effort. Therefore, well-functioning reward structures are essential for keeping customers active in the LP and preventing them from leaving. We also extensively discuss rewarding customer behavior in Sect. 5.3.

Finally, firms can decide to terminate the LP. The reasons to terminate could include marketing budget cuts (e.g. during bad economic times; Van Heerde et al. 2013), re-allocation of marketing budgets across instruments or strategies (e.g. acquisition vs. retention; Reinartz et al. 2005), or negative returns on the LP due to insufficient behavioral loyalty effects and/or high LP maintenance costs. LP termination by firms has received scant attention, with only Melnyk and Bijmolt (2015) considering customer reactions to LP termination. Their study suggests that termination of an LP can have considerable negative consequences for customer retention and that the competitive environment and duration of membership in an LP are the primary drivers of customer reactions to LP termination.

In the next sections, we discuss specific aspects on which researchers have worked extensively—namely, LP effectiveness, personalization, design, and analytics. We do not discuss less researched topics, such as firm LP introduction and termination decisions or customer LP enrollment decisions. For a more extensive review of these topics, we refer readers to Bijmolt et al. (2011).

5.3 Effect Mechanisms of Loyalty Programs

5.3.1 Customer Responses to LPs

The essential purpose of an LP is to enhance customer loyalty. In general, customer loyalty consists of two interrelated dimensions: attitudinal loyalty and behavioral loyalty (Dick and Basu 1994). Positive effects of LPs on attitudinal loyalty (i.e. affective commitment) have been demonstrated in the literature (for a review, see Sect. 6 of Bijmolt et al. 2011). Customers' participation in an LP can enhance their sense of gratitude, belonging, status, prestige, or recognition; these effects depend on the LP design and customer satisfaction with the LP (e.g. Keh and Lee 2006). Attitudinal loyalty can reduce customers' responsiveness to competitive actions and enhance their word-of-mouth and other beneficial-to-the-firm behaviors, and therefore attitudinal loyalty is critical for achieving sustainable long-term customer loyalty. As such, an LP may have a positive impact on attitudinal loyalty and thereby an indirect effect on behavioral loyalty. In the rest of this section, we focus on the behavioral responses of customers to the LP.

Three mechanisms may cause an LP to influence customer behavior (Fig. 5.1). First, the design of the LP itself, regardless of past purchases and rewards, may affect customer purchase behavior. A program may contain direct benefits, such as immediate discounts for LP members. In addition, in most LPs, customers can

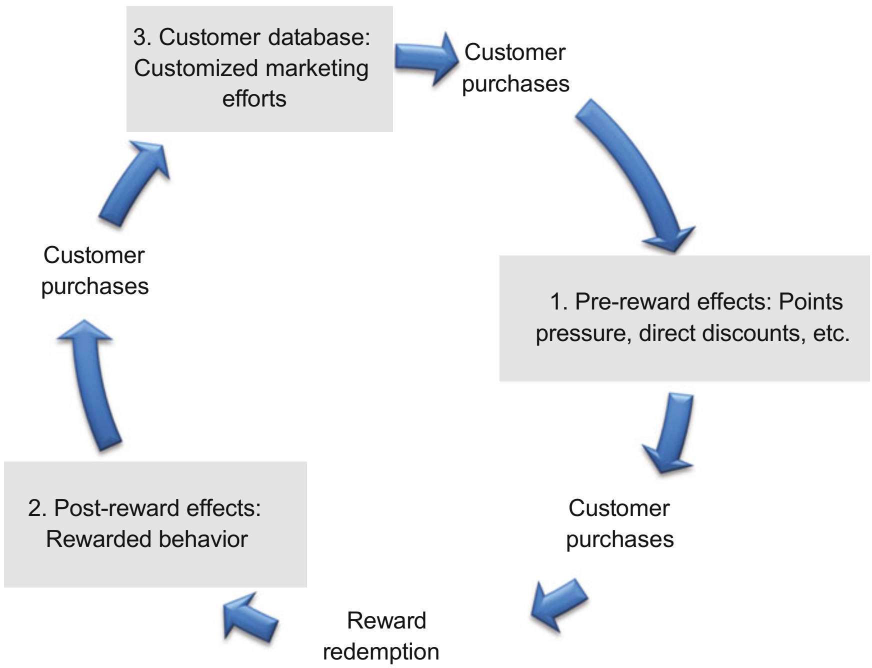


Fig. 5.1 Three effect mechanisms of the loyalty program

collect points or some other type of currency. Collecting these points and aiming to reach certain thresholds may affect customer behavior, also known as the so-called points pressure effect (Kivetz et al. 2006; Taylor and Neslin 2005). The second effect mechanism plays a role during and after reward redemption. After collecting points for a certain period, customers can redeem a reward, and receiving the reward may subsequently lead to lower churn probabilities or higher repeat purchases. This mechanism has been labeled as the rewarded behavior effect (Dorotic et al. 2014; Palmatier et al. 2009; Taylor and Neslin 2005). Third, customers' behavior will be registered by the LP system. Firms can exploit their customer databases to personalize marketing efforts and other "big data"—related issues, which again may affect customer purchase behavior. We discuss these three mechanisms next.

5.3.2 Pre-rewarding Effects: Direct Discounts and Points Pressure

Before LP members obtain a reward, and independent of past purchases (mechanism 1 in Fig. 5.1), the LP may affect customer purchase behavior. First, direct monetary benefits for LP members constitute an important element in the design of many LPs. Such benefits can be provided by means of price discounts, coupons, cash rebates, and so on. If customer involvement is low and customers are not intrinsically motivated to build a relationship with the firm, they often prefer immediate to delayed benefits, even if the immediate rewards are of lower monetary value (Keh and Lee 2006; Yi and Jeon 2003). This preference would favor direct monetary discounts over a points-saving format. Van Heerde and Bijmolt (2005) find that sales to LP members indeed increase considerably if such direct discounts are communicated through direct mail or store flyers.

Second, the most traditional element of LPs is a reward system based on past purchases. Customers collect points, stamps, or other LP currencies and then exchange these for rewards at a later stage. Customers may increase their purchase levels to earn sufficient points to be able to redeem the reward (Taylor and Neslin 2005): this is called the points pressure effect. Taylor and Neslin (2005) demonstrate the points pressure effect for an LP with a finite horizon (see Sect. 5.4.2), and Dorotic et al. (2014) do the same for a continuous LP. In the latter case, the points pressure effect is smaller, because LP members are not confronted with an external deadline and can personally decide whether, when, and how much to redeem; however, the anticipated redemption now causes a positive impact on purchase levels (Dorotic et al. 2014). In the context of a so-called item-based LP, in which customers obtain additional points for purchases of specific products, next to overall spending, Zhang and Breugelmans (2012) demonstrate that consumers were more responsive to reward point promotions than to direct price discounts of the same monetary value, implying that an item-based LP can be more effective in attracting new customers and retaining existing customers.

5.3.3 Post-rewarding Effects: Rewarded Behavior

After collecting a sufficient number of points, a customer can use these points to obtain a reward. Next, the customer may decide to maintain or increase purchase levels (mechanism 2 in Fig. 5.1) because of increased attitudinal loyalty (Smith and Sparks 2009) or because purchases lead to rewards (Taylor and Neslin 2005) or the purchase behavior itself has become a habit (Henderson et al. 2011). Reward redemption may help firms build relationships and achieve long-term customer loyalty, and therefore firms should stimulate reward redemption by LP members. Empirical research has found increased purchase levels after reward redemption in short-term programs (Lal and Bell 2003; Taylor and Neslin 2005) and continuous LPs (Dorotic et al. 2014; Drèze and Nunes 2011; Kopalle et al. 2012). However, the rewarded behavior effect appears to have a relatively short duration, often just a few weeks (Kopalle et al. 2012; Meyer–Waarden and Benavent 2009; Taylor and Neslin 2005). Moreover, the strength and the duration of the reward redemption effect depend on previous levels of customer loyalty and the type of reward (Dorotic et al. 2014; Keh and Lee 2006; Kopalle et al. 2012; Lal and Bell 2003; Liu 2007;

Taylor and Neslin 2005; Wirtz et al. 2007). The long-term impact of LPs due to the rewarded behavior mechanism and moderators of this effect warrant further research.

5.3.4 LP-Based Personalization

There is an increasing amount of literature in marketing on personalization, in which marketing efforts are adapted to individual customers or small groups of customers. In marketing practice, personalization has become prominent, owing to the increasing availability of data as a result of increased digitalization and digitalized services, such as Spotify and Netflix (Chung et al. 2009; Chung and Wedel 2014). Personalization in the area of LPs is typically more traditional and involves selecting customers according to their expected response behavior to an offer or so-called look-a-like analysis (customers who buy product X also buy product Y). LPs create customer databases with transaction data; especially for retailers, this provides opportunities to apply CRM and more individualized marketing aimed to enhance behavioral loyalty (mechanism 3 in Fig. 5.1).

A well-known example of LP-based personalization is Tesco's use of its Clubcard data to learn more about its customers and their buying behavior, to segment customers, and to provide these customer segments with offers through direct mailings and coupons (Humby et al. 2008). In the early stages of this program, the firm used the LP member database to segment customers on the basis of their life stage. Next, targeted promotions were provided to specific segments, and simple errors were avoided (e.g. offering coupons for Coca-Cola to 'Tea Drinking pensioners'; Humby et al. 2008, p. 109). In later stages of the Clubcard, Tesco distinguished segments on the basis of shopping habits, on which specific assortment decisions could be made. For example, by analyzing the purchase behavior of committed organic shoppers, the firm decided to introduce organic products in specific categories (Humby et al. 2008, p. 161).

In the digital era, adaptive personalization systems have developed, in which customers receive real-time offers based on their most recent purchase behavior. For this purpose, firms are using complex statistical algorithms—for example, Bayesian models. Chung and Wedel (2014) provide an overview of different personalization methods and discuss two major types of systems: recommendation systems and personalization systems. Recommendation systems have been around since the start of Internet retailing. The key idea of these systems is that, based on a customer's characteristics and characteristics of other customers, individual-specific recommendations can be given (e.g., 'This product might be something you would enjoy, because customers similar to you have also purchased this product'). For this purpose, firms may use content-filtering systems, collaborative-filtering systems, or a combination of the two. Personalization systems adapt the offering to customer needs; that is, firms tailor the marketing mix to the customer according to available customer information. For example, Last.fm might play music in line with a customer's prior selection of songs and likes/dislikes of songs (Verhoef et al. 2016, p. 210).

In a retail setting, the benefits of personalization for firms are not as obvious. Zhang and Wedel (2009) provide the results of personalized offers for a traditional offline supermarket and an online supermarket. They show that the incremental payoffs of personalized promotions compared with segment- and market-level customized promotions were greater for online than offline stores and that the increments for offline stores were minimal, due to low promotion redemptions in that channel. This might have occurred because many decisions are made in-store, resulting in a weaker effect of out-of-store communication. In an online environment, personalization can be offered on the website, and thus redemption rates can be higher online than offline. These results suggest that in an offline retail environment, personalized marketing can be less rewarding for firms. One way to make this environment more rewarding is to use less costly digital media, such as e-mail. For example, the Dutch retailer Albert Heijn now provides personalized offers to its offline loyalty card members by e-mail. For its online operations, personalized offers are provided in the online shop as well as by e-mail. Whether these actions solve the redemption problem offline remains to be seen, as response rates on e-mails are relatively low. One potential problem is that personalization is more costly offline because more expensive communication methods are typically used. The use of new digital channels, such as mobile phones (Fang et al. 2015), may eliminate some of these issues.

The studies on personalization have mainly considered behavioral responses and specifically considered redemption of the personalized offers. Personalization might also have other positive and negative consequences. On the positive side, personalization may result in more customer-centric offers and stronger customer attitudes toward the firm, inducing long-term loyalty outcomes. On the negative side, personalization may be perceived as intrusive and may result in privacy concerns as well as distrust in the firm (Van Doorn and Hoekstra 2013). LP managers should carefully weight these potential benefits and costs. Research has also shown that developing a clear privacy policy and specifically providing a simple control function to protect one's privacy can help ensure personalization and result in positive outcomes (Tucker 2014).

5.4 LP Design

5.4.1 The General LP Design

The design of LPs can feature several key elements: (1) the overall program structure, (2) participation requirements, (3) the point structure, and (4) the reward structure (Breugelmans et al. 2015). Across all LPs, the most common types are those that firms introduce by themselves versus with partners. Here, customers sign up for the program and collect points according to their purchase behavior. LP members can also receive additional benefits from the program, such as direct discounts, personalized communication, improved service levels, and so on. In principle, the program is initiated without an ending date. The role and importance of the three effect mechanisms (Fig. 5.1) may vary considerably with the LP design. Prior research has examined a broad range of these design elements and the consequences for LP effectiveness (for detailed reviews, see Bijmolt et al. 2011; Breugelmans et al. 2015). Here, we discuss three important LP types that differ in the overall program structure from this basic LP design and which have been used in practice and studied in the literature: (1) programs that have an explicit ending date, (2) programs that are a collaboration between multiple firms, and (3) hierarchical loyalty programs (HLPs) that classify customers into tiers.

5.4.2 Short-Term Programs

In recent years, many retailers, especially supermarket chains, have used programs that run for a limited time (often several months) instead of for infinity. This limited period length is a key element of the program and explicitly communicated to customers. To some extent, these short-term programs lie between the regular long-term, continuous LP design and traditional promotions that typically last for one or two weeks. However, such short-term LPs do reward and stimulate repeat purchasing, so they may have an impact on customer loyalty. Two types of short-term programs can be distinguished.

First, firms can use a points-saving format. Here, however, customers are confronted with a deadline for collecting sufficient points to redeem the rewards. This time limit causes additional pressure for customers to meet the threshold criteria and to adjust their purchase behavior accordingly, which we previously referred to as the points pressure mechanism (see Sect. 5.3.2). The shorter, finite horizon of the program will make the points pressure mechanism more imperative for short-term programs than regular, continuous LPs. Empirical studies have found a relatively strong impact on sales during the period the program is running, and retailers tend to value such a short-term boost in sales. For example, Taylor and Neslin (2005) report a 6% increase in average expenditure levels during the eight-week period of the program, and Drèze and Hoch (1998) show an increase in category sales of 25%. The effect is often relatively small for the best customers, because purchase levels of these customers offer little to no room for improvement, even though these customers are most likely to redeem the reward (Lal and Bell 2003; Liu 2007). However, the effect after the program period might be limited, or even negative—comparable to a sales dip after a temporary price discount in sales promotion settings. Repeating such short-term programs might be a fruitful approach, but the dynamic effects of repeated programs on customer loyalty are still unclear.

Future research should assess the sales effect after the program period and especially the long-term impact of short-term programs.

Second, a rapidly growing form of short-term LPs rewards consumers instantly with small premiums per fixed spending. Often these premiums are part of a set of collectibles. Such programs have been labeled as instant reward programs (IRPs; Minnema et al. 2017). Minnema et al. (2017) examine various IRPs related to the world cup soccer in 2010 and run by supermarket chains in the Netherlands. The IRPs rewarded consumers directly at the point of purchase, with premiums at fixed spending levels. These premiums were small toy figures related to the world cup soccer and had little to no monetary value but formed a set of collectibles. The results showed that the IRP led to incremental shopping trips. Furthermore, the IRP was more effective for households that reported collecting the premiums. Further research is necessary to assess the moderating role of IRP design elements, such as the number of collectibles in the set, different types of collectibles, and length of the program period. Furthermore, the effect outside the program period might be limited, and future research should examine such effects for IRPs over time.

5.4.3 Multi-vendor Programs

Firms' relationship marketing investments, such as LPs, are primarily intended to enhance customer loyalty to an individual firm. Yet coalitions are formed in so-called multi-vendor LPs, with tens or even hundreds of firms involved (e.g. Nectar in the United Kingdom, Payback in Germany, Airmiles in Canada and the Netherlands), often from diverse sectors (e.g. groceries, gasoline, apparel, airlines, credit cards). Partnerships in multi-vendor LPs are becoming more prominent and are likely to increase over time (Capizzi and Ferguson 2005).

For LP members, multi-vendor LPs offer convenience, faster point accumulation, and more redemption options than single-vendor LPs, because customers who are LP members can earn and/or redeem LP rewards from multiple firms while carrying only a single card. Thus, purchasing from multiple partners brings customers economic benefits of increased points collection and redemption opportunities.

Firms are attracted to multi-vendor LPs because such programs can offer strategic networking and cost advantages over establishing and running a regular, sole-proprietary LP. In addition, a multi-vendor LP may provide spillover effects between firms from their affiliation with the partnership LP, as consumers are likely to cross-patronize firms within the LP partnership. With the broad set of firms participating in the LP, each firm will have access to a wider customer base that may provide them with new customers and increase the engagement of their current customers. Rese et al. (2013) indeed find a positive effect of a multi-vendor program on the attraction of new customers from the customer database of partnering firms. The effect of the multi-vendor program on customer retention, however, is not significant, whereas a single-vendor LP has positive effects on retention.

Lemon and von Wangenheim (2009) show that customer usage and satisfaction with the core service increase cross-buying from complementary partners, which reinforces usage of the core service. However, these spillover effects are limited to highly complementary partners. Dorotic et al. (2011) show no significant spillover effects of promotional mailings of one partner on sales at partner firms in a multi-vendor LP. In addition to behavioral effects, multi-vendor LPs may have an impact on customer attitudes toward partners through spillover of a positive (or negative) image or satisfaction between partners and between the program and its partners (Lemon and von Wangenheim 2009; Schumann et al. 2014). With increasing digitalization and collaboration in marketing (see Sect. 5.6), we expect that multi-vendor LPs will increase in importance. Whether and how these new developments will help improve effectiveness of multi-vendor LPs in terms of increasing attitudinal and behavioral loyalty requires detailed insights from scientific research.

5.4.4 Hierarchical Loyalty Programs

A hierarchical loyalty program (HLP) explicitly stratifies the customer base into a hierarchical set of status tiers, to affect the behavior and attitudes of customers. Typically, customers are classified into these status tiers on the basis of past purchase behavior (Blattberg et al. 2008). The status levels are often labeled with classifications such as bronze, silver, gold, and platinum to further reinforce the notion of a hierarchy and provide observable indicators of status (Drèze and Nunes 2009; Melnyk and van Osselaer 2012). The thresholds for the hierarchical status tiers consist of predefined requirements and are communicated to the customer base. Next, an HLP provides benefits to the customers, and these benefits increase between the customer tiers. Thus, higher-status levels are accompanied by preferential treatments, such as special offers, larger rewards, lower prices, better service, and so on, with the idea to reward and stimulate high customer loyalty levels.

An HLP can have a positive impact on customer behavior and attitudes, with the impact increasing across levels, for several reasons. Status-related (soft) benefits are powerful instruments for stimulating customer loyalty (Henderson et al. 2011). Therefore, acknowledging the special status of top customers reinforces their customer loyalty (Drèze and Nunes 2009), and this effect may be stronger for men than for women (Melnyk and van Osselaer 2012). In addition, the increasing levels of hard benefits, such as larger rewards, more points, better service, or lower prices, for customers in higher-status tiers should encourage customers to try to reach and maintain a high-status tier. Yet results from empirical studies on HLP effects are mixed; studies find evidence for positive effects (Kopalle et al. 2012; Wang et al. 2016), negative effects (Steinhoff and Palmatier 2016; Wagner et al. 2009), or both depending on the context (Eggert et al. 2015).

Negative effects of an HLP may arise from several potential pitfalls. First, explicit preferential treatment implies that some customers receive greater benefits while others are treated less favorably. This treatment can be perceived as unfair and may lead to negative bystander effects for those not receiving the special treatment (Steinhoff and Palmatier 2016). Here, it is important to communicate the rules for differential customer treatment clearly and explicitly. Second, a customer can move up or down the hierarchy of the status levels, such that going down in status might have a negative impact on customer behavior and attitudes. These status dynamics can be due to changes in the customer's own behavior (Van Berlo et al. 2014; Wagner et al. 2009), to changes in hierarchical program design (e.g. status thresholds; Drèze and Nunes 2009), or to a "free" promotion or demotion induced by the firm (Eggert et al. 2015). The negative effects of a demotion may even be larger than the positive effects, and thus the final outcome for a customer who has been promoted and then demoted will be lower than for customers whose status never changed (Van Berlo et al. 2014; Wagner et al. 2009). Finally, acknowledging that a customer is a "gold" member may increase customer expectations (Von Wangenheim and Bayon 2007), as the customer may anticipate being treated as such. In the well-known expectation–disconfirmation framework, satisfaction is the outcome of a comparison between expectations and perceived performance levels. Thus, if the actual service levels do not change or change less than the customer expected, customer satisfaction will decrease. As a consequence, customers in higher tiers tend to be particularly sensitive to service failures and loss of status (Von Wangenheim and Bayon 2007). All these pitfalls do not match with the original intention of introducing the HLP. Future research should address HLPs' potential advantages and disadvantages to explore how changes in status levels may affect customer attitudes and behavior and how modern designs of HLPs (see Sect. 5.6) can enhance the positive aspects and circumvent the negative ones.

5.5 Modeling LP Effects

With the increasing complexity of LP designs (see Sect. 5.4) and trends such as digitalization (see Sect. 5.6), it has become more relevant to know whether and how a broad range of LP features will lead to positive or negative synergies in terms of enhancing attitudinal and behavioral loyalty. Furthermore, the three effect mechanisms (Fig. 5.1) do not act independent of one another and should be examined in an integrated framework. Fortunately, LPs produce massive databases with details on whether and when customers have received what offers and communications, whether and when they have redeemed a reward, whether and when they have purchased, and, most important, what purchases they have made. Such databases can be exploited for two main purposes. First, firms can assess the effects of their LPs on customer behavior through the three mechanisms mentioned previously (see Fig. 5.1), possibly moderated by the LP design. Second, they can use LP analytics to optimize marketing-mix decisions.

With a few exceptions (e.g. Dorotic et al. 2014; Kopalle et al. 2012; Taylor and Neslin 2005), empirical studies have examined a limited number of effect mechanisms and design components of LPs. Future research should assess simultaneously the role of multiple effect mechanisms, possibly moderated by various LP design elements. Doing so will require complex, longitudinal databases and advanced econometric methods, along the lines of other "big data" challenges. In particular, the analysis model should accommodate the dynamics of the underlying processes, in which purchase behavior, points collection, reward redemption, and marketing communication are interrelated.

As an example of such a comprehensive model, we discuss the work of Dorotic et al. (2014). The researchers account for the dependencies between various behaviors: purchasing and collecting of points and reward redemptions. As such, they formulate a simultaneous model for purchase and redemption decisions of customers and apply Bayesian estimation techniques to estimate the model.

Dorotic et al. (2014) use the database of a large multi-vendor LP in the Netherlands. They examine a period of 3.5 years, and weekly information is available for more than 3,000 LP members making purchases, collecting points, and redeeming rewards. Data are aggregated across the coalition partners. Customer who sign up for the program can collect points at a broad range of coalition partners, including online and offline retailers. The points collected are a direct function of the purchase amount (one point per euro spent) and therefore form a good proxy for customer purchase behavior. In addition, the LP members receive coupons, direct mailings (intended to stimulate purchases and/or redemptions), and occasionally bonus points for specific types of purchase behavior. An important feature of the programs is that, in principle, the points do not expire (they do so only if program management does not observe any activity by an LP member for a period of more than one year). In addition, many LP members have earned a large number of points over the years and can decide themselves whether and when to redeem a reward. Furthermore, a variety of rewards are available, from somewhat low to very large numbers of points: the database contains redemptions of 100–60,000 points. On average, a customer redeems approximately 25% of his or her points when redeeming; in only 3% of the cases is the redemption more than 90% of the points collected by that customer. Thus, the traditional points pressure argument for pre-redemption effects (i.e. just before redemption a customer increases purchases to attain the required number of points) is not likely to hold.

Dorotic et al. (2014) build a comprehensive model for members' purchase and redemption decisions. In particular, they model four dependent variables: purchase incidence, purchase amount, redemption incidence, and redemption fraction. The redemption fraction is computed on a weekly basis by dividing the number of points redeemed that week by the number of points available for the LP member; the latter is computed by the previous balance of points plus the points collected that week less the points redeemed that week less the points deducted from product returns. Purchase incidence is modeled by means of a Probit equation and purchase amount (after an Ln-transformation) by a regression equation with a Normal distribution. Similarly, redemption incidence is modeled by means of a Probit model and the redemption fraction (after a Logit-transformation) by a regression equation with a Normal distribution. Furthermore, the purchase amount and redemption fraction models are estimated conditional on the corresponding incidence equations.

Dependencies between the purchase and redemption decisions are achieved in several ways. First, the balance of points connects both decisions, because purchases add points to the balance and redemption depends on this balance of points. Second, whether or not the LP member has decided to redeem points is included as an explanatory variable in the purchase equations. This captures potential redemption momentum effects: do the LP members change their purchase behavior after making the decision to redeem? Third, all parameters are treated as random effects, and this heterogeneity is modeled using member-specific variables and normally distributed error terms, where these error terms are correlated across the four model equations. Because of model complexity and the inclusion of many member-specific parameters, parameter estimates are obtained by means of Bayesian methods.

The model includes the number of direct mailings as one of the main explanatory variables. LP management sends these mailings to a selection of LP members to stimulate purchases and redemptions. Thus, this variable should be treated as endogenous. The decision of which LP members receive the mailings is based on specific (unknown) criteria, but the timing of the mailings is less systematic. Therefore, Dorotic et al. (2014) deal with this type of endogeneity by including the average number of mailings as an additional explanatory variable, also known as the Mundlak correction in panel data econometrics (Mundlak 1978).

The results show that direct mailings have a significant, positive effect on purchase incidence and amount, such that the effect on purchase amount increases with the duration of LP membership. Similarly, direct mailings have positive effects on redemption incidence and fraction. A key finding of the article is that for the majority of LP members (approximately 70%), the redemption decision precedes the purchase decision. The actual redemption typically follows shortly after the LP member has decided to redeem, but Dorotic et al. (2014) find support for the redemption momentum effects, because if an LP member has decided to redeem, his or her purchase probability and purchase amount both increase. These positive effects on purchase behavior also hold for the week after the redemption. The authors find significant, positive pre- and post-redemption effects on purchase behavior of members of continuous LP, where the LP member decides whether, when, and how much to redeem.

In addition to examining the LP effect mechanisms, firms can exploit the LP-based customer database to assess the effectiveness of marketing-mix instruments and to improve (personalized) marketing efforts. Again, advanced analytical approaches are required. As the dependent variables of these models will reflect whether, when, and how much each customer will purchase, a broad range of econometric methods can be applied. In particular, hierarchical Bayesian estimations methods will prove useful because customer- or store-specific predictions are required. The studies we discuss in Sect. 5.3.4 on LP-based personalization provide excellent examples of such models.

Here, as an example of using LP databases to assess marketing-mix effects, we discuss the work of Van Heerde and Bijmolt (2005), in which the researchers examine the effects of sales promotions on the number of buyers and shopping basket sizes. Van Heerde and Bijmolt examine six outlets from a chain of clothing stores. The chain runs an LP in which approximately 70% of sales come from LP members. LP members can collect points (worth 5% of the purchase value), receive discounts through coupons, and receive targeted direct mailings. Van Heerde and Bijmolt combine data from the LP-based customer database, data from an infrared traffic counter at the store entrance, sales registered at the counter, and marketing-mix data. The marketing-mix data consist of two parts: direct mail (DM) targeted to LP members and door-to-door flyers (DtD) distributed to all customers within the store trade area.

Van Heerde and Bijmolt (2005) decompose total daily sales at the store level, S_{it}, into sales to LP members, SL_{it}, and sales to non-members, SN_{it}. Next, they split each of the two components further into number of buyers (CL_{it} and CN_{it}) and average purchase amount (EL_{it} and EN_{it}), as follows:

$$S_{it} = SL_{It} + SN_{it} = CL_{it} \times EL_{it} + SN_{It} \times EN_{it}. \tag{5.1}$$

Furthermore, they consider the number of store visitors who do not make a purchase, which is approximated by the number of visitors registered by the traffic counter at the store entrance less the number of buyers (CL_{it} and CN_{it}). This variable is a potential source of additional sales, but it may also have negative effects, due to in-store crowding effects. Therefore, this variable is included as an additional dependent variable and as an explanatory variable for the other four dependent variables. Van Heerde and Bijmolt model each of the final five components, after an Ln-transformation, as a function of environmental factors, time-related variables, marketing-mix variables, and other explanatory variables. Furthermore, they model the parameters as random effects to account for heterogeneity and estimate the four equations simultaneously in a system of equations. They obtain model estimates by means of hierarchical Bayesian methods. The model allows for assessing differential effects of DM and DtD flyers on the number of buyers versus expenditure per buyer, for both customer groups (LP members and non-members). These DM and DtD campaigns typically run for two weeks and have an average discount of approximately 30%.

The results show that the number of non-buyers on day $t - 1$ leads to a small but statistically significant increase in the number of buyers on the next day, both CL_{it} and CN_{it}. However, the results do not support crowding effects, because more store traffic does not lead to lower (or higher) average expenditures.

The promotion variables (DM to LP members and DtD flyers) serve as dummy variables in the model. Van Heerde and Bijmolt (2005) find that DM has a considerable impact on the number of LP members who make a purchase (+30%). The number of non-members who make a purchase increases (+16%), which might be due to in-store communication supporting the DM campaign. The effect of DtD flyers is even larger: +73% for LP members and +75% for non-members.

The decomposition of sales and profit effects yields several valuable insights. First, the discount percentage is a major driver of the effect size. The number of buyers increases considerably with increasing discount percentages, but the average expenditures decrease considerably as well. Second, for comparable discount levels, the DtD campaigns contribute more to the increase in sales than the DM campaigns. Third, a large proportion of the additional sales comes from non-members, and this proportion increases with the discount percentage. Finally, the net profit contribution is positive for the DtD campaigns and negative for the DM campaigns. For both types of campaigns, the optimal discount level is at 20%, which is lower than the regular discount levels used by the retailer.

5.6 Emerging Trends

5.6.1 Digitalization of LPs

One emerging trend in business and marketing is the increasing digitalization, which provides both challenges and opportunities for firms (e.g. Leeflang et al. 2014), including in the area of LPs. LPs have typically been developed in non-digital environments to stimulate, for example, offline store loyalty. LPs were also a response of firms in many mature markets to the relatively low growth levels, with many firms moving from a focus on growth through customer acquisition to a more defensive loyalty focus aimed to keeping customers (Reichheld and Thomas 1996).

Digital market growth rates are still very high, and thus many online firms still focus on customer acquisition. In digital environments, there has been a strong focus on more short-term outcomes. Online retailers aim to influence the purchase funnel and mainly focus on website visits and converting these visits to purchases (conversion rate). In addition, firms are using attribution models to understand the effects of different touchpoints on these outcomes (Li and Kannan 2014). There is less focus on loyalty, though consultants from McKinsey have strongly advocated the experience-loyalty loop (Edelman and Singer 2015; Lemon and Verhoef 2016). Given the lack of LPs in online environments, researchers also lack knowledge on the effects of these programs in an online environment. LPs might be less effective for online retailers, given the presence of more information on prices, assortments, and quality ratings online and customers becoming more experienced with online purchasing (Melis et al. 2015). Moreover, the data benefits of LPs are absent, given that online retailers have already build up customer databases through their direct business models. Further research is required on LPs in the online setting, especially on whether and how effectiveness may depend on the actual implementation of these programs.

Needless to say, traditional offline LPs have been challenged to include digital elements in their design and execution. Recently, Cap Gemini concluded that firms offering LPs have a lot of catching up to do in the digital area.[2] What we typically observe is that the digital channel is mainly used as a communication channel to members. For example, the Flying Blue LP of Air France/KLM uses websites and apps for members to log into and sends (personalized) e-mails with specific offers and status updates (e.g. so many flights required to become gold members). However, this is combined with offline communication as well, such as a magazine. Thus, this LP uses a typical multi-channel approach (Neslin et al. 2006). There are also examples of firms that integrate new media into their LPs. For example, the telecom provider O_2 provides location-based offers to its LP members on their phone. The Turkish telecom Turkcell Gnctrkcll uses social media to listen to members and to improve its reward system (Cap Gemini 2015).

5.6.2 Aligning LPs with the Customer Experience

Building on the digital trend, a new emerging development is the use of LPs to foster the total customer experience. Traditional programs were not developed to manage the total customer experience across touchpoints. Today, many firms strive to create memorable experiences for their customers (Schmitt 2003; Lemon and Verhoef 2016). LPs have moved from a typical transaction-oriented program to an integrated system intended to provide a strong customer experience and to influence repeat purchase behavior and cross-selling. Technological advances allow for highly individualized elements in the LP design and communication, throughout the individual-specific customer journey. Furthermore, to keep customers active in the LP, firms integrate customer behavior other than purchases—for example, online word of mouth—and use the LP to stimulate customer engagement.

Specifically, the leisure and entertainment industry is full of new, strongly digital-based programs, in which firms use LPs to enhance customer experiences. For example, in Slovakia a ski resort has developed a program called GoPass, which entails a loyalty card as well as a kind of payment device that members can use to pay for skiing, hotels, and other services. The system uses digital communication in which customers can observe their skiing performance. The program initiator Milan Schnorrer of PriceWise argues that the LP should provide three general benefits[3]: (1) rational, (2) comfort, and (3) emotional. For example, this program offers not only regular rational benefits (e.g. reduced price) but also improved comfort, as it makes life much easier for tourists (e.g. less waiting time at the cashier). For the emotional element, motivation schemes are used to obtain emotional rewards for gaining a stronger skiing performance by participating in a

[2]See https://www.capgemini-consulting.com/reinventing-loyalty-programs.

[3]We gratefully thank Milan Schnorrer, a past graduate of the M.Sc. Marketing program at the University of Groningen, for sharing this case, which he discussed in a guest lecture in the M.Sc. Marketing course customer management (Schnorrer 2015).

skiing challenge (e.g. achieving 'King of the Mountain Status' based on total number of miles skied). The most motivated members are likely to return to move from one status to the next. These customer experience-oriented digital solutions do not need to be combined with an LP. For example, Disney provides the magic band, which uses RFID technology, to visitors to the park, which improves the comfort of visiting Disney (i.e. payment) and may also improve emotions through, for example, the provision of personalized pictures of customers' visit to the park afterward. However, as the GoPass example indicates, these systems can be easily combined with a loyalty scheme, because these systems are personally linked to a provider.

5.6.3 Reviving Existing LPs

Many LPs are now almost 20 years old, originating, as mentioned, in the 1990s. To date, LP research has focused on traditional LPs, but, as discussed previously, these programs need to adapt to the new digital environment. Many of these old programs are also confronted with an aging membership. Moreover, younger consumers may perceive LPs differently than the early members of these programs. To remain attractive and effective in influencing loyalty, the old programs need to be renewed. So far, we have observed incremental changes in saving rules and rewards; however, more radical changes are likely required for these old programs to survive. LP research could focus on which strategies LP managers could use to renew existing LPs and make them attractive for consumers in modern, digitalized markets. Furthermore, research should investigate the effectiveness of these new programs, specifically examining multiple outcome measures (i.e. both experience measures and purchase behavior).

5.7 Conclusion

In this chapter, we have discussed LPs as one of the most prominent instruments to improve customer relationships and foster customer loyalty. In the past decades, we have observed a growing stream of research on LPs, due to the increasing use of LPs in practice and data availability. Moreover, there were doubts about the effectiveness of LPs (Dowling and Uncles 1997), which called for insights based on research. In this chapter, we did not aim to provide a comprehensive overview of the literature (see Bijmolt et al. 2011; Breugelmans et al. 2015), but rather to specifically discuss some important topics. LP research is now in its mature stage. Much is known about LPs' effectiveness and design, and thus future research might contribute by filling in any remaining gaps or building on what is known so far.

In this chapter, we also observe some areas in which research can still contribute (see Table 5.2). In general, these topics are a bit more specific (i.e. short-term LPs,

Table 5.2 Summary of future research areas for LPs

Topic	Research questions
Customer responses	• How do various LP features lead to positive or negative synergies in terms of enhancing attitudinal and behavioral loyalty?
Reward effects	• What is the long-term impact of (post-)reward effects? • What are specific moderators of (post-)reward effects?
LP analytics	• How should LPs personalize their offers to be more effective, though not intrusive?
LP design	• What is the effect of more complex integrated LPs on customer loyalty and the entire customer experience? Are there positive synergies between various elements of LPs? • How can LPs be combined or even integrated with other marketing-mix instruments, and what is the effect of such combinations?
Short-term LP	• What is the long-term impact of short-term programs, and what is the impact of repeating short-term programs over time? • What is the impact of IRP design elements, such as the number of collectibles in the set, different types of collectibles, and length of the program period, on the effectiveness of short-term programs?
LP and digitalization	• How can LPs be effective in online environments? • What are the actual benefits of LPs for online retailers, and do they differ from those of LPs of offline retailers? • How can firms use LPs in multi-channel integration?
LP and customer experience	• How can firms adapt programs to improve the customer experience? • What are effects of customer experience—aligned programs on customer experience measures as well as purchase behavior?
Reviving LPs	• Which strategies should firms use to successfully renew existing LPs? • How do today's consumers view LPs in general and how do new generations of consumers (i.e. generation Y) perceive the potential benefits of LPs in particular?

multi-vendor LPs, and HLPs) or consider emerging themes (e.g. digitalization, customer experience). We contend that the emerging trends on digitalization and those related to new and more integrated LP designs are the most fruitful research areas on which to focus. Digitalization will definitely affect LP effectiveness. Moreover, integrated customer programs will add new dimensions to traditional card-based LPs. We hope to see contributions on these important emerging developments.

References

Berry, L.L. 1995. Relationship marketing of services—Growing interest, emerging perspectives. *Journal of the Academy of Marketing Science* 23 (4): 236–245.

Bijmolt, T.H.A., M. Dorotic, and P.C. Verhoef. 2011. Loyalty programs: Generalizations on their adoption, effectiveness and design. *Foundations and Trends® in Marketing* 5 (4): 197–258.

Blattberg, R.C., B.D. Kim, and S.A. Neslin. 2008. *Database marketing: Analyzing and managing customers*. New York: Springer.

Breugelmans, E., T.H.A. Bijmolt, J. Zhang, L.J. Basso, M. Dorotic, P. Kopalle, A. Minnema, W. J. Mijnlieff, and N.V. Wünderlich. 2015. Advancing research on loyalty programs: A future research agenda. *Marketing Letters* 26 (2): 127–139.

Cap, Gemini. 2015. Fixing the cracks: Reinventing loyalty programs for the digital age. https://www.capgemini-consulting.com/reinventing-loyalty-programs. Assessed 1 May 2016.

Capizzi, M.T., and R. Ferguson. 2005. Loyalty trends for the twenty-first century. *Journal of Consumer Marketing* 22: 72–80.

Chung, T.S., R.T. Rust, and M. Wedel. 2009. My mobile music: An adaptive personalization system for digital audio players. *Marketing Science* 28 (1): 52–68.

Chung, T.S., and M. Wedel. 2014. Adaptive personalization of mobile information services. In *Handbook of service research*, ed. R. Rust and M.H. Huang, 395–412. Cheltenham, UK: Edward Edgar Publishers.

Dick, A.S., and K. Basu. 1994. Customer loyalty: Towards an integrated conceptual framework. *Journal of the Academy of Marketing Science* 22 (2): 99–113.

Dorotic, M., P.C. Verhoef, T.H.A. Bijmolt, and D. Fok. 2011. Do vendors benefit from promotions in a multi-vendor loyalty program?. *Marketing Letters* 22 (4): 341–356.

Dorotic, M., P.C. Verhoef, D. Fok, and T.H.A. Bijmolt. 2014. Reward redemption effects in a loyalty program when customers choose how much and when to redeem. *International Journal of Research in Marketing* 31 (4): 339–355.

Dowling, G.R., and M. Uncles. 1997. Do customer loyalty programs really work? *Sloan Management Review* 38 (4): 71–82.

Drèze, X., and S.J. Hoch. 1998. Exploiting the installed base using cross-merchandising and category destination programs. *International Journal of Research in Marketing* 15 (5): 459–471.

Drèze, X., and J.C. Nunes. 2009. Feeling superior: The impact of loyalty program structure on consumers' perceptions of status. *Journal of Consumer Research* 35 (6): 890–905.

Drèze, X., and J.C. Nunes. 2011. Recurring goals and learning the impact of successful reward attainment on purchase behavior. *Journal of Marketing Research* 48 (April): 268–281.

Edelman, D.C., and M. Singer. 2015. Competing on customer journeys. *Harvard Business Review* (November): 88–94.

Eggert, A., L. Steinhoff, and I. Garnefeld. 2015. Managing the bright and dark sides of status endowment in hierarchical loyalty programs. *Journal of Service Research* 18 (2): 210–228.

Fang, Zheng, X. Luo, and M. Keith. 2015. how effective is location-targeted mobile advertising? *MIT Sloan Management Review* 56 (2): 14–15.

Gable, M., S.S. Fiorito, and M.T. Topol. 2008. an empirical analysis of the components of retailer customer loyalty programs. *International Journal of Retail and Distribution Management* 36 (1): 32–49.

Henderson, C.M., J.T. Beck, and R.W. Palmatier. 2011. Review of the theoretical underpinnings of loyalty programs. *Journal of Consumer Psychology* 21 (3): 256–276.

Humby, C., T. Hunt, and T. Phillips. 2008. *Scoring points. How Tesco continues to win customer loyalty*. London: Kogan Page.

Keh, H.T., and Y.H. Lee. 2006. Do reward programs build loyalty for services? The moderating effect of satisfaction on type and timing of rewards. *Journal of Retailing* 82 (2): 127–136.

Kivetz, R., O. Urminsky, and Y. Zheng. 2006. The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research* 43 (1): 39–58.

Kopalle, P.K., S.A. Neslin, B. Sun, Y. Sun, and V. Swaminathan. 2012. the joint sales impact of frequency reward and customer tier components of loyalty programs. *Marketing Science* 31 (2): 216–235.

Lal, R., and D.E. Bell. 2003. The impact of frequent shopper programs in grocery retailing. *Quantitative Marketing and Economics* 1 (2): 179–202.

Leeflang, P.S.H., P.C. Verhoef, P. Dahlström, and T. Freundt. 2014. Challenges and solutions for marketing in a digital era. *European Management Journal* 32 (1): 1–12.

Leenheer, J., and T.H.A. Bijmolt. 2008. Which retailers adopt a loyalty program? An empirical study. *Journal of Retailing and Consumer Services* 15 (6): 429–442.

Leenheer, J., H.J. van Heerde, T.H.A. Bijmolt, and A. Smidts. 2007. Do loyalty programs really enhance behavioral loyalty? An empirical analysis accounting for self-selecting members. *International Journal of Research in Marketing* 24 (1): 31–47.

Lemon, K.N., and P.C. Verhoef. 2016. Understanding the customer experience and customer journey. *Journal of Marketing* 80 (6): 69–96.

Lemon, K., and F. von Wangenheim. 2009. The reinforcing effects of loyalty program partnerships and core service usage. *Journal of Service Research* 11 (4): 357–370.

Li, H., and P.K. Kannan. 2014. Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research* 51 (1): 40–56.

Liu, Y. 2007. The long-term impact of loyalty programs on consumer purchase behavior and loyalty. *Journal of Marketing* 71 (October): 19–35.

Liu, Y., and R. Yang. 2009. competing loyalty programs: Impact of market saturation, market share, and category expandability. *Journal of Marketing* 73 (1): 93–108.

Melis, K., K. Campo, E. Breugelmans, and L. Lamey. 2015. The impact of the multi-channel retail mix on online store choice: Does online experience matter? *Journal of Retailing* 91 (2): 272–288.

Melnyk, V., and T.H.A. Bijmolt. 2015. The effects of introducing and terminating loyalty programs. *European Journal of Marketing* 49 (3/4): 398–419.

Melnyk, V., and S.M.J. van Osselaer. 2012. Make me special: Gender differences in consumers' responses to loyalty programs. *Marketing Letters* 23 (3): 545–559.

Meyer-Waarden, L., and C. Benavent. 2009. Retail loyalty program effects: Self-selection or purchase behavior change?. *Journal of the Academy of Marketing Science* 37 (3): 345–358.

Minnema, A., T.H.A. Bijmolt, and M.C. Non. 2017. The impact of instant reward programs and bonus premiums on consumer purchase behavior. *International Journal of Research in Marketing* 34 (1): 194–211.

Mundlak, Y. 1978. On the pooling of time series and cross section data. *Econometrica* 46 (1): 69–85.

Neslin, S.A., D. Grewal, R. Leghorn, V. Shankar, M.L. Teerling, J.S. Thomas, and P.C. Verhoef. 2006. Challenges and opportunities in multichannel customer management. *Journal of Service Research* 9 (2): 95–112.

Palmatier, R.W., C.B. Jarvis, J.R. Bechkoff, and F.R. Kardes. 2009. The role of customer gratitude in relationship marketing. *Journal of Marketing* 73 (5): 1–18.

Payne, A., and P. Frow. 2005. A strategic framework for customer relationship management. *Journal of Marketing* 69 (4): 167–176.

Reichheld, F., and T.E.A.L. Thomas. 1996. *The loyalty effect: The hidden force behind growth, profits, and lasting value*. Boston: Harvard Business School Press.

Reinartz, W., M. Krafft, and W.D. Hoyer. 2004. The customer relationship management process: Its measurement and impact on performance. *Journal of Marketing Research* 41 (3): 293–305.

Reinartz, W., J.S. Thomas, and V. Kumar. 2005. Balancing acquisition and retention resources to maximize customer profitability. *Journal of Marketing* 69 (1): 63–79.

Rese M., A. Hundertmark, H. Schimmelpfennig, and L.M. Schons. 2013. Loyalty program types as drivers of customer retention: A comparison of stand-alone programs and multi-vendor loyalty programs through the lens of transaction cost economics. *International Review of Retail Distribution and Consumer Research* 23 (3): 305–323.

Schmitt, B. 2003. *Customer experience management: A revolutionary approach to connecting with your customers*. New York: Wiley.

Schnorrer, M. 2015. Slides guest lecture customer management, Nov 2015, University of Groningen, The Netherlands. http://www.pricewise.sk/en/gopass. Assessed 1 May 2016.

Schumann, J.H., N.V. Wünderlich, and H. Evanschitzky. 2014. spillover effects of service failures in coalition loyalty programs: The buffering effect of special treatment benefits. *Journal of Retailing* 90 (1): 111–118.

Sheth, J.N., and A. Parvatiyar. 1995. Relationship marketing in consumer markets: Antecedents and consequences. *Journal of the Academy of Marketing Science* 23 (4): 255–271.

Smith, A., and L. Sparks. 2009. It's nice to get a wee treat if you've had a bad week: Consumer motivations in retail loyalty scheme points redemptions. *Journal of Business Research* 62: 542–547.

Smith, A., L. Sparks, S. Hart, and N. Tzokas. 2004. Delivering customer loyalty schemes in retailing: Exploring the employee dimension. *International Journal of Retail and Distribution Management* 32 (4): 190–204.

Steinhoff, L., and R.W. Palmatier. 2016. Understanding the effectiveness of loyalty programs: Managing target and bystander effects. *Journal of the Academy of Marketing Science* 44: 88–107.

Stourm, V., E.T. Bradlow, and P.S. Fader. 2015. Stockpiling points in linear loyalty programs. *Journal of Marketing Research* 52 (2): 253–267.

Taylor, G.A., and S.A. Neslin. 2005. The current and future impact of a retail frequency reward program. *Journal of Retailing* 81 (4): 293–205.

Tucker, C.E. 2014. Social networks, personalized advertising, and privacy controls. *Journal of Marketing Research* 51 (5): 546–562.

Van Berlo, G., J. Bloemer, and V. Blazevic. 2014. customer demotion in hierarchical loyalty programmes. *The Service Industries Journal* 34 (11): 922–937.

Van Doorn, J., and J.C. Hoekstra. 2013. Customization of online advertising: The role of intrusiveness. *Marketing Letters* 24 (4): 339–351.

Van Heerde, H.J., and T.H.A. Bijmolt. 2005. Decomposing the promotional revenue bump for loyalty program members versus nonmembers. *Journal of Marketing Research* 17: 443–457.

Van Heerde, H.J., M.J. Gijsenberg, M.G. Dekimpe, and J.-B.E.M. Steenkamp. 2013. Price and advertising effectiveness over the business cycle. *Journal of Marketing Research* 50 (2): 177–193.

Verhoef, P.C. 2003. Understanding the effect of customer relationship management efforts on customer retention and customer share development. *Journal of Marketing* 67 (4): 30–45.

Verhoef, P.C., P.K. Kannan, and J.J. Inman. 2015. From multi-channel retailing to omni-channel retailing: Introduction to the special issue on multi-channel retailing. *Journal of Retailing* 91 (2): 174–181.

Verhoef, P.C., E. Kooge, and N. Walk. 2016. *Creating value with big data analytics: making smarter marketing decision*. London: Routledge.

Von Wangenheim, F., and T. Bayon. 2007. Behavioral consequences of overbooking service capacity. *Journal of Marketing* 71 (4): 36–47.

Wagner, T., T. Hennig-Thurau, and T. Rudolph. 2009. Does customer demotion jeopardize loyalty? *Journal of Marketing* 73 (3): 69–85.

Wang, Y., M. Lewis, C. Cryder, and J. Sprigg. 2016. Enduring effects of goal achievement and failure within customer loyalty programs: A large-scale field experiment. *Marketing Science* 35 (4): 565–575.

Wirtz, J., A.S. Mattila, and M.O. Lwin. 2007. How effective are loyalty reward programs in driving share of wallet? *Journal of the Academy of Marketing Science* 9 (4): 327–334.

Yi, Y., and H. Jeon. 2003. Effects of loyalty programs on value perception, program loyalty, and brand loyalty. *Journal of the Academy of Marketing Science* 31 (3): 229–240.

Zeithaml, V.A., R.T. Rust, and K.N. Lemon. 2001. The customer pyramid: Creating and serving profitable customers. *California Management Review* 43 (4): 118–142.

Zhang, J., and E. Breugelmans. 2012. The impact of an item-based loyalty program on consumer purchase behavior. *Journal of Marketing Research* 49 (1): 50–65.

Zhang, J., and M. Wedel. 2009. The effectiveness of customized promotions in online and offline stores. *Journal of Marketing Research* 46 (2): 190–206.

Ziliani, C., and S. Bellini. 2004. From loyalty cards to micro-marketing strategies: Where is Europe's retail industry heading? *Journal of Targeting, Measurement and Analysis for Marketing* 12: 281–289.

Chapter 6
Structural Models in Marketing: Consumer Demand and Search

Pradeep Chintagunta

Over the past two decades structural models have come to their own in empirical research in marketing. The basic notion of appealing to economic theory when building models of consumer (e.g., Guadagni and Little 1983) and firm behavior (Horsky 1977; Horsky and Nelson 1992) in marketing has been around for much longer than that. Yet, this idea has come to the forefront as authors have confronted the challenges associated with drawing inferences from purely statistical relationships governing the behaviors of the agents of interest. While these relationships provide important insights into the correlational structure underlying the data, they are less useful when one is interested in quantifying the consequences of a change in either the structure of the market (e.g., what happens when a retailer closes down its bricks and mortar operations to focus solely on online sales) or in the nature of conduct of one or more players in that market (e.g., what happens to prices of car insurance when consumers change the ways in which they search for these prices). Since the economics underlying the conduct, or the behavior of agents in the presence of the structure, are not explicitly built into models that only focus on describing statistical relationships between agents' actions and outcomes, it is difficult if not impossible for those models to provide a prediction when one of these dimensions actually changes in marketplace.

As marketers move away from being focused only on "local" effects of marketing activities, e.g., what happens when I change price by 1%, in order to better

I thank Anita Rao and S Sriram for their useful comments on an earlier version. My thanks to the Kilts Center at the University of Chicago for financial support. Note that parts of this chapter appear elsewhere in "Handbook of Marketing Analytics with Applications in Marketing, Public Policy, and Litigation": (Edward Elgar; Natalie Mizik & Dominique M. Hanssens, Editors).

P. Chintagunta (✉)
Booth School of Business, The University of Chicago, 5807 S Woodlawn Ave,
60637 Chicago, IL, USA
e-mail: Pradeep.Chintagunta@chicagobooth.edu

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_6

understand the consequences of broader shifts in policy, the need for structural models has also grown. In this chapter, I will focus on a small subset of such "structural models" and provide brief discussions of what we mean by structural models, why we need them, the typical classes of structural models that we see being used by marketers these days, along with some examples of these models. My objective is not to provide a comprehensive review. Such an endeavor is far beyond my current purview. Rather, I would like to provide a basic discussion of structural models in the context of the marketing literature. In particular, to keep the discussion focused, I will limit myself largely to models of demand rather than models of firm behavior.

6.1 What Is a Structural Model?

The definition and key elements of a structural model have been well established, at least since the important chapter by Reiss and Wolak (2007). Other papers by Kadiyali et al. (2001), Chan et al. (2009), Chintagunta et al. (2004, 2006), Keane (2010) have also stayed close to this early work. And I will not depart in any significant way from the previous work that precedes this chapter and will draw heavily from that work. In simple terms, a structural model is an empirical model; one that can be taken to the data. But it is not any empirical model—since an equation that establishes a statistical relationship between a set of explanatory variables and an outcome variable is also an empirical model. What distinguishes a structural model is that the relationship between explanatory and outcome variables is based on theory—most often in economic theory—although it is not limited just to economic principles and can encompass theories from other disciplines such as psychology as well (Erdem et al. 2005; Teixeira et al. 2010; Sahni 2015). The theory for its part makes a prediction about the behavior of some set of economic agents (consumers, firms, etc.) and thereby governs how the outcome variable of interest is influenced by the explanatory variables. Thus the key ingredients of the model are the (economic) agents involved; the nature of their behavior (optimizing, satisficing, etc.); the interactions among the agents (e.g., competitive, collaborative, etc.) and the relationships between explanatory and outcome variables ensuing from such behavior. These ingredients stem from the researcher's beliefs about how they map onto the specific context of interest.[1]

Since theories make specific predictions, it is unlikely that these predictions about the explanatory and outcome variables can perfectly rationalize the actual data one observes on these variables in the market. The link between the predictions of the model and the observed outcome data is provided by the "unobservables" in the model. These unobservables essentially allow us to convert the economic

[1]A point to emphasize here relates to causality. If the researcher is interested only in establishing causality then a structural model per se may not be required (see e.g., Goldfarb and Tucker 2014).

(or some other theory-based) model into an econometric model, i.e., the final empirical model that we take to the data. These unobservables get their nomenclature from variables that are *unobserved to us as researchers but are, in general, known to the agents whose behavior is being modeled*. As we shall discuss below however, sometimes there are factors that are, a priori, not always known to the agents as well.

As Reiss and Wolak point out these unobservables can be of different forms. First, we have "structural" error terms—variables that belong to the set of explanatory variables in the economic model but constitute the subset that we do not observe as researchers. For example, we know that shelf space and shelf location are important determinants of a brand's market share in addition to price and advertising. But in many situations we do not have access to data on these variables. In such situations they become part of the unobservables and constitute "structural" error in the sense that they are directly related to the theory we are trying to create an empirical model for.

The second set of unobservables has a very long history in the marketing literature—unobserved heterogeneity. These unobservables help explain differences in the relationship between the explanatory and outcome variables across different agents whose behavior is being characterized by the structural model. For instance when looking at brand choice behavior, data patterns might reveal that one consumer is very price sensitive whereas another consumer is not. By allowing the consumers' utility parameters to differ from one another we can capture some of the deviations between the theoretical model and the data on hand across consumers in the market.

The third set of unobservables comes about in structural models that involve agent uncertainty about a specific parameter in the model. In these models, agents learn about the parameter they are uncertain about over time but usually have some prior belief about the parameter, often characterized via a distribution. Learning is a consequence of "signals" received by the agent (say a consumer) from another agent (say a firm) or from the environment that allows the former agent to update his/her belief about the uncertain parameter. As the agent receives more signals, the uncertainty gets resolved over time. While there exist instances where the researcher also observes the signals received by the agent, in most instances this is not the case. In such situations the signals received become part of the set of unobservables from the researcher's perspective.

Agents may also be uncertain about the values of specific attributes that drive their choices. Consider for example, a consumer who wants to renew his or her automobile insurance. The consumer may have well-formed preferences for the various options; hence does not have the unobservables described in the previous paragraph. Nevertheless, prior to choosing, the consumer may not know the prices charged by the various insurance companies. In this case the consumer may have to engage in (costly) search to uncover the values of these unobserved attributes (see e.g., Mehta et al. 2003). Based on whether the consumer has low or high search costs, (s)he may end up finding out the prices of many or few insurance companies in the market. This set of companies can be viewed as the consumer's "consideration set." Given knowledge of prices for these alternatives, the consumer then makes a choice from the subset for which the uncertainty is resolved. Thus, a priori, the consumer does not know the prices. (S)he then searches for these prices and forms a consideration set. This situation is then identical to that we described earlier where a subset of attributes that are observed by the consumer continue to be unobserved by the researcher.

A final set of unobservables comes from measurement error. For instance one might be interested in studying the relationship between the level of advertising received by a consumer and the purchases that might be caused by this advertising. In these cases, one observes advertising at a level different from the exact exposure that the consumer members receive. Rather, one might have proxies for advertising such as the expenditure on that activity in the market where the consumer resides or the average exposure of the specific demographic profile to which the consumer belongs. Such errors in the measurement of variables constitute another unobservable from the researcher's perspective.

The remainder of this chapter is organized as follows. First, I begin by giving a very simple example of a structural model that most marketers would be familiar with—the multinomial logit model of brand (or alternative) choice. I then walk the reader through each of the unobservables I describe above. A majority of this discussion is based on the assumption that researchers have access to data at the individual consumer level to estimate the model parameters.

1. Researcher unobserved but known-to-the-consumer unobservables.
2. Consumer heterogeneity in their preferences and responsiveness to marketing activities. Here I also make a short detour into the realm of aggregate data and how these models can be estimated with such data as well.
3. Consumer uncertainty about certain model parameters—e.g., their preferences for various brands in a category that might be new to them. This is applicable for the class of "experience" goods (defined later).
4. Consumer uncertainty about certain drivers of their choices prior to purchase—e.g., prices for the various alternatives.

I then discuss some applications of structural models and point to possible future directions of research.

6.2 Structural Models: A Simple Illustration

I begin with the classic brand choice model that has been ubiquitous in marketing and that is based on the model of consumer utility maximization. I use the framework from Deaton and Muellbauer (1980) or Hanemann (1984), for a consumer i on purchase occasion t choosing from among J brands in a category ($j = 1, 2, \ldots, J$). The consumer starts out with a bivariate direct utility function; with one argument being a "quality" $\left(\psi_{ijt}\right)$ weighted sum of the quantities $\left(x_{ijt}\right)$ of each of the brands in the category $\left(\sum_{j=1}^{J} \psi_{ijt} x_{ijt}\right)$ and the other being the quality weighted quantity of an "outside" good. When the consumer maximizes this utility function subject to a budget constraint; the condition under which a single brand, j is picked from the category is give by the following expression (see e.g., Hanemann 1984):

$$\frac{p_{ijt}}{\psi_{ijt}} = min\left\{ \begin{array}{c} min \\ k=1,2,\ \ldots, J \end{array} \left(\frac{p_{ijt}}{\psi_{ijt}}\right), \frac{1}{\psi_{i0t}} \right\} \tag{6.1}$$

where p_{jt} denotes the price of brand j and the price of the outside good has been normalized to 1 and ψ_{i0t} denotes the quality of the outside good. This model and its underpinnings have a long and rich history in marketing due largely to the widespread available of scanner panel data from the 1980s and the associated interest among researchers in trying to understand what drives consumer choices in these packaged goods categories (i.e., the long history stemming from Guadagni and Little 1983).

6.2.1 The First Unobservable: An Aspect of Quality Known to the Consumer, But Not Observed by the Researcher

Since the quality is a positive quantity, we can define the quality as $\psi_{ijt} = \exp\left(\tilde{\alpha}_j + Z_{jt}\widetilde{\beta} + \varepsilon_{ijt}\right)$ where $\tilde{\alpha}_j$ denotes the intrinsic utility that consumers have for brand j; Z_{jt} are the marketing variables (other than price) associated with brand j on occasion t; $\widetilde{\beta}$ is the vector denoting the effects of these marketing variables on the indirect utility; and ϵ_{ijt} denotes other factors that are observed by the consumer but not by the researcher that affect quality for the brand at that occasion for the consumer (one of the "unobservables" referred to earlier). Further, I write the quality of the outside good or the "no purchase" alternative as: $\psi_{i0t} = \exp(\epsilon_{i0t})$. Now taking the logarithm of both sides of Eq. (6.1) and simplifying we can write u_{ijt} as the following equation (see the Appendix for a derivation of this expression):

$$u_{ijt} = \tilde{\alpha}_j + Z_{jt}\widetilde{\beta} - \ln\left(p_{jt}\right) + \epsilon_{ijt} \tag{6.2}$$

Following in the long tradition of logit models starting with McFadden (1974) in economics and Guadagni and Little (1983) in marketing, I make the assumption that the ϵ_{ijt} terms (for alternatives 0 through J) have the i.i.d extreme value distribution with scale factor θ. We can therefore obtain the probability that the consumer i chooses brand j on purchase occasion t $Pr_{ijt} = P(u_{ijt} \geq u_{ikt}, \forall k = 0, 1, 2, 3, \ldots, J)$ as follows:

$$Pr_{ijt} = \frac{\exp\left(\alpha_j + Z_{jt}\beta - \theta \ln\left(p_{jt}\right)\right)}{1 + \sum_{k=1}^{J} \exp(\alpha_k + Z_{kt}\beta - \theta \ln(p_{kt}))} \tag{6.3}$$

where α_j (referred to as the "intrinsic preference parameter) and β (referred to as the responsiveness parameters) are scaled versions of the original parameters in the quality functions.

Why is the logit model as described above a "structural" model? Recall, a key ingredient of a structural model is the presence of an economic agent—in this case it the consumer. Further, the consumer engages in optimizing behavior—in this case that of utility maximization. Based on this behavior we have obtained a relationship between outcomes that we observe in the data (purchases of the different brands) and the various explanatory variables such as prices and other marketing variables is obtained as a consequence of such behavior.

Estimation of the parameters, $\Theta = \{\alpha_j, j = 1, \ldots, J; \beta; \theta\}$ of the above model proceeds usually with consumer-level choice data over time. While other approaches have been used as well, a popular means of estimating the model parameters is via maximum likelihood estimation. First, we write out the joint likelihood of purchases across purchase occasions, brands and consumers that corresponds to the actual purchases one observes in the data and then choosing the Θ to maximize this likelihood function. The model parameters are identified as follows. The share of purchases in the data corresponding to each brand and to the outside good identifies the α_j parameters; whereas the $\{\beta, \theta\}$ parameters are identified off how the choices made by consumers vary with changes in the prices and other marketing activities across consumers, brands and purchase occasions. Even in the absence of panel data, i.e., only with consumer choices at 1 purchase occasion the parameters are identified due to variation across brands and consumers.

6.2.2 The Second Unobservable: Consumers Are Heterogeneous in Their Preferences and How They Respond to Marketing Activities

The next set of unobservables that we can introduce into the above model corresponds to the heterogeneity across consumers in their preference and responses to marketing activities. Accordingly, several researchers, e.g., Kamakura and Russell (1989), Chintagunta et al. (1991), Gonul and Srinivasan (1993), Rossi et al. (1996), among many others have allowed Θ to vary across consumers following some distribution (either discrete or continuous) across consumers such that $\Theta_i \sim f(\Theta)$; where $f(.)$ represents the density of a specified multivariate distribution. Specifically, when the parameters are heterogeneous, the consumer's probability of choosing brand j can be written as:

$$Pr_{ijt} = \frac{\exp\left(\alpha_{ij} + Z_{jt}\beta_i - \theta_i \ln\left(p_{jt}\right)\right)}{1 + \sum_{k=1}^{J} \exp(\alpha_{ik} + Z_{kt}\beta_i - \theta_i \ln(p_{kt}))} \tag{6.3'}$$

Such random effects models require the assumption that α_{ij} and Z_{jt} are uncorrelated. A popular choice for Θ is the multivariate normal distribution such that $\Theta_i \sim MVN(\Theta, \Omega)$ where Θ denotes the mean vector of the multivariate normal distribution (MVN) and Ω is the associated covariance matrix. The specific origins of the choice of the MVN are unclear, researchers have over the years, considered other alternatives such as discrete distributions (Kamakura and Russell 1989), Gamma distributions on $\exp\left(\alpha_{ij}\right)$ (Chintagunta et al. 1991), etc. The key insight comes from Dalal and Klein (1988) who show that the "mixed" logit model can approximate any flexible distribution used to characterize consumer choice behavior.

Identification of the parameters of this model, in contrast with those from the previous model, requires the presence of panel data. Why? As before the mean parameters of the heterogeneity distribution Θ requires, in principle, only data such as those required for model (6.3). However, the identification of the parameters of the Ω matrix comes from how an individual consumer's purchase shares of the various brands varies across consumers (for the α_j parameters); and how that consumer's purchases change with changes in prices and other marketing activities vis-à-vis the purchases of other consumers. The more the variation in within-consumer behavior across consumers in the sample, larger is the estimated heterogeneity across consumers. However if the nature of variation for one consumer is very much like that for any other consumer then there is little to distinguish between the behaviors of the different consumers leading to the finding of limited heterogeneity in the data.

The estimation of the model parameters once again proceeds by constructing the likelihood function. Since a given consumer is assumed to carry the same vector of Θ parameters across purchase occasions, the likelihood function is first constructed for an individual consumer across his or her purchases, conditional on the parameters for that consumer (which represents a draw from the heterogeneity distribution). The unconditional likelihood for the consumer is then just the conditional likelihood integrated over the distribution of heterogeneity across consumers. The sample likelihood will then be the product of the individual unconditional likelihoods across consumers.

An important point to note is that for the model in (6.2) either with or without heterogeneity, the model prediction for a given set of marketing variables and prices will be a probability that a consumer purchases a brand at that purchase occasion. This prediction will not be perfect since we as researchers never observe the error term or unobservable, ϵ_{ijt}.

6.3 A Detour: Discrete-Choice Demand Models for Aggregate Data

More recently, the logit model has been used in conjunction with aggregate data—store or market (e.g., Berry 1994; Berryet al. 1995; Nevo 2001; Sudhir 2001) level data. Assuming for now that there is no heterogeneity in the intrinsic preference or the responsiveness parameters, the probability of a consumer purchasing brand j is once again given by the expression in Eq. (6.3). Aggregating these probabilities across all consumers (N) visiting the store or purchasing in that market in a given time period t (say a week) we can obtain the market share as follows:

$$S_{jt} = \left(\frac{1}{N}\right) \sum_{i=1}^{N} Pr_{ijt} = Pr_{ijt} = \frac{\exp\left(\alpha_j + Z_{jt}\beta - \theta \ln\left(p_{jt}\right)\right)}{1 + \sum_{k=1}^{J} \exp(\alpha_k + Z_{kt}\beta - \theta \ln(p_{kt}))} \tag{6.4}$$

The sampling error associated with the share in Eq. (6.4) is then given as follows:

$$se_{jt} = \sqrt{\frac{S_{jt}(1 - S_{jt})}{N}} \tag{6.5}$$

It is clear that as the number of consumers in the market becomes "large," the sampling error shrinks to zero. And Eq. (6.4) will represent the market share of the brand in that week. At the aggregate level however, S_{jt} represents a deterministic relationship between the various explanatory variables (prices and other marketing variables) and the outcome variable—market share. Recall that this was not the case at the individual level. So although the expressions in the two cases are identical, the nature of the outcome variable has different implications.

At issue is that if the expression in Eq. (6.4) is to be used as a predictor of the outcome variable—share, then it implies that given a set of parameters and a set of observable variables, researchers will be able to predict market shares perfectly, i.e. with no error associated with this. Clearly such a claim would be inappropriate as one cannot perfectly predict shares. This brings up a need for another error that can explain the discrepancies between the model prediction and what we observe in the data in terms of the brand shares for different time periods. An easy way in which these errors can be introduced is additively in Eq. (6.4). In other words we can write the share expression as:

$$S_{jt} = \frac{\exp(\alpha_j + Z_{jt}\beta - \theta \ln\left(p_{jt}\right))}{1 + \sum_{k=1}^{J} \exp(\alpha_k + Z_{kt}\beta - \theta \ln(p_{kt}))} + e_{jt} \tag{6.5}$$

But would such an error term be viewed as being "structural"? Perhaps the error can be viewed as measurement error in shares. However, the source of the deviation is unclear.[2]

6.4 Unobserved Demand Factors at the Aggregate Level (I.e., Common Across Consumers)

One can alternatively argue that these are brand-level factors that have not been included as part of vector $\{p_{jt}, Z_{jt}\}$ that we have already introduced into the model. So these are unobservables like shelf space and shelf location that are common across consumers who visit a store, are brand specific, influence shares, but are not observed by us as researchers (in most cases). So if the error term captures such factors that have been omitted in the model, where would they belong? It appears that they should be included as a brand- and week-specific measure of quality when one is looking at store share data. Denoting these factors as ξ_{jt} for brand j in week t, the share equation in (6.5) can instead be written as:

$$S_{jt} = \frac{\exp(\alpha_j + Z_{jt}\beta - \theta \ln(p_{jt}) + \xi_{jt})}{1 + \sum_{k=1}^{J} \exp(\alpha_k + Z_{kt}\beta - \theta \ln(p_{kt}) + \xi_{kt})} \tag{6.6}$$

Since the ξ_{jt} are not observed by us as researchers they qualify for inclusion as unobservables. Further, since they are integral to the utility maximization problem considered earlier, they can also be viewed as being structural in nature.

So the (observed) explanatory variables are the same as those in Eq. (6.2) but the outcome variable is the shares of the different brands in a given market- and time-period. Per se, estimation of the model in Eq. (6.6) is straightforward since it can be "linearized" as follows:

$$\ln\left(\frac{S_{jt}}{S_{0t}}\right) = \alpha_j + Z_{jt}\beta - \theta \ln(p_{jt}) + \xi_{jt} \tag{6.7}$$

In general, given the observables in the above model it would appear that the unknown parameters can be estimated within a regression framework. Indeed that is the case. The structural error term ξ_{jt} plays the role of the error term in this regression.

One issue to be cognizant of when estimating the parameters using the aggregate data is to make sure that one understands how managers are setting their levels of

[2]The source of measurement error may be clear or unclear, depending on the researcher's understanding of the measurement technology. For example, if measurement comes from an unbiased survey and the researcher knows the sample size, we might be able to specify the distribution of measurement error exactly.

Z_{jt}, p_{jt} and ξ_{jt}. Consider a store manager that provides prime shelf space for a product that it then wants to charge a premium price for. In this case, p_{jt} is being set based on the ξ_{jt} for that brand. In such a situation, one of the explanatory variables in the model, i.e., price is going to be correlated with the error term in the model, ξ_{jt}. In other words, in this case, prices are being set "endogenously" and one must address the associated endogeneity issue.

I will not go into the issue of endogeneity and how one goes about resolving endogeneity in such a model. Suffice to say that others have tackled this issue (Berry 1994; Berry et al. 1995; Rossi 2014). There are several approaches to addressing the problem although consensus about a universal best approach is lacking. There are of course pros and cons associated with each approach and each context within which it is applied.

While the presence of the structural error term ξ_{jt} in Eq. (6.7) addresses the issue of variability of shares from observed outcomes, there is another form of variability that the model does not account for. Specifically, the model in Eq. (6.6) suffers from the Independence from Irrelevant Alternatives (or IIA) problem. In particular what that means is that if brand j changes its prices then the shares of the other brands will change proportional to those brands' market shares (i.e., in a manner consistent with the IIA assumption). In reality of course, careful inspection of the share data in conjunction with changes in prices (for example) might reveal to the researcher that the IIA assumption is inconsistent with the data on hand. In such instances the logical question that arises it: how can I modify the model to be able to accommodate deviations from the IIA?

The answer to this stems from one of the unobservables we have already introduced—that of heterogeneity in the preferences and response parameters. The presence of "heterogeneity" in preferences and responsiveness parameters results in an aggregate share model that no longer suffers from the IIA problem. This is how. Recall, that in the context of consumer data, we allowed these consumers to have parameters Θ that vary according to a multivariate normal distribution. The question then becomes, if such heterogeneity exists at the consumer level, what does the aggregate share of brand j look like in week t? If the consumer level probability is given by the expression in Eq. (6.3') then the aggregate share of brand j in week (or some other time period) t requires us to integrate out the heterogeneity distribution in that week. This yields the following expression.

$$\begin{aligned} S_{jt} &= \int \frac{\exp\left(\alpha_{ij} + Z_{jt}\beta_i - \theta_i \ln\left(p_{jt}\right) + \xi_{jt}\right)}{1 + \sum_{k=1}^{J} \exp(\alpha_{ik} + Z_{kt}\beta_i - \theta_i \ln(p_{kt}) + \xi_{kt})} f(\Theta_i) d\Theta \\ &= \int \frac{\exp\left(\left[\alpha_j + Z_{jt}\beta - \theta \ln\left(p_{jt}\right) + \xi_{jt}\right] + \left[\Delta\alpha_{ij} + Z_{jt}\Delta\beta_i - \Delta\theta_i \ln\left(p_{jt}\right)\right]\right)}{1 + \sum_{k=1}^{J} \exp([\alpha_k + Z_{kt}\beta - \theta \ln(p_{kt}) + \xi_{kt}] + [\Delta\alpha_{ik} + Z_{kt}\Delta\beta_i - \Delta\theta_i \ln(p_{kt})])} f(_i) d \end{aligned} \tag{6.8}$$

In Eq. (6.8), $\alpha_{ij} = \alpha_j + \Delta\alpha_{ij}$, where the first term on the right-hand-side, α_j is the mean of that parameter across consumers and the second term is the deviation of consumer i's preference from the mean. The second line of Eq. (6.8) separates the part that is not consumer-specific from the part that is; so the heterogeneity distribution only pertains to the distribution of consumer deviations $_i$ from the overall mean. Thus, $_i \sim MVN(0,\ \Omega)$.

From the above expression it is clear that the ratio of the shares of 2 brands j and k depends on the levels of the explanatory variables of all other brands and hence free from the effects of the IIA property. A clear downside to the model in (6.8) is that it is no longer linearizable as it once was. Hence other approaches need to be employed to address the unobservability of ξ_{jt}. In particular, Berry (1994) proposed the contraction mapping procedure to isolate the component $\alpha_j + Z_{jt}\beta - \theta \ln(p_{jt}) + \xi_{jt}$ (or the "linear utility" component in the language of Berry and BLP) in the first square bracket in the numerator and denominator from (6.8) above; conditional on a chosen set of parameters for the "nonlinear" part, i.e., that corresponding to the heterogeneity distribution. This restores the linearity we saw in (6.7) and regression methods can once again be employed. An alternative approach that has been proposed more recently is that by Dube et al. (2012) using an MPEC (Mathematical Programming with Equilibrium Constraints) approach. The identification of the parameters of this model was implicit in my discussion for the motivation of including the "additional" error term (to better fit share variability over time) and heterogeneity in the parameters (to better account for deviations from IIA). Small deviations from IIA will result in finding low variances for the heterogeneity distribution, $\Delta\Theta_i \sim MVN(0, \Omega)$.

6.5 Back to the Consumer Demand Model

The above discussion covers the first two types of unobservables identified earlier. It also introduced a third unobservable identified in the context of aggregate demand data.

6.5.1 A Third Unobservable: Consumption (and Other) Signals Received by Consumers As They Seek to Learn About the Quality of a Product

The third set alluded to previously involves agent uncertainty about a specific parameter in the model. In the logit model, this is often assumed to be the preference for a product, i.e., α_j. What is a context within which such uncertainty could occur? One obvious case would be when a consumer encounters purchasing in a new category (s)he has not purchased from before. Take for example, a consumer who has newly become a first-time parent and has never made a purchase of diapers before. In this instance, the consumer might not know the quality of each of the brands of diapers available in the marketplace. When this happens, α_j is not known to the consumer and can hence be thought of as a random variable from the consumer's perspective, $\breve{\alpha}_j$. Now, if we assume that the consumer is risk-neutral and maximizes expected utility then the probability of the consumer purchasing brand j will be:

$$Pr_{ijt} = \frac{\exp\left(E\left(\breve{\alpha}_j\right) + Z_{jt}\beta - \theta \ln\left(p_{jt}\right)\right)}{1 + \sum_{k=1}^{J} \exp\left(E\left(\breve{\alpha}_k\right) + Z_{kt}\beta - \theta \ln\left(p_{kt}\right)\right)} \tag{6.9}$$

where $E(.)$ is the expectation operator.

The question is: what happens when the consumer does not know the mean of the distribution of $\breve{\alpha}_j$? In such a situation, does the consumer seek to resolve his or her uncertainty regarding this quality, and if so how does (s)he do it? (The following discussion draws heaving from Sriram and Chintagunta 2009). Here we consider the case in which the consumer learns about the unknown quality. The typical assumption is that consumers learn in a Bayesian fashion over time. Let α_j be the true quality of the brand j. Consumers do not know this true quality. And while they know that it comes from a distribution, unlike the case above, they do not know the mean of that distribution. In period 0, the consumer starts with a prior belief that the quality is normally distributed with mean α_{0j} and variance σ_{0j}^2, i.e.,

$$\breve{\alpha}_{0j} \sim N\left(\alpha_{0j}, \sigma_{0j}^2\right) \tag{6.10}$$

For now we assume that the above prior belief is common across consumers. In period 1, the consumer would make a purchase decision based on these prior beliefs for each of the J brands. If consumer i, $i = 1, 2, \ldots, I$, purchases brand j, she can assess the quality of the product from her consumption experience, α_{Eij1}. If we assume that the consumer always derives the experience of quality that is equal to the true quality, then this one consumption experience is sufficient to assess the true quality of the product.[3] However, in reality, this experienced quality might differ from the true quality, because of (a) intrinsic product variability and/or (b) idiosyncratic consumer perceptions and usage contexts. Hence, researchers typically assume that these experienced quality signals are draws from a normal distribution whose mean equals the true quality, i.e., that these are unbiased signals. Thus, we have

$$\alpha_{Eij1} \sim N\left(\alpha_j, \sigma_j^2\right) \tag{6.11}$$

[3]Such products are referred to as "experience goods." These are products or services where product characteristics are difficult to observe in advance but can be ascertained upon consumption or usage "experience."

where σ_j^2 captures the extent to which the signals are noisy. Thus, for learning to extend beyond the initial purchase, we need $\sigma_j^2 > 0$. In (6.11) consumers do not know the mean but are assumed to know the variance.

Subsequent to the first purchase (and consumption experience) the consumer has some more information than the prior she started with. Consumers use this new information along with the prior to update their beliefs about the true quality of the product in a Bayesian fashion. Specifically, since both the prior and the signal are normally distributed, conjugacy implies that the posterior belief at the end of period 1 would also follow a normal distribution with mean $\bar{\alpha}_{ij1}$ and variance σ_{ij1}^2 such that

$$\begin{aligned}
\bar{\alpha}_{ij1} &= \theta_{ij1}\alpha_{0j} + \varpi_{ij1}\alpha_{Eij1} \\
\sigma_{ij1}^2 &= \frac{1}{\frac{1}{\sigma_{0j}^2} + \frac{1}{\sigma_j^2}} \\
\theta_{ij1} &= \frac{\sigma_j^2}{\sigma_{0j}^2 + \sigma_j^2} \\
\varpi_{ij1} &= \frac{\sigma_{0j}^2}{\sigma_{0j}^2 + \sigma_j^2}
\end{aligned} \tag{6.12}$$

If none of the other brands is purchased in the first period, the posterior distributions for those brands will be the same as the prior distributions as there is no additional information to update the consumer's beliefs about these brands.

This posterior belief at the end of period 1 acts as the prior belief at the beginning of period 2. Thus, when the consumer makes a purchase decision in period 2, she would expect her quality experience to come from the distribution

$$\breve{\alpha}_{ij2} \sim N\left(\bar{\alpha}_{ij1}, \sigma_{ij1}^2\right)$$

On the other hand, a consumer who does not make a purchase in period 1 will use the same prior in period 2 as she did in period 1. Hence, we can generalize the above equations for any time period t, $t = 1, 2, \ldots, T$, as follows

$$\begin{aligned}
\bar{\alpha}_{ijt} &= \theta_{ijt}\bar{\alpha}_{ij(t-1)} + \varpi_{ijt}\alpha_{Eijt} \\
\sigma_{ijt}^2 &= \frac{1}{\frac{1}{\sigma_{ij(t-1)}^2} + \frac{I_{ijt}}{\sigma_j^2}} = \frac{1}{\frac{1}{\sigma_{0j}^2} + \frac{\sum_{\tau=1}^{t} I_{ij\tau}}{\sigma_j^2}} \\
\theta_{ij1} &= \frac{\sigma_j^2}{I_{ijt}\sigma_{ij(t-1)}^2 + \sigma_j^2} \\
\varpi_{ij1} &= \frac{I_{ijt}\sigma_{ij(t-1)}^2}{I_{ijt}\sigma_{ij(t-1)}^2 + \sigma_j^2}
\end{aligned} \tag{6.13}$$

where I_{ijt} is an indicator variable that takes on the value 1 if consumer i makes a purchase of brand j in period t and 0 otherwise. Similarly, when the consumer makes a purchase in period $t + 1$, she would assume that the quality of the product comes from the posterior distribution at the end of period t. The above equations also imply that as the number of consumption experiences increase, the consumer learns more and more about the true quality of the product. As a result, her posterior mean would shift away from her initial prior and move closer to the true mean quality. Similarly, as she receives more information, her posterior variance would decrease. It is in this sense that the consumer "learns" about quality in this model.

In learning models as described above, the consumer actually observes the signals α_{Eijt} in each time period; so this quantity is known to the consumer. However, the signal observed by the consumer is seldom observed by the researcher (for an exception see Sriram et al. 2015). Thus in such situations the signals received by consumers become part of the set of unobservables from the researcher's perspective. Researchers typically assume, as above that the signals come from a known distribution with unknown parameters and then simulate these signals over the course of the estimation. Accordingly, identification in learning models poses a challenge. One needs to observe a pattern in the data that suggests that behavior evolves over time consistent with converging towards some preference level if indeed there is support for the Bayesian updating mechanism described above.[4]

For example, one implication of the expression in Eq. (6.13) is that if the variance of the received signals σ_j^2 is high then learning will be slower than when the variance is low. As an example of identification using this idea, Sriram et al. (2015) look at a situation where the variance of signals received by consumers can be high or low with these variances being observed by researchers. The context is that of consumers deciding whether to continue subscribing to a video-on-demand service. Consumers who receive high (low) quality service are more likely to continue (stop) subscribing but consumers are uncertain about their quality. They learn about this quality based on the signals received. If the signals consumers receive have low variance then consumers receiving either high or low quality of service learn about this quality quickly; those with high quality continue with the firm and those with low quality leave, i.e., terminate the service. But if signals have a high variance, learning is slow and consumers receiving low quality service may continue with the service. Indeed, the patterns in the data suggest precisely this nature of behavior.

Given the nonlinearity associated with learning models, one often finds evidence of learning even when it is unclear whether such learning is going on in the data. For example, consider a situation where there might be (temporally local) trends in the share of a brand over time. These trends may to entirely attributable to either a change in the characteristics of the product or changes in the firm's marketing activities over

[4]Ultimately, structural empirical parameters are typically identified both by (1) functional form assumptions and (2) data. As researchers we should be concerned about identification that comes largely from the former.

time. However, if these factors are not adequately controlled for, such trends may be reflected as learning in a model that imposes such behavior on agents in the marketplace. Thinking about the sources of identification prior to estimation makes for good practice not just with these models but with all econometric models in general.

6.5.2 A Fourth Unobservable: The Consumer is Uncertain About the Value of an Attribute (Say Price) and Engages in Costly Search to Resolve This Uncertainty

The particular model we will consider here is of the following type: Prior to visiting a store, a consumer does not know about the prices of various brands of tuna. (S)he however, knows the distribution of prices in the market. So the consumer's indirect utility for brand j is given as $u_{ij} = \alpha_{ij} + X_j\beta_i + \gamma p_{ij} + \epsilon_{ij}$. Here we are assuming that the data are cross-sectional so the heterogeneity in α and β are only due to observable sources.

Consumers do not know p_{ij} but believe that prices come from a Type-I Extreme Value distribution with location η_{ij} and scale parameter μ. So consumers know η_{ij} and μ. We as researchers also know η_{ij} and μ (This can be relaxed although we will assume that researchers can estimate these parameters.). Search is costly and "costs" consumer c_i per search.

Note: We have also assumed here that the scale parameter is common across j. This is an important assumption and is critical for characterizing subsequent choice probabilities. Before proceeding let us be clear about what the researcher sees and what the consumer sees.

As researchers we also do not observe all components of the utility function—specifically we do not observe ϵ_{ij}. But we observe the consumer's consideration set i.e. the set of all brands that the consumer searched as well as all the prices of the alternatives searched. Finally researchers also observe the brand chosen and the corresponding price.

The first thing we should ask ourselves as researchers is what can we infer from observing the consumers' consideration set (CS)?

Let s_i denote the CS of consumer i. There are two ways in which the consumer may arrive at this CS.

1. **Simultaneous** or fixed sample search: Here the consumer decides a priori to search for s_i brands' prices by trading off the costs and benefits of the search. Once this is fixed, the consumer looks for s_i prices and does not stop even if (s)he encounters a really low price with e.g., the first search.
2. **Sequential Search**: The consumer starts with the "best" option, searches for its price and then decides whether or not to make another search. In general this yields a smaller s_i than simultaneous search.

6.6 Simultaneous Search

So how does consumer decide on s_i? In general this is a tricky problem. On the one hand, there might be some alternatives with a high mean price and a low variance and another with a low mean price and a higher variance. Suppose we assume that the variances are the same what happens? It now makes sense for the consumer to pick alternatives with higher means in order to search for their prices. This is the main insight from Chade and Smith (2005, 2006). While their proof is more general, equal variances is one version of this assumption. Given this assumption how does search progress (Honka 2014 provides more details)?

Step 1: Prior to search, consumer only knows the price distributions. So (s)he rank orders options by their expected utilities: $E(u_{ij}) = \alpha_{ij} + X_j\beta_i + \gamma\eta_{ij} + \epsilon_{ij}$.

Step 2: If the s_i has only 1 element that would be the one with the highest expected utility $E(u_{ij})$. The consumer will add a second element to the consideration set only if the net benefit of searching for the prices of the "top 2" exceeds the net benefit of only searching once. So how do we define the net benefit? The net benefit associated with a set of size k is given as follows:

$$\Gamma_{ik} = E\left[\underset{j\epsilon R_{ik}}{max} u_{ij}\right] - kc_i$$

where R_{ik} denotes the set of top k companies that consumer ranks highest according to their **expected** utilities. What is the intuition for the above criterion? In essence, the idea is that when considering k companies, what matters to the consumer is the "best" option among those considered. However, since the consumer does not know prices at this stage, (s)he is interested in the expectation of the utility of the "best" alternative. Since we have assumed that prices are Type-I Extreme Value distributed, we know that:

$$\max_{j\in R_{ik}} \quad u_{ij} \sim EV\left(\frac{\gamma}{\mu}\ln\sum_{j\in R_{ik}}\exp\left(\frac{\mu}{\gamma}\left(\alpha_{ij} + x_j\beta_i + \gamma\eta_{ij} + \epsilon_{ij}\right)\right), \frac{\mu}{\gamma}\right)$$

So

$$E\left[\max_{j\in R_{ik}}\right] = \frac{\gamma}{\mu}\ln\sum_{j\in R_{ik}}\exp\left(\frac{\mu}{\gamma}\left(\alpha_{ij} + x_j\beta_i + \gamma\eta_{ij} + \epsilon_{ij}\right)\right) + \frac{e_c\gamma}{\mu}$$

where e_c is the Euler constant = 0.5772. It is important to note the following property of $E\left[\underset{j\in R_{ik}}{max} u_{ij}\right]$. As the consumer adds more elements with declining mean utilities to the expression, it increases at a *decreasing*

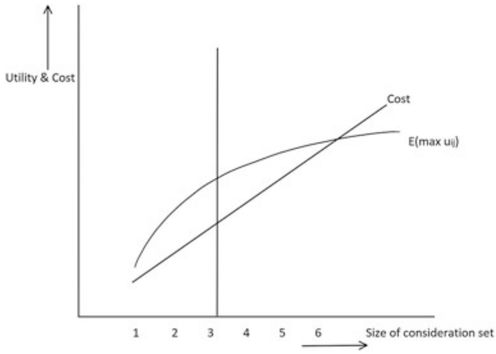


Fig. 6.1 Tradeoff in search costs and expected utility from consideration sets of different sizes

rate. However costs are increasing linearly. So the trade off looks like that in the graph (Fig. 6.1).

So prior to commencing her/his search, the consumer is going to pick a consideration set of size 3 since it maximizes the net benefit of searching (the gap between the curved and straight line) before actually searching. And the 3 brands (s)he will search are the top 3 ranked by their expected indirect utilities.

Step 3: Once the CS size and composition is known to the consumer, (s)he then goes out and searches the prices for the specific alternatives included in the CS.

Step 4: Once the prices are known, the consumer can fully evaluate the options since (s)he is back in the situation where from her/his perspective there are no unobservables. The only unobservable is that facing the researcher, (the first unobservable in our discussion above) where the consumers know all the elements of the utility function. So (s)he picks the alternative j **in the CS** that maximizes her/his utility $u_{ij} > u_{ik}$ for all the elements other than j in the CS.

Example

To make this more concrete, let us look at a simple case with 3 brands. For the consumer

$$u_1 = \alpha_1 + \gamma p_1 + \epsilon_1$$
$$u_2 = \alpha_2 + \gamma p_2 + \epsilon_2$$
$$u_3 = \alpha_3 + \gamma p_3 + \epsilon_3$$

The consumer knows the α's, γ and ε's; but not p's. (S)he knows $p_j \sim EV(\eta_j, \mu)$. What we observe as researchers in the data, for example, are:

1. Consumer's consideration set is {2, 3}
2. Consumer chooses brand 3.

From the consumer's perspective, (s)he first evaluates their expected utilities

$$\begin{aligned} E(u_1) &= \alpha_1 + \gamma\eta_1 + \gamma\tfrac{0.5772}{\mu} + \epsilon_1 \\ E(u_2) &= \alpha_2 + \gamma\eta_2 + \gamma\tfrac{0.5772}{\mu} + \epsilon_2 \\ E(u_3) &= \alpha_3 + \gamma\eta_3 + \gamma\tfrac{0.5772}{\mu} + \epsilon_3 \end{aligned}$$

Note, for example, that (s)he *could* have ordered these expected utilities as follows: $E(u_2) > E(u_3) > E(u_1)$. Let C be the consumer's search cost; we are going to assume that C is small enough to fulfill searching at least once. Next, (s)he computes the net benefits of a consideration set of size 1.

$$\begin{aligned} \Gamma_1 &= E[\max(u_2)] - C \\ &= \alpha_2 + \gamma\eta_2 + \gamma\frac{0.5772}{\mu} + \epsilon_2 - C \end{aligned}$$

Since the set is a singleton the max operator does not bite. Next she looks at a 2 element CS.

$$\begin{aligned} \mathit{\Gamma}_2 &= E[\max(u_2, u_3)] - 2C \\ &= \frac{\gamma}{\mu}\ln\left(e^{\frac{\mu}{\gamma}(\alpha_2 + \gamma\eta_2 + \epsilon_2)} + e^{\frac{\mu}{\gamma}(\alpha_3 + \gamma\eta_3 + \epsilon_3)} + \frac{\gamma 0.5572}{\mu} - 2C\right) \end{aligned}$$

Since the consumer finds that $\mathit{\Gamma}_2 > \mathit{\Gamma}_1$, she goes on and computes $\mathit{\Gamma}_3$.

$$\begin{aligned} \mathit{\Gamma}_3 &= E[\max(u_2, u_3, u_1) - 3C] \\ &= \frac{\gamma}{\mu}\ln\left[e^{\frac{\mu}{\gamma}(\alpha_2 + \gamma\eta_2 + \varepsilon_2)} + e^{\frac{\mu}{\gamma}(\alpha_3 + \gamma\eta_3 + \varepsilon_3)} e^{\frac{\mu}{\gamma}(\alpha_1 + \gamma\eta_1 + \varepsilon_1)}\right] + \frac{\gamma 0.5572}{\mu} - 3C \end{aligned}$$

(S)he now finds that $\mathit{\Gamma}_2 > \mathit{\Gamma}_3$. So the optimal CS is of size 2 and has the brands 2 and 3 as elements of the CS. Now the consumer looks for the prices of these two brands and picks the one that maximizes her utility, say brand 3.

However, the tricky part of this computation is from the perspective of the researcher. The researcher observes that the consumer has chosen brands 2 and 3 to be in her/his consideration set but does not observe the ϵ's. Since these ϵ's are not appearing linearly in the above expression for the researcher to "integrate" out. Honka (2014) shows how one can use the knowledge of the consideration sets to, from the researchers' perspective, compute the probability of observing a particular CS size and its membership.

Given knowledge of the CS of a consumer, the researcher can infer the following 2 points:

1. The minimum expected utility of the brands in the consideration set exceeds the maximum expected utility of the brands not in the consideration set. In the above example it must be the case that $\min\{E(u_2), E(u_3)\} > E(u_1)$.
2. The benefit from the set {2, 3} is greater than the benefit associated from any other set.

The problem is that as researchers, since we do not know the $\epsilon^{'}$s, we cannot rank order the expected utility levels for the consumer. So for estimation, one has to construct a simulation estimator that tries to "mimic" the steps the consumer goes through. One way of going about this is as follows (note that this method has not been subject to testing; for an approach that has, the reader can refer to Honka 2014). The steps below are intended to provide the intuition behind how one deals with the estimation problem.

Step 1: Make R draws for each brand from the EV distribution for the $\epsilon^{'}$s.

Step 2: Make initial guesses for the parameters α's, γ. Since the parameters of the price distributions are not known to the researcher, they need to be estimated from the data on prices (see Honka 2014). With the additional knowledge of these parameters, the researcher can now compute, for each draw r (i.e., each of the R draws made previously)

$$E(u_{jr}) = \alpha_j + \gamma\hat{\eta}_j + \frac{\gamma}{\mu}0.5772 + \varepsilon_{jr}$$

Step 3: Score a 1 if brands 2 and 3 are the top 2 brands in terms of expected utilities and if adding brand 1 to the consideration set lowers the benefit of searching and if removing either 2 or 3 from the CS also lowers the benefit of searching and if u_3 (at the observed price) is greater than u_2 (at the observed price).

Step 4: Repeat steps 2 and 3 for all R draws and compute the **proportion** of 1's in the draws. This gives us an estimate of the joint CS and choice probability at the chosen set of parameters.

Step 5: Construct the likelihood function for the sample of consumers for whom we have the observed data.

Step 6: Iterate over the values of the parameters to maximize the likelihood function on consideration set and choices.

It is important to note that the probability computed in step 4 is likely to be very "lumpy". To avoid this problem one can use a kernel smoothed simulation estimator (see for example, Hajivassiliou 2000).

6.7 Sequential Search

As mentioned before, the consumer in this case starts with the "best" alternative. This requires some *a priori* ranking of the options. In the case of sequential search, Weitzman (1979) has shown that this ranking is based on computing a **reservation utility** for each option (along the lines of the expected utility for simultaneous search). This reservation utility is given as follows.

$$c_i = \int_{u_{ij}^*}^{\infty} \left(u_{ij} - u_{ij}^*\right) f(u_{ij}) du_{ij}$$

$$c_i = \int_{u_{ij}^*}^{\infty} \left(\left(\alpha_{ij} + x_{ij}\beta + \gamma p_{ij} + \epsilon_{ij}\right) - u_{ij}^*\right) \gamma f(p_{ij}) dp_{ij}$$

The reservation utility is the level of utility that makes the consumer indifferent between searching and not searching for j's prices. Weitzman characterizes sequential search in terms of the following 3 rules.

1. *Selection Rule*: Order the alternatives based on their reservation utilities starting with the highest reservation utility. Select the product to be searched as the one with the highest reservation utility among the products not yet searched. Obtain the price of the product to be searched. Once the price is searched, the consumer knows the realized utility for the option searched.
2. *Stopping Rule*: If the maximum **realized** utility among alternatives searched so for is **greater** than the highest reservation utility among unsearched alternatives, stop. Else go to the next ranked alternative by reservation utility.
3. *Choice Rule*: Pick the alternative with the highest realized utility among alternatives searched at the time of stopping.

Let us revisit our 3 alternative example.

1. The consumer rank orders the 3 options based on reservation utilities; for example, for the consumer, a possible ordering would be $u_3^* > u_2^* > u_1^*$.
2. Given the above ordering, the consumer looks for the price of brand 3. This informs the consumer of that brands price. The consumer now observes the realized utility of brand 3, u_3.
3. If $u_3 > u_2^*$, the consumer stops searching and chooses option 3. If not, the consumer looks for the price of brand 2. This gives the consumer the value of u_2.
4. Compare the maximum of the realized utilities u_3 and u_2 with the reservation utility u_1^*. If the former is bigger then stop and pick the alternative corresponding to the higher of u_3 and u_2.

The above description of sequential search is from the consumer's perspective. What about the scenario from the researcher's perspective? Let the researcher once again **only** observe the consideration set and the final choice; consideration set {2, 3}, and choice of 3. The researcher can infer the following:

1. $u_3 > u_2$ (Choice Rule)
2. min $\left(u_3^*, u_2^*\right) > u_1^*$; (Selection Rule)
3. max(u_3 and u_2) > u_1^*; (Stopping Rule)

The key problem with only imposing the above conditions in the estimation, without observing the search sequence, is that we will not be able to impose enough discipline on the parameter space in the estimation. As Honka and Chintagunta (2016) point out, other restrictions may need to be imposed to consistently estimate the model parameters.

The above description indicates that with limited data available from consumers —i.e., with only information on the consideration set and its membership as well as the final choice, one might be able to shed some light on search costs although more work is required in order to obtain robust inferences from these models.

6.8 Why Do We Need Structural Models?

Structural models are useful in many contexts; I highlight two of them here. The first is in quantifying the effects of various marketing interventions by estimating the underlying parameters of the structural model of interest. The second is using the estimated parameters from the model to assess the consequences of changing one of the ingredients of the model. For example, one might be interested in understanding the consequences of changing the nature of interactions among the agents involved in the structural model. I will now illustrate these two types of applications and explain why it might be difficult to make the same assessments sans the structural model.

In the Sriram et al. paper (2015) mentioned above, some consumers are exposed to signals about the quality they receive that have high variance whereas the signals that others receive have low variance. The latter are able to learn about the true quality they receive quicker than those with high variance. An implication of this is that when consumers are uncertain about the quality they experience, those experiencing low temporal variability in quality are likely to be more responsive (in terms of termination) to the average quality level compared to those experiencing high variability. Specifically, if at the time of signing up for the service, if a consumer has a high prior belief on the quality, then it becomes more difficult for the consumer to learn that the quality is actually low when the variance of signals received is high. As a consequence these consumers will respond less, in terms of termination, to the quality they receive. On the other hand, for consumers receiving higher quality than their prior belief, high variability will interfere with such learning so termination may be higher than for those with high quality but low signal variability. In other words, we would see an interaction effect between average quality and variability on termination in the data. Indeed, the authors find such an interaction effect in the data. Interestingly, the data also reveal that the main effect of variability is negative which is indicative of a form of "risk aversion" among the consumers. Such a risk aversion effect would also translate to higher termination at high levels of variability. To quantify the level of quality sensitivity and risk aversion however, requires a model that also controls for other factors that could be affecting termination behavior. This is the role that the structural model plays in that paper. Estimating the quality effect for different consumers in such a model provides insights to managers interested in lowering termination rates for their service.

Now consider the case when one did not use a structural model based on the data patterns but instead specified a functional relationship between termination behavior and the level of quality received by a consumer. Such a model would be entirely plausible for the data on hand since the interest would be on quantifying the effects of raising or lowering quality on termination behavior. While such a model can be made extremely flexible, it is unclear whether it would have included variability as a covariate. Suppose the researcher chooses to include variability, the likely conclusion would have corresponded to the main effect of variability mentioned above—that of higher variability leading to a higher termination rate. What would have been critical to include would be the interaction effect. Even if the researcher chooses to include an interaction effect, it would be unclear where such an effect would be coming from and what the consequences of such an effect would be for a manager trying to change the level of quality available in the marketplace. As the structural model reveals, variability aids retention at low quality levels so the manager would have to assess the consequence of affecting quality in such a scenario. The structural model is useful in assessing what would happen in this context. Of course, structural models are not infallible—an incorrectly specified model would lead to incorrect inferences being drawn about the behavior of consumers. Hence it is crucial to base the model on patterns observed in the data and to then check for robustness of the results to alternative specifications that might also be consistent with patterns in the data.

Next, I turn to an example where the structural model can help answer a question dealing with a change in agent interaction or the structure of the market in which the agents make decisions. An important paper in this area that showcases this role of structural models is Misra and Nair (2011). The paper looks at the topic of salesforce compensation and asks the question: what is the likely consequence of modifying the compensation scheme provided to the salesforce? Companies may be interested in answering this question but may be reluctant to experiment with alternative schemes for several reasons. First, changing the compensation scheme could be, at least in the short-run, a very expensive proposition for the firm. Second, an inappropriate change in schemes might have a negative impact on the morale of the salespeople. Thus, if there is a way for the firm to understand the consequences of changing the compensation scheme, such an approach would be very valuable to the firm. This is where the Misra and Nair paper comes in.

The authors have access to rich individual salesperson level performance data (in terms of sales calls made and sales generated) from a specific firm. This allows them to build a rich dynamic structural model of agent behavior that captures the specifics of the compensation scheme that the firm had in place as well as the data patterns that characterize the behavior of the salespeople. Next, Misra and Nair estimate the model parameters (using recent techniques for the estimation of such dynamic models). The important aspect of this paper is what it does next. It does not content itself by simply estimating the model parameters; rather, the authors first conduct counterfactuals with alternative compensation schemes to understand specific schemes in which firm profits would go up. Next, and then implement a new compensation scheme for the employees of firm. The behavior of the salespeople as well as their output levels change in a manner as predicted by the counterfactual analysis under this new compensation plan. The new plan resulted in a 9% improvement in overall revenues. Such an increase corresponds to about $12 million incremental revenues annually. In addition, the paper shows an improvement in performance and satisfaction among the salespersons after the implementation of the new program. This provides a very strong vindication of the use of structural models to improve outcomes for firms as well as their employees. Further, the insights from the structural model are critical for identifying and evaluating alternative schemes and their consequences. In general, while structural models often are accompanied by counterfactuals and "policy simulations," external validation is a topic that deserves greater attention in the future.

Clearly, a field implementation of the output of a structural model is quite novel; indeed, this is a direction in which the literature in structural models appears to be progressing. In addition to the above study, there are a few other studies that have assessed the external validity of predictions from structural models—Cho and Rust (2008), in the context of implementing new auto rental policies and Bajari and Hortacsu (2005), in the context of estimating bidder valuations in auctions to name a couple. My expectation is that such studies will gather steam in the future.

Next, I discuss two recent papers, Rossi and Chintagunta (2015a, b), where the context is more slanted towards public policy. The idea behind the first paper is as follows. On the Italian highway, drivers are faced with the problem of not knowing the prices at the gasoline stations that are located on the highway. Price information can only be obtained by getting off the highway and driving to the rest stop. Drivers in other countries face a similar problem, i.e., while information on the location of the next station is posted on the highway, prices at the station are not known to the drivers. To engender price transparency and make the information more accessible to drivers, the Italian government required the management of the highway system to install price signs on the highway. These signs located every four stations were required to provide the prevailing prices at the 4 gas stations following the sign in the direction of travel. The signs were installed in the period from July 2007 to 2009. What is of interest here is whether the introduction of the signs resulted in a change in prices charged by the stations whose prices are posted on the signs relative to those whose prices are not posted.

In order to measure the impact of the price signs, it is important to control for a variety of confounding factors that might affect the identification and estimation of the effect of signs on prices. Rossi and Chintagunta (2015a) find that the installation of signs indeed lowers prices charged by stations whose prices are posted on the signs. Curiously however, the level in dispersion across prices on a given sign does not diminish significantly as a consequence of sign installation. A potential explanation for this is that while 94% of those driving past the first station on the sign also drive past the sign, the number drops to 64% for the second station, 49% for the third station and only 39% for the fourth station. This means that having signs every fourth station does not inform a majority of consumers driving past a station about prices at that station. A question that then arises is: by how much further would prices at the stations fall if drivers were informed about prices at all stations. Such a scenario can, e.g., occur if signs were installed prior to each and every station on the highway. Since there is a cost associated with installing these signs, a related question that arises is whether the benefits outweigh the costs in this situation and whether we can determine this even prior to the installation of the signs. This is where the structural model comes in.

Rossi and Chintagunta (2015b) develop a structural model that incorporates consumers' uncertainty about prices when driving down the motorway. Resolving the uncertainty in the absence of price signs requires consumers to engage in costly search, i.e., they need to drive to the gas station to obtain price information. This could lead to higher prices at the pump since the gas station recognizes that if the consumer leaves without filling gas, they will need to expend the search cost again to visit another station. For drivers transiting in front of the sign, price uncertainty is resolved due to the presence of the sign. The authors then leverage the difference in pre- and post-disclosure prices to recover the cost that a fraction of consumers (who are exposed to the price signs and whose data are available to the authors) incur to obtain price information before the signs are installed. A second component of the structural model that Rossi and Chintagunta propose involves the oligopolistic price-setting behavior of gas stations given the above demand model. This component of the model allows them to predict the level of prices that would prevail if all consumers have access to price information in the counterfactual scenario. The authors find that, compared with the case of perfect price information, in the absence of mandatory price disclosure, gas stations increase their margins by about 31% thereby indicating the benefits of installing the signs. This approach therefore provides valuable input to policy makers considering the costs and benefits of installing additional signs on the highway.

6.9 Looking Back and Looking Ahead

A large fraction of structural models in marketing has tended to fall into three main buckets. The first of these is models of "demand" and "supply". Such models have a long association in the economics literature. According to Reiss and Wolak (2007), such models have been popular since the time of the Cowles Commission for Research in Economics—an economic research institute founded by the businessman and economist Alfred Cowles in Colorado Springs in 1932. The commission that also had a home at the University of Chicago from 1932 to 1955 and now is located at Yale University emphasized econometrics in the context of "economic equilibrium." It is in this light that a vast majority of early structural models in marketing developed and flourished (see e.g., Bronnenberg et al. 2005 for a discussion of models built in this tradition). The typical structure of these models entails a demand specification derived from the underlying utility behavior of consumers; and a supply model of firm behavior that characterizes firms' actions for a variety of marketing mix decisions—prices, advertising, etc. In this bucket I also include studies that focus on simple and more complex demand models (e.g., Berry et al. 2014) that explicitly account for supply-side considerations in the estimation of demand parameters (e.g., Nevo 2001).

As a second bucket of structural models that have been popular in marketing, I include those in the dynamic structural tradition (Dube et al. 2005). On the demand side, dynamics can arise for several reasons (see Chintagunta and Nair 2010 for a discussion)—storability, durability, and experience goods, etc. Why does storability result in dynamics in behavior? The main reason is a purchase today by a consumer increases his or her inventory for the product. In turn, this makes the consumer less likely to buy the product tomorrow. Thus a marketer who encourages a customer to make a purchase today needs to explicitly take into account the consequences of this purchase for the future. Some examples of studies in this vein in marketing are Erdem et al. (2003) and Sun (2005). Durable good demand, at least as it refers to the first time adoption of a product, on the other hand is a dynamic problem because if a consumer makes a purchase today it implies that the consumer is out of the market tomorrow. The consumer in this case is explicitly trading off making a purchase today (at a potentially higher price and lower quality) and enjoying the utility from consuming the product for one day with waiting till tomorrow and buying the product at a potentially lower price and higher quality. A good exemplar of this research in marketing is Nair (2007). Experience goods I have referred to previously under the nomenclature of learning models. Experience goods are therefore characterized by ex ante uncertainty about some aspect of the product (say its quality). This uncertainty is then resolved by consumption. In this case, dynamics arise because if a consumer makes a purchase today, it provides that customer with a signal of the uncertain aspect (quality), which provides the consumer with new information when (s)he goes to make the next purchase. This provides an explicit link between purchasing today and purchasing tomorrow (see Ching et al. 2013). Note however, that the model I described previously was a "myopic" model of learning since it did not fully consider this intertemporal link.

The third bucket includes models that have recently seen an interest in marketing—those involving uncertainty, not about the parameters of the utility function as in learning models, but about some feature or characteristic of the product itself. Here I am referring to the models of search discussed above but with the variation that consumers search for a product that best matches their preferences as in shopping online for a digital camera that best suits ones needs (e.g., Kim et al. 2010). In particular as online browsing behavior, visit and purchase information become more widely available I expect these models to see increasing application in marketing.

Structural models have certainly made an impact in the field of marketing. While diffusion has taken awhile, today they are considered an integral part of the marketers' toolbox. Looking ahead there appear to be three principal domains in which the research seems to be progressing. I will very briefly mention each of them in turn.

(1) **Combining multiple data sources**: This is a topic I had alluded to earlier in the paper with Harikesh Nair (Chintagunta and Nair 2010). As structural models get more complicated, they place an increasingly bigger burden on the data used for parameter identification and estimation. While one ideally seeks patterns in the data that can identify the key parameters of interest (see Einav and Levin 2010), researchers in marketing are increasingly recognizing that one can leverage multiple sources of data—outcomes data from the marketplace, survey data on consumers, experimental data from the lab—to improve the credibility of estimates and to relax assumptions made by structural models. For example, in the context of dynamic structural models it is notoriously difficult to identify the discount factor of consumers (separately from the other parameters in the model). Dube et al. (2014) show how we can combine inputs from conjoint analysis to better inform the estimates of such models (see also Rao 2015).

(2) **Combining multiple methods**: Continuing with the above theme, in instances where identification depends critically on some variation in the data, it may make sense to first establish that such a variation actually exists before constructing a complicated structural model. Often the presence of the variation can be established via other methods, say a difference-in-differences analysis in the first stage as a prelude to the estimation of the structural model. Previously, I described the Rossi and Chintagunta (2015a and b) papers. A key parameter of interest in the latter paper is the search cost incurred by customers when shopping for gasoline. This parameter is identified off the change in prices charged by gas stations after information provision via price signs. So it was important to first establish that prices did change with the introduction of the signs before attempting to identify the search costs from the structural model. This required a "pre-analysis" using a different approach.

A similar strategy is employed by Daljord (2015); first using retail sales data from the Norwegian book market under two regimes—with and without resale price maintenance—he shows that in the absence of price maintenance, it is difficult to sustain a price skimming strategy. Next, he formulates and estimates a dynamic demand model and evaluates the impact of various vertical contracts. My sense is that going forward there will be a bigger need to bringing multiple methods to bear when dealing with increasingly more complex structural models. Tuchman (2016) parallels such work as well; she is interested in studying the impact of e-cigarette advertising on the sales of traditional cigarettes in the U.S. market.

(3) **Using field experiments to validate and implement recommendations based on counterfactuals**. The real power of structural models as a useful tool to improve managerial practice is only now being seen. As field implementations of recommendations from these models such as the one carried out by Misra and Nair become more widespread, the power of structural models to aid decision-making will increasingly become clear. Such implementations are however, not without associated costs. Consequently the availability of a company willing to field-test model based counterfactuals should not be a substitute for carefully thought out structural models to obtain these counterfactuals.

To summarize, I feel that while we have come a long way, there is still much to be discovered in the realm of structural models in marketing. Points (2) and (3) above make me particularly optimistic about bridging the gap between the more economics-oriented researchers in marketing and the more psychology-oriented researchers for whom laboratory and field experiments are the methodologies of choice. As I alluded to earlier, the models underlying structural methods can draw from beyond the discipline of economics. As an illustration consider Sahni (2015). The study collected data in collaboration with one of the largest restaurant search websites in the world. In particular, Sahni designed a field experiment to randomize the advertising treatments that the site visitors were exposed to in the form of sponsored banners. The experimental design generates random variation in whether site visitors were treated with restaurant ads, random variation in the frequency of ad exposure conditional on the total number of pages on the site viewed, and random variation in the spacing (temporal difference) between ad exposures on the site.

Using these data, the paper documents a novel stylized-fact on the "spacing effect" of advertising. Consider an individual who visited the site several times to search at least three times. As expected, the effect of advertising at the first visit on sales leads in the third session diminishes in the time elapsed between the first and third visits. However, the effect of advertising exposure at the second visit on sales in the third visit *increases* if the time between the first and the second sessions is large. In other words, the effect on future sales from additional advertising is smaller if the spacing between the initial and the additional advertising is small.

The spacing effect is not easily reconcilable with the established view of how advertising works. The most common approach to measure the long-run effect of advertising is based on the Arrow-Nerlove model, which postulates that sales are affected by a "goodwill" stock, defined as a distributed lag of advertising exposures. In the Arrow-Nerlove model, the effect of past advertising on sales is unambiguously smaller if an advertising episode occurred in the more distant past. Hence, recognizing that the standard theory cannot explain the spacing patterns documented in his data, Sahni introduces a new memory-based advertising model to marketing, based on the memory module of the ACT-R model of the mind that was developed in cognitive psychology and can be thought of as an optimal information retrieval system. He then uses this memory model to explain the spacing effect found in the data.

Next, to formally test the new memory-based model of advertising the paper develops and estimates a structural consumer choice model. In the baseline specification, advertising influences the current purchase probability in the Arrow-Nerlove distributed lag form with exogenous depreciation factors. In a specification based on the memory model the depreciation factor associated with a past advertising episode is affected by the goodwill level at the time of exposure. These two specifications are then estimated with the data, controlling for other factors influencing the probability of purchase such as the number of pages visited and consumer-specific factors. The results reveal that the goodwill stock at a past exposure event decreases the depreciation factor and that the effect is statistically significant. Thus, the data reject the established Arrow-Nerlove model in favor of the memory-based model of advertising. This paper therefore is in the finest tradition of combining structural models with field experiments and with theories from psychology.

It is also clear from the above illustration that knowledge and implementation of experimental methods will likely enrich our understanding of markets using structural methods. A recent example of research in this vein is Dube et al. (2016). Structural models therefore, provide an excellent platform for researchers with economics and psychology backgrounds to come together to make contributions to the field of marketing.

Appendix: Deriving the Indirect Utility Function (Equation (6.15))

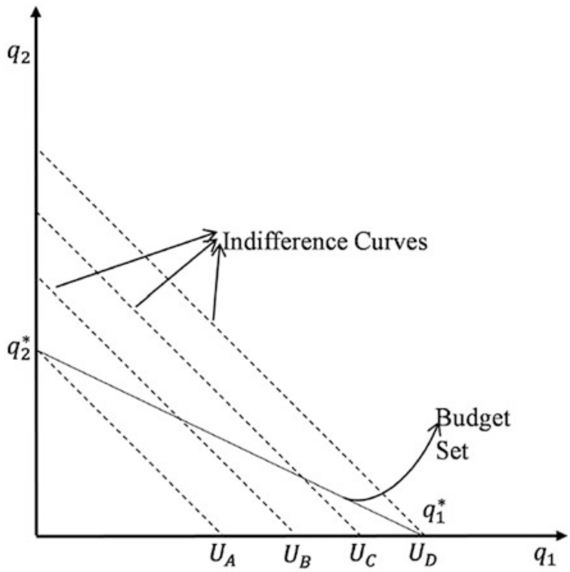
In the marketing literature, it is typical to characterize a consumer's utility function as follows

$$u_{ijt} = \alpha_{ij} + x_{jt}\beta_i + \epsilon_{ijt} \tag{6.14}$$

i: consumer, j: brand, t: time period. And one of the x_{jt}'s was price, p_{jt}. Strictly speaking, however, since the above equation has prices embedded in it, it is better referred to as an "indirect utility" function that is obtained from a direct utility function that is being maximized subject to a budget constraint. Here we show how the direct utility function as in Eq. (6.14) of the chapter may lead to an indirect utility function like Eq. (6.14) above. Consider a world where there are only 2 "goods"—Strawberry yogurt and raspberry yogurt, represented by their respective quantity q_1 and q_2. The consumer derives utility from these two goods according to the relationship:

$$u(q_1, q_2) = \psi_1 q_1 + \psi_2 q_2$$

where ψ_1, $\psi_2 > 0$, ψ_1 and ψ_1 are the "quality" indices of the 2 flavors. Then the consumers' indifference curves would look like the dotted lines in the figure here.



Now the consumer's budget constraint for these goods can be written as $p_1q_1 + p_2q_2 \leq B$ where p_1 and p_2 are the prices of the two goods. We represent the budget set by the continuous line in the figure.

Given the linear indifference curves and budget set, the utility maximizing condition for the consumer lies in a "corner" i.e. to spend all the money (budget) on either strawberry or on raspberry. In the above case spending all the money on raspberry (good2) yields lower utility to the consumer than the alternative. So the consumer makes the "discrete" choice of picking only strawberry yogurt. This is because $u_D > u_A$ in the figure. At the q_{1^*} "corner", $q_{2^*} = 0$ and the consumer obtains utility u_D. At the q_{2^*} "corner", $q_{1^*} = 0$ and the consumer obtains utility u_A. So $u_D = \psi_1 q_1$ (i.e. utility when $q_2 = 0$) and $u_A = \psi_2 q_2$ (i.e. when $q_1 = 0$). From the budget constraint we know that when, $q_2 = 0$; $q_1 = (B/p_1)$, when $q_1 = 0$; $q_2 = (B/p_2)$. So the "indirect utility" $V_D = (\psi_1 B)/(p_1)$ and $V_A = (\psi_2 B)/(p_2)$.

Since we know $V_D > V_A$ in the above case, we can write $(\psi_1 B)/(p_1) > (\psi_2 B)/(p_2)$ and since $B > 0$ then $(\psi_1)/(p_1) > (\psi_2)/(p_2)$. We can now characterize ψ_1 and $\psi_2 > 0$ by writing them as:

$$\psi_1 = \exp(x_1\widetilde{\beta} + e_1) \quad \psi_2 = \exp(x_2\widetilde{\beta} + e_2)$$

where x_1 and x_2 are observable attributes (to the researcher) and e_1 is unobservable as is e_2. So,

$$(\exp(x_1\widetilde{\beta} + e_1))/(p_1) = (\exp(x_2\widetilde{\beta} + e_2))/(p_2)$$

Taking logs on both sides:

$$x_1\widetilde{\beta} + e_1 - \ln p_1 > x_2\widetilde{\beta} + e_2 - \ln p_2 \tag{6.15}$$

References

Bajari and Hortacsu. 2005. Are structural estimates of auction models reasonable? Evidence from experimental data. *Journal of Political Economy* 113 (4): 703–741.

Berry, S. 1994. Estimating discrete-choice models of product differentiation. *The Rand Journal of Economics* 25 (2): 242–262.

Berry, S., Levinsohn and A. Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* 60 (4): 841–890.

Berry, S., A. Khwaja, V. Kumar, B. Anand, A. Musalem, K.C. Wilbur, G. Allenby, and P. Chintagunta. 2014. Structural models of complementary choices. *Marketing Letters*.

Bronnenberg, B.J., P.E. Rossi, and N.J. Vilcassim. 2005. Structural modeling and policy simulation. *Journal of Marketing Research* 42: 22–26.

Chade, H., and L. Smith. 2006. Simultaneous search. *Econometrica* 74 (5): 1293–1307.

Chade, H., and L. Smith. 2005. Simultaneous search. Working Paper.

Chan, T., V. Kadiyali, and P. Xiao. 2009. Structural models of pricing. In *Handbook of pricing research in marketing*. Elgar Publishing.

Ching., A.T., T. Erdem and M.P. Keane. 2013. Learning models: An assessment of progress, challenges and new developments. *Marketing Science* 32 (6): 913–938.

Chintagunta, P.K., D.C. Jain, and N.J. Vilcassim. 1991. Investigating heterogeneity in brand preferences in logit models for panel data. *Journal of Marketing Research* 417–428.

Chintagunta, P.K., V. Kadiyali, N. Vilcassim, and J. Naufel. 2004. Structural models of competition: A marketing strategy perspective. In *Assessing marketing strategy performance*, ed. Christine Moorman, Donald R. Lehmann. Marketing Science Institute.

Chintagunta, P.K., T. Erdem, P.E. Rossi, and M. Wedel. 2006. Structural modeling in marketing: review and assessment. *Marketing Science* 25 (6): 604–616.

Chintagunta, P.K., and H. Nair. 2010. Discrete choice models of consumer demand in marketing. *Marketing Science* 30 (6): 977–996.

Cho, S., and J. Rust. 2008. Is econometrics useful for private policy making? A case study of replacement policy at an auto rental company. *Journal of Econometrics* 145 (1–2): 243–257.

Dalal, S.R., and R.W. Klein. 1988. A flexible class of discrete choice models. *Marketing Science* 7 (3): 232–251.

Daljord, O. 2015. Commitment, vertical contracts and dynamic pricing of durable goods. Working paper, University of Chicago, Booth School of Business.

Deaton, A., and J. Muellbauer. 1980. *Economics and consumer behavior*. New York: Cambridge University Press.

Dube, J.-P., K. Sudhir, A. Ching, G.S. Crawford, M. Draganska, J.T. Fox, W. Hartmann, G. J. Hitsch, V.B. Viard, M. Villas-Boas, and N. Vilcassim. 2005. Recent advances in structural econometric modeling: Dynamics product positioning and entry. *Marketing Letters* 16 (3/4): 209–224.

Dube, J.-P., J. Fox, and C.-L. Su. 2012. Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation. *Econometrica* 80 (5): 2231–2267.

Dube, J.-P., G. Hitsch, and P. Jindal. 2014. The joint identification of utility and discount functions from stated choice data: An application to durable goods adoption. *Quantitative Marketing and Economics* December 12 (4): 331–377.

Dube, J.-P., Z. Fang, N. Fong, and X. Luo. 2016. Competitive price targeting with smartphone coupons. Working paper, University of Chicago, Booth School of Business.

Einav, L., and J. Levin. 2010. Empirical industrial organization: A progress report. *Journal of Economic Perspectives* 24 (2): 145–162.

Erdem, T., S. Imai, and M.P. Keane. 2003. Brand and quantity choice dynamics under price uncertainty. *Quantitattive Marketing and Economics* 1 (1): 5–64.

Erdem, T., K. Srinivasan, W. Amaldoss, P. Bajari, H. Che, Teck H. Ho, W. Hutchinson, M. Katz, M.P. Keane, R. Meyer, and P. Reiss. 2005. Theory-driven choice models. *Marketing Letters* 16 (3): 225–237.

Goldfarb, A., and C.E. Tucker. 2014. Conducting research with quasi-experiments: A guide for marketers. Rotman School Working Paper, Toronto, CA.

Gonul, F., and K. Srinivasan. 1993. Modeling multiple sources of heterogeneity in multinomial logit models: Methodological and managerial issues. *Marketing Science* 12 (3): 213–229.

Guadagni, P., and J.D.C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Science* 2 (3): 203–238.

Hajivassiliou, V. 2000. Some practical issues in maximum simulated likelihood. In *Simulation-based inference in econometrics: Methods and applications*, ed. R. Mariano, T. Schuermann, and M. Weeks. Cambridge: Cambridge University Press.

Hanemann, M.W. 1984. Discrete/continuous models of consumer demand. *Econometrica* 52: 541–561.

Honka, E. 2014. Quantifying search and switching costs in the U.S. auto insurance industry. *RAND Journal of Economics.* 45 (4): 847–884.

Honka, E., and P.K. Chintagunta. 2016. Simultaneous or sequential? Search strategies in the US auto insurance industry. *Marketing Science* (forthcoming).

Horsky, D. 1977. An empirical analysis of the optimal advertising policy. *Management Science* 23 (10): 1037–1049.

Horsky, D., and P. Nelson. 1992. New brand positioning and pricing in an oligopolistic market. *Marketing Science* 11 (2): 133–153.

Kadiyali, V., K. Sudhir, and V.R. Rao. 2001. Structural analysis of competitive behavior: New empirical industrial organization methods in marketing. *International Journal of Research in Marketing* 18 (1): 161–186.

Kamakura, W.A., and G.J. Russell. 1989. A probabilistic model for market segmentation and elasticity structure. *Journal of Marketing Research* 26: 279–390.

Keane, M.P. 2010. Structural vs. atheoretic approaches to econometrics. *Journal of Econometrics* 156: 3–20.

Kim, J., P. Albuquerque, and B. Bronnenberg. 2010. Online demand under limited consumer search. *Marketing Science.* 29 (6): 1001–1023.

McFadden, Daniel. 1974. Conditional logit analysis of qualitative choice behavior. In *Frontiers in econometrics*, ed. P. Zarembda. New York: Academic Press.

Mehta, N., S. Rajiv, and K. Srinivasan. 2003. Price uncertainty and consumer search: A structural model of consideration set formation. *Marketing Science* 22 (1): 58–84.

Misra, S., and H. Nair. 2011. A structural model of sales-force compensation dynamics: Estimation and field implementation. *Quantitative Marketing and Economics* 211–257.

Nair, H. 2007. Intertemporal price discrimination with forward-looking consumers: Application to the US market for console video-games. *Quantitative. Marketing and Economics* 5 (3): 239–292.

Nevo, A. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* 69 (2): 307–342.

Rao, A. 2015. Online content pricing: Purchase and rental markets. *Marketing Science* 34 (3): 2015.

Reiss, P.C., and F.A. Wolak. 2007. Structural econometric modeling: Rationales and examples from industrial organization. In *Handbook of Econometrics* ed. J.J. Heckman, E.E. Leamer, vol. 6A, 4277–4415. North-Holland, Amsterdam.

Rossi, P.E., R. McCulloch, and G.M. Allenby. 1996. The value of purchase history data in target marketing. *Marketing Science* 15 (4): 321–340.
Rossi, F., and P.K. Chintagunta. 2015a. Price transparency and retail prices: Evidence from fuel price signs in the Italian motorway. *Journal of Marketing Research* (forthcoming).
Rossi, F., and P.K. Chintagunta. 2015b. Price uncertainty and market power in retail gasoline. Working paper, University of Chicago.
Rossi, P.E. 2014. Even the rich can make themselves poor: A critical examination of IV methods in marketing applications. *Marketing Science* 33 (5): 655–672.
Sahni, N. 2015. Effect of temporal spacing between advertising exposures: Evidence from online field experiments. *Quantitative Marketing and Economics* 13 (3): 203–247.
Sriram, S., and P.K. Chintagunta. 2009. Learning models. *Review of Marketing Research* 6: 63–83.
Sriram, S., P.K. Chintagunta, and P. Manchanda. 2015. Service quality variability and termination behavior. *Management Science* (forthcoming).
Sudhir, K. 2001. Competitive pricing behavior in the auto market: A structural analysis. *Marketing Science* 20 (1): 42–60.
Sun, B. 2005. Promotion effect on endogenous consumption. *Marketing Science* 24 (3): 430–443.
Teixeira, T., M. Wedel, and R. Pieters. 2010. Moment-to-moment optimal branding in TV commercials: Preventing avoidance by pulsing. *Marketing Science* 29 (5): 783–804.
Tuchman, A.E. 2016. Advertising and demand for addictive goods: The effects of e-cigarette advertising. Working paper, Stanford GSB.
Weitzman, M. 1979. Optimal search for the best alternative. *Econometrica* 47 (3): 641–654.

Chapter 7
Economic Models of Choice

Greg M. Allenby, Jaehwan Kim and Peter E. Rossi

7.1 Introduction

Models of choice are fundamental to the field of marketing because they represent the culmination of marketing efforts embedded in the 4P's (product, price, promotion and place). It is by understanding how people make purchase decisions that we can inform firms on the success of their efforts in each of the functional areas of marketing. Choice models quantify the process of exchange, allowing us to understand the origins of preference and the determinants of costs in a transaction, including the time, money and other resources needed to acquire and use an offering.

Choice is complex. It involves our resources, perceptions, memory and other factors as we acquire and use products to improve our lives. Our goal in this chapter is to provide a review of direct utility choice models that attempt to rationalize choice. We do this within an economic framework of demand where people are assumed to be constrained utility maximizers. We take this view because marketplace data supports the concept of constrained maximization as evidenced by the large proportion of zero's in disaggregate marketing data, coupled with the observation that people are sensitive to price and demands on their time. That is, marketing data is overwhelmingly characterized by sparse demand where most people don't purchase most of the

G.M. Allenby (✉)
Fisher College of Business, Ohio State University, Columbus, OH, USA
e-mail: allenby.1@osu.edu

J. Kim
Korea University Business School, Seoul, Korea
e-mail: jbayes@korea.ac.kr

P.E. Rossi
Anderson School of Managment, UCLA, Los Angeles, CA, USA
e-mail: perossichi@gmail.com

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_7

products available for sale, don't frequent most websites available to them, and don't read most of the literature published on topics of interest. Instead, they select what they consume in a manner that suggests people are resource conserving.

We acknowledge that our treatment of choice models is selective and related to our own research agenda. We believe it is not possible to provide a comprehensive survey of choice models in marketing, as evidenced by the presence of dedicated conferences and journal volumes to the issue of choice. Choice encompasses a vast domain of economic, psychologic and social subject matter, and our chapter provides a narrow emphasis in an area of choice which we hope to popularize and expand.

An advantage of rationalizing decisions within an economic framework is that it can lead to interventions and policy recommendations that improve the profitability of the firm. Measuring the impact of product quality on demand often requires models capable of dealing with more than simple discrete choices, where only one unit of one product is chosen. Multi-part pricing, time, space and other constraints likewise impact the attainable utility consumers can achieve. Firms considering changes to their product line require measures of consumer satisfaction and compensating values based on flexible patterns of substitution that are not pre-ordained by properties such as IIA (Allenby 1989). A potential disadvantage is that economic models of choice and demand can be too simplistic, as often pointed out by behavioral decision theorists. Our view is that much of the criticism of standard economic models of choice results from simplistic model assumptions rather than a defect in the fundamental paradigm of constrained choice. Our view is that the solution is to develop richer specifications of utility and constraints than to reject an economic formulation.

This chapter in the *Handbook of Marketing Decision Models* starts with a simple discrete choice model that has become a workhorse model in marketing over the last 25 years. The simple discrete choice model allows us to introduce terminology and basics of choice modeling. We then expand our discussion by considering direct utility models that allow the study of utility formation separate from the role of constraints, which provide insights into what people give up in an exchange. The direct utility formulation also allows us to model choices in the context of continuous demand where more than one item may be selected. Throughout our analysis, we offer a critical assessment of model assumptions and point to future directions for additional research.

7.2 A Simple Model of Discrete Choice

The discrete choice model is characterized by one and only one choice alternative being selected at a time. This model of choice is applicable to a wide variety of product categories, ranging from automobiles to smartphones to cordless power drills. Consumers are assumed to be endowed with a budget for the purchase that determines the upper limit of expenditure they are willing to make. If the good costs less than the budgeted amount, which we denote "E" for expenditure, then the remainder of the unspent money $(E-p)$ can be used for other purposes. Thus, we conceive of choice as between a number of different "inside" goods and an "outside" good that represents money unspent in the product category. The utility function for this situation can be represented as:

$$u(x,z) = \sum_{k=1} \psi_k x_k + \psi_z z$$

where x is a vector of demand for the inside goods, z is the demand for the outside good or non-purchase, ψ_j is the marginal utility of choosing the j^{th} good and ψ_z is the marginal utility for money. It is customary in choice models to include an error term to allow for unobserved factors affecting choice. We can then consider the utility from selecting each of the alternatives as:

$$\begin{aligned} u\left(x_1 = 1, z = E - p_1\right) &= \psi_1 + \psi_z\left(E - p_1\right) + \varepsilon_1 \\ u\left(x_2 = 1, z = E - p_2\right) &= \psi_2 + \psi_z\left(E - p_2\right) + \varepsilon_2 \\ &\vdots \\ u\left(x = 0, z = E\right) &= \psi_z\left(E\right) + \varepsilon_z \end{aligned}$$

The utility for each of the inside goods is comprised of three terms (i) the marginal utility for the inside good; (ii) utility for the unspent money that can be put to other use; and (iii) an error term. The utility for the outside good is just the utility for the budgeted allotment, E, plus error.

Consumers are assumed to select the choice alternative that provides them greatest utility. The choice model becomes:

$$\begin{aligned} \Pr(j) &= \Pr\left(\psi_j + \psi_z\left(E - p_j\right) + \varepsilon_j > \psi_k + \psi_z\left(E - p_k\right) + \varepsilon_k \text{ for any } k \neq j\right) \\ &= \Pr\left(V_j + \varepsilon_j > V_k + \varepsilon_k \text{ for any } k \neq j\right) \\ &= \Pr\left(\varepsilon_k < V_j - V_k + \varepsilon_j \text{ for any } k \neq j\right) \\ &= \int_{-\infty}^{+\infty} \left[\int_{-\infty}^{V_j - V_1 + \varepsilon_j} \cdots \int_{-\infty}^{V_j - V_k + \varepsilon_j} \phi\left(\varepsilon_k\right) \cdots \phi\left(\varepsilon_1\right) \right] \phi\left(\varepsilon_j\right) d\varepsilon_k \cdots d\varepsilon_1 d\varepsilon_j \\ &= \int_{-\infty}^{+\infty} \prod_{k \neq j} \Phi\left(V_j - V_k + \varepsilon_j\right) \phi\left(\varepsilon_j\right) d\varepsilon_j \end{aligned}$$

where Φ denotes the cdf and ϕ denotes the pdf of the distribution of ε. Distributional assumptions play a role in determining the functional form of the choice model, with extreme value errors leading to the logit model and normally distributed errors giving rise to the probit model. Assuming standard extreme value errors (i.e., $EV(0,1)$) results in the following logit expression for the choice probability:

$$
\begin{aligned}
Pr(j) &= \frac{\exp\left[V_j\right]}{\exp\left[V_z\right] + \sum_k \exp\left[V_k\right]} \\
&= \frac{\exp\left[\psi_j + \psi_z\left(E - p_j\right)\right]}{\exp\left[\psi_z(E)\right] + \sum_k \exp\left[\psi_k + \psi_z\left(E - p_k\right)\right]} \\
&= \frac{\exp\left[\psi_j - \psi_z p_j\right]}{1 + \sum_k \exp\left[\psi_k - \psi_z p_k\right]}
\end{aligned}
\tag{7.1}
$$

Thus, the choice probability is a function of a choice-specific intercept $\left(\psi_k\right)$ and a price term with a coefficient that is common across the choice alternatives.

This simple model of choice is used extensively in marketing because of its computational simplicity. Demand is restricted to two points for each of the inside goods, $\{0, 1\}$, with only one good allowed to be chosen and the remaining budget allocated to the outside good, z. Moreover, because utility is specified as linear, the budgetary allotment, E, cancels out of the expression and does not figure into the choice probability specification in a meaningful way, other than to remove choice options from the denominator of the probability expression when the price is too high, i.e., $p_k > E$.

The discrete choice model requires an additional constraint to be statistically identified, or estimable, from a choice dataset. It is traditionally assumed that the scales of the error terms (ε) are set to one $\sigma = 1$. The likelihood function for the standard discrete choice model is therefore equal to the product of the individual choice probabilities:

$$
\pi\left(y_t|\psi\right) = \prod_t \Pr(1)_t^{y_{1t}} \Pr(2)_t^{y_{2t}} \cdots \Pr(k)_t^{y_{kt}} \Pr(z)_t^{y_{zt}}
$$

where ψ represents the model parameters and $y_{i,t}$ are the multinomial choices with one element equal to one and the rest equal to zero. Maximum likelihood estimates of the model parameters are the parameters that maximize the joint probability of the observed choices. Bayesian estimates of the model parameters introduce a prior distribution, $\pi(\psi)$, that is combined with the likelihood to derive the posterior distribution $\pi(\psi|y) \propto \pi(y|\psi)\,\pi(\psi)$ (see Rossi et al. 2005). In a Bayesian analysis, point estimates of model parameters are typically taken as the mean of the posterior distribution.

7.2.1 Applications and Extensions

Thousands of research articles have been written that have extended and/or applied the logit choice model to choice data. Groundbreaking work on the logit model and transportation choice can be traced back to the work of McFadden (1973, 1986), which was applied to marketing demand data by (Guadagni and Little 1983) and others. The linearity of the utility function results in linear indifference curves and corner solutions, where only one of the choice alternatives is selected.

The introduction of Bayesian statistical methods into marketing (Rossi et al. 2005) led to the incorporation of respondent heterogeneity in choice models, which greatly increased their popularity, especially in the context of conjoint analysis (Green and Srinivasan 1978). Allenby and Ginter (1995) examined the use of a binary logit model in market segmentation to understand which respondents are most likely to respond to a product reformulation, and (Lenk et al. 1996) studied the use of partial factorial designs and their ability to inform the parameters of this class of models. The results of these and subsequent studies demonstrated that choice models were useful for accurately representing preferences for a heterogeneous set of consumers. Moreover, empirical results over the years have supported the use of economic models to represent consumer preferences and sensitivities to variables like prices. Many models of choice specified without heterogeneity employed many interaction terms to represent demand. These interaction terms have largely disappeared from the published literature in the presence of heterogeneity.

The simple logit model of discrete choice described above has been extended in many ways. Allenby and Rossi (1991) propose a model that maintains linearity of the indifference curves but allows them to rotate in the positive orthant to represent goods of different levels of quality. As the budgetary allotment E is relaxed, their model predicts that consumers would trade-up to higher quality offerings. Berry et al. (1995) introduce demand shocks into a logit model to accommodate factors that shift the utility of all consumers in a market while parsimoniously representing a demand system. Their model has become a standard in the empirical industrial organization literature. The logit model has been generalized by Marshall and Bradlow (2002) to accommodate a variety of preference measures beside simple choice, Edwards and Allenby (2003) discuss multivariate extensions, and Chandukala et al. (2007) provide an review of choice models in marketing. Most recently, Allenby et al. (2014a) and Allenby et al. (2014b) discuss using a simple choice model to estimate the economic value of product features. Proceedings of the triennial Invitational Choice Symposium, in the journal *Marketing Letters*, provides an ongoing review of innovations and applications of the simple discrete choice model.

7.3 A General Model for Choice

We now consider a general model for choice that allows for the possibility that more than one offering may be selected, often referred to as models of multiple discreteness (Kim et al. 2002). The demand for multiple offerings is common in the purchase of goods offering different varieties, such as flavors of a good, and whenever more than one unit is purchased at a time. Allowing for the possibility of purchasing multiple units require us to employ a calculus-based approach to associate observed choices to constrained utility maximization. It is not feasible to search over a continuous demand space to find the utility maximizing solution. Instead, first-order conditions (i.e., setting derivatives to zero) are used to connect utility maximization to observed demand.

We begin with a utility specification that leads to a version of the standard discrete choice model discussed earlier. Consumers are assumed to be utility maximizers subject to a budgetary constraint. Utility is specified logarithmically for the inside goods, linearly for the outside good, and with a parameter γ that introduces flexibility in the rate of satiation (see Bhat 2005). The assumption of a linear outside good is almost universally made in quantitative choice models, and, as shown below, significantly degrades the fit of models relative to a non-linear specification:

$$\text{Max } u(x, z) = \sum_k \frac{\psi_k}{\gamma} \ln\left(\gamma x_k + 1\right) + z \quad \text{subject to} \quad p'x + z \le E \tag{7.2}$$

where x is a vector of demand of dimension k, $\{\psi_k\}$ are baseline utility parameters, γ is a satiation parameter constrained to be positive, p is a vector of prices, z is an outside good with price equal to one, and E is the expenditure allocation. Equation (7.2) is additively separable and therefore assumes the goods are substitutes. The form of the utility function is selected because of the simplicity of the expression for marginal utility:

$$u_k = \frac{\partial u(x, z)}{\partial x_k} = \frac{\psi_k}{\gamma x_k + 1}$$
$$u_z = \frac{\partial u(x, z)}{\partial z} = 1$$

Marginal utility for the inside goods diminishes as quantity $\{x_k\}$ increases, and is equal to ψ_k when $x_k = 0$. The rate of satiation, or the rate at which marginal utility decreases, is governed by the satiation parameter γ. A plot of marginal utility as a function of quantity (x_k) is provided in Fig. 7.1.

We solve for the utility maximizing solution by the method of Lagrangian multipliers that combines the constraint and utility function by introducing a parameter λ that ensures their slopes are proportional, or that the utility function and budget constraint are tangent, at the point of constrained maximization:

$$\text{Max } L = \sum_k \frac{\psi_k}{\gamma} \ln\left(\gamma x_k + 1\right) + z + \lambda(E - p'x - z)$$

Setting partial derivatives to zero we obtain the optimality conditions:

$$\frac{\partial L}{\partial x_k} = \frac{\psi_k}{\gamma x_k + 1} - \lambda p_k = 0$$
$$\frac{\partial L}{\partial z} = 1 - \lambda = 0$$

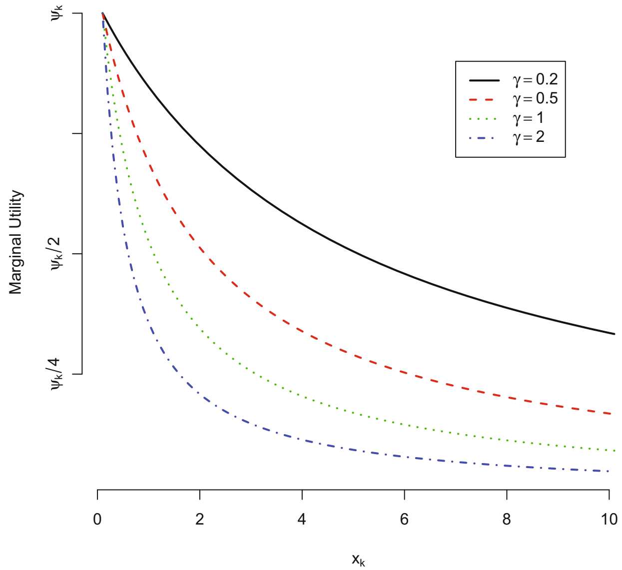


Fig. 7.1 Marginal Utility

From the second equation we see that $\lambda = 1$, and we can substitute for λ in the first equation to obtain:

$$\frac{\psi_k}{\gamma x_k + 1} = p_k$$

This expression holds whenever demand is positive, or $x_k > 0$, indicating that marginal utility is equal to price for positive demand, i.e., the "bang" is equal to the "buck." When demand is observed to be zero, we have the condition that marginal utility is less than the price, or that the bang is less than the buck. Re-arranging terms results in an explicit expression for observed demand x:

$$x_k = \frac{\psi_k - p_k}{\gamma p_k} \quad \text{for} \quad \psi_k > p_k \quad \text{else} \quad x_k = 0$$

A plot of demand is provided in Fig. 7.2 for $\psi_k = 8$ and $E = \$8.00$.

Demand is increasing in the level of baseline marginal utility (ψ_k) and decreasing in price and the satiation parameter γ. An advantage of this expression is that demand is declining in prices, which is sometimes violated in regression-based models of demand. A disadvantage is that cross-effects are not present. Below we show that this is due to assuming that the utility in (7.2) is additively separable and the outside good (z) does not satiate. Non-satiation of the outside good results in the utility maximizing solution unaffected by the budgetary allotment, or expenditure (E), similar to that encountered with the simple model of discrete choice discussed earlier.

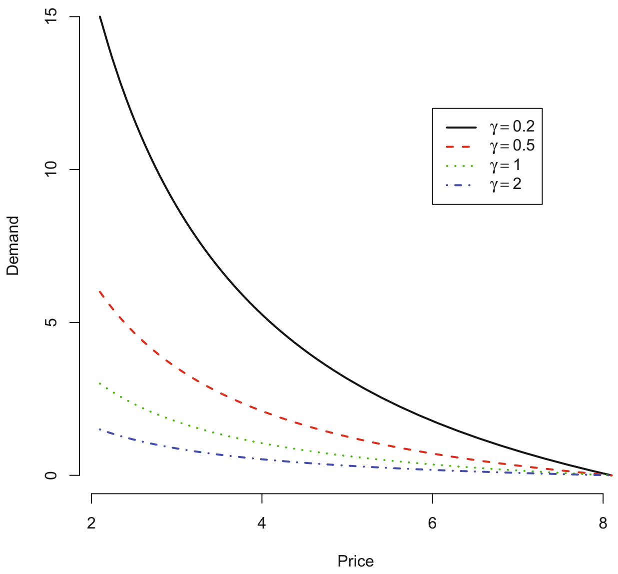


Fig. 7.2 Demand Curves

The general form of the above solution is referred to as the Kuhn-Tucker (KT) conditions of utility maximizing demand:

$$\text{if} \quad x_k > 0 \quad \text{and} \quad x_j > 0 \quad \text{then} \quad \lambda = \frac{u_k}{p_k} = \frac{u_j}{p_j} \quad \text{for all } k \text{ and } j$$

$$\text{if} \quad x_k > 0 \quad \text{and} \quad x_j = 0 \quad \text{then} \quad \lambda = \frac{u_k}{p_k} > \frac{u_j}{p_j} \quad \text{for all } k \text{ and } j$$

where λ is equal to the marginal utility of the outside good u_z because its price is normalized to equal one. We regard the above utility model as a basic structure from which to build various models of demand. The simplicity of the model leads to closed-form expressions for demand forecasting, and nests the standard discrete choice model where the data indicate the most preferred choice option instead of demand quantities. The KT condition for preference data, where respondents are asked to indicate their preference without reference to quantities (i.e., $x = 0$) reduces to:

$$\text{if } k \text{ preferred then} \quad \frac{\psi_k}{p_k} > \frac{\psi_j}{p_j} \quad \text{for all } j$$

Taking logarithms leads to an expression similar to Equation (1) except that price is replaced by $\ln(p_j)$. The analysis of volumetric demand data (i.e., $x_k > 0$), however, is more informative of model parameters because the equality restrictions in the KT conditions are more informative than inequality restrictions.

7.3.1 Statistical Specification

Variation in observed demand for a respondent often requires the introduction of error terms to rationalize choice. It is convenient to introduce error terms in the baseline utility parameters by specifying a functional form that ensures that marginal utility is always positive:

$$\psi_{kt} = \exp\left[a'_{kt}\beta + \varepsilon_{kt}\right]$$

where a_{kt} is a vector of attributes of the k^{th} good and the error term is allowed to vary over time (t). The parameters β are sometimes referred to as "part-worths" in conjoint analysis, reflecting the partial worth of product attributes and benefits. Substituting the expression for ψ_{kt} into the expression that equates the Lagrange multiplier (λ) to the ratio of marginal utility to price, and recalling that $\lambda = u_z/1 = 1$ results in the expression:

$$\frac{\exp\left[a'_{kt}\beta + \varepsilon_{kt}\right]}{\gamma x_{kt} + 1} = p_{kt}$$

Solving for ε_{kt} results in the following expression for the KT conditions:

$$\varepsilon_{kt} = g_{kt} \quad \text{if} \quad x_{kt} > 0 \tag{7.3}$$

$$\varepsilon_{kt} < g_{kt} \quad \text{if} \quad x_{kt} = 0 \tag{7.4}$$

where

$$g_{kt} = -a'_{kt}\beta + \ln(\gamma x_{kt} + 1) + \ln(p_{kt})$$

The assumption of *i.i.d.* extreme-value errors, i.e., EV(0,σ), results in a closed-form expression for the probability that R_t of N goods are chosen. Indexing the chosen goods by $n_{1,t}$ and the remainder by $n_{2,t}$ results in the following expression for the likelihood:

$$\begin{aligned}
\Pr(x_t) &= \Pr(x_{n_{1,t}} > 0, \quad x_{n_{2,t}} = 0, \quad n_{1,t} = 1, \ldots, R_t, \quad n_{2,t} = R_t + 1, \ldots, N) \\
&= |J_{R_t}| \int_{-\infty}^{g_N} \cdots \int_{-\infty}^{g_{R_t+1}} f(g_{1t}, \ldots, g_{R_t}, \varepsilon_{R_t+1}, \ldots, \varepsilon_N) d\varepsilon_{R_t+1}, \ldots, d\varepsilon_N \\
&= |J_{R_t}| \left\{ \prod_{i=1}^{R_t} \frac{\exp(-g_{it}/\sigma)}{\sigma} \exp\left(-e^{-g_{it}/\sigma}\right) \right\} \left\{ \prod_{j=R_t+1}^{N} \exp\left(-e^{-g_{jt}/\sigma}\right) \right\} \\
&= |J_{R_t}| \left\{ \prod_{i=1}^{R_t} \frac{\exp(-g_{it}/\sigma)}{\sigma} \right\} \exp\left\{ -\sum_{j=1}^{N} \exp(-g_{jt}/\sigma) \right\}
\end{aligned}$$

where $f(\cdot)$ is the joint density distribution for ε and $|J_{R_t}|$ is the Jacobian of the transformation from random-utility error (ε) to the likelihood of the observed data (x). For this model, the Jacobian is equal to:

$$\left|J_{R_t}\right| = \prod_{i=1}^{R_t} \frac{\gamma}{\gamma x_{i,t} + 1}$$

The expression for the probability of the observed demand vector x_t is seen to be the product of R_t "logit" expressions multiplied by the Jacobian, where the purchased quantity, x_{it} is part of the value (g_{it}) of the choice alternative. For the standard discrete choice model, $g_{kt} = -a'_{kt}\beta + \ln(p_{kt})$ and the Jacobian is equal to one because demand (x) enters the KT conditions through the conditions $x_{kt} > 0$ or $x_{kt} = 0$ only in Eqs. (7.3) and (7.4). The price coefficient in the standard choice model is the scale value of the Extreme Value error ($1/\sigma$). Variation in the specification of the choice model utility function and budget constraint results in different values of (g_{kt}) and the Jacobian $|J_{R_t}|$, but not to the general form of the likelihood, i.e.,

$$\Pr(x_t) = |J_{R_t}| \left\{ \prod_{i=1}^{R_t} f(g_{it}) \right\} \left\{ \prod_{j=R_t+1}^{N} F(g_{jt}) \right\}$$

7.3.2 Non-linear Outside Good

The assumption that utility is linear in the outside good (z) results in KT conditions that do not involve the budgetary allotment E. A non-linear specification for the outside good leads to a demand model where the budgetary allotment plays a role in identifying the utility maximizing solution. This is important as it allows for the presence of cross-price effects on the demand for each of the items. For example, the utility function with logarithmic specification for the quantity of the outside good leads to:

$$u(x, z) = \sum_k \frac{\psi_k}{\gamma} \ln\left(\gamma x_k + 1\right) + \ln(z)$$

has marginal utility for the outside good equal to:

$$u_z = \frac{\partial u(x, z)}{\partial z} = \frac{1}{z}$$

and the KT condition $\lambda = u_z/1 = u_k/p_k$ leads to new expressions for g_{kt} and the Jacobian:

$$g_{kt} = -a'_{kt}\beta + \ln(\gamma x_{kt} + 1) + \ln\left(\frac{p_{kt}}{E - p'_t x_t}\right) \tag{7.5}$$

$$
\left|J_{R_t}\right| = \det\left[\frac{\partial g_{R_t}}{\partial x_{R_t}{}'}\right] = \det\begin{bmatrix} \frac{\gamma}{\gamma x_{1t}+1}+\frac{p_{1t}}{E-p_t{}'x_t} & \frac{p_{2t}}{E-p_t{}'x_t} & \cdots & \frac{p_{R_t}}{E-p_t{}'x_t} \\ \frac{p_{1t}}{E-p_t{}'x_t} & \frac{\gamma}{\gamma x_{2t}+1}+\frac{p_{2t}}{E-p_t{}'x_t} & \cdots & \frac{p_{R_t}}{E-p_t{}'x_t} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{p_{1t}}{E-p_t{}'x_t} & \frac{p_{2t}}{E-p_t{}'x_t} & \cdots & \frac{\gamma}{\gamma x_{R_t}+1}+\frac{p_{R_t}}{E-p_t{}'x_t} \end{bmatrix}
$$

$$
= \prod_{k=1}^{R_t}\left(\frac{\gamma}{\gamma x_{kt}+1}\right)\left\{\sum_{k=1}^{R_t}\frac{\gamma x_{kt}+1}{\gamma}\cdot\frac{p_{kt}}{E-p_t{}'x_t}+1\right\}
$$

The off-diagonal elements of the Jacobian are non-zero because of the right-most term in the expression for g_{kt}. The expenditure allotment E can be treated as a parameter and is statistically identified through the KT condition associated with positive demand that result in the equality restrictions $\varepsilon_{kt} = g_{kt}$. However, its estimated value will depend on the degree of assumed concavity of the utility function for the outside good (e.g., logarithm versus a power function).

Predicted demand for the model with non-linear outside good does not have a closed form because the KT conditions leads to an implicit function for x:

$$
x_k = \frac{\psi_k - \lambda p_k}{\lambda\gamma p_k} \quad \text{for} \quad \psi_k > \lambda p_k \quad \text{else} \quad x_k = 0
$$

where $\lambda = \frac{1}{E-p'x}$. However, there are methods for obtaining demand estimates. A general solution is using standard constrained optimization routines such as `constrOptim` in the R statistical package that directly maximizes the utility function subject to the budget constraint.

7.3.3 Applications and Extensions

Volumetric (non-binary) demand is common in marketing, occurring in the consumption of most packaged goods and services. The quantity purchased is often not restricted to just a single unit of a good, and the development of a quantity-based model reduces the number of distinct choice alternatives that need to be modeled. For example, in the beverage category, soda is routinely sold as 6-packs and 12-packs of 12 ounce cans. The decision to purchase a 12-pack of Coke reflects both an item and quantity decision, and economic models used to rationalize this choice treat the demand quantities as the outcome of a constrained utility problem. The advantage of this is that it requires fewer parameters and error terms than in models that treat different package sizes as having unique intercept and error terms. Moreover, it can handle zero demand quantities and is especially well suited for sparse data environments.

Models of discrete and continuous demand were pioneered by Hanemann (1984) who coupled a discrete choice (logit) model with a conditional demand model. The horizontal variety literature on multiple discreteness models (e.g., Kim et al. 2002; Bhat 2005, 2008) extends these models to allow for the selection of more than one choice alternative. In addition, the direct utility model described above is flexible with regard to the utility function that is employed. Lee et al. (2013), for example, study the presence of complements where goods have a super-additive effect on consumer utility.

The direct utility approach has been usefully extended to the areas where data are from non-consumer product categories. Luo et al. (2013), for example, investigated how consumers allocate time resources among leisure activities over time based on dynamic version of direct utility specifications above. In similar vein, Lin et al. (2013) study consumers' media consumptions such as TV, radio, and internet accounting for substitution and complementarities in multiplexing activities.

7.4 Constraints

An advantage of using a direct utility model to study consumer purchase decisions is that it separates what is gained in an exchange from what is given up. Consumers give up various resources for the right to acquire and use marketplace offerings that provide them with utility. These resources include money, time, attention, and any other constraint on their lives. Dieters, for example, pay attention to the caloric content of the food they consume and may not purchase items on sale if they are high in calories. Consumers constrained by space may not purchase large package sizes, and consumers may not purchase goods with large access costs, such as the fixed costs associated with learning to play a new sport. In many cases, consumer choice is governed more by constraints on available options than by the utility afforded by different offerings. We begin this section with a discussion of choice models with multiple constraints (Satomura et al. 2011) and then examine models with non-linear constraints (Howell et al. 2015; Howell and Allenby 2015).

7.4.1 Multiple Constraints

We develop our model of multiple constraints for consumers constrained by money and quantity. Quantity constraints arise when consumers have limited storage space in their homes that they wish to develop to a product category. Space constraints are represented as Q denoting the upper limit of quantity. Consumers are assumed to make choices that maximize utility subject to multiple constraints:

$$\text{Max } u(x,z,w) = \sum_k \frac{\psi_k}{\gamma} \ln\left(\gamma x_k + 1\right) + \ln(z) + \ln(w)$$

$$\text{subject to} \quad p'x + z \le E \quad \text{and} \quad q'x + w \le Q$$

The utility maximizing solution is found by forming the auxiliary function L, but this time with two multipliers λ and μ:

$$\text{Max } L = u(x, z, w) + \lambda\{M - p'x - z\} + \mu\{Q - q'x - w\}$$

resulting in the following first-order conditions for constrained utility maximization:

$$\begin{aligned}
&\varepsilon_{kt} = g_{kt} \quad \text{if} \quad x_{kt} > 0 \\
&\varepsilon_{kt} < g_{kt} \quad \text{if} \quad x_{kt} = 0 \\
&g_{kt} = -a'_{kt}\beta + \ln(\gamma x_{kt} + 1) + \ln\left(\frac{p_{kt}}{E - p'_t x_t} + \frac{q_k}{Q - q' x_t}\right)
\end{aligned}$$

that differs from the earlier expression in Equation (5) in that the last term involves both the budget and quantity restrictions. As either $p'_t x_t$ approaches E, or $q' x_t$ approaches Q, the last term on the right side becomes large, making it less likely to observe positive demand ($x_{kt} > 0$) and more likely to observe zero demand ($x_{kt} = 0$). Thus, goods that tend to exhaust either of the allocated budgets E and Q are less likely to be selected.

The Lagrangian multipliers λ and μ can be shown to be the expected change in attainable utility for a unit change in the constraint (see Sydsæter et al. 2005, Chap. 14). Thus, one can evaluate the impact of the constraints on choice and utility, and determine which is more profitable for the consumer to relax. Budgets (E) can be relaxed by endowing consumers with greater wealth by the use of coupons and other means of temporary price reductions, and quantity constraints (Q) can be relaxed by improvements in packaging and other forms of space saving. By comparing the cost of these changes to the expected increase in utility allows firms to determine which constraint to relax.

7.4.2 *Non-linear Constraints*

Non-linear constraints arise when costs, viewed in a broad sense, do not scale in proportion to the quantity consumed. Examples include fixed costs that are incurred just once for the first unit of demand (Howell and Allenby 2015), access costs that arise as consumers transform market-place goods for consumption (Kim et al. 2015b), and when unit prices depend on the quantity purchased. An example of a fixed cost is the cost of a coffee maker, while an access cost for coffee involves the purchase of coffee beans and its daily preparation. Access costs might be shared among different choice alternatives that affects the variety of goods consumed (Kim et al. 2015a). Quantity-dependent pricing, often referred to as multi-part pricing, has been studied extensively in analytic models but are often difficult to implement in practice because of the different prices that consumers face. Non-linear pricing can result in irregular budget sets, where the budgetary constraint having kink points and possibly points

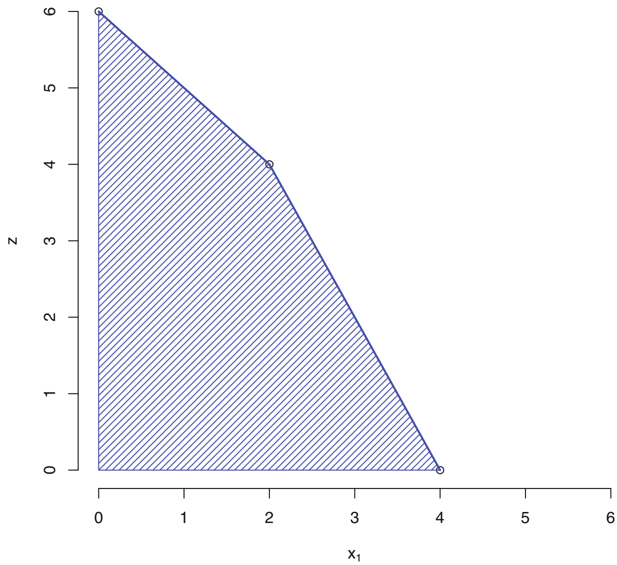


Fig. 7.3 Irregular budget set

of discontinuity. When this occurs, it is not possible to use first-order conditions to find a global optimal quantity of demand that maximizes constrained utility.

Figure 7.3 displays the budget constraint for two goods, x_1 and z, when $E = \$6.00$ and the price for the first two units of x is \$1.00, and then the price rises to \$2.00 per unit. The budgetary constraint has a kink point at $x_1 = 2$ units that will result in a build-up of mass in the likelihood of demand. That is, many consumers will want to purchase two units of x because of the low price. The solution to modeling demand when the budgetary constraint has a kink point is to employ first-order conditions to find optimal demand quantities within regions of the budget set that are linear, and then to compare the solutions to find the utility maximizing demand.

Consider the case where there are two inside goods, x_1 and x_2 with kink points at τ_1 and τ_2. Price is assumed to take on a low value (p_ℓ) below the kink point and a higher value above the kink point (p_h). We can partition the demand space into four regions:

$$\begin{aligned}
\mathbb{P}_1 : \ & \text{Max } u(x_{1t}, x_{2t}, z_t) \\
& \text{s.t. } p_{\ell 1}x_{1t} + p_{\ell 2}x_{2t} + z_t = E \\
& \quad 0 \le x_{1t} \le \tau_1, 0 \le x_{2t} \le \tau_2 \\
\mathbb{P}_2 : \ & \text{Max } u(x_{1t}, x_{2t}, z_t) \\
& \text{s.t. } p_{\ell 1}\tau_1 + p_{h1}(x_{1t} - \tau_1) + p_{\ell 2}x_{2t} + z_t = E \\
& \quad \tau_1 < x_{1t}, 0 \le x_{2t} \le \tau_2
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}_3 :\ & \text{Max } u(x_{1t}, x_{2t}, z_t) \\
& \quad \text{s.t. } p_{\ell 1}\tau_1 + p_{\ell 2}x_{2t} + p_{h2}(x_{2t} - \tau_2) + z_t = E \\
& \quad\quad 0 \le x_{1t} \le \tau_1, \tau_2 < x_{2t} \\
\mathbb{P}_4 :\ & \text{Max } u(x_{1t}, x_{2t}, z_t) \\
& \quad \text{s.t. } p_{\ell 1}\tau_1 + p_{\ell 2}\tau_2 + p_{h1}(x_{1t} - \tau_1) + p_{h2}(x_{2t} - \tau_2) + z_t = E \\
& \quad\quad \tau_1 < x_{1t}, \tau_2 < x_{2t}.
\end{aligned}$$

Then, the first-order conditions associated with observed demand are:

$$\begin{aligned}
&\varepsilon_{kt} < g_{\ell kt} && \text{if } x_{kt} = 0 \\
&\varepsilon_{kt} = g_{\ell kt} && \text{if } 0 < x_{kt} < \tau_k \\
&g_{\ell kt} < \varepsilon_k < g_{hkt} && \text{if } x_{kt} = \tau_k \\
&\varepsilon_{kt} = g_{hkt} && \text{if } x_{kt} > \tau_k
\end{aligned}$$

where

$$\begin{aligned}
g_{\ell kt} &= -a_k'\beta + \ln(\gamma x_{kt} + 1) + \ln\left(\frac{p_{\ell kt}}{z_t}\right) \\
g_{hkt} &= -a_k'\beta + \ln(\gamma x_{kt} + 1) + \ln\left(\frac{p_{hkt}}{z_t}\right)
\end{aligned}$$

when the outside good is specified logarithmically. We define the set $A = \{k : x_{kt} = 0\}$, the set $B = \{k : 0 < x_{kt} < \tau_k\}$, the set $C = \{k : x_{kt} = \tau_k\}$, and the set $D = \{k : x_{kt} > \tau_k\}$. The likelihood is therefore:

$$\begin{aligned}
Pr(x_{kt}) &= Pr(x_{At} = 0, 0 < x_{Bt} < \tau_k, x_{Ct} = \tau_k, x_{Dt} > \tau_k) \\
&= |J_{B\cup D}| Pr(\varepsilon_{At} < g_{\ell At}, \varepsilon_{Bt} = g_{\ell Bt}, g_{\ell Ct} < \varepsilon_{Ct} < g_{hCt}, \varepsilon_{Dt} = g_{hDt}) \\
&= |J_{B\cup D}| \times F(g_{\ell At}) \times f(g_{\ell Bt}) \times (F(g_{hCt}) - F(g_{\ell Ct})) \times f(g_{hDt})
\end{aligned}$$

where f is the pdf of the error distribution and F is the CDF of that distribution. $J_{B,D}$ is the Jacobian of $\varepsilon_{C\cup D}$ and is defined as:

$$J_{ij} = \frac{\partial g_{\cdot i}}{\partial x_{jt}} = \frac{1}{x_{it} + 1} I(i = j) + \frac{p_{\cdot jt}}{z_t}$$

with $\cdot = \ell$ if $i \in B$, and $\cdot = h$ if $i \in D$.

The above framework can be extended to an arbitrary number of kink points and an arbitrary number of goods. The challenge in applying the model is in keeping track of the number of possible outcome regions (e.g., regions A–D above) and correctly computing the likelihood.

More generally, constraints limit the region over which utility is maximized and often results in the need for estimators that do not rely solely on first-order conditions to associate observed demand and choices to model parameters. Demand, for example, could take the form of a mixture of demand types, with it being discrete for some decision variables (e.g., the decision to stream music) and continuous for others (the number of songs to download) (Kim et al. 2015a). Another example involves the use of screening rules and the presence of consideration sets in consumer decision making (Kim et al. 2015c) that involves a subset of the items available for sale. Although such constraints complicate the estimation of direct utility models, many creative procedures have been proposed to relate observed demand to constrained utility maximization for model estimation.

7.5 Error Specification

We considered the form of the utility function and the nature of constraints in developing variants of direct utility choice models in the discussion above. We now consider the influence of the error term in the model. The error term plays an important role in models of choice because of what it implies about factors that are not explicitly present in the utility function or the budget constraint. In general, if an iid (independent and identically distributed) additive error term is used in the model, there are some realizations of the error that guarantee that a particular good will be chosen. That is, there will not exist dominated alternatives for which demand is zero for any respondent. While this assumption might seem harmless for the analysis of small choice sets, it is problematic whenever the choice set becomes large and every alternative is assigned some positive purchase probability that is independent of a brand's attributes. Most product categories consist of dozens, and sometimes hundreds, of goods for sale. Assuming iid error terms for each alternatives leads to demand predictions that are not sufficiently sensitive to changes in prices and other product features, particularly when there are some brands that compete at a heightened level with other brands.

Another problem with standard error assumptions is that observed demand is often discrete, not continuous, and is often constrained by packaging decisions made by firms. Consumers, for example, may not be able to purchase five cans of soda, or eight eggs at a grocery store. Marketplace demand must therefore be viewed as a censored outcome of a constrained model of choice, where the offerings made available reside on a coarse grid dictated by the packaging. Choice models must therefore be modified to acknowledge this restriction when inferring about model parameters and when predicting future demand.

7.5.1 Correlated Errors

One solution to recognizing groups of similar products is to allow the random utility errors to be correlated. The nested logit model (Hausman and McFadden 1984), for example, is motivated by the presence of correlated errors for goods that are similar. Correlated errors are most flexibly introduced into a choice model with Normally distributed error terms. Dotson et al. (2015) discuss a parsimonious model that allows for error covariances that are related to product features. The effect of the model is to relax the assumption of IIA associated with the standard logit model, where the ratio of choice probabilities for any two alternatives does not involve any of the other choice alternatives. With IIA, the introduction of a good similar to an existing good draws share proportionally to the baseline shares of all other goods and does not favor other, similar goods.

It is desirable that the correlations among the error terms of similar goods be large, and the correlation among goods that are dissimilar be small. In the extreme, if two goods are identical, the errors should be perfectly collinear and would split the demand associated with them. A simple model of correlated errors for N choice alternatives is:

$$\Sigma = \begin{bmatrix} 1 & \cdots & \sigma_{1N} \\ \vdots & \ddots & \vdots \\ \sigma_{N1} & \cdots & 1 \end{bmatrix}$$

where

$$\sigma_{kj} = \exp\left[\frac{-d_{kj}}{\theta}\right]$$

and d_{kj} is a measure of perceptual distance between goods k and j. Thus, if two goods are perceived to be nearly alike, then d_{kj} is close to zero and σ_{kj} is close to one. They investigate alternative parameterizations of the distance measure, and find that one based on baseline utility consistently fits the data best:

$$d_{kj} = \left|\psi_k - \psi_j\right|$$

where ψ is the marginal utility of the offering. Thus, the ψ parameters appear in the mean and covariance of the utility specification, which necessitates the need for customized software for model estimation. Dotson et al. (2015) discuss estimation of this model as a hierarchical Bayes model.

7.5.2 Indivisible Demand

Restriction on demand caused by the discreteness of packaging results in an additional constraint on choice that censors the model error term to produce integer demand:

$$x_{kt} \in \{0, 1, 2, \cdots\}, \quad \forall k \in \{1, \cdots, N\}$$

Thus, instead of consumers purchasing the alternative with greatest utility among those that satisfy the budgetary allotment, they are assumed to choose from among the available alternatives that maximize utility.

Lee and Allenby (2014) show how to deal with this constraint in model estimation and prediction. The model likelihood become a product of mass points, as there are multiple realizations of each model error term that can correspond to observed demand on the package grid. A variant of Bayesian data augmentation (Tanner and Wong 1987) is used to evaluate the likelihood at feasible points on the package grid:

$$\begin{aligned} &U^* \left(x_{1t}^*, \cdots, x_{nt}^*\right) \\ &\geq \max \left\{U^* \left(x_{1t}^* + \varDelta_1, \cdots, x_{nt}^* + \varDelta_n\right) | \left(x_{1t}^* + \varDelta_1, \cdots, x_{nt}^* + \varDelta_n\right) \in F\right\}_{\varDelta_i \in \{-1,0,1\}} \end{aligned} \tag{7.6}$$

and use the inequality relationship to determine ranges of the error term that are consistent with the observed demand being utility maximizing. Their analysis indicates that data corresponding to the corner solutions are most affected by the presence of packaging constraints, where zero demand should be interpreted as not liking an offering enough to buy one unit, as opposed to not liking it enough to buy any.

7.6 Indirect Utility Models

Models of discrete/continuous choice have a long history in the marketing and economics literature beginning with the work of Hanemann (1984) and extended by many that build on a framework where demand is positive for just one of many different choice alternatives. Krishnamurthi and Raj (1988) discuss the estimation of models where the continuous component of demand is driven by a set of covariates different from the covariates used to form utility in the discrete choice portion of the model, and rely on a an indirect utility specification to provide a theoretical basis for demand quantities. Similarly, Harlam and Lodish (1995) introduce variables into their model specification that provide summary measures of merchandising activity that is difficult to interpret in terms of a direct utility model. Chintagunta (1993), Dubé (2004) and Song and Chintagunta (2007) also motivate their model specification using the concept of an indirect utility function without explicitly relating it to a specific direct utility model. An indirect utility function is defined as the maximal attainable utility as a function of prices and expenditure (E). Indirect utility models, however, are not amenable to disaggregate demand analysis in marketing where the attributes of products can change and there exist mass points of demand at particular prices.

For example, products in a conjoint analysis change across choice sets as product attributes and levels are experimentally manipulated. As the attribute-levels of the choice alternatives change, so should their degree of substitution and level of price interaction, indicating that many of the parameters of a indirect utility model would also need to be functions of product characteristics. It is difficult to implement a characteristics model of demand within an indirect utility model because indirect utility models often have many parameters. In addition, it is not clear how to incorporate model error into an indirect utility specification to allow for mass points of demand that occur at corners, kink points and due to packaging constraints. As discussed above, these issues can be addressed through a direct utility specification where corner solutions give rise to inequality constraints in the likelihood through the Kuhn-Tucker conditions.

To illustrate, consider a constrained maximization problem involving the utility function similar to (2):

$$\max \; u(x) = \sum_k \frac{\psi_k}{\gamma} \ln(\gamma x + 1) \quad \text{subject to} \;\; p'x \leq E$$

where the ψ's are assumed to sum to one (i.e., $\sum \psi_k = 1$). We can solve for the utility maximizing quantities, x^* and obtain the optimal demand function (see appendix):

$$x_k^* = \frac{1}{\gamma}\left(\frac{\psi_k}{\lambda p_k} - 1\right)$$

where $\lambda = \frac{1}{\gamma E + \sum_k p_k}$ is a Lagrangian multiplier. By substituting the demand function (x^*) into the utility function, one can obtain the expression for indirect utility (V) as follows:

$$V \equiv u(x^*) = \sum_k \frac{\psi_k}{\gamma}\left(\ln \psi_k - \ln p_k + \ln\left(\gamma E + \sum_k p_k\right)\right)$$

Details are provided in the appendix. While this formulation provides an elegant solution to optimal demand and indirect utility, it depends critically on equality constraints between the ψ parameters and optimal demand quantities x^* that are associated with interior solutions, not corner solutions. Corner solutions result in inequality restrictions, not equality restrictions, in the Kuhn-Tucker conditions.

The indirect utility function is often expressed as in terms of a Taylor series approximation to an unspecified utility function such as translog, and includes pairwise price interactions among the choice options so that a flexible pattern of substitution can be achieved (Pollak and Wales 1992). For example, consider a generalized quadratic indirect utility function often referred to as a translog indirect utility (Christensen et al. 1975):

$$\ln V = \alpha_0 + \sum_k \alpha_k \ln \frac{p_k}{E} + \frac{1}{2}\sum_k \sum_j \beta_{kj} \ln \frac{p_k}{E} \ln \frac{p_j}{E}$$

where $(\alpha, \beta)'$ are parameters that capture the substitution among product offerings. Recently, Mehta (2015) proposes an indirect utility model for a general demand model based on Kuhn-Tucker conditions (Wales and Woodland 1983) that employs virtual prices to deal with corner solutions (see Lee and Pitt 1987). Virtual prices are the prices at which demand is expected to be exactly equal to zero given the parameters of the indirect utility function. The virtual prices are then substituted into the demand system as if they were observed. The problem with this formulation is that it assumes more than what is observed, by conditioning on latent quantities, and overstates the value of information coming from zero demand by assuming a density contribution to the likelihood rather than a mass contribution.

The generalized quadratic indirect utility function is over-parameterize and, in general, not valid unless monotonicity and concavity constraints are imposed. There is no general theory of the extent to which this quadratic approximation provides a uniform functional approximation. Therefore, there is no reason to believe that the even a "regular" generalized quadratic indirect utility function can approximate any indirect utility for the purpose of demand specification. This would require a proof of global approximation not only of the indirect utility function but also of its derivatives. But, our principal objection to the indirect utility formulation is that it obscures the process of formulated direct utility models and associated constraints that embody the reality of the consumer purchase process. For example, if the consumer faces a non-linear budget set due to pack-size discounts, the researcher should write down the direct utility model and constraints rather than choosing an arbitrary indirect utility function which may not be consistent with this situation.

For virtually all situations in consumer choice modeling, there will be no closed form expression for the indirect utility function. Moreover, given corners and kinks, the indirect utility function may not be differentiable everywhere, eliminating the convenience of Roy's identity for deriving demand. Indirect utility functions are useful for welfare computations but are not of practical value in the specification of consumer choice models.

7.7 Conclusion

This chapter has reviewed an emerging area of analysis in marketing decision models that rationalizes choice from principles of constrained utility maximization. We advocate for a direct utility specification of choice models for a variety of reasons. Models of constrained utility maximization reflect goal-directed behavior on the part of consumers, which is overwhelmingly supported in disaggregate marketing data by the prevalence of zero demand for most offerings. By far, the most frequently observed number in disaggregate marketing datasets is the number zero, implying that consumers are resource conserving and not acting randomly.

The direct utility formulation separates that which is gained in an exchange (utility) from that which is given up (constraint)—i.e., the resources needed to acquire and use a marketplace offering. Understanding the determinants of utility is useful for product strategy as marketers advocate for making what people will want to buy. Quantifying the relationship between product attributes and the benefits, and the resulting utility afforded by a competitive set of products, is one of the most important tasks of marketing research. Likewise, understanding the impediments to acquiring and using a product is useful for driving sales and effectively communicating with prospects.

We examine three aspects of direct utility models—the utility function, constraints and error—that are combined to form the likelihood for the data. Our treatment of these constructs is structural in nature, as we avoid the temptation of simply adding an error term to a flexible model that combines brands, attributes and prices. The problem with taking this flexible approach is that it is not consistent with the lumpiness of marketing data, where corner solutions are prevalent, where demand is often constrained to lie on a grid of available package sizes, and where pricing discounts lead to a mass buildup in the likelihood even for interior solutions. Throughout our development above we stress the importance of deriving the likelihood function from principles of constrained optimization, and show how these realities of marketing data can be accommodated within the framework of constrained utility maximization.

We also advocate against the use of indirect utility models for data that contain corner solutions, mass buildup at specific demand quantities, or multiple constraints. While an indirect utility function can always be defined in these nonstandard situations, it may be difficult, if not impossible, to express in closed form. More importantly, the indirect utility function will not be useful in deriving the associated demand system and its associated likelihood.

Additional research is needed to develop and apply a broader class of utility functions, constraints, and error specifications for marketing analysis. For more than 50 years, marketing has embraced the notion of extended models of behavior where needs, wants, beliefs, attitudes, consideration and perceptions have been shown to be determinants of demand (Howard and Sheth 1969). Often, a large battery of variables is used to represent each of these constructs. Mapping these variables to one another and to marketplace demand within a principled structure is a worthy endeavor.

Appendix

Consider the direct utility model:

$$\max u(x) = \sum_k \frac{\psi_k}{\gamma} \ln\left(\gamma x_k + 1\right) \text{ subject to } p'x \le E$$

where the $\psi' s$ are assumed to sum to one (i.e., $\sum \psi_k = 1$) and $\gamma > 0$. Solving for the utility maximizing quantities, x^*, by the Lagrangian method will give rise to the following objective function:

$$L = \sum_k \frac{\psi_k}{\gamma} \ln\left(\gamma x_k + 1\right) - \lambda\left(p'x - E\right)$$

The FOC's for optimal demand (x^*) are:

$$
\begin{aligned}
0 = \frac{\partial L}{\partial x_1} = \frac{\psi_1}{\gamma x_1 + 1} - \lambda p_1 \qquad &\Leftrightarrow \qquad \psi_1 = \lambda p_1 \left(\gamma x_1 + 1\right) \\
0 = \frac{\partial L}{\partial x_2} = \frac{\psi_2}{\gamma x_2 + 1} - \lambda p_2 \qquad &\Leftrightarrow \qquad \psi_2 = \lambda p_2 \left(\gamma x_2 + 1\right) \\
&\vdots \\
0 = \frac{\partial L}{\partial x_k} = \frac{\psi_k}{\gamma x_k + 1} - \lambda p_k \qquad &\Leftrightarrow \qquad \psi_k = \lambda p_k \left(\gamma x_k + 1\right)
\end{aligned}
$$

Using $\sum \psi_k = 1$, we solve for λ:

$$
1 = \sum_{k=1}^{K} \psi_k = \sum_{k=1}^{K} \lambda p_k \left(\gamma x_k + 1\right)
$$

or

$$
\lambda = \frac{1}{\sum_{k=1}^{K} p_k \left(\gamma x_k + 1\right)} = \frac{1}{\gamma E + \sum_{k=1}^{K} p_k}
$$

Substituting λ back into FOCs yields optimal demand equations:

$$
\begin{aligned}
x_k^* &= \frac{1}{\gamma}\left(\frac{\psi_k}{\lambda p_k} - 1\right) \\
&= \frac{1}{\gamma}\left(\frac{\psi_k}{p_k}\left(\gamma E + \sum_k p_k\right) - 1\right)
\end{aligned}
$$

Substituting x^* into the direct utility function allows us to obtain the expression for indirect utility (V):

$$
\begin{aligned}
V &\equiv u\left(x^*\right) \\
&= \sum_k \frac{\psi_k}{\gamma} \ln\left(\gamma x_k^* + 1\right) \\
&= \sum_k \frac{\psi_k}{\gamma} \ln\left(\gamma\left(\frac{1}{\gamma}\left(\frac{\psi_k}{p_k}\left(\gamma E + \sum_k p_k\right) - 1\right)\right) + 1\right) \\
&= \sum_k \frac{\psi_k}{\gamma} \ln\left(\frac{\psi_k}{p_k}\left(\gamma E + \sum_k p_k\right)\right) \\
&= \sum_k \frac{\psi_k}{\gamma}\left(\ln \psi_k - \ln p_k + \ln\left(\gamma E + \sum_k p_k\right)\right)
\end{aligned}
$$

References

Allenby, Greg M. 1989. A unified approach to identifying, estimating and testing demand structures with aggregate scanner data. *Marketing Science* 8 (3): 265–280.

Allenby, Greg M., Jeff Brazell, John R. Howell, and Peter E. Rossi. 2014a. Valuation of patented product features. *Journal of Law and Economics* 57 (3): 629–663.

Allenby, Greg M., Jeff D. Brazell, John R. Howell, and Peter E. Rossi. 2014b. Economic valuation of product features. *Quantitative Marketing and Economics* 12 (4): 421–456.

Allenby, Greg M., James L. Ginter. 1995. Using extremes to design products and segment markets. *Journal of Marketing Research* 392–403.

Allenby, Greg M., and Peter E. Rossi. 1991. Quality perceptions and asymmetric switching between brands. *Marketing science* 10 (3): 185–204.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* 63 (4): 841–890.

Bhat, C.R. 2005. A multiple discrete-continuous extreme value model: Formulation and application to discretionary time-use decisions. *Transportation Research* 39 (8): 679–707.

Bhat, C.R. 2008. The multiple discrete-continuous extreme value (mdcev) model: Role of utility function parameters, identification considerations, and model extensions. *Transportation Research* 42 (3): 274–303.

Chandukala, Sandeep, Jaehwan Kim, Thomas Otter, Peter Rossi, and Greg Allenby. 2007. Choice models in marketing: Economic assumptions, challenges and trends. *Foundations and Trends in Marketing* 2 (2): 97–184.

Chintagunta, Pradeep K. 1993. Investigating purchase incidence, brand choice and purchase quantity decisions of households. *Marketing Science* 12 (2): 184–208.

Christensen, L.R., D.W. Jorgenson, and L.J. Lau. 1975. Transcendental logarithmic utility functions. *The American Economic Review* 3: 367–383.

Dotson, Jeff P., John R. Howell, Jeff D. Brazell, Thomas Otter, Peter J. Lenk, Steven N. MacEachern, and Greg M. Allenby. 2015. A probit model with structured covariance for similarity effects and source of volume calculations. *Working paper, Ohio State University*.

Dubé, Jean-Pierre. 2004. Multiple discreteness and product differentiation: Demand for carbonated soft drinks. *Marketing Science* 23 (1): 66–81.

Edwards, Yancy D., and Greg M. Allenby. 2003. Multivariate analysis of multiple response data. *Journal of Marketing Research* 40 (3): 321–334.

Green, Paul E., and Venkatachary Srinivasan. 1978. Conjoint analysis in consumer research: issues and outlook. *Journal of consumer research* 103–123.

Guadagni, Peter M., and John D.C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing science* 2 (3): 203–238.

Hanemann, W.M. 1984. Discrete/continuous models of consumer demands. *Econometrica* 52: 541–561.

Harlam, Bari A., and Leonard M. Lodish. 1995. Modeling consumers' choices of multiple items. *Journal of Marketing Research* 404–418.

Hausman, Jerry, Daniel McFadden. 1984. Specification tests for the multinomial logit model. *Econometrica: Journal of the Econometric Society* 1219–1240.

Howard, John A, Jagdish N Sheth. 1969. *The theory of buyer behavior*, vol. 14. Wiley New York.

Howell, John R, Greg M Allenby. 2015. Choice models with fixed costs. *Fisher College of Business Working Paper* .

Howell, John R, Sanghak Lee, and Greg M. Allenby. 2015. Price promotions in choice models. *Marketing Science* forthcoming.

Kim, Dongsoo, Sanghak Lee, Jaehwan Kim, and Greg M. Allenby. 2015a. A direct utility model for economies of scope and access. *Working paper, Ohio State University* .

Kim, Hyowon, Sanghak Lee, Jaehwan Kim, and Greg M. Allenby. 2015b. A direct utility model with discrete and continuous decision variables. *Working paper, Ohio State University* .

Kim, Jaehwan, Greg M. Allenby, and Peter E. Rossi. 2002. Modeling consumer demand for variety. *Marketing Science* 21 (3): 229–250.

Kim, Youngju, Jaehwan Kim, and Greg M. Allenby. 2015c. Conjunctive screening in models of multiple discreteness. *working paper, Ohio State Univeristy* .

Krishnamurthi, Lakshman, and S.P. Raj. 1988. A model of brand choice and purchase quantity price sensitivities. *Marketing Science* 7 (1): 1–20.

Lee, Lung-Fei, and Mark M. Pitt. 1987. Microeconometric models of rationing, imperfect markets, and non-negativity constraints. *Journal of Econometrics* 36 (1): 89–110.

Lee, Sanghak, and Greg M. Allenby. 2014. Modeling indivisible demand. *Marketing Science* 33 (3): 364–381.

Lee, Sanghak, Jaehwan Kim, and Greg M. Allenby. 2013. A direct utility model for asymmetric complements. *Marketing Science* 32 (3): 454–470.

Lenk, Peter J., Wayne S. DeSarbo, Paul E. Green, and Martin R. Young. 1996. Hierarchical bayes conjoint analysis: Recovery of partworth heterogeneity from reduced experimental designs. *Marketing Science* 15 (2): 173–191.

Lin, Chen, Sriram Venkataraman, and Sandy D. Jap. 2013. Media multiplexing behavior: Implications for targeting and media planning. *Marketing Science* 32 (2): 310–324.

Luo, Lan, Brian T. Ratchford, and Botao Yang. 2013. Why we do what we do: A model of activity consumption. *Journal of Marketing Research* 50 (1): 24–43.

Marshall, Pablo, and Eric T. Bradlow. 2002. A unified approach to conjoint analysis models. *Journal of the American Statistical Association* 97 (459): 674–682.

McFadden, Daniel. 1973. Conditional logit analysis of qualitative choice behavior. *Frontiers of econometrics*. New York: Academic Press.

McFadden, Daniel. 1986. The choice theory approach to market research. *Marketing Science* 5 (4): 275–297.

Mehta, N. 2015. A flexible yet globally regular multi-good demand system. *Marketing Science* forthcoming.

Pollak, R.A., and T.J. Wales. 1992. *Demand System Specification and Estimation*. Oxford, UK.

Rossi, P.E., G.M. Allenby, and R. McCulloch. 2005. *Bayesian Statistics and Marketing*. Wiley.

Satomura, Takuya, Jaehwan Kim, and Greg M. Allenby. 2011. Multiple constraint choice models with corner and interior solutions. *Marketing Science* 30 (3): 481–490.

Song, I., and P.K. Chintagunta. 2007. A discrete-continuous model for multicategory purchase behavior of households. *J. Marketing Res*. 44: 595–612.

Sydsæter, Knut, Arne Strøm, Peter Berck, Arne Strøm, and Peter Berck. 2005. *Economists' mathematical manual*, vol. 3. Springer.

Tanner, Martin A., and Wing Hung Wong. 1987. The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association* 82 (398): 528–540.

Wales, Terence J., and Alan Donald Woodland. 1983. Estimation of consumer demand systems with binding non-negativity constraints. *Journal of Econometrics* 21 (3): 263–285.

Chapter 8
Empirical Models of Learning Dynamics: A Survey of Recent Developments

Andrew T. Ching, Tülin Erdem and Michael P. Keane

8.1 Introduction

There is now a very large literature on dynamic models in marketing. In a narrow sense, dynamics can be understood as a mechanism whereby past product purchases affect a person's current evaluation of the utility he/she will obtain from buying a product. Most of the prior literature has focussed on three mechanisms that may generate such a causal link from past to current purchases: learning, habit persistence, and inventory dynamics.[1] This work has been reviewed extensively in papers by Ching et al. (2013) and Keane (2015).

However, dynamics can be more broadly defined as encompassing any process whereby the prior history of a consumer or market affects current utility evaluations. For example, it is clear that, aside from past purchase, a consumer's perception of a product may be influenced by experiences of friends or other social network members ("social learning"), experience with related products ("correlated learning" or "information spillovers"), examination of publicly available information or expert opinion ("search"), inferences about product attributes that may be drawn from the

[1]Given habit persistence or learning, past purchase creates exposure to a product, which directly affects a consumer's perceived utility of a product. In an inventory model, past purchase matters because it affects current inventory, but also, more subtly, because the prices at which past purchases are made affect the reference price of the product.

A.T. Ching (✉)
University of Toronto, Toronto, Canada
e-mail: Andrew.Ching@Rotman.Utoronto.Ca

T. Erdem
New York University, New York, USA

M.P. Keane
University of Oxford, Oxford, UK

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_8

purchase decisions of others, etc. In the present chapter we focus on the rapidly growing literature that deals with this broader view of dynamics and learning.

We begin in Sect. 8.2 with a description of the basic structural learning model developed by Erdem and Keane (1996). In this model, consumer trial and advertising are the only sources of information about a product. It is important to understand this model before we proceed, as most of the subsequent work in this area can be understood as extending Erdem-Keane to allow for additional information sources. Then, in Sects. 8.3 through 8.6, we focus on (i) learning from others (social learning), (ii) learning and strategic interactions, (iii) information spillovers and correlated learning, and (iv) learning and search. In Sect. 8.7 we consider the use of heuristics to capture aspects of learning. In Sect. 8.8 we consider recent work that uses exogenous events and policy changes to study learning behavior. Section 8.9 discusses directions for future research, in particular attempts to relax some of the strong assumptions of the Bayesian learning model. Section 8.10 concludes.

8.2 The Basic Bayesian Learning Model

In this section, in order to provide a background for the subsequent discussion, we describe a basic consumer learning model, similar to Erdem and Keane (1996). Of course, there were important antecedents to their paper. In particular, the seminal papers by Roberts and Urban (1988) and Eckstein et al. (1988) also developed structural learning models. Roberts-Urban modelled learning by risk-averse but myopic consumers, while Eckstein et al. modelled learning by forward-looking but risk-neutral consumers.[2] The key innovation of Erdem and Keane (1996) was to develop a framework that could accommodate both risk aversion with respect to product attributes and forward-looking behavior.

The key feature of the Erdem-Keane model is that consumers do not know the attributes of brands with certainty. Before receiving any information via use experience, consumers have a normal prior on brand quality:

$$Q_j \sim N\left(Q_{j1}, \sigma_{j1}^2\right), \quad j = 1, \ldots, J. \tag{8.1}$$

This says that, prior to any use experience, consumers perceive that the true quality of brand j (Q_j) is distributed normally with a mean of Q_{j1} and variance σ_{j1}^2. The values of Q_{j1} and σ_{j1}^2 may be influenced by many factors, such as reputation of the manufacturer, advice from friends, etc.

[2]In order to implement their model, Erdem and Keane (1996) used the approximate solution methods for dynamic programming models developed in Keane and Wolpin (1994). See Ching et al. (2013) for a complete explanation of the procedure. The simplifying assumptions in Eckstein et al. (1988) allowed them to use the Gittin's index to find the solution of their model (see Appendix A of Ching et al. 2013).

Use experience does not fully reveal quality because of "inherent product variability." This has two main sources: First, quality of different units of a product may vary. Second, and more importantly, a consumer's experience of a product will vary across use occasions.[3]

Given inherent product variability, there is a difference between "experienced quality" by consumer i for brand j on purchase occasion t, which we denote Q^E_{ijt}, and true quality Q_j. Assume the experienced quality delivered by use experience is a noisy signal of true quality, as in:

$$Q^E_{ijt} = Q_j + \varepsilon_{ijt} \text{ where } \varepsilon_{ijt} \sim N(0, \sigma^2_\varepsilon). \tag{8.2}$$

Here σ^2_ε is the variance of inherent product variability, which we call "experience variability." It should be noted that all brands have experience variability, so (Eq. 8.2) holds for all j.

Note that we have conjugate priors and signals, as both the prior on quality (Eq. 8.1) and the noise in the quality signals (Eq. 8.2) are assumed to be normal. The posterior for perceived quality at $t = 2$, after a single use experience signal is received at $t = 1$, is given by the updating formulas:

$$Q_{ij2} = \frac{\sigma^2_{j1}}{\sigma^2_{j1} + \sigma^2_\varepsilon} Q^E_{ij1} + \frac{\sigma^2_\varepsilon}{\sigma^2_{j1} + \sigma^2_\varepsilon} Q_{j1}, \tag{8.3}$$

$$\sigma^2_{ij2} = \frac{1}{\left(1/\sigma^2_{j1}\right) + (1/\sigma^2_\varepsilon)}. \tag{8.4}$$

Equation (8.3) describes how a consumer's prior on quality of brand j is updated as a result of the experience signal Q^E_{ij1}. Note that the extent of updating is greater the more accurate is the signal (i.e., the smaller is σ^2_ε). Equation (8.4) describes how a consumer's uncertainty declines as he/she receives more signals. The variable σ^2_{ijt} is often referred to as the "perception error variance."

Equations (8.3) and (8.4) generalize to any number of signals received. Let $N_{ij}(t)$ denote the total number of use experience signals received by person i before he/she makes a purchase decision at time t. Then we have that:

$$Q_{ijt} = \frac{\sigma^2_{j1}}{N_{ij}(t)\sigma^2_{j1} + \sigma^2_\varepsilon} \sum_{s=1}^{t} Q^E_{ijs} d_{ijs} + \frac{\sigma^2_\varepsilon}{N_{ij}(t)\sigma^2_{j1} + \sigma^2_\varepsilon} Q_{j1}, \tag{8.5}$$

$$\sigma^2_{ijt} = \frac{1}{\left(1/\sigma^2_{j1}\right) + N_{ij}(t)(1/\sigma^2_\varepsilon)}, \tag{8.6}$$

[3]For instance, a diaper may hold all of a baby's urine on some occasions but not on others (depending on how much milk the baby drank), so one use may not fully reveal its quality.

where d_{ijt} is an indicator for whether brand j is bought/consumed at time t by person i.

In Eq. (8.5), the perceived quality of brand j at time t, Q_{ijt}, is a weighted average of the prior and all quality signals received up through time t, $\sum_{s=1}^{t} Q_{ijs}^{E} d_{ijs}$. Perceived quality is random across consumers, as some receive, by chance, better quality signals than others. So the learning model endogenously generates heterogeneity across consumers in their perceptions of products.

Let $I_{it} = \{Q_{it}, \boldsymbol{\sigma}_{it}^2\}$ denote consumer i's information set (i.e., all the past signals he/she has received), where $Q_{it} = \{Q_{i1t}, \ldots, Q_{iJt}\}$ and $\boldsymbol{\sigma}_{it}^2 = \{\sigma_{i1t}^2, \ldots, \sigma_{iJt}^2\}$.[4] As Eq (8.6) indicates, the variance of perceived quality around true quality declines as more signals are received, and in the limit perceived quality converges to true quality.[5]

The Erdem and Keane (1996) model generalizes the model in Eqs. (8.1)–(8.6) by including advertising as a second signal of quality. It is fairly easy to modify Eqs. (8.1)–(8.6) to accommodate two (or more) signals. We do not need to go into such complications here, as our goal is just to explain the basics of the framework. But it is important to note that many extensions of the Erdem-Keane model that we discuss below rely on the fact that it is fairly straightforward to extend the Bayesian learning framework to accommodate multiple sources of information.[6] We give some concrete examples in subsequent sections.

Finally, to complete the model we need to assume a particular functional form for utility. For instance, we could assume consumer i's (conditional indirect) utility of consuming brand j is:

$$U_i\left(Q_{jt}^{E}, P_{jt}\right) = f\left(Q_{ijt}^{E}\right) - w_P P_{ijt} + e_{ijt}, \tag{8.7}$$

where P_{ijt} is price, w_P is marginal utility of income and e_{ijt} is an idiosyncratic brand, time and person specific error, distributed *iid* extreme value.[7] If we then assume that $f\left(Q_{ijt}^{E}\right)$ takes the constant absolute risk aversion form (CARA), then the expected utility is given by:

$$E\left[U\left(Q_{ijt}^{E}, P_{ijt}\right)|I_{it}\right] = -\exp\left(-r\left(Q_{ijt} - \frac{r}{2}\left(\sigma_{ijt}^2 + \sigma_{\varepsilon}^2\right)\right)\right) - w_P P_{ijt} + e_{ijt}, \tag{8.8}$$

[4]Note that this is equivalent to say I_{it} consists of all past signals consumer i has received before he/she makes the purchase at time t, given the Bayesian learning framework.

[5]Still, heterogeneity in S_{it} may persist over time, because: (i) both brands and consumers are finitely lived, (ii) as people gather more information the value of trial purchases diminishes, and so eventually learning about unfamiliar products will become slow; (iii) there is a flow of new brands and new consumers entering a market.

[6]Ching et al. (2013) explain how to extend the above basic framework to allow consumers to learn from multiple information sources (such as advertising, word-of-mouth and so on).

[7]In almost all cases no purchase is also an option. Erdem and Keane (1996) denote the no purchase option as $j = 0$, and simply set the expected utility of no purchase to $E[U_{i0}|S_{it}] = e_{i0t}$.

where $r > 0$ captures risk aversion with respect to variation in product quality.

One can see from Eq. (8.8) that a higher perception error variance σ^2_{ijt} reduces the expected utility of purchase of brand j, *ceteris paribus*. Thus, purchase of an unfamiliar brand is risky, which lowers its expected utility for a risk averse consumer. But on the other hand, trial of an unfamiliar brand has the benefit that it generates new information that lowers σ^2_{ijt}.

Finally we note that Bayesian learning models can be classified as "forward-looking" or "myopic" depending on whether consumers take the benefit of future information into account when they make current purchase decisions. Erdem and Keane (1996) found that accounting for forward-looking behavior led to only modest improvement in fit in the detergent category, but Ching et al. (2014a) find a more important role for forward-looking behavior in the diaper category. The literature we review below contains many models of both types.

8.3 Learning from Others (Social Learning)

Learning from others encompasses a wide range of activities. Some examples are "word-of-mouth" learning, the source of which is typically friends and relatives, consulting expert opinions (as in movie reviews, magazine articles, and news coverage), or reading online reviews by other consumers. Erdem et al. (2008) allow for four sources of information when studying scanner panel data for consumer package goods, and Erdem et al. (2005) allow for five sources (including word-of-mouth, salespeople, and articles in computer magazines and other magazines) when studying the PC purchase decisions.

Alternatively, one may consider an environment where there exists a special agent (or information source) who serves as an information aggregator who pools consumers' experiences together. For instance, in the context of choosing between brand-name drugs or generic counterparts, Ching (2010a, b) assumes that a random sample of individual's experiences can be observed by everyone via physician networks or consumer watch groups. The Ching (2010a, b) model considers a representative physician who learns about the average quality of a generic drug (relative to the quality of its brand-name originator).

Specifically, let S_t be the random sample of individual experience signals that are revealed to the representative physician at time t. Let q_t be the quantity of the generic drug sold at time t and let κ be the proportion of individual experience signals revealed in each period. Then $card(S_t) = \kappa q_t$.[8] Assume the experience

[8] *card*(.) is the cardinality of the set in question. It measures the number of elements in the set.

signals are distributed *iid* across patients with mean Q_j and variance σ_ε^2. Finally, let $\bar{Q}_{ijt}^E$ denote the mean of the κq_t individual experience signals observed by the representative physician regarding drug j at time t. The representative physician's updating process can be described by extending Eqs. (8.3) and (8.4) as follows:

$$Q_{jt+1} = \frac{\sigma_{jt}^2}{\sigma_{jt}^2 + \sigma_\varepsilon^2/(\kappa q_t)} \bar{Q}_{ijt}^E + \frac{\sigma_\varepsilon^2/(\kappa q_t)}{\sigma_{jt}^2 + \sigma_\varepsilon^2/(\kappa q_t)} Q_{jt}, \tag{8.9}$$

$$\sigma_{jt+1}^2 = \frac{1}{\left(1/\sigma_{jt}^2\right) + \left((\kappa q_t)/\sigma_\varepsilon^2\right)}, \tag{8.10}$$

where $\bar{Q}_{ijt}^E | \kappa q_t \sim N\left(Q_j, \frac{\sigma_\varepsilon^2}{\kappa q_t}\right)$ by the central limit theorem.

Another very interesting problem that has received attention lately is learning about product quality in environments where consumers only observe the choices of others (without specifically observing their experience signals). This is known as "observational learning." To our knowledge, the first structural empirical model of observational learning is Zhang (2010), who extends observational learning to a dynamic setting in order to explain consumers' decisions to accept a donated organ for transplant—specifically a kidney.

Zhang (2010) considers an environment where patients wait in line to receive a kidney for transplant. However, it is not uncommon for a patient to choose not to accept a kidney and wait for a better match. Hence, in the model, when a patient receives a kidney offer, he/she needs to choose whether to accept it or decline it and continue to wait.

Prior to making this accept vs. decline-and-wait decision, the first patient in the line "examines" his match with the kidney and obtains a noisy signal distributed around the true quality of the kidney. When making his/her decision, all the first patient can rely on is his/her signal and initial prior belief. Hence, the posterior expected value of the kidney is simply given by Eq. (8.3). If the expected utility of receiving the kidney is higher than the expected future value of waiting, the first patient will accept. Otherwise, he/she will decline-and-wait.

However, the decision facing the second patient is more complex. When he/she decides whether to accept the kidney or not, the second patient must take into account not only his/her own signal, but also the fact that the first patient declined. As the first patient's decision was a function of the signal that he/she observed, the first patient's choice reveals that the first signal must have been below a certain cutoff (or reservation) value. When the second patient updates his/her belief, he/she should take this fact into consideration, rather than relying purely on his/her own personal signal. Similarly, when it is the third patient's turn, he/she takes into account that the first two patients have declined the kidney.

A simplified version of the Zhang (2010) model can be described as follows. Let Q_i^P be the private signal received by i-th consumer, where i indexes the position of the consumer on the waiting list. Then,

$$Q_i^P = Q + \varepsilon_i \text{ where } \varepsilon_i \sim N(0, \sigma_\varepsilon^2). \tag{8.11}$$

For i = 1 (the first patient in line), the decision problem is the same as in the Erdem-Keane set up. Specifically, patient 1 uses his/her own noisy signal Q_1^P to update his/her belief about Q. If the expected utility of accepting the kidney is higher than that of declining (and waiting for the next offer), then patient 1 accepts. It is easy to show that $E[Q|Q_1^P]$ is monotonically increasing in Q_1^P. Hence, the first patient's decision rule can be characterized by a cutoff rule – i.e., there exists a B_1 such that if $Q_1^P \geq B_1$, patient 1 accepts (i.e., $d_1 = 1$); otherwise he/she declines (i.e., $d_1 = 0$).

But for the second patient in line (i = 2) the situation is more complex. Specifically, for the second patient the information set is given by $I_2 = \{d_1 = 0; Q_2^P\} = \{Q_1^P < B_1; Q_2^P\}$. The key is to find the conditional distribution, $p(Q|I_2) = p(Q|Q_1^P < B_1; Q_2^P)$. Zhang (2010) assumes that each patient draws an independent signal. Hence, it follows from the Bayes' rule that,

$$p(Q|Q_1^P < B_1; Q_2^P) \propto p(Q_1^P < B_1; Q_2^P|Q) \cdot p(Q), \tag{8.12}$$

where $p(Q)$ is the initial prior belief about Q. Moreover,

$$p(Q_1^P < B_1; Q_2^P|Q) = \Phi\left(\frac{B_1 - Q}{\sigma_\varepsilon}\right) \cdot \phi\left(\frac{Q_2^P - Q}{\sigma_\varepsilon}\right), \tag{8.13}$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are the cumulative distribution function and probability density function of standard normal, respectively. One can then use Eqs. (8.12) and (8.13) to obtain

$$E(Q|Q_1^P < B_1; Q_2^P) = \frac{\int p(Q|Q_1^P < B_1; Q_2^P) \cdot Q \cdot dQ}{\int p(Q|Q_1^P < B_1; Q_2^P) \cdot dQ}, \tag{8.14a}$$

where the denominator is a normalizing factor to ensure that the posterior density of Q is proper. Note that it is straightforward to extend the logic above to the decision problem of the i-th consumer, for i > 2.[9] This leads to higher order conditioning, of the form:

[9]Hendrick et al. (2012) propose a similar framework to study how consumers choose a product among J > 2 alternatives. Newberry (2016) extends this framework to study the role of pricing in observational learning using data from an online market for music.

$$E(Q|Q_1^P < B_1, \ldots, Q_{i-1}^P < B_{i-1}; Q_i^P) = \frac{\int p(Q|Q_1^P < B_1, \ldots, Q_{i-1}^P; Q_i^P) \cdot Q \cdot dQ}{\int p(Q|Q_1^P < B_1, \ldots, Q_{i-1}^P; Q_i^P) \cdot dQ}. \tag{8.14b}$$

Then, solving for this expectation typically requires using Monte Carlo simulation methods, such as the recursive conditioning simulator developed in Keane (1994).

Many other problems are similar in structure to the observational learning problem discussed here. Examples are decisions of whether or not to accept a job offer when the decisions of prior individuals who were offered the same position can be observed. This tends to be the case for high-profile positions such as deanships, coaching positions, key executive positions and so on. Alternatively, one might consider the decision of whether or not to extend a job offer to an applicant, given knowledge of the set of offers and rejections that he/she has so far received (as is often the case in organized job markets like that for assistant professors).

Returning to the medical example, Ching and Ishihara (2010) consider an alternative situation where patients can only obtain qualitative information about a signal. Often times, a product review may simply reveal whether product A is better product B, without revealing the exact realization of the quality signal. In this case, consumers can only infer that the signal lies within a certain range.

Of course, social learning has become an area of great applied interest in recent years. Thus, many other new papers on social learning are notable for their substantive (as opposed to methodological) contributions.[10] An important substantive topic that has recently received attention is how consumers may learn from each other through online reviews. Zhao et al. (2013) estimate a Bayesian learning model with myopic consumers which allows for consumers to learn about product quality both through their own experiences with the same type of product (e.g., a book genre), as well as through product reviews posted by other consumers. Furthermore, in addition to learning from others about the same product, they also allow for what is known as "correlated learning." That is, other consumers' experiences with books of the same genre can allow one to update his/her belief about other books in that genre. The Zhao et al. (2013) model also incorporates learning about the credibility of product reviews posted by others (captured as the precision of the information provided). The model is estimated on book purchases of a panel of consumers. The results indicate that consumers learn more from online reviews of book titles than their own experiences with other books of the same genre.

Similarly, Wu et al. (2015) study the economic value of online reviews to consumers, as well as to restaurants, using a dataset from Dianping.com, a leading Chinese website providing user-generated reviews. The proposed Bayesian learning model with myopic consumers allows for different reviews to be of different

[10]The models in these papers largely adopt the framework discussed so far, and hence we will not devote space to explicitly discussing their structure.

informational value to different consumers. It also allows consumers to learn about their own preferences for multiple product attributes, as well as learn the mean and variance of consumption experiences in the population. The findings indicate that the majority of the created value comes from reviews on the quality of the restaurants, and that contextual comments are more valuable than numerical ratings in reviews.

Using field experiments, Godlonton and Thornton (2013) study the impact of others' testing on individual perceptions of AIDS risk and subsequent decisions to practice safe sex in rural Malawi. In this context, it appears that individuals tend to overestimate the underlying prevalence of HIV incidence. Godlonton-Thornton measure the response to others' HIV testing, which alters individuals' beliefs about the underlying prevalence. They measure the causal effect of others' testing by utilizing an experiment that randomly offered incentives to individuals to learn about their HIV test results at randomly located results centers. They use the village-level average of these incentives and distance from results centers to instrument for the proportion of community testing. They find robust evidence of downward revision of beliefs about HIV prevalence, and subsequent changes in sexual behavior (e.g., reduced condom use).

Knight and Schiff (2010) develop and estimate a social learning model to study voters' decisions in US presidential primary elections—a system under which States vote sequentially, so voters in later State elections can learn about candidates from earlier State results. The advantage of a sequential system is that it provides late voters with valuable information, but its drawback is that it exaggerates the influence of early States.

Hummel and Knight (2015) use this model to conduct a counterfactual experiment to compare outcomes under simultaneous vs. sequential elections. The advantage of a simultaneous election is that it weighs States equally. However, it also places great weight on voter priors, creating a large advantage for front-runners. Thus, simultaneous (sequential) elections are preferred if the front-runner advantage is small (large). The quantitative welfare analysis of presidential primaries reported in Hummel and Knight (2015) suggests that simultaneous systems would slightly outperform sequential systems.

Lee and Bell (2013) model social learning through neighbors' past purchases. They estimate a Bayesian learning model with myopic consumers on combined data from consumer purchases on Bonobos, a leading online fashion retailer in the US, and the Social Capital Community Benchmark Survey (SCCBS). SCCBS has data relating to two dimensions of social capital, that is, trust among local neighbors and the frequency of interaction. Utilizing these data, they estimate a model where consumers update their beliefs about experience attributes through a social learning mechanism where local neighbors' purchases serve as information signals. The results indicate that social capital improves the learning process and therefore indirectly drives sales when the information that is communicated is favorable.

Huang et al. (2015) model social learning in the context of crowdsourcing new product ideas. They propose a Bayesian learning model that accounts for consumers' learning about the potential of their ideas, as well as the cost structure of the firm. The model is estimated on data from IdeaStorm.com (a crowdsourced ideation initiative affiliated with Dell). The findings suggest that individuals over-estimate the potential of their own ideas and underestimate the firm's costs. They learn about both, but they learn a lot faster about the potential of their own ideas than about the firm's cost structure.

In a reduced form study, Ching et al. (2016) find evidence that in the prescription drug market, where physicians can learn from different sources, the interaction between detailing (i.e., pharmaceutical firms send sales representatives to visit doctors) and publicity (news coverage) could have non-trivial outcomes, depending on the complexity of the information. In the setting of anti-cholesterol drugs, some information is simple to describe (e.g., side-effects or the extent to which a drug can reduce cholesterol), and some can be much more complicated (e.g., a drug's ability to reduce heart disease risks). They argue that news coverage is subject to tight space constraints and hence, even though the source is credible, it cannot report all the relevant information related a clinical trial that documents reducing heart disease risks.

But physicians may treat news coverage as corroborative evidence that supports what sales representatives claim. In particular, when physicians see that sales representatives' claims are consistent with news coverage, they may give more time to the sales rep and let them explain the details of new clinical trials. This may cause detailing and publicity about complex information to be complements. On the other hand, simple information is much easier for physicians to verify, and hence credibility is less of an issue. As a result, different sources of simple information are substitutes. This research suggests the importance of distinguishing the complexity of information, and modeling the idea of corroborative evidence in the context of learning—areas that the structural learning literature has not tackled yet.

Several recent papers also focus on learning about entertainment products. Lovett and Staelin (2016) model social learning in the context of TV shows. They focus on how paid (e.g., advertising) and owned media (e.g., a TV network's own website) differ from social media in influencing consumers' utility of watching a TV show. Lovett and Staelin (2016) decompose these influences into three channels: learning, reminding and social engagement. The unique feature of their data is that it contains consumer's stated expectations about TV shows. They use these to calibrate the informative effect of these media, and allow the rest of their impacts to be picked up by their reminding and social engagement functions.

Liu and Ishihara (2015) study consumer learning about new video games, using product level sales data and critic and user reviews from the US video game market. Their model allows heterogeneity in consumer tastes (leading to horizontal differentiation of games), and controls for the spurious correlation that is likely to exist between review ratings and demand. Specifically, they use the market shares of pre-order data (period 0) to measure consumers' initial priors (about games). After critic reviews become available in period 1, consumers' update their priors about games. Then, the market share in period 1 shows the impact of critic ratings.

The Liu-Ishihara model departs from standard learning models by also incorporating the psychological theory of reference points. Consumers use critic and user reviews over time to update their prior and form a reference quality for games. Consumer reviews then depend on how experienced quality compares to the reference point. For instance, if experience is worse than expected it may lead to a very poor review, while if experience is good but similar to what was expected it may lead to only a mildly positive review.

Liu and Ishihara (2015) find that a 20% discount for pre-orders will increase immediate profits, but reduce future profits via lower consumer review ratings. The latter happens because the pre-order discount attracts consumers with lower product valuations, and these consumers' reviews pull down the average consumer review rating (compared to the situation where only consumers with high product valuation purchase and write consumer reviews). Ignoring such consumer review endogeneity when conducting profit simulations might cause one to overstate the effectiveness of the pre-order discount in improving profitability.

Wei (2015) models how movie studios decide which potential projects to invest in when facing uncertainty about the potential of novel types of movies. He hypothesizes that in addition to observable characteristics (e.g., budget, actors, director, genre), studios can rely on the past performance of a new line of movies to update their beliefs about its latent quality distribution. This in turn will help the studios to compute the expected profits of adding such a project to their existing production portfolio.

In a different type of application, Liu et al. (2012) model learning from others using panel data obtained from a series of laboratory experiments. Specifically, they study agent's strategic behavior in an entry limit pricing game where firms use price to signal costs (Milgrom and Roberts 1982). In experimental economics, researchers often provide subjects with choices made by others as feedback, in order to speed up the convergence of one's strategy. Thus, Liu et al. (2012) argue that peer group effects should be very pronounced in experiments run in this style (which is typical). To capture learning from peers, they extend dynamic discrete choice panel data models (Heckman, 1981) by introducing a time-lagged social interactions variable.[11] Their results indicate that learning from peers is important in this experiment.

Like Liu et al. (2012) and Chan et al. (2014) also model learning from peers, but they use field data. They make use of a unique data set which consists of all sales persons' performance (i.e., their actual weekly sales) from all cosmetic product counters in a department store for a period of four years. The authors use this data set to measure the impacts of learning from others within the same counter, and from adjacent counters. They are able to identify these effects because workers are assigned to different shifts (there are three shifts) "randomly" based on fairness instead of their productivity. This provides an excellent opportunity to measure the

[11]The likelihood of their model involves multiple integrals because the explanatory variables include lagged latent dependent variables and serially correlated errors, but they show that the GHK simulator remains tractable for this generalized framework (see Keane 1994).

effects of learning from others based on different sources, and disentangle them from self-learning. They argue that learning from peers can be fundamental to knowledge spillovers and explaining organizational learning curves.

8.4 Learning and Strategic Interaction

Ching (2010b) is the first paper that incorporates consumer and firm learning in a dynamic oligopoly structural model. In the model, firms are forward-looking and set price in each period. The demand side of the model is taken from Ching (2010a). The dynamic oligopoly model is developed to study the competition between a brand-name drug and its generic counterparts. The key innovation here is that firms are also uncertain about the true quality of generic drugs, and they can use price to control the rate of learning. In particular, assuming consumers are risk-averse, generic firms may have an incentive to price low to encourage more consumers to try their products and resolve the uncertainty (measured by the variance of their posterior belief) sooner.

To explain the main features of the Ching (2010b) model, we make the simplifying assumption that there is one brand-name firm and one generic firm.[12] Let p_{bt} and p_{gt} be the brand-name price and generic price at time t, respectively. The per period profit for firm $j \in \{b, g\}$ is $\pi_j = (p_j - mc) \cdot q_j(p_b, p_g)$, where $q_j(p_b, p_g)$ is determined by the discrete choice model described in Eqs. (8.1)–(8.8).[13] The model assumes that in each period the brand-name firm acts as the Stackelberg leader, and the generic firm is the follower. Let $S_t = (Q_{gt}, \sigma^2_{gt})$ be the state variables at time t, which evolve according to Eqs. (8.3) and (8.4), respectively. The generic firm's dynamic problem can be characterized using dynamic programming as follows:

$$V_g(S_t) = \max_{p_{gt}} [\pi_g(S_t, p_{bt}, p_{gt}) + \beta E[V_g(S_{t+1}) | S_t, q_{gt}(p_{bt}, p_{gt})], \text{ for } t < T; \tag{8.15}$$

$$V_g(S_T) = \max_{p_{gT}} \pi_g(S_T, p_{bT}, p_{gT}). \tag{8.16}$$

Similarly, the brand-name firm's dynamic problem can be characterized as follows.

$$V_b(S_t) = \max_{p_{bt}} [\pi_b\left(S_t, p_{bt}, p^*_{gt}(p_{bt})\right) + \beta E[V_b(S_{t+1}) | S_t, q_{gt}(p_{bt}, p_{gt})], \text{ for } t < T; \tag{8.17}$$

[12] Ching (2010b) allows for multiple generic firms. In addition, generic firms' entry decisions are also endogenous. But since the focus of this chapter is learning, we abstract away the entry decisions when describing the model.

[13] It is the choice probability of choosing product j multiplied by the total number of potential patients in this market.

$$V_b(S_T) = \max_{p_{bT}} \pi_b\left(S_T, p_{bT}, p^*_{gT}(p_{bt})\right). \tag{8.18}$$

Note that the main difference between the problem faced by the brand-name and generic firms is that the brand-name firm takes into account that its price will influence the generic price; on the contrary, the generic firm simply takes the brand-name price as given. The equilibrium concept is Markov Perfect Nash equilibrium (i.e., the pricing function only depends on payoff relevant state variables contained in S_t). Because this model has a final period, T, it can be solved using backward induction. Although this model is conceptually tractable, it is computationally very challenging to solve. This is because the state space is continuous, and unlike other discrete choice problems, the firms are choosing a continuous variable to maximize their total discounted profits.

Zou (2014) develops and estimates an equilibrium model of intertemporal pricing of new products. The model extends Ching (2010b) by allowing consumers to be heterogeneous in their information sets. The learning process in the model naturally generates such an outcome, similar to Erdem and Keane (1996). In contrast, Ching (2010b) assumes there is a representative consumer and hence there is only one information set. Similar to Ching (2010b), this model allows for forward-looking firms but myopic consumers. In addition, Zou (2014) also allows for additional state dependence beyond what consumer learning implies. He applies his model to data from the Yogurt category, for a time period covering the entry of Chobani. Using counterfactual analysis, Zou finds evidence that the firm's introductory pricing strategy is mainly driven by positive state dependence rather than learning.

Huang et al. (2015) use a structural learning model to study how dealers set prices for used cars over time. They argue that used cars are hard to price because, unlike new cars, they differ in multiple dimensions (depending on mileage, year, and maintenance), and it is not clear how consumers trade-off these dimensions a priori. Facing this uncertainty about the unobserved demand factor, in every period a dealer sets the price for a used car and then consumers decide whether to buy it. If consumers choose not to buy the used car, this gives the dealer a signal about its unobserved demand component, and he then updates his belief accordingly. If a dealer is forward-looking, he has an incentive to price the used car high early on, because not selling the car gives him an opportunity to obtain better information about the demand for the car.

Using a panel dataset of used-car sales from CarMax, Huang et al. (2015) find that their structural model can explain demand and pricing patterns well. As in Ching (2010b), learning is the main source of dynamics that determines how a firm sets their prices dynamically over time in this paper. But Huang et al. (2015) focus on a dynamic monopoly problem, while Ching (2010b) studies a dynamic oligopoly problem. Moreover, Ching (2010b) has a symmetric two-sided learning environment (both firms and consumers are equally uncertain about the quality of generic drugs), while Huang et al. (2015) have an asymmetric one-sided learning environment (consumers know their demand, and only firms are uncertain about demand conditions).

A recent paper by Chen et al. (2009) investigates the interaction between learning and addiction in the tobacco market by estimating a forward-looking structural learning model. They model addiction using the brand loyalty variable of Guadagni and Little (1983) (GL), i.e., exponential smoothing of the past choices. They use their model to study the effects of Marlboro's permanent price cut that happened on April 2, 1993 (Marlboro Friday) as a reaction to the continuous loss of market share to generic brands.[14] They find that by permanently lowering the price, consumers who previously bought only generics became willing to experiment. Their estimation results suggest that there is positive interaction between expected quality and the GL variable, and that Marlboro has a higher quality than generic brands. This implies that the brand loyalty effect due to GL is stronger for Marlboro. As a result, when the permanent price cut induces consumers to try Marlboro, most of these new consumers stay with it because of Marlboro's stronger GL effect.

It is interesting to compare the Marlboro permanent price cut strategy with how brand-name drugs increase prices in response to generic competition (Ching 2010a, b). At first, these two situations seem similar (both of them face generic competition). But a closer examination reveals some key differences of the environments considered in Chen et al. (2009) and Ching (2010b). The paper by Ching (2010b) assumes that by the time the patent expires, most consumers know the quality of the branded product. But they need to learn about the quality of generics. The story is that consumers who are price-sensitive would slowly switch to generics as they learn and become increasingly more confident that they are safe over time. But this implies that the demand faced by the brand-name firm becomes more price inelastic over time. Of course, this effect cannot last forever, and the theory suggests that at some point, the price for the brand-name drug should come down. But, as long as there is a mass of consumers loyal to the brand-name drug who do not update their belief at all, that may be sufficient to keep the brand-name drug price high permanently.

On the contrary, in the tobacco case, Chen et al. (2009) hypothesize that some consumers have not tried Marlboro (or other premium brands) before. So a permanent price cut allows Marlboro to regain some market share by gaining new customers. Quality here refers to taste (i.e., Marlboro could taste better than other generic brands). So in the tobacco case, some consumers may discover that they actually like Marlboro more in a complete information situation.

Structural learning has also been introduced in other problems of strategic interactions. Yang (2016) introduces learning from others in an incomplete information discrete dynamic game with entry and exit (similar to Aguirregabiria and Mira 2007). The model captures the idea that firms have uncertainty about the

[14]Although this paper does not explicitly model a dynamic game, the dynamic demand model is very useful in evaluating the consequences of Marlboro's strategic response to the competition of generic brands.

market potential of their products. There are two ways a firm can resolve this uncertainty: (i) learn directly by entering the market; or (ii) learn from other firms which are operating in the market. The learning mechanism is similar to Ching (2010a). Yang (2016) finds that learning from others can partially offset the negative business-stealing effects of rivals (because learning from others helps a firm find out sooner if it should exit the market).

Ho, Park and Su (2015) expand on standard models of iterative thinking by introducing a Bayesian level-k model,[15] in which players perform Bayesian updating of their beliefs about opponents' rule levels, and best-respond with different rule levels over time. The authors apply this sophisticated learning model to experimental data on *p*-beauty contest and price matching games and find evidence for this type of sophisticated learning.

8.5 Information Spillovers and Correlated Learning

As we noted in the introduction, information spillovers and correlated learning refer to situations where one can learn about a given product via experience with related products. The standard learning model can be easily extended to study such environment. We first specify a more general prior belief to capture the idea that consumers may believe the qualities of products are correlated. Using vector notation, one can modify Eq. (8.1) as:

$$Q \sim N(Q_{t=1}, \Sigma_{t=1}), \tag{8.19}$$

where $Q_{t=1}$ is the $J \times 1$ initial prior mean vector and $\Sigma_{t=1}$ is the $J \times J$ initial prior variance-covariance matrix. Consider a two-product case. With off-diagonal elements greater than zero, an information signal for product 1 will be used to update one's belief about product 2, and vice versa. The updating formula can be generalized as follows.

$$\Sigma_t = \begin{bmatrix} \sigma^2_{1,t} & \pi_t \\ \pi_t & \sigma^2_{2,t} \end{bmatrix}. \tag{8.20}$$

When receiving an information signal for product 1 at time t, the updating for $Q_{1,t+1}$ and $\sigma^2_{1,t+1}$ will be the same as the standard learning model explained in

[15]The basic idea of this model is that players in a game vary in their depth of strategic thinking. A completely naïve player will choose actions by completely ignoring the presence of other players (level zero). A level one player believes that other players will not react to his choice, and his action is the best response with respect to this belief. A level two player believes that all other players are level one, and so on and so forth. This model captures bounded rationality, and can explain players' behavior in games that cannot be rationalized by standard game theory.

Eqs. (8.3) and (8.4). But the information signal for product 1 will also be used to update the consumer's belief about product 2 as follows:

$$Q_{2,t+1} = Q_{2t} + \frac{\pi_t}{\sigma_{2t}^2 + \sigma_\varepsilon^2}\left(Q_{1t}^E - Q_{1t}\right). \tag{8.21}$$

The variance-covariance matrix for posterior beliefs becomes:

$$\sigma_{2,t+1}^2 = \sigma_{2,t}^2 - \frac{\pi_t^2}{\sigma_{2t}^2 + \sigma_\varepsilon^2}, \tag{8.22}$$

$$\pi_{t+1} = \frac{\pi_t \sigma_\varepsilon^2}{\sigma_{2t}^2 + \sigma_\varepsilon^2}. \tag{8.23}$$

Erdem (1998) was the first paper to apply this framework to show that, in the case of umbrella brands, consumers learn about brand quality in one category through their experiences with the same brand in another category. Recent years have witnessed a marked increase in papers that focus on this topic. For instance, expanding on Erdem (1998)'s idea about cross-brand learning, Szymanowski and Gijsbrechts (2012) study whether experiences with a private label affects consumer quality perceptions about other private labels. They estimate a Bayesian learning model with myopic consumers on household scanner panel data on dish soap and breakfast cereals. Their results indicate there is cross-retailer learning among standard private labels regardless of their name and quality differences.

Ching and Lim (2016) significantly extend Erdem's framework to an environment where firms are selling similar but differentiated competing products. To explain why sometimes late entrants can easily surpass incumbents, they propose a new consumer theory of correlated learning and indirect inference. They apply their model to study the market for anti-cholesterol drugs, where the late entrant, Lipitor, overtook long-time incumbents within six quarters of its entry. They argue that physicians use each drug's ability to lower cholesterol, and their prior belief about its efficiency ratio,[16] to infer each drug's ability to reduce heart disease risks. They argue that correlated learning happens when new clinical trials provide evidence about another drug's efficiency ratio. Therefore, even without any direct clinical evidence to show it can reduce heart disease risks, Lipitor was perceived to be the best drug for this purpose because of its superior ability to lower cholesterol. The physicians act as if they believe that evidence on an individual drug's ability to reduce heart disease risks can be generalized to the whole class of statins. This is the first paper that provides a structural explanation for a *late mover advantage*.

[16]The efficiency ratio measures how well a drug converts reduced *cholesterol* levels to reduced *heart disease* risks.

Finally, Che et al. (2015) use a forward-looking dynamic demand model to examine how brand preferences evolve when consumers are new to a market and their needs change over time. They allow for strategic sampling behavior of consumers under quality uncertainty, and they also allow for strategic sampling to increase periodically when consumers' needs change. The proposed model differs from previous work on forward-looking consumer Bayesian learning by allowing for (1) spill-over learning effects across different versions of products (or products in different product categories that share a brand name); and (2) duration-dependence in utility for a specific version of a product or product class to capture systematic periodic changes in consumer utility. Che et al. estimate their model using scanner data for the disposable diaper category. Here, it is likely that use experience with a particular size of a brand provides noisy information about another size of the same brand. And consumers' size needs change exogenously over time as the baby grows older and needs to change diaper sizes. The proposed model is useful in assessing the extent of use experience spillover effects and the degree to which information from past use experience is retained when consumers migrate across classes, versions, and the like.

8.6 Models Incorporating Both Learning and Search

Both the learning literature and the search literature focus on consumer choice under uncertainty. Search models are usually applied to explain dispersion in prices (or wages). In its simplest form, this class of models usually assumes there are a large number of retailers that sell the same product. Conditioning on a consumer who has already made up his mind to buy this product, his/her objective is to buy it from a retailer that offers the lowest price. But before visiting a retailer, he/she does not know what price it offers. However, it is typically assumed that consumers know the distribution of prices in the market. There is a cost to visiting a retailer (search cost). Finally, assume that consumers are forward-looking and conduct a sequential search. Then the dynamic programming problem solution implies that consumers' decision whether to continue to search (or to make a purchase now) is an optimal stopping problem. The solution is characterized by a reservation price: a consumer's decision rule is to reject any price above the reservation price, and accept any price that is below the reservation price.

Several papers have tried to relax the assumption that the price distribution is known. Rothschild (1974) proposes the first theoretical model to characterize the decision rule under such an environment. Recently, Koulayev (2013) and De los Santos et al. (2013) have extended Rothschild's model to an empirical setting. Koulayev (2013) uses Dirchlet distribution priors to model the uncertainty about the price distribution, while De los Santos et al. (2013) use Dirchlet process priors.

With Dirchlet distribution priors, Koulayev (2013) is able to derive closed form choice probabilities, and the characterization only relies on: (i) the identity of the second-best product among the discovered set; (ii) the number of searches to date. This allows Koulayev (2013) to estimate the model using only market share data. The Dirchlet process considered in De los Santos et al. (2013) is more general (i.e., it is the infinite dimensional version of the Dirchlet prior). Thus, to estimate their model, they need to observe price histories of each consumer.[17] They employ a moment inequality approach to estimate bounds on the parameters.

Interestingly, while the standard search model implies that consumers always buy on the last search, these learning and search models are able to explain why some consumers return to a previous search (as is supported by empirical patterns). This is because, in a learning and search model, reservation price is decreasing with the number of searches (conditional on continuing search). The reason is that, if a consumer continues to search, the price he/she just sampled must be higher than his reservation price. Thus, when the consumer uses this last observed price to update his prior, it must drive up his perceived "average" price (based on the prior belief he/she held right before seeing the last observation). The price distribution thus shifts up slightly, resulting in a higher reservation price. But with a higher reservation price, it is possible that some previously seen prices actually fall below it. That's why it may make sense for consumers to return to a previous search.

Roos et al. (2015) develop a structural model of hyper-media search and consumption while accounting for features unique to this consumption context: the rapid refresh of information (and consumers' concomitant uncertainty about its relevance and availability), the role played by linked excerpts in signaling the relevance and availability of new information on other sites, and the potential novelty or redundancy of information across sites. They estimate this new search model using panel data on consumer celebrity blog browsing and information scraped from sites regarding the links between them. Their results indicate that celebrity blogs are differentiated horizontally by their degree of sexually-oriented content and that links are a useful signal of the linked sites' content. Moreover, in many cases, links decrease (increase) visits to the linked (linking) sites.

[17]Note that the basic idea of search models is that consumers need to compare the expected gain from searching vs. the cost of search. In contrast, standard choice models with learning assume that consumers learn about an attribute by buying the product multiple times because information signals are noisy. These models also assume there are a fixed number of alternatives to choose from. In search models with an unknown price (or attribute) distribution, consumers learn about the parameters that characterize the distribution. For a normal distribution, that would be simply learning its mean and standard deviation. But, for a Dirichlet distribution, a consumer needs to use the whole history of price realizations and the initial parameters that characterize the prior to construct his posterior.

8.7 Heuristic and Approximation Approaches to Study Consumer Incentives to Explore

In most structural models of learning, the literature uses a dynamic programming approach to capture the idea that consumers make their choices while taking into account the benefit of exploration (or experimentation). In order to illustrate the set up, note that one can express the value of choosing alternative j as follows:

$$V(j,t|I_t) = E[U(Q_{jt}^E, P_{jt})|I_t] + \beta EV(I_{t+1}|I_t, j) \quad \text{for} \quad j = 0, \ldots, J. \tag{8.24}$$

Here we supress the person subscript i for notational convenience. As one will recall from the setup in Eqs. (8.1)–(8.8), experienced utility depends on experienced quality, Q_{jt}^E. This may depart from the true quality due to experience variability. Furthermore, the consumer must also account for the fact that he/she does not know the true quality of brand j with certainty. Rather, the consumer uses his/her information set I_t to infer the subjective distribution of brand j quality. In forming expected utility, $E[U(Q_{jt}^E, P_{jt})|I_t]$, the consumer must account for both of these sources of uncertainty.

Finally, $EV(I_{t+1}|I_t, j)$ is the expected future value of choosing product j, which takes into account how the choice of j changes the information set at $t + 1$. The parameter β is the discount factor. The "alternative specific value function" $V(j,t|I_t)$ simply adds together the current and discounted future payoffs from choosing brand j. A complete solution of the consumers dynamic optimization problem would give the values of the $EV(I_{t+1}|I_t, j)$ at every possible state point. This would enable a researcher to construct the alternative specific value functions $V(j,t|I_t)$, from which one could construct the choice probabilities and the likelihood function.[18]

To be more concrete, if we substitute using Eq. (8.7) we obtain:

$$V(j,t|I_t) = Ef\left(Q_{jt}^E|I_t\right) - w_P P_{jt} + e_{jt} + \beta EV(I_{t+1}|I_t, j). \tag{8.25}$$

This expression makes clear that the current payoff depends on (i) the subjective distribution of quality, which depends on the information set I_t, (ii) price (as well as other possible covariates like promotion that we might choose to add), and (iii) transitory taste shocks.

Now, suppose we compare the value of choosing two brands j and k. We obtain:

$$V(j,t|I_t) - V(k,t|I_t) = \left[Ef\left(Q_{jt}^E|I_t\right) - w_P P_{jt} + e_{jt}\right] - \left[Ef\left(Q_{kt}^E|I_t\right) - w_P P_{kt} + e_{kt}\right] + G(j,k,I_t), \tag{8.26}$$

[18]This is in contrast to static models, where the current expected utilities, $E[U(Q_{jt}^E, P_{jt})|I_t]$, alone determines choice probabilities.

where:

$$G(j,k,I_t) \equiv \beta[EV(I_{t+1}|I_t,j) - EV(I_{t+1}|I_t,k)]. \tag{8.27}$$

Equations (8.26)–(8.27) make clear that the value of choosing j over k can be decomposed into (i) the difference in expected current payoffs (which is all that matters in a static model) and (ii) the information advantage of choosing j over k, which we denote by $G(j,k,I_t)$. Intuitively, if k is a very familiar brand while j is new, we would expect $G(j,k,I_t) > 0$ as there is more information to be gained by trying j, which might turn out to be better than k. The existence of the G function is what generates the incentive for strategic trial in dynamic learning models.

Erdem and Keane (1996), along with most subsequent dynamic structural learning models, obtain the expected future value functions $EV(I_{t+1}|I_t,j)$ by solving a dynamic programming problem. However, it is not feasible to solve for $EV(I_{t+1}|I_t,j)$ at every possible state point (I_t,j), so the usual approach is to approximate the solution. For instance, the Keane and Wolpin (1994) approximation technique involves solving for the $EV(I_{t+1}|I_t,j)$ at a subset of state points, and interpolating to the other points (see Ching et al. (2013) for further details). But we now discuss a number of alternative heuristic and approximation approaches that have been proposed more recently.

A recent trend in the economics/marketing literatures has been to model consumer learning without using the full dynamic optimization and Bayesian updating framework. This may be done either by assuming (or allowing) that consumers use heuristics, or that consumers solve their "true" underlying dynamic optimization problem by approximation using heuristic methods. Gabaix and Laibson (2000) argue that since cognition is costly, sophisticated decision-makers should adopt heuristics or short cuts to reduce cognitive burden. They use a decision tree model where agents systematically prune away low probability paths. They estimate the model on lab data, using students as subjects. The model successfully captures the decisions students made in the experiment. The authors note that further research should attempt to identify a parsimonious set of parameterized algorithms, and provide a theory that describes how the parameters adjust across problems. Natural adjustment candidates include reinforcement learning and expected-payoff maximization subject to constraints on calculation and memory.

Geweke and Keane (2000) (hereafter, GK) develop a method to approximate the solution to DP problems by replacing the "future component" of the value function with a flexible function of the state variables. Specifically, they rewrite (Eq. 8.24) as:

$$V(j,t|I_t) = E[U(Q^E_{jt},P_{jt})|I_t] + F[I_{t+1}(I_t,j)|\pi_t] \quad \text{for} \quad j = 0,\ldots,J. \tag{8.28}$$

Here $F[I_{t+1}(I_t,j)|\pi_t]\approx\beta EV(I_{t+1}|I_t,j)$ is a flexible polynomial in the state variables that approximates the "future component" of the value function. And π_t is a vector of reduced form parameters that characterize the future component. The structural parameters of the current payoff function are then estimated jointly with the reduced form parameters of this polynomial approximation. GK showed that this method, which involves no greater computational burden than estimating a static discrete choice model, uncovers estimates of the structural parameters that exhibit negligible bias.

It is interesting to note that the GK approach is equivalent to directly assuming a simple parameterization of the G function in Eq. (8.27). For instance, one might assume that the information gain from choosing j over k is an increasing function of the perception error variance for j relative to that of k (see Eq. (8.6)).

Houser et al. (2004) extend GK to introduce a new Bayesian procedure for drawing inferences about both the nature and number of decision rules that are present in a population of subjects, where each subject is confronted with a dynamic decision problem. More specifically, a game is designed for an experiment in which participants play for money. Participants are allowed to practice to learn the game before playing. Data from the experiment shows that some participants make close to optimal decisions, while others appear to use simple heuristic rules (some of which are less accurate than others). The main take-away of this paper is that there is significant heterogeneity in how consumers solve dynamic problems.

Ching et al. (2014a) use the GK approach to study consumer's incentive to experiment with unfamiliar brands in the diaper category. They also provided some new insights on identification in the GK framework. Notice that, as F in Eq. (8.28) is just a flexible function of the state variables, all that is assumed in the GK approach is that consumers understand the laws of motion of the state variables (i.e., how $(I_{t+1}\ |I_t,\ j)$ is formed). They need not form expectations based on the true model. The approach is also agnostic about whether consumers use Bayesian updating or some other method. In general, identification of π_t requires either: (i) observing current payoffs,[19] or (ii) exclusion restrictions, such that some variables enter the future component F but not current utility. Ching et al. (2014a) point out that such exclusion restrictions arise naturally in dynamic learning models, because the updated perception error variances $\sigma^2_{ij,t+1}$ only affect future payoffs, not current utility.

As one can see from Eq. (8.25), when the full structure is not imposed, we will not be able to identify the discount factor. The β is subsumed as a scaling factor for the parameters π_t of the F function. However, one can test whether $\pi_t = 0$, which is a test for forward-looking behavior (or "strategic trial"). Although this test makes weak assumptions about F, it is *not* non-parametric, as a functional form must be

[19]In labor economics, researchers may argue that wages capture much of the current payoff. Or, researchers can control current payoffs in a lab experiment (e.g., Houser et al. 2004).

chosen for the current payoff function. As Ching et al. (2014a) show, given the current payoff function, the π_t are identified in the learning model because different current choices lead to different values of next period's state variables (e.g., the posterior variances in Eq. (8.6)). Ching et al. (2014a) find evidence of forward looking behavior in the diaper category.

Another way to reduce computational burden in dynamic models is to rely on "index strategies." These include the Gittin's index developed by Gittins and Jones (1974) and Whittle's index developed by Whittle (1988). To understand how these approaches work, return to Eqs. (8.26)–(8.27), but now assume the choice is between brand j and a hypothetical certain alternative (denoted by 0) that delivers a fixed payoff λ_j and that, when chosen, leads to no gain of information. Then we can write that:

$$V(j,t|I_t) - V(0,t|I_t) = \left[Ef\left(Q_{jt}^E|I_t\right) - w_P P_{jt} + e_{jt}\right] - \lambda_j + G(j,0,I_t) \tag{8.29}$$

where:

$$G(j,0,I_t) \equiv \beta[EV(I_{t+1}|I_t,j) - EV(I_t)]. \tag{8.30}$$

Here $EV(I_t)$ is simply the expected value of arriving in the next period with no more information than one has today. One can see that the consumer will be indifferent between alternative j and the hypothetical certain alternative if:

$$\lambda_j \equiv \left[Ef\left(Q_{jt}^E|I_t\right) - w_P P_{jt} + e_{jt}\right] + G(j,0,I_t). \tag{8.31}$$

That is, the sure payoff λ_j from choosing the hypothetical alternative is equal to the expected payoff from alternative j plus the value of information gained by choosing j. The value of λ_j in Eq. (8.31) is known as Whittle's index.

Whittle's (1988) result is that, under certain conditions, it is optimal in each period to choose the brand j that has the highest λ_j. This greatly simplifies the dynamic optimization problem because, instead of a dynamic problem with J choices, one only has to solve a set of J simple and independent optimal stopping problems. In each of those simple problems, agents chose in each period between a single brand j (for $j = 1, \ldots, J$) and the certain option 0.[20] However, for this simplification to work, the key condition that must be satisfied is that, if brand j is

[20]Each sub-problem can be characterized as follows. A consumer either chooses a fixed reward in each period forever, or chooses brand j this period. If he/she chooses brand j this period, a noisy quality signal about brand j will be revealed, and then the consumer faces these two choices again next period. The reasons the index method provides significant computational gains are: (a) it reduces the size of the state space from N^J to $J \times N$, where N is the number of state points associated with each alternative, and (b) solving for the index strategy for J optimal stopping problems is much less costly compared with solving one J dimensional dynamic programming problem.

chosen at time t, the information sets for all brands $k \neq j$ must remain unchanged from period t to $t + 1$. This rules out exogenous sources of information (such as the advertising signals in Erdem and Keane (1996), word of mouth, etc.), as well as correlated learning across brands.

The Gittin's index is basically a simplified version of the Whittle index that rules out exogenous shocks to current utilities of the various brands. The first paper in marketing to use an index strategy to solve a dynamic model was Eckstein et al. (1988), who used the Gittin's index. When consumers are risk-neutral (i.e., Q_{jt}^{E} enters the utility function linearly) and there are no random shocks to the values of the alternatives, a solution based on choosing the alternative with the highest Gittin's index is exactly the same as solving a full-fledged dynamic programming problem. However, in the typical random utility framework used in marketing and economics, a solution based on Gittin's index may not be optimal. Moreover, if consumers are risk-averse, the Gittin's index may not exist. Even if it exists, it does not necessarily lead to the optimal choice.

Lin et al. (2015) show how to use the Whittle index to deal with dynamic random utility models that include learning through use experience as well as risk aversion and both observed and unobserved shocks to the utilities of alternatives (the literature classifies these as "restless-bandit" problems). They allow for direct persuasive effects of advertising, but do not allow advertising to convey information about brand quality (which, as noted above, would violate the Whittle index assumptions). Lin et al. (2015) apply their model to IRI diapers data and find evidence for forward-looking behavior (i.e., strategic trial). That is, a forward-looking learning model fits the data better than a myopic learning model.

Lin et al. (2015) also show that the Whittle index provides a solution to the DP program that is very close to the Keane and Wolpin (1994) approximate solution, but at lower computational cost. They also argue that an index strategy would be intuitive to consumers, so that it is plausible that consumers follow heuristics that are close to an index strategy. This argument can be understood by looking at Eqs. (8.26)–(8.27). Clearly, the optimal decision rule is equivalent to the static decision rule except that a value of the gain from acquiring information is added on to the value of each alternative. It seems intuitive that consumers understand there is some value to the information gained by trying out unfamiliar brands, and that they would try to take this into account when making purchase decisions.

Notably, the GK approach is equivalent to a simplified index strategy where the analyst directly choses a functional form for the gain from gathering information.[21] The GK method then involves directly inferring from the data the (possibly sub-optimal) index rule or heuristic that rationalizes consumer choice behavior.

Sauer (2015) develops what he calls a "hybrid" approach that combines GK with Keane and Wolpin (1994). He assumes that consumers can look one period ahead —that is, they can backsolve a dynamic programming problem optimally from one

[21]To see this, compare Eq. (8.28) with Eqs. (8.29)–(8.30). Clearly, the GK approach amounts to choosing a parameterization for the $G(j, 0, I_t)$ function.

period ahead—but at $t + 2$ they use the GK approach to approximate the future expected value functions as simple functions of the state variables. An advantage of this approach is that it allows one to estimate the discount factor.

Tehrani and Ching (2016) propose another heuristic concept called the Value of Perfect Information (VPI), which dates back Howard (1966). The basic idea of this concept is to capture the expected gain of finding out the true value of choosing alternative j. To illustrate how to obtain the VPI for alternative j, let's first reorder the alternatives such that the best myopic choice is 1 based on $I(t)$, that is,

$$E[U_1|I(t)] > E[U_2|I(t)] > \cdots > E[U_J|I(t)]. \tag{8.32}$$

Let's consider alternative $j = 1$ first. Suppose the true quality is Q_j^*. This knowledge is valuable if it reveals that the original best myopic choice is no longer the best, i.e., $U_1(Q_1^*) < E[U_2|I(t)]$ (because that will lead the consumer to choose alternative 2 instead of 1); otherwise, the new knowledge does not change choice and hence renders no gain. Similarly, for alternatives $j > 1$, the knowledge of true Q_j^* is valuable only if $U_j\left(Q_j^*\right) > E[U_1|I(t)]$ as this will trigger the consumer to choose alternative j over the original best myopic choice. To illustrate how to obtain VPI, let's define a gain function as follows:

$$Gain_{jt}(Q_j^*) = \begin{cases} E[U_2|I(t)] - U_1\left(Q_1^*\right) & \text{if } j = 1 \parallel \text{and} \quad U_1\left(Q_1^*\right) < E[U_2|I(t)]; \\ U_j\left(Q_j^*\right) - E[U_1|I(t)] & \text{if } j > 1 \parallel \text{and} \quad U_j\left(Q_j^*\right) > E[U_1|I(t)]; \\ 0 & \text{otherwise.} \end{cases} \tag{8.33}$$

However, the consumer is uncertain about Q_j^*. Hence, he can only compute the expected $Gain_{jt}(Q_j^*)$ based on his/her prior belief at t, $f_{jt}(.)$, and this gives us the VPI associated with alternative j:

$$VPI_{jt} = \int_{-\infty}^{\infty} Gain_{jt}(x) f_{jt}(x) dx. \tag{8.34}$$

Tehrani and Ching (2016) propose including VPI_{jt} as an additively separable variable to the current expected utility associated with alternative j. In a sense, VPI_{jt} is a replacement for the expected future value, which we normally obtain by solving a dynamic programming problem.

One way to interpret VPI_{jt} is that consumers look one period ahead and assume that all uncertainty will be resolved by one trial. It does not take into account that learning could be slow (as when signals are noisy, so that one-trial-learn-everything cannot be achieved). This shortcoming can be addressed by modifying the definition of VPI_{jt} above by using the Bayesian updating formula to take the variance of the noisy signals into account. It is worth highlighting that the computational burden of solving for VPI_{jt} is relatively light, because it only involves solving a one-dimensional integration instead of a dynamic programming problem. In fact, it is much easier to implement the VPI approach compared with the index strategy which still requires solving for J dynamic programming problems (optimal stopping problems). Therefore, it is conceivable that consumers may adopt a heuristic like VPI. Tehrani and Ching (2016) provide evidence that the VPI approach can explain the brand choice dynamics well in the diaper category.

The concept of learning has also been treated from the viewpoint of a company, which learns how to match the "look and feel" of a web site to the cognitive styles of consumers (a process known as "website morphing"). Hauser et al. (2009) use clickstream data to infer cognitive styles. Their proposed model balances exploration (learning how morphing affects consumer choice probabilities) with exploitation (maximizing short term sales) by solving a dynamic program using Gittin's index. The authors apply their Bayesian updating and dynamic programming model to an experimental British Telecom web site. Findings reveal that adaptation of such approaches can lead to substantial additional revenues.

Last but not least, Dzyabura and Hauser (2011) develop and test an active machine-learning model to identify heuristic decision-making. They illustrate their algorithm using data from a web-based survey conducted by an American automotive manufacturer to study vehicle consideration. The conjoint experiment included 872 respondents and 53 feature-levels. The authors conclude that active machine learning is an effective methodology to select questions adaptively in a conjoint context in order to identify consideration heuristics. But many challenges remain open in this area.

One point that should be stressed about all heuristic approaches is they are subject to the Lucas-Marschak critique of using reduced form models to predict the effect of policy changes. This is because a heuristic that works well in one environment may perform poorly in another. Thus, if the policy environment changes, people may change the heuristics that they use.

For example, consider demand for diapers. In an environment where stores keep prices fairly stable, except they put diapers on sale on most Fridays, consumers may well hit on a (close to cost minimizing) heuristic that simply says "Buy diapers on Friday." However, if stores start to randomize the day of sales, consumers will presumably change their heuristic. One would need a structural model to forecast the new heuristic that consumers adopt in the new context.

Similarly, several papers we have discussed are motivated by the idea that consumers use heuristics to reduce information processing costs. But such costs are a function of the market environment. For example, say a consumer faces a choice between only two product varieties. In this case, he/she may compare all their attributes carefully. But if the variety in the market expands to 50 items this becomes infeasible. Then, the consumer may adopt a lexicographic rule (e.g., first screen by price range, then by color, etc.) as a simplifying heuristic. Again, one would need a structural model to predict the point at which the choice environment gets sufficiently complex that people switch from a compensatory to a lexicographic decision rule.[22]

Similar arguments apply to the use of approximate/heuristic approaches to solving dynamic optimization problems. An approximation that is accurate in one context may be inaccurate in another. For example, as noted by Keane and Wolpin (1994), the expected maximum of several alternative specific value functions (the “E max”) is well approximated by maximum of the expected values of those same value functions (the “max E”) *provided* that the choice specific error variances are small. But this approximation breaks down in an environment with more uncertainty. So, while the use of heuristic-based models may ease the burden of econometric estimation, they are unlikely to substitute for more structural approaches when it comes to the problem of predicting behavior under very different policy regimes.

These caveats should not be taken to diminish the potential importance of heuristic-based models in providing valuable information about how consumers actually behave in particular real world choice environments. Our point is simply that heuristic-based models (like reduced from models) should only be used to predict behavior in response to policy changes if it is plausible to maintain that the choice heuristic is invariant to the policy change.[23]

8.8 Using Exogenous Events and Policy Changes to Study Learning

In recent years, there has been an increased interest in studying the nature of consumer learning by exploring the impact of exogenous events or policy changes. For example, one event that has generated a great deal of attention is the

[22]Another example is that the advent of internet retailing has made it possible to do comparison shopping from home, thus arguably reducing the costs of gathering information. An interesting hypothesis is that this change in the environment may have caused consumers to engage in more comparison shopping.

[23]One way to interpret our argument is that only structural models *attempt* to predict what decision rules consumers will adopt in a new environment (indeed, this is precisely what structural models are designed to do). But that doesn’t mean their predictions will necessarily be correct. It is important to keep in mind the point that a structural model is only invariant to all conceivable environmental changes if it is perfectly correctly specified—that is, if it is in fact the “true model.” As all models are ultimately false (as they are simplifications), a completely policy invariant model is an aspirational goal, not a reality. The best we can do in practice is to incrementally validate a structural model by showing that it predicts well across a range of policy environments. This may give us confidence in using the model to predict in a new environment. But we can never be certain that the new environment won’t be the one that reveals the flaw in the model! As a practical matter, the best we can hope for is to build structural models we are confident in using for certain types of policy predictions, but perhaps not for others (i.e., it is perfectly possible that a structural model can reliably predict responses to some types of policy changes but not others—just as we see with commonly used models in the physical sciences and engineering). See Keane (2010) for further discussion of these issues.

implementation of the Medicare prescription drug plan in the US (known as Medicare "Part D"). This program went into effect in 2006, at which point the federal government created (via heavy subsidies) a new private market in prescription drug plans for people 65 and over. This event created a unique opportunity to study consumer learning behavior, because, essentially over-night, a large number of alternative prescription drug plans were suddenly made available to senior citizens.

On average, each senior citizen faced a choice among roughly 50 drug plans, offered by 20 different insurers. These plans vary in terms of premium, out-of-pocket costs, formulary, the ease of seeking reimbursement, and customer service. With such a degree of complexity, it is likely that a significant portion of consumers make uninformed decisions. Other than uncertainty about plan's attributes, consumers may also be uncertain about their state of the world in the coming year (i.e., how sick he/she could become and hence the type of drugs needed). Because the timing of decisions is clear (open enrolment happens once a year), this market provides an excellent opportunity to study how consumers decide to consider switching (Ching et al. 2009; Ketcham et al. 2015a; Ching and Lim 2016). Moreover, the data on their initial choice, subsequent choices, and actual spending patterns potentially provide researchers with information about how consumers learn about which plan fits them best over time.

Ketcham et al. (2012) find evidence that consumers are in fact "learning" about these plans over time. They study whether Medicare Part D enrollees improved over time in terms of reducing overspending. They find that the mean cost difference between individuals' actual choices and their cheapest option fell by about $330 from 2006–2007 and this average reduction in overspending is in part due to individuals who chose to switch plans. The likelihood of switching plans for 2007 increased substantially with the amount of overspending in 2006. They attribute these results to participants learning about the costs and benefits of different plans.

However, given the complexity of the market, it is possible that a simple Bayesian learning model may still miss many important features in the data. As pointed out in Ching et al. (2013), having stated preference data could potentially enhance researchers' ability to build a better structural model as a closer proxy to actual behavior.

To address this research agenda, Ketcham et al. (2015b) have linked the claim data made available by The Center of Medicare and Medicaid Services with the Medicare Current Beneficiary Survey conducted three times per year. Using this data set, Ketcham et al. (2015b) are able to separate informed and uninformed consumers. Their main research question is to estimate how consumers may change their choices and the welfare consequences under several counterfactual policies. But this data set can also potentially tell us how confused or uninformed consumers are. However, so far researchers in this area have mainly relied on a simple static multinomial logit model to draw inferences. It would be very interesting to develop a model with limited foresight and formation of consideration sets to understand how consumers choose in such a complicated environment.

Turning to a different type of example, Sudhir and Yang (2015) explore free upgrade events in car rentals, and argue that they provide exogenous "random" assignment of car types to consumers, independent of their preferences. As a result, they argue that stickiness of preferences after such consumption experiences can be used to draw causal inferences about state dependence.

Similarly, Larcom et al. (2015) use another exogenous event to study the stickiness of choice that leads to suboptimal experimentation. The February 2014 London underground train strike temporarily shut down some stations, and that forced some commuters to look for alternative routes to get to their destination. Larcom et al. (2015) argue that this provides them with an opportunity to investigate whether commuters were using the best route prior to the strike, or whether they had not explored all the options yet and were using a suboptimal route. The strike forced some commuters to experiment, but because not all commuters were affected (some stations operated during the strike), Larcom et al. (2015) are able to use the unaffected commuters as a control group, and apply the difference-in-difference approach to test their hypothesis. Interestingly, they find that a majority of commuters return to the original routes after the strike; but a small percentage switched to the new routes, suggesting they were using a suboptimal route before.

Gallaghier (2014) uses an event study framework to estimate the effect of large regional floods on the take-up of flood insurance. He finds that insurance take-up spikes the year after the flood and then declines steadily to baseline. Residents in non-flooded communities in the same TV media market increase take-up at one third rate of the flooded communities. Gallaghier's findings are consistent with a Bayesian learning model with forgetting and/or incomplete information about past floods. Thus, the form of belief updating is an area where it may be important to relax or generalize the behavioral assumptions of standard Bayesian learning models (a point we return to in Sect. 8.9).

Finally, Davis (2004) measures the impact of an outbreak of pediatric leukemia on local housing values. A model of location choice is used to describe conditions under which the gradient of the hedonic price function with respect to pediatric leukemia risk is equal to household marginal willingness to pay to avoid risk. The equalizing differential is estimated using property-level sales records from a county of Nevada where residents experienced a severe increase in pediatric leukemia. Housing prices are compared before and after the increase with a nearby county acting as a control group. The results indicated that housing values decreased 15.6% during the period of maximum risk. Using lifetime estimates of risk derived from a Bayesian learning process, the results imply the statistical value of pediatric leukemia is $5.6 million. The approach adopted in this study suggests avenues for future research on quantifying the trade-offs between money and various risks, such as health risks.

8.9 Future Research Directions

In this section we discuss some largely unexplored territories for future research and/or areas that are under-researched. In Sect. 8.7, we discussed papers that approximate agents' optimization decisions with heuristic methods, as well as papers that attempt to capture agents' use of heuristics. Such use of heuristics and/or behavioral phenomena that deviate from typical rational decision-making may be important for understanding choice in many contexts.

One important way that people may deviate from the behavioral assumptions of standard Bayesian learning models is in how they update beliefs in response to new information. There is, for example, evidence that consumers either over or under-react to information. There is also evidence suggesting that consumers may over-react to experience signals when learning opportunities happen infrequently, and then "forget" that experience rather quickly over time. Agarwal et al. (2013) measure learning and forgetting in the credit card market, using a panel with four million monthly credit card statements. They find that paying a fee last month, which is a negative feedback, reduces fee payment in the current month 40%. However, the authors also find evidence for recency effects and a 10% or more depreciation of knowledge per month.[24] The paper by Gallaghier (2014) on flood insurance that we discussed in Sect. 8.8 also presents results that appear consistent with forgetting.

Another area of particular interest is choice in very complex situations. What we mean by a "complex" choice situation includes cases where:

(a) The object under consideration is complex, in that it has many attributes, or some attributes that are difficult to understand or evaluate;
(b) The choice set is complex because there are a very large number of alternatives;
(c) Choice requires evaluating probabilities and/or making intertemporal allocations.

Good examples of what we mean are choices in areas such as health/life insurance, retirement plans or investments. These types of choices all arise in the area of optimal life-cycle planning, and they exhibit all three factors that contribute to complexity of the decision task as described above. However, optimal life-cycle planning is an area largely ignored in the standard learning literature, and the marketing literature in general.[25]

Optimal life-cycle planning requires the solution of a complex dynamic programming (DP) problem. But actual decision making in the domain of such planning (e.g., retirement planning, saving for college education of children, etc.) often departs in obvious ways from this normative principle, and people often seem

---

[24]Interestingly, higher-income borrowers learn twice as fast, and forget twice as slowly, as lower-income borrowers.

[25]An exception is Yang and Ching (2014) who develop and estimate a consumer life-cycle model to explain the adoption decision of a new technology.

to react to the difficulty of the problem with the use of simple heuristics, or even with delays and procrastination (see Keane and Thorp 2016 for a review).

Interestingly, methods that appear to be relevant for such planning problems have already been developed for the closely related problem of inventory planning. Specifically, Ching et al. (2009, 2014a)—henceforth CEK—develop a model of consumer demand for a storable (or quasi-durable) branded commodity. In this context, optimal behavior involves: (i) checking the prices of all brands of a product in every period, and (ii) solving a DP problem to determine both (ii-a) the reservation price for purchase of each brand, and (ii-b) the optimal quantity to buy in the event that the price of a brand is below its reservation price. Of course, the reservation prices and optimal purchase quantities both evolve in a complex way with inventories.

CEK argue that a normative model is unrealistic for two reasons: (i) For most products, consumers presumably do not have the time, interest or mental capacity to check all prices in every period, and (ii) in those periods when consumers do pay close attention to a category, they presumably make decisions using more or less sophisticated rules of thumb, not by literally solving a DP problem. Thus, CEK develop a two-stage model of demand for a storable branded product. In the first stage, consumers decide whether or not to pay attention to the product category.[26] If they do decide to pay attention then, in stage two, they use a rule of thumb that may or may not provide a good approximation to the DP problem (depending on parameter estmates), as in Geweke and Keane (2000, 2001). In CEK's empirical applications, the decision whether to consider a category is modelled as a simple probit or logit discrete choice model, where the factors that drive consideration are cues like advertising, displays and low inventory.[27]

It seems fairly clear how one might apply the CEK framework to financial products like annuities, life insurance or choice of retirement plans. As discussed in Keane and Thorp (2016), there is clear evidence that most consumers are averse to thinking about these products on a regular basis. For example, as is well-known, the typical consumer does not engage in a frequent re-balancing of his/her stock portfolio as the state of the world changes. It is natural to think of a framework where, in a first-stage, consumers decide on, say, a quarterly or annual basis whether to consider financial products in a certain category. The decision to consider could be driven by advertising cues, as well as by major life events such as retirement, children leaving home, a spouse passing away, selling a house and/or moving house, or reaching a milestone birthday. In the second stage, it would again

[26] CEK's work was originally motivated by the observation that brand choice *conditional* on category purchase is very sensitive to price, while the decision to make a purchase in a category is quite insenitive to price. CEK showed that these seemingly contradictory facts could be explained if consumers only occasionally look at (i.e., consider) a category.

[27] In the optimal solution consumers should consider a category in every period regardless of their inventory. Even if inventory is high, a low enough price would make it optimal to stock up even more.

be optimal to estimate a behavioral rule of thumb from the data, rather than imposing an optimal DP solution.

Given such an estimated model, one could simulate behavior under the model versus under a normative solution to the planning problem. One could then evaluate whether or not the wealth losses from following the simplified decision process rather than the optimal DP solution are substantial (as in Houser et al. 2004).

Another under-researched area is the endogenous formation of preferences. Malmendier and Nagel (2011) present evidence that risky asset returns experienced over the course of an individual's life have significant effects on the willingness to take financial risks. People who have experienced high stock market returns report lower aversion to financial risks, are more likely to participate in the stock market, and allocate a higher proportion of their liquid asset portfolio to risky assets. While individuals put more weight on recent returns than on more distant realizations, the impact fades only slowly with time. Malmendier and Nagel remain agnostic whether the experience effects on risk taking arise from experience-dependent beliefs or from endogenous risk preferences. Their results show, however, that there is dependence on "experienced data"—as opposed to "available data"—in standard rational and boundedly rational learning models. The implications of this in a variety of contexts (from exploring micro issues such as how heterogeneity arises among agents to macro issues, such as the dynamics of asset prices in this specific case) are another interesting avenue for future research.

Finally, another avenue for future research is to bring new data to bear on the problem of identifying learning processes—that is, information that goes beyond just the history of signals received and choice made. For example, Erdem et al. (2005), who combined psychometric data on product ratings with revealed preference to help identify parameters of the learning process. In another example, Ching et al. (2014b) develop a novel way to measure the extent to which consumers are uncertain about the true quality of quasi-durable goods. Using data on diapers, they argue that data on consumer buying decisions reveals their subjective perceived qualities, while data on inter-purchase spells reveals the objective quality (at least in terms of durability). The difference of these two measures tells us to what extent consumers are uncertain about the quality of the products.

8.10 Conclusion

In this review paper, we discussed recent developments in the literature on consumer choice dynamics and learning. Erdem and Keane (1996) showed that a simple Bayesian learning model was quite successful in explaining the observed dynamics in consumer choice behavior. Their application was to learning via use experience and advertising about frequently purchased consumer goods. Since that time, a large literature has developed that extends learning models both in terms of (i) the sources of information and types of learning that take place and (ii) the nature of the objects that agents are learning about.

In particular, recent years have seen learning models extended to include learning from experiences of friends or other social network members ("social learning"), experience with related products ("correlated learning" or "information spillover"), examination of publicly available information or expert opinion ("search"), and inferences about product attributes from purchase decisions of others ("observational learning"). Similarly, learning models have been successfully applied to many different types of products and activities including high-tech durables, drugs, medical procedures, movies, books, video games, restaurants, sexual behaviors, new product ideas, health insurance plans, and so on.

In this review we have summarized papers that find evidence of learning as one of the main mechanisms that explains dynamics in consumer choice behavior for a wide range of products or activities in a wide variety of settings. We have also summarized papers that find evidence of learning from many different sources. Learning models have also been applied to model firm learning about consumer willingness to pay for goods and market entry potential.

One important recent development has been the adoption of heuristic approaches to model consumer and firm learning and decision-making in complex environments. The basic idea is to relax some of the normative assumptions of Bayesian learning models to allow for cognitive limitations and/or behavioral biases on the part of consumers. While we have discussed some work in this area, it remains an under-researched topic that should be a fruitful avenue for future research.

Another key area for future research is to develop models that integrate learning with other potentially important sources of choice dynamics, such as inventories, switching costs, habit persistence, as well as behavioral factors like inattention/procrastination. For example, in Ching et al. (2014a) we develop a choice framework that integrates learning, inventories and rational inattention, while relaxing some assumptions of the Bayesian learning model. But much more work remains to be done in this area.

In conclusion, learning models have proven to be a fruitful area of research activity for the past 20 years, and the level of activity in this area has been growing substantially. This is reflected in the fact that the bulk of our references are from just the past few years. We expect this literature to continue to grow rapidly as learning models are applied to more and more domains, and as researchers continue to generalize the structure of the original Bayesian learning models in Roberts and Urban (1988), Eckstein et al. (1988) and Erdem and Keane (1996) in many interesting new directions.

Acknowledgements Keane's work on this project has been supported by Australian Research Council grants FF0561843 and FL110100247. Ching's work has been supported by SSHRC. But the views expressed are entirely our own. We thank two anonymous referees and Ralf van der Lans for their helpful comments. We also thank Shervin Shahrokhi Tehrani for providing excellent research assistance. The article was previously circulated under the title "Models for Marketing Dynamics and Learning: A Survey of Recent Developments."

References

Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David Laibson. 2013. Learning in the credit market. Working paper.

Aguirregabiria, V., and P. Mira. 2007. Sequential estimation of dynamic discrete games. *Econometrica* 75 (1): 1–53.

Chan, T., J. Li, and L. Pierce. 2014. Learning from peers: Knowledge transfer and sales force productivity growth. *Marketing Science* 33 (4): 463–484.

Che, Hai, Tülin Erdem, and T. Sabri Öncü. 2015. Consumer learning and evolution of consumer brand preferences. *Quantitative Marketing and Economics* 13 (3): 173–202.

Chen, T., B. Sun, and V. Singh. 2009. An empirical investigation of the dynamic effect of marlboro's permanent pricing shift. *Marketing Science* 28 (4): 740–758.

Ching, A.T. 2010a. Consumer learning and heterogeneity: dynamics of demand for prescription drugs after patent expiration. *International Journal of Industrial Organization* 28 (6): 619–638.

Ching, A.T. 2010b. A dynamic oligopoly structural model for the prescription drug market after patent expiration. *International Economic Review* 51 (4): 1175–1207.

Ching, A., T. Erdem, and M. Keane. 2009. The Price Consideration Model of Brand Choice. *Journal of Applied Econometrics* 24 (3): 393–420.

Ching, A.T., T. Erdem, and M. Keane. 2013. Learning Models: An Assessment of Progress, Challenges and New Developments. *Marketing Science* 32 (6): 913–938.

Ching, A.T., T. Erdem, and M. Keane. 2014a. A Simple method to Estimate the Roles of Learning, Inventories and Category Consideration in Consumer Choice. *Journal of Choice Modelling* 13: 60–72.

Ching, A.T., T. Erdem, and M. Keane. 2014b. How much do consumers know about the quality of products? Evidence from the diaper market. Working paper, Rotman School of Management, University of Toronto.

Ching, A.T., R. Clark, I. Horstmann, and H. Lim. 2016. The effects of publicity on demand: the case of anti-cholesterol drugs. *Marketing Science* 35 (1): 158–181.

Ching, A.T., and M. Ishihara. 2010. The effects of detailing on prescribing decisions under quality uncertainty. *Quantitative Marketing and Economics* 8 (2): 123–165.

Ching, A.T., and H. Lim 2016. A Structural analysis of correlated learning and late-mover advantages: The case of statins. Working paper, Rotman School of Management, U of Toronto.

Davis, L. 2004. The effect of health risk on housing values: evidence from a cancer cluster. *American Economic Review* 94 (5): 1693–1704.

De los Santos, B., A. Hortacsu, M. R. Wildenbeest. 2013. Search with learning. Working paper, Kelley School of Business, Indiana University.

Dzyabura, D., and J.R. Hauser. 2011. Active Machine Learning for Consideration Heuristics. *Marketing Science* 30 (5): 801–819.

Eckstein, Z., D. Horsky, and Y. Raban. 1988. An empirical dynamic model of brand choice. Working paper 88, University of Rochester.

Erdem, T. 1998. An empirical analysis of umbrella branding. *Journal of Marketing Research* 35 (3): 339–351.

Erdem, T., and M. Keane. 1996. Decision-making under uncertainty: capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing Science* 15 (1): 1–20.

Erdem, T., M. Keane, S. Oncu, and J. Strebel. 2005. Learning about computers: an analysis of information search and technology choice. *Quantitative Marketing and Economics* 3 (3): 207–246.

Erdem, T., M. Keane, and B. Sun. 2008. A dynamic model of brand choice when price and advertising signal product quality. *Marketing Science* 27 (6): 1111–1125.

Gabaix, X., and D. Laibson. 2000. A boundedly rational decision algorithm. *American Economic Review* 90 (2): 433–438.

Gallaghier, J. 2014. Learning about an infrequent event: evidence from flood insurance take-up in the United States. *American Economic Journal: Applied Economics* 6 (3): 206–233.

Geweke, J., and M. Keane. 2000. Bayesian inference for dynamic discrete choice models without the need for dynamic programming. In *Simulation based inference and econometrics: methods and applications*, eds. Mariano, Schuermann and Weeks, 100–131. Cambridge University Press.

Geweke, J., and M. Keane. 2001. Computationally Intensive Methods for Integration in Econometrics. In *Handbook of econometrics:* vol. 5, Heckman, J.J., and E.E. Leamer eds., Elsevier Science B.V., 3463–3568.

Gittins, J., and D. Jones. 1974. A dynamic allocation index for the sequential design of experiments. In *Progress in Statistics*, ed. J. Gani, K. Sarkadi, and I. Vince, 241–266. Amsterdam: North-Holland.

Godlonton, S., and R.L. Thornton. 2013. Learning from others' HIV testing: updating beliefs and responding to risk. *American Review Papers and Proceedings* 103 (3): 439–444.

Guadagni, P.M., and John D.C. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Science* 2 (3): 203–238.

Hauser, J.R., G.L. Urban, G. Liberali, and M. Braun. 2009. Website Morphing. *Marketing Science.* 28 (2): 202–223.

Heckman, J.J. 1981. Statistical models for discrete choice panel data. In *Structural analysis of discrete data with econometric applications*, eds. Manski, C.F., and D. McFadden, 114–178. Cambridge, MA: MIT Press.

Hendrick, K., A. Sorensen, and T. Wiseman. 2012. Observational learning and demand for search goods. *American Economic Journal: Microeconomics* 4 (1): 1–31.

Ho, T.-H., S.-E. Park, and X. Su. 2015. A bayesian level-k model in n-person games. Working paper.

Houser, D., M. Keane, and K. McCabe. 2004. Behavior in a dynamic decision problem: an analysis of experimental evidence using a bayesian type classification algorithm. *Econometrica* 72 (3): 781–822.

Howard, R.A. 1966. Information Value Theory. *IEEE Transactions on System Science and Cybernetics* 2 (1): 22–26.

Huang, G., H. Luo, and J. Xia. 2015. Invest in information or wing it? A model of dynamic pricing with seller learning. Working paper, Carnegie Mello University.

Hummel, P., and B. Knight. 2015. Sequential or simultaneous elections? A Welfare Analysis. *International Economic Review* 56 (3): 851–887.

Keane, M. 1994. A computationally practical simulation estimator for panel data. *Econometrica* 62 (1): 95–116.

Keane, M. 2010. Structural versus atheoretic approaches to econometrics. *Journal of Econometrics* 156 (1): 3–20.

Keane, M. 2015. Panel data discrete choice models of consumer demand. In *Oxford handbook of panel data*, ed. Baltagi, B., Chapter 18, pp. 548–582 Oxford University Press.

Keane, M., and S. Thorp. 2016. Complex decision making: the roles of cognitive limitations, cognitive decline and aging. In *Handbook of the economics of population aging*, Vol. 1B, J. Piggott, and A. Woodland, eds., Elsevier Science B.V., pp. 661-709.

Keane, M., and K. Wolpin. 1994. The solution and estimation of discrete choice dynamic programming models by simulation: monte carlo evidence. *Review of Economics and Statistics* 76 (4): 648–672.

Ketcham, J.D., C. Lucabell, E.J. Miravete, and M.C. Roebuck. 2012. Sinking, swimming, or learning to swim in medicare Part D. *American Economic Review* 102 (6): 2639–2673.

Ketcham, J.D., C. Lucabell, and C.A. Powers. 2015a. Paying attention or paying too much in medicare Part D. *American Economic Review* 105 (1): 204–233.

Ketcham, J.D., N.V. Kuminoff, and C.A. Powers. 2015b. Estimating the heterogeneous welfare effects of choice architecture: an application to the medicare prescription drug insurance market. Working paper, Arizona State University.

Knight, B., and N. Schiff. 2010. Momentum and social learning in presidential primaries. *Journal of Political Economy* 118 (6): 1110–1150.

Koulayev, S. 2013. Search with dirichlet priors: estimation and implications for consumer demand. *Journal of Business and Economic Statistics* 31 (2): 226–239.

Larcom, S., F. Rauch, and T. Willems. 2015. The benefits of forced experimentation: striking evidence from the London underground network. Working paper, University of Oxford.

Lee, J.Y., and D.R. Bell. 2013. Neighborhood social capital and social learning for experience attributes of products. *Marketing Science* 32 (6): 960–976.

Lin, S., J. Zhang, and J.R. Hauser. 2015. Learning from experience, Simply. *Marketing Science* 34 (1): 1–19.

Liu, Y., and M. Ishihara. 2015. Causal impact of online reviews on durable goods adoption: the case of the U.S. Video-Game market. Working paper, Stern School of Business, NYU.

Liu, X., J.H. Kagel, and L.F. Lee. 2012. Learning from peers in signaling game experiments. *Journal of Applied Econometrics* 27: 1037–1058.

Lovett, M.J., and R. Staelin. 2016. The role of paid, earned, and owned media in building entertainment brands: reminding, informing, and enhancing enjoyment. *Marketing Science* 35 (1): 142–157.

Malmendier, U., and S. Nagel. 2011. Depression babies: do microeconomic experiences affect risk-taking? *The Quarterly Journal of Economics* 126 (1): 373–416.

Milgrom, P., and J. Roberts. 1982. Limit pricing and entry under incomplete information: an equilibrium analysis. *Econometrica* 50 (2): 443–459.

Newberry, P.W. 2016. An empirical study of observational learning. *RAND Journal of Economics* 47 (2): 394–432.

Roberts, J., and G. Urban. 1988. Modelling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice. *Management Science* 34 (2): 167–185.

Roos, J.M.T., C.F. Mela, and R. Shachar. 2015. The effect of links and excerts on internet news consumption. Working paper, Rotterdam School of Management, Erasmus University.

Rothschild, M. 1974. Searching for lowest price when the distributino of prices is unknown. *Journal of Political Economy* 82 (4): 689–711.

Sauer, Robert M. 2015. Does it pay for women to volunteer? *International Economic Review* 56 (2): 537–564.

Sudhir, K., and N. Yang. 2015. Exploiting the choice-consumption mismatch: A new approach to disentagle state dependence and heterogenity. Working paper.

Szymanowski, M., and E. Gijsbrechts. 2012. Consumption-based cross-brand learning: are private labels really private? *Journal of Marketing Research.* XLIX (April), 231–246.

Tehrani, Shervin S., and A.T. Ching. 2016. A heuristic approach to explore: value of perfect information. Working paper, Rotman School of Management, University of Toronto.

Wei, Y. 2015. Imitation versus innovation: product similarity network in the motion picture industry. Working paper, University of Pennsylvania.

Whittle, P. 1988. Restless bandits: activity allocation in a changing world. *Journal of Applied Probability* 25 (2): 287–298.

Wu, C., H. Che, T.Y. Chan, and X. Lu. 2015. The economic value of online reviews. *Marketing Science* 34 (5): 739–754.

Yang, B., and A.T. Ching. 2014. Dynamics of consumer adoption of financial innovation: the case of ATM cards. *Management Science* 60 (4): 903–922.

Yang, N. 2016. Learning in retail entry. Working paper, McGill University.

Zhang, J. 2010. The sound of silence: observational learning in the U.S. kidney market. *Marketing Science* 29 (2): 315–335.

Zhao, Y., S. Yang, V. Narayan, and Y. Zhao. 2013. Modeling consumer learning from online product reviews. *Marketing Science* 32 (1): 153–169.

Zou, D. 2014. Intertemporal pricing of new products: incentivizing consumer learning and inertia. Working paper, Booth School of Business, University of Chicago.

Chapter 9
Measurement Models for Marketing Constructs

Hans Baumgartner and Bert Weijters

9.1 Introduction

Researchers who seek to understand marketing phenomena frequently need to measure the phenomena studied. However, for a variety of reasons, measuring marketing constructs is generally not a straightforward task and often sophisticated measurement models are needed to fully capture relevant marketing constructs. For instance, consider brand love, a construct that has become a common theme in advertising practice and academic marketing research, but which has proven hard to measure adequately. In an effort to assess brand love, Batra et al. (2012) develop a measurement model that comprises no fewer than seven dimensions of the construct (passion-driven behaviors, self-brand integration, positive emotional connection, long-term relationship, anticipated separation distress, overall attitude valence, and attitude strength in terms of certainty/confidence). The hierarchical model they propose can assist marketing executives in showing how to influence a consumer's feeling of brand love by targeting the lower-level, concrete subcomponents through product and service design and/or marketing communications (e.g., by providing trusted expert advice on a website a company can leverage a feeling of anticipated separation distress).

But even for marketing constructs that seem more concrete and for which well-established measures are readily available, researchers face important challenges in terms of measurement modeling. For instance, validly measuring satisfaction is often more challenging than it may seem. Consider a researcher who is interested in consumers' satisfaction with a firm's offering and the determinants of

H. Baumgartner (✉)
Smeal College of Business, Penn State University, University Park, USA
e-mail: HansBaumgartner@psu.edu

B. Weijters
Ghent University, Ghent, Belgium

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_9

their satisfaction (e.g., both proximal determinants such as their satisfaction with particular aspects of the product and more distal determinants such as prior expectations). It is well-known that the responses provided by consumers may not reflect their "true" satisfaction with the product or other product-related characteristics because of various extraneous influences, both random and systematic, related to the respondent (e.g., acquiescence, social desirability), the survey instrument (e.g., wording of the items, response format), and situational factors (e.g., distractions present in the survey setting). Furthermore, it may not be valid to assume that the responses provided by consumers can be treated at face value and used as interval scales in analyses that require such an assumption. Because of these problems, both in the way respondents provide their ratings and in how researchers treat these ratings, comparisons across individuals or groups of individuals may be compromised. As a case in point, Rossi et al. (2001) demonstrate that a model in which scale usage differences and the discrete nature of the rating scale are taken into account explicitly leads to very different findings about the relationship between overall satisfaction and various dimensions of product performance compared to a model in which no corrections are applied to the data. In another study using satisfaction survey data, Andreassen et al. (2006) illustrate how alternative estimation methods to account for non-normality may lead to different results in terms of model fit, model selection, and parameter estimates and, as a consequence, managerial priorities in the marketing domain.

Measurement is the process of quantifying a stimulus (the object of measurement) on some dimension of judgment (the attribute to be measured). Often, a measurement task involves a rater assigning a position on a rating scale to an object based on the object's perceived standing on the attribute of interest. In a marketing context, objects of measurement may be individuals (consumers, salespeople, etc.), firms, or advertisements, to mention just a few examples, which are rated on various individual differences, firm characteristics, and other properties of interest. For example, raters may be asked to assess the service quality of a particular firm (specifically, the reliability dimension of service quality) by indicating their agreement or disagreement with the following item: "XYZ [as a provider of a certain type of service] is dependable" (Parasuraman et al. 1988). Although the use of raters is common, the quantification could also be based on secondary data and other sources.

Three important concepts in the measurement process have to be distinguished (Groves et al. 2004): construct, measure, and response. A *construct* is the conceptual entity that the researcher is interested in. Before empirical measurements can be collected, it is necessary that the construct in question be defined carefully. This requires an explication of the essential meaning of the construct and its differentiation from related constructs. In particular, the researcher has to specify the domain of the construct in terms of both the attributes (properties, characteristics) of the intended conceptual entity and the objects to which these attributes extend (MacKenzie et al. 2011; Rossiter 2002). For example, if the construct is service quality, the attribute is ‘quality’ (or specific aspects of quality such as reliability) and the object is ‘firm XYZ’s service’.

Both the object and its attributes can vary in how concrete or abstract they are and, generally, measurement becomes more difficult as the abstractness of the object and/or attributes increases (both for the researcher trying to construct appropriate measures of the construct of interest and the respondent completing a measurement exercise). In particular, more abstract constructs generally require multiple measures.

One important question to be answered during the construct specification task is whether the object and/or the attributes of the object should be conceptualized as uni- or multidimensional. This question is distinct from whether the items that are used to empirically measure a construct are uni- or multidimensional. If an object is complex because it is an aggregate of sub-objects or an abstraction of more basic ideas, or if the meaning of the attribute is differentiated and not uniformly comprehended, it may be preferable to conceptualize the construct as multidimensional. For multidimensional objects and attributes, the sub-objects and sub-attributes have to be specified and measured separately. For example, if firm XYZ has several divisions or provides different types of services so that it is difficult for respondents to integrate these sub-objects into an overall judgment, it may be preferable to assess reactions to each sub-object separately. Similarly, since the quality of a service is not easily assessed in an overall sense (and an overall rating may lack diagnosticity at any rate), service quality has been conceptualized in terms of five distinct dimensions (tangibles, reliability, responsiveness, assurance, and empathy) (Parasuraman et al. 1988).

Once the construct has been defined, *measures* of the construct have to be developed. Under special circumstances, a single measure of a construct may be sufficient. Specifically, Bergkvist and Rossiter (2007) and Rossiter (2002) argue (and present some supporting evidence) that a construct can be measured with a single item if “in the minds of raters … (1) the object of the construct is “concrete singular,” meaning that it consists of one object that is easily and uniformly imagined, and (2) the attribute of the construct is “concrete,” again meaning that it is easily and uniformly imagined” (Bergkvist and Rossiter 2007, p. 176). However, since either the object or the attribute (or both) tend to be sufficiently abstract, multiple measures are usually required to adequately capture a construct.

An important consideration when developing measures and specifying a measurement model is whether the measures are best thought of as manifestations of the underlying construct or defining characteristics of it (MacKenzie et al. 2005). In the former case, the indicators are specified as effects of the construct (so-called reflective indicators), whereas in the latter case they are hypothesized as causes of the construct (formative indicators). Mackenzie et al. (2005) provide four criteria that can be used to decide whether particular items are reflective or formative measures of a construct. If an indicator is (a) a manifestation (rather than a defining characteristic) of the underlying construct, (b) conceptually interchangeable with the other indicators of the construct, (c) expected to covary with the other indicators, and (d) hypothesized to have the same antecedents and consequences as the other indicators, then the indicator is best thought of as reflective. Otherwise, the indicator is formative.

Based on the measures chosen to represent the intended conceptual entity, observed *responses* of the hypothesized construct can be obtained. For "constructs" for which both the object and the attribute are relatively concrete (e.g., a person's chronological age, a firm's advertising spending), few questions about the reliability and validity of measurement may be raised. However, as constructs become more abstract, reliability and validity assessments become more important. Depending on how one specifies the relationship between indicators and the underlying construct (i.e., reflective vs. formative), different procedures for assessing reliability and validity have to be used (MacKenzie et al. 2011).

Constructing reliable and valid measures of constructs is a nontrivial task involving issues related to construct definition and development of items that fully capture the intended construct. Since several elaborate discussions of construct measurement and scale development have appeared in the recent literature (MacKenzie et al. 2011; Rossiter 2002), we will not discuss these topics in the present chapter. Instead, we will focus on models that can be used to assess the quality of measurement for responses that are already available. We will start with a discussion of the congeneric measurement model in which continuous observed indicators are seen as reflections of an underlying latent variable, each observed variable loads on a single latent variable (provided multiple latent variables are included in the model), and no correlations among the unique factors (measurement errors) are allowed. We will also contrast the congeneric measurement model with a formative measurement model, consider measurement models that incorporate a mean structure (in addition to a covariance structure), and present an extension of the single-group model to multiple groups.

We will then discuss three limitations of the congeneric model. First, it may be unrealistic to assume that each item loads on a single latent variable and that the loadings on non-target factors are zero (provided the measurement model contains multiple latent variables). Second, often the observed variables are not only correlated because they load on the same factor or because the factors on which they load are correlated. There may be other sources of covariation (due to various "method" factors) that require the specification of correlations among the unique factors or the introduction of method factors. Third and finally, although the assumption of continuous, normally distributed indicators, which is probably never strictly satisfied, may often be adequate, sometimes it is so grossly violated that alternative models have to be entertained. Below we will discuss the three limitations in greater detail and consider ways of overcoming these shortcomings. Throughout the chapter, illustrative examples of the various models are presented to help the reader follow the discussion more easily.

9.2 The Congeneric Measurement Model

9.2.1 *Conceptual Development*

The so-called congeneric measurement model is a confirmatory factor model in which I observed or manifest variables x_i (also called indicators), contained in an $I \times 1$ vector $\mathbf{x}$, are a function (i.e., reflections of) J latent variables (or common factors) ξ_j (included in a $J \times 1$ vector $\boldsymbol{\xi}$) and I unique factors δ_i (summarized in an $I \times 1$ vector $\boldsymbol{\delta}$). The strength of the relationship between the x_i and ξ_j is expressed by an $I \times J$ matrix of factor loadings $\boldsymbol{\Lambda}$ with typical elements λ_{ij}. In matrix form, the model can be written as follows:

$$\mathbf{x} = \boldsymbol{\Lambda\xi} + \boldsymbol{\delta} \tag{9.1}$$

For now, we assume that $\mathbf{x}$ and $\boldsymbol{\xi}$ are in deviation form (i.e., mean-centered), although this assumption will be relaxed later. Assuming that $E(\boldsymbol{\delta}) = \mathbf{0}$ and $Cov(\boldsymbol{\xi}, \boldsymbol{\delta}') = \mathbf{0}$, this specification of the model implies the following structure for the variance-covariance matrix of $\mathbf{x}$, which is called $\boldsymbol{\Sigma}$:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}(\boldsymbol{\Lambda}, \boldsymbol{\Phi}, \boldsymbol{\Theta}) = \boldsymbol{\Lambda\Phi\Lambda}' + \boldsymbol{\Theta} \tag{9.2}$$

where $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$ are the variance-covariance matrices of $\boldsymbol{\xi}$ and $\boldsymbol{\delta}$, respectively (with typical elements φ_{ij} and θ_{ij}), and the symbol $'$ is the transpose operator. In a congeneric measurement model, each observed variable is hypothesized to load on a single factor (i.e., $\boldsymbol{\Lambda}$ contains only one nonzero entry per row) and the unique factors are uncorrelated (i.e., $\boldsymbol{\Theta}$ is diagonal). For identification, either one loading per factor has to be fixed at one, or the factor variances have to be standardized to one. If there are at least three indicators per factor, a congeneric factor model is identified, even if there is only a single factor and regardless of whether multiple factors are correlated or uncorrelated (orthogonal). If there are only two indicators per factor, a single-factor model is not identified (unless additional restrictions are imposed), and multiple factors have to be correlated for the model to be identified. If there is only a single indicator per factor, the associated unique factor variance cannot be freely estimated (i.e., has to be set to zero or another assumed value). A graphical representation of a specific congeneric measurement model with 6 observed measures and 2 factors is shown in Fig. 9.1.

Factor models are usually estimated based on maximum likelihood (which assumes multivariate normality of the observed variables and requires a relatively large sample size), although other estimation procedures are available. To evaluate the fit of the overall model, one can use a likelihood ratio test in which the fit of the specified model is compared to the fit of a model with perfect fit. A nonsignificant χ^2 value indicates that the specified model is acceptable, but often the hypothesized

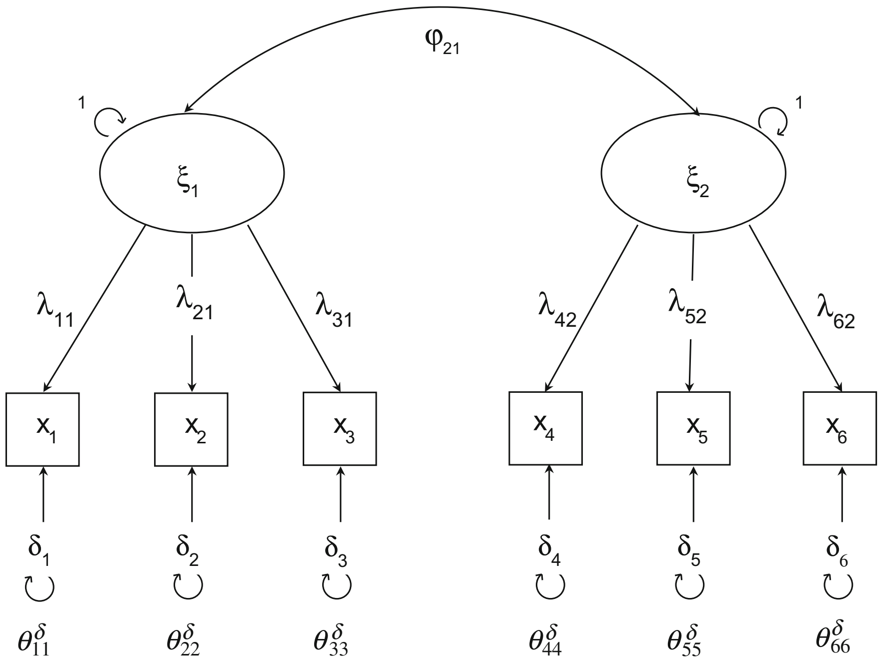


Fig. 9.1 A congeneric measurement model. Note for Fig. 9.1: In the illustrative example of Sect. 2.2, x_1–x_3 are regularly worded environmental concern items; x_4–x_6 are regularly worded health concern items; ξ_1 refers to environmental concern, and ξ_2 refers to health concern (see Table 9.2 for the items)

model is found to be inconsistent with the data. Since models are usually not meant to be literally true, and since at relatively large sample sizes the χ^2 test will be powerful enough to detect even relatively minor misspecifications, researchers frequently use alternative fit indices to evaluate whether the fit of the model is "good enough" from a practical perspective. Among the more established alternative fit indices are the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), the comparative fit index (CFI), and the Tucker-Lewis fit index (TLI). For certain purposes, information theory-based fit indices such as BIC may also be useful. Definitions of these fit indices, brief explanations, and commonly used cutoff values are provided in Table 9.1. In our experience, researchers are often too quick to dismiss a significant χ^2 value based on the presumed limitations of this test (i.e., a significant χ^2 test does show that there are problems with the specified model and the researcher should investigate potential sources of this lack of fit), but it is possible that relatively minor misspecifications lead to a significant χ^2 value, in which case a reliance on satisfactory alternative fit indices may be justified.

If a model is deemed to be seriously inconsistent with the data, it has to be respecified. This is usually done with the help of modification indices, although

Table 9.1 A summary of commonly used overall fit indices

Index	Definition of the index	Interpretation and use of the index
Minimum fit function chi-square (χ^2)	*(N–1) f*	Tests the hypothesis that the specified model fits perfectly (within the limits of sampling error); the obtained χ^2 value should be smaller than χ^2_{crit}; note that the minimum fit function χ^2 is only one possible chi-square statistic and that different discrepancy functions will yield different χ^2 values
Root mean square error of approximation (RMSEA)	$\sqrt{\frac{(\chi^2 - df)}{(N-1)df}}$	Estimates how well the fitted model approximates the population covariance matrix per *df*; Browne and Cudeck (1992) suggest that a value of 0.05 indicates a close fit and that values up to 0.08 are reasonable; Hu and Bentler (1999) recommend a cutoff value of 0.06; a *p*-value for testing the hypothesis that the discrepancy is smaller than 0.05 may be calculated (so-called test of close fit)
(Standardized) Root mean squared residual (S)RMR)	$\sqrt{\frac{2(s_{ij} - \hat{\sigma}_{ij})^2}{(p)(p+1)}}$	Measures the average size of residuals between the fitted and sample covariance matrices; if a correlation matrix is analyzed, RMR is "standardized" to fall within the [0, 1] interval (SRMR), otherwise it is only bounded from below; a cutoff of 0.05 is often used for SRMR; Hu and Bentler (1999) recommend a cutoff value close to 0.08
Bayesian information criterion (BIC)	$[\chi^2 + r \ln N]$ or $[\chi^2 - df \ln N]$	Based on statistical information theory and used for testing competing (possibly non-nested) models; the model with the smallest BIC is selected
Comparative Fit Index (CFI)	$1 - \frac{\max(\chi^2_t - df_t, 0)}{\max(\chi^2_n - df_n, \chi^2_t - df_t, 0)}$	Measures the proportionate improvement in fit (defined in terms of noncentrality, i.e., $\chi^2 - df$) as one moves from the baseline to the target model; originally, values greater than 0.90 were deemed acceptable, but Hu and Bentler (1999) recommend a cutoff value of 0.95
Tucker-Lewis nonnormed fit index (TLI, NNFI)	$\frac{\frac{\chi^2_n - df_n}{df_n} - \frac{\chi^2_t - df_t}{df_t}}{\frac{\chi^2_n - df_n}{df_n}}$	Measures the proportionate improvement in fit (defined in terms of noncentrality) as one moves from the baseline to the target model, per *df*; originally, values greater than 0.90 were deemed acceptable, but Hu and Bentler (1999) recommend a cutoff value of 0.95

Notes N = sample size; f = minimum of the fitting function; df = degrees of freedom; p = number of observed variables; r = number of estimated parameters; χ^2_{crit} = critical value of the χ^2 distribution with the appropriate number of degrees of freedom and for a given significance level; the subscripts n and t refer to the null (or baseline) and target models, respectively. The baseline model is usually the model of complete independence of all observed variables other tools are also available (such as an analysis of the residuals between the observed and model-implied covariance matrices). A modification index is the predicted decrease in the χ^2 statistic if a fixed parameter is freely estimated or an equality constraint is relaxed. For example, a significant modification index for a factor loading that is restricted to be zero suggests that the indicator in question may have a non-negligible loading on a non-target factor, or maybe that the observed variable was incorrectly assumed to be an indicator of a certain construct (particularly when the loading on the presumed target factor is small). Associated with each modification index is an expected parameter change (EPC), which shows the predicted estimate of the parameter when the parameter is freely estimated. Although models can be brought into closer correspondence with the data by repeatedly freeing parameters based on significant modification indices, there is no guarantee that the respecified model will be closer to "reality", or hold up well with new data (MacCallum 1986).

Once a researcher is satisfied that the (respecified) model is reasonably consistent with the data, a detailed investigation of the local fit of the model can be conducted. From a measurement perspective, three issues are of paramount importance. First, the items hypothesized to measure a given construct have to be substantially related to the underlying construct, both individually and collectively. If one assumes that the observed variance in a measure consists of only two sources, substantive variance (variance due to the underlying construct) and random error variance, then convergent validity is similar to reliability (some authors argue that reliability refers to the convergence of measures based on the same method, whereas different methods are necessary for convergent validity); henceforth, we will therefore use the term reliability to refer to the relationship between measures and constructs (even though the unique factor variance usually does not only contain random error variance). Individually, an item should load significantly on its target factor, and each item's observed variance should contain a substantial amount of substantive variance. One index, called individual-item reliability (IIR) or individual-item convergent validity (IICV), is defined as the squared correlation between a measure x_i and its underlying construct ξ_j (i.e., the proportion of the total variance in x_i that is substantive variance), which can be computed as follows:

$$IIR_{x_i} = \frac{\lambda_{ij}^2 \varphi_{jj}}{\lambda_{ij}^2 \varphi_{jj} + \theta_{ii}} \tag{9.3}$$

Ideally, at least half of the total variance should be substantive variance (i.e., IIR $\geq$ 0.5), but this is often not the case. One can also summarize the reliability of all indicators of a given construct by computing the average of the individual-item reliabilities. This is usually called average variance extracted (AVE), that is,

$$AVE = \frac{\sum IIR_{x_i}}{K} \tag{9.4}$$

where K is the number of indicators for the construct in question. A common rule of thumb is that AVE should be at least 0.5.

Collectively, all measures of a given construct combined should be strongly related to the underlying construct. One common index is composite reliability (CR), which is defined as the squared correlation between an unweighted sum (or average) of the measures of a construct and the construct itself. CR is a generalization of coefficient alpha to a situation in which items can have different loadings on the underlying factor and it can be computed as follows:

$$CR_{\sum x_i} = \frac{\left(\sum \lambda_{ij}\right)^2 \varphi_{jj}}{\left(\sum \lambda_{ij}\right)^2 \varphi_{jj} + \sum \theta_{ii}} \tag{9.5}$$

CR should be at least 0.7 and preferably higher.

Second, items should be primarily related to their underlying construct and not to other constructs. In a congeneric model, loadings on non-target factors are set to zero a priori, but the researcher has to evaluate that this assumption is justified by looking at the relevant modification indices and expected parameter changes. This criterion can be thought of as an assessment of discriminant validity at the item level.

Third, the constructs themselves should not be too highly correlated if they are to be distinct. This is called discriminant validity at the construct level. One way to test discriminant validity is to construct a confidence interval around each construct correlation in the Φ matrix (the covariances are correlations if the variances on the diagonal have been standardized to one) and to check whether the confidence interval includes one (in which case a perfect correlation cannot be dismissed). However, this is a weak criterion of discriminant validity because with a large sample and precise estimates of the factor correlations, the factor correlations will usually be distinct from one, even if the correlations are quite high. A stronger test of discriminant validity is the criterion proposed by Fornell and Larcker (1981). This criterion says that each squared factor correlation should be smaller than the AVE for the two constructs involved in the correlation. Intuitively, this rule means that a construct should be more strongly related to its own indicators than to another construct from which it is supposedly distinct.

It is easy to test alternative assumptions about how the measures of a given latent variable relate to the underlying latent construct. In the congeneric model, the observed measures of a construct have a single latent variable in common, but both the factor loadings and unique factor variances are allowed to differ across the indicators. In an essentially tau-equivalent measurement model, all the factor loadings for a given construct are specified to be the same (Traub 1994). This means that the scale metrics of the observed variables are identical. In a parallel measurement model, the unique factor variances are also specified to be the same. This means that the observed variables are fully exchangeable. A χ^2 difference test can be used to test the relative fit of alternative models. If, say, the model positing equality of factor loadings does not show a significant deterioration in fit relative to a model in which the factor loadings are freely estimated, the hypothesis of tau-equivalence is consistent with the empirical evidence.

In measurement analyses the focus is generally on the interrelationships of the observed variables. However, it is possible to incorporate the means into the model. When data are available for a single group only, little additional information is gained by estimating means. If the intercepts of all the observed variables are freely estimated, the means of the latent constructs have to be restricted to zero in order to identify the model and the estimated intercepts are simply the observed means. Alternatively, if the latent means are to be estimated, one intercept per factor has to be restricted to zero. If this is done for the indicator whose loading on the underlying factor is set to one (the so-called marker variable or reference indicator), the latent factor mean is simply the observed mean of the reference indicator.

One can also test more specific hypotheses about the means. For example, if it is hypothesized that the observed measures of a given latent construct all have the same relationship to the underlying latent variable, one could test whether the measurement intercepts are all the same, which implies that the means of the construct indicators are identical. Of course, one can also compare the means of observed variables across constructs, although this is not very meaningful unless the scale metrics are comparable.

9.2.2 Empirical Example

Congeneric measurement models are very common in marketing research. For example, Walsh and Beatty (2007) identify dimensions of customer-based corporate reputation and develop scales to measure these dimensions (Customer Orientation, Good Employer, Reliable and Financially Strong Company, Product and Service Quality, and Social and Environmental Responsibility). An important advantage of confirmatory factor analysis is that hierarchical factor models can be specified, where a second-order factor has first-order factors as its indicators (more than two levels are possible, but two levels are most common). For instance, Yi and Gong (2013) develop and validate a scale for customer value co-creation behavior. The scale comprises two second-order factors (customer participation behavior and customer citizenship behavior), each of which consists of four first-order factors:

Table 9.2 Empirical illustration—a congeneric factor model for environmental and health concern

			Standardized loading	IIR	AVE	CR
Environmental concern	item 1	I would describe myself as environmentally conscious	0.762	0.581	0.671	0.860
	item 2	I take into account the environmental impact in many of my decisions	0.862	0.743		
	item 3	My buying habits are influenced by my environmental concern	0.830	0.689		
Health concern	item 1	I consider myself as health conscious	0.842	0.709	0.575	0.796
	item 2	I think that I take health into account a lot in my life	0.824	0.679		
	item 3	My health is so valuable to me that I am prepared to sacrifice many things for it	0.581	0.338		

Note IIR—individual-item reliability; AVE—average variance extracted; CR—composite reliability

information seeking, information sharing, responsible behavior, and personal interaction for customer participation behavior; and feedback, advocacy, helping, and tolerance for customer citizenship behavior.

To illustrate the concepts discussed so far, we present an empirical example using data from N = 740 Belgian consumers, all with primary responsibility for purchases in their household, who responded to a health consciousness scale and an environmental concern scale. The scale items were adapted from Chen (2009) and translated into Dutch. For now, we will only use the first three items from each scale (see Table 9.2). Later, we will analyze the complete scales, including the reversed items. In Mplus 7.3, we specify a congeneric two-factor model where each factor has three reflective indicators. The χ^2 test is significant (χ^2 (8) = 33.234, p = 0.0001), but the indices of local misfit do not indicate a particular misspecification (the modification indices point toward some negligible cross-loadings and residual correlations, although the modification indices are all smaller than 10). Based on the alternative fit indices, the model shows acceptable fit to the data: RMSEA = 0.065 (with a 90% confidence interval [CI] ranging from 0.043 to 0.089), SRMR = 0.031, CFI = 0.986, and TLI = 0.974. Table 9.2 reports the IIRs for all items, and the AVE and CR for both factors. One of the health concern items shows an unsatisfactory IIR (below 0.50), probably because it is worded more extremely and somewhat more verbosely than the other two health concern items. Nevertheless, the AVE for both factors is above 0.50 and the CR for both factors is above 0.70, which indicates acceptable reliability. The correlation between health concern and environmental concern is 0.38 (95% CI from 0.31 to 0.46). Since the shared variance of 0.15 is smaller than the AVE of either construct, the factors show discriminant validity.

9.3 Multi-sample Congeneric Measurement Models with Mean Structures

9.3.1 Conceptual Development

One advantage of using structural equation modeling techniques for measurement analysis is that they enable sophisticated assessments of the measurement properties of scales across different populations of respondents. This is particularly useful in a cross-cultural context, where researchers are often interested in either assessing the invariance of findings across countries or establishing nomological differences between countries. If cross-cultural comparisons are to be meaningful, it is first necessary to ascertain that the constructs and measures are comparable.

A congeneric measurement model containing a mean structure for group g can be specified as follows:

$$\boldsymbol{x}^g = \boldsymbol{\tau}^g + \boldsymbol{\Lambda}^g \boldsymbol{\xi}^g + \boldsymbol{\delta}^g \tag{9.6}$$

where $\boldsymbol{\tau}$ is an I $\times$ 1 vector of equation intercepts, the other terms were defined earlier, and the superscript g refers to group g. Under the assumptions mentioned earlier (although $\mathbf{x}$ and $\boldsymbol{\xi}$ are not mean-centered in the present case), the corresponding mean and covariance structures are:

$$\boldsymbol{\mu}^g = \boldsymbol{\tau}^g + \boldsymbol{\Lambda}^g \boldsymbol{\kappa}^g \tag{9.7}$$

$$\boldsymbol{\Sigma}^g = \boldsymbol{\Lambda}^g \boldsymbol{\Phi}^g \boldsymbol{\Lambda}'^g + \boldsymbol{\Theta}^g \tag{9.8}$$

where $\boldsymbol{\mu}$ is the expected value of $\mathbf{x}$ and $\boldsymbol{\kappa}$ is the expected value of $\boldsymbol{\xi}$ (i.e., the vector of latent means of the constructs). To identify the covariance part, one loading per factor should be set to one (the corresponding indicator is called the marker variable or reference indicator); the factor variances should not be standardized at one since this would impose the assumption that the factor variances are equal across groups, which is not required and which need not be the case. To identify the means part, the intercepts of the marker variables have to be set to zero, in which case all the latent means can be freely estimated, or one latent mean (the latent mean of the reference group) has to be set to zero and the intercepts of the marker variables are specified to be invariant across groups. In the latter case, the latent means in the remaining groups express the difference in latent means between the reference group and the other groups.

In the model of Eqs. (9.7) and (9.8), five different types of parameters can be tested for invariance. Two of these are of substantive interest ($\boldsymbol{\kappa}^g$, $\boldsymbol{\Phi}^g$); the remaining ones are measurement parameters ($\boldsymbol{\tau}^g$, $\boldsymbol{\Lambda}^g$, $\boldsymbol{\Theta}^g$). In order for the comparisons of substantive interest to be meaningful, certain conditions of measurement invariance have to be satisfied. To begin with, the same congeneric factor model has to hold in each of the g groups; this is called configural invariance, and it is a minimum condition of comparability (e.g., if a construct is unidimensional in one group and multi-dimensional in another, meaningful comparisons are difficult if not impossible). Usually, more specific comparisons are of interest and in this case more stringent forms of invariance have to be satisfied. Specifically, Steenkamp and Baumgartner (1998) show that if relationships between constructs are to be compared across groups, metric invariance (equality of factor loadings) has to hold, and if latent construct means are to be compared across groups, scalar invariance (invariance of measurement intercepts) has to hold as well. For example, consider the relationship between the mean of variable x_i and the mean on the underlying construct ξ_j, that is, $\mu_i^g = \tau_i^g + \lambda_{ij}^g \kappa_j^g$. The goal is to compare κ_j^g across groups based on x_i^g. Unfortunately, the comparison on μ_i^g depends on τ_i^g, λ_{ij}^g, and κ_j^g. Inferences about the latent means will only be unambiguous if τ_i^g and λ_{ij}^g are the same across groups. For certain purposes, one may also want to test for the invariance of unique factor variances across groups, but usually this comparison is less relevant.

The hypothesis of full metric invariance can be tested by comparing the model with invariant loadings across groups to the model in which the loadings are freely estimated in each group. If the deterioration in fit is nonsignificant, metric invariances is satisfied. Similarly, the hypothesis of full scalar invariance can be tested by comparing the model with invariant loadings and intercepts to the model with invariant loadings. Metric invariance should be established before scalar invariance is assessed.

In practice, full metric and scalar invariance are frequently violated (esp. the latter). The question then arises whether partial measurement invariance is sufficient to conduct meaningful across-group comparisons. Note that in the specification of the model in Eqs. (9.7) and (9.8), one variable per factor was already assumed to have invariant loadings and intercepts (because one loading per factor was set to one and the corresponding intercept was fixed at zero). However, these restrictions are necessary to identify the model and do not impose binding constraints on the model (this can be seen by the fact that regardless of which variable is chosen as the marker variable, the fit of the model will always be the same). In order to be able to test (partial) metric or scalar invariance, at least two items per factor have to have invariant loadings or intercepts (see Steenkamp and Baumgartner 1998). This is a minimum requirement; ideally, more indicators per factor will display metric and scalar invariance. Asparouhov and Muthén (2014) have recently proposed a new procedure called the alignment method, in which these strict requirements of measurement invariance are relaxed, but their method is beyond the scope of this chapter.

Although the tests of (partial) metric and scalar invariance described previously are essential, one word of caution is necessary. These tests assume that any biases in $\boldsymbol{\tau}$ and $\boldsymbol{\Lambda}$ that may distort comparisons across groups are nonuniform across items. If the bias is the same across items (e.g., $\boldsymbol{\tau}$ is biased upward or downward by the same amount across items), the researcher may wrongly conclude that measurement invariance is satisfied and mistakenly attribute a difference in intercepts to a difference in latent means (Little 2000).

9.3.2 Empirical Example

Multi-sample congeneric measurement models (with or without mean structures) are commonly applied in cross-national marketing research. Some examples are the following. Strizhakova et al. (2008) compare branded product meanings (quality, values, personal identity, and traditions) across four countries based on newly developed measures for which they demonstrate cross-national measurement invariance. Their results show that identity-related and traditions-related meanings are more important in the U.S. than in three emerging markets (Romania, Ukraine, and Russia). Singh et al. (2007) test models involving moral philosophies, moral intensity, and ethical decision making across two samples of marketing practitioners from the United States and China. Their measurement models show partial metric and scalar invariance. In a similar way, using multi-group confirmatory factor analysis, Schertzer et al. (2008) establish configural, metric and partial scalar invariance for a gender identity scale across samples from the U.S., Mexico, and Norway.

To further illustrate measurement invariance testing, we analyze data from an online panel of respondents in two countries, Slovakia (N = 1063) and Romania (N = 970), using four bipolar items to measure attitude toward the brand Coca-Cola. The four items (unpleasant-pleasant, negative-positive, unattractive-attractive, and low quality-high quality) were translated (and back-translated) by professional translators into respondents' native languages. Respondents provided their ratings on seven-point scales. The samples from both countries were comparable in social demographic makeup for reasons of comparability.

The four items are modeled as reflective indicators of one latent factor, using a two-group congeneric model with a mean structure in Mplus 7.3. We test a sequence of nested models, gradually imposing constraints that reflect configural, metric, and scalar invariance. The fit indices are reported in Table 9.3. In the configural model (Model A), the same congeneric model is estimated in the two groups (this is the so-called configural model), but the loadings and intercepts are estimated freely in each group. The exception is the marker item, which is specified to have a loading of one and an intercept of zero in both groups. The model shows acceptable fit to the data. With sample sizes around 1000, the χ^2 test is sensitive to even minor misspecifications, and the modification indices do not indicate a specific

Table 9.3 Model fit indices for measurement invariance tests of brand attitude in two countries

	Absolute fit			Difference test				Alternative fit indices				
	χ^2	df	*p*	Reference model	$\Delta\chi^2$	df	*p*	RMSEA (90% CI)	CFI	TLI	SRMR	BIC
A. Configural invariance	31.93	4	<0.001					0.083 (0.058, 0.111)	0.995	0.986	0.009	26760.7
B. Metric invariance	34.91	7	<0.001	A	2.99	3	0.394	0.063 (0.043, 0.084)	0.995	0.992	0.018	26740.8
C. Scalar invariance	52.76	10	<0.001	B	17.85	3	0.001	0.065 (0.048, 0.083)	0.993	0.991	0.014	26735.8
D. Partial scalar invariance	37.33	9	<0.001	B	2.42	2	0.299	0.056 (0.038, 0.075)	0.995	0.994	0.018	26728.0

Note See Table 9.1 for an explanation of these fit indices misspecification that is serious. The RMSEA will evolve toward more acceptable levels when more constraints are added (the reason being that this fit index imposes a substantial penalty for the number of freely estimated parameters).

Model B specifies metric invariance by restricting the factor loadings of all items (not only the marker item) to equality across the two groups (since there are three non-marker items, metric invariance is tested based on a χ^2 difference test with three degrees of freedom). In support of metric invariance, the χ^2 difference test is nonsignificant. Moreover, the alternative fit indices show acceptable fit and the RMSEA and BIC values (which impose a penalty for estimating many free parameters) even show a clear improvement in fit.

Model C imposes scalar invariance by additionally restricting all item intercepts to be equal across groups (scalar invariance is tested with a χ^2 difference with three degrees of freedom as well). The evidence for scalar invariance is somewhat mixed. The BIC improves, while the RMSEA and the other alternative fit indices remain almost stable. However, the χ^2 difference test is statistically significant ($p < 0.001$), so the hypothesis of scalar invariance is rejected. Closer inspection of the modification indices (focusing on MI's > 10, given the large sample size) indicates that the intercept for item 4 is non-invariant. We therefore estimate an additional model, model D, to test for partial scalar invariance, in which scalar invariance for item 4 is relaxed (i.e., the intercepts for all items except item 4 are set to equality across groups) and test the deterioration in fit relative to the metric invariance model (model D is nested in model B). The χ^2 difference test is statistically nonsignificant, in support of partial scalar invariance. Since there are still three items that are scalar invariant, the latent means can now be compared across the two groups. The means are 5.165 (Standard error = 0.047) for the Slovakian sample and 5.207 (standard error = 0.061) for the Romanian sample. A χ^2 difference test for latent mean equality shows that the difference is non-significant ($\Delta\chi^2$ (1) = 0.313, p = 0.576).

9.4 The Formative Measurement Model

9.4.1 Conceptual Development

Sometimes it is not meaningful to assume that an observed measure is a reflection of the operation of an underlying latent variable. For example, assume that job satisfaction is measured with items assessing satisfaction with various aspects of the job, such as satisfaction with one's supervisor, co-workers, pay, etc. Satisfaction with each facet of the job is presumably a contributing factor to overall job satisfaction, not a reflection of it. Jarvis et al. (2003) reviewed the measurement of 1,192 constructs in 178 articles published in four leading Marketing journals and found that 29% of constructs were modeled incorrectly; the vast majority of measurement model misspecifications was due to formative indicators being modeled as reflective (see also MacKenzie et al. 2005). This practice is problematic because simulations have demonstrated that if the measurement model is misspecified, this will bias estimates of structural paths (see Diamantopoulos et al. 2008 for a review of the evidence).

In a formative measurement model, the direction of causality goes from the indicator to the construct, so the observed measures are also called cause indicators (rather than effect indicators). This reversal of causality has several implications. First, error does not reside in the indicators, but in the construct. Since the variance of the construct is a function of the variances and covariances of the formative indicators, plus error, the construct is not a traditional latent variable and should be more accurately referred to as a composite variable (MacCallum and Browne 1993). Second, in general the variance of the error term of a construct that is a function of its indicators is not identified. In order for the model to be identified, directed paths have to go from the formative construct to at least two other variables or constructs. Often, two global reflective indicators are used for this purpose (MacKenzie et al. 2011), but in our experience these reflective indicators are usually not very sophisticated and well-developed measures of the underlying construct. Furthermore, this type of model is empirically indistinguishable from a model in which a reflectively measured construct is related to various antecedents, so the question arises whether the formative measurement model is a measurement model at all. Third, formative measures need not be positively correlated (e.g., dissatisfaction with one's supervisor does not mean that one is also dissatisfied with one's co-workers), so that conventional convergent validity and reliability assessment based on internal consistency is not applicable. Instead, formative measures should have a significant effect on the construct, and collectively the formative measures should account for a large portion of the construct's variance (e.g., at least 50%). Unfortunately, formative measures are frequently quite highly correlated, which leads to multicollinearity problems and the likely non-significance of some of the relationships between the indicators and the construct. This then raises the thorny question of whether an item that may be conceptually important but happens to be empirically superfluous should be retained in the model. An additional difficulty is that because of the correlations among the formative indicators, no firm conclusions about the measurement quality of individual indicators can be drawn. Fourth, formative models assume that the formative measures are error-free contributors to the formative construct, which seems unrealistic. To circumvent this problem, multiple reflective measures can be used to correct for measurement error, which makes the formative measures first-order factors and the formative construct a second-order construct. Unfortunately, such models are quite complex.

In sum, while it is certainly true that formative measures should not be specified as reflective indicators, formative measurement faces a range of formidable difficulties, and several authors have recommended that formative measurement be abandoned altogether (Edwards 2011). Formative measurement is sometimes equated with the partial least squares (PLS) approach, which has also seen increased criticism in recent years (McIntosh et al. 2014), but formative models can be estimated using traditional structural equation modeling techniques, as shown below. Still, the meaningfulness of formative measurement is a topic of active debate and it remains to be seen whether better alternatives can be formulated.

9.4.2 Empirical Example

The marketing research literature offers several recent examples of formative measurement models. Coltman et al. (2008) propose that market orientation viewed from a behavioral perspective (where market orientation is the result of allocating resources to a set of specific activities) can be conceptualized as a formative construct with a reactive and a proactive component. Dagger et al. (2007) develop a multidimensional hierarchical scale for measuring health service quality and validate it in three field studies. In their formative model, nine subdimensions (interaction, relationship, outcome, expertise, atmosphere, tangibles, timeliness, operation, and support) drive four primary dimensions (interpersonal quality, technical quality, environment quality, and administrative quality), which in turn drive service quality perceptions.

To further illustrate the formative measurement model, we will use data from 497 respondents who indicated their attitude toward self-service technologies (SSTs), that is, self-scanning in grocery stores. Specifically, respondents rated the perceived usefulness, perceived ease of use, reliability, perceived fun, and newness of the technology on three items each (e.g., “Self-scanning will allow me to shop faster” for perceived usefulness, and “Self-scanning will be enjoyable” for fun, rated on 5-point agree-disagree scales). The items within each of the five factors were averaged, and the five averages will be used as formative indicators of attitude toward SST. Three overall measures of attitude are also available (i.e., How would you describe your feelings toward using self-scanning technology in this store”, rated on 5-point favorable-unfavorable, I like it-I dislike it, and good-bad scales), which will be used as reflective indicators to identify the model. The model is shown graphically in Fig. 9.2.

The estimated model fit the data well: $\chi^2(10) = 19.693$, p = 0.03; SRMR = 0.01; RMSEA = 0.044 (with a 90% CI ranging from 0.012 to 0.073);

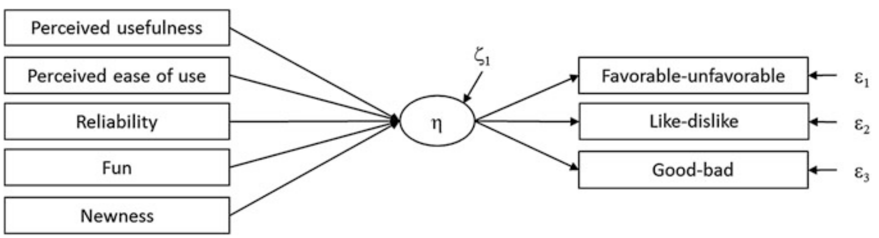


Fig. 9.2 Formative measurement model for attitude towards self-scanning

CFI = 0.994; TLI = 0.990. There were no significant modification indices exceeding 3.84. Perceived usefulness (estimate of 0.294, 95% CI = 0.212 to 0.376), ease of use (0.374, 95% CI = 0.279 to 0.469), reliability (0.128, 95% CI = 0.028 to 0.227) and fun (0.237, 95% CI = 0.163 to 0.310) all had a significant effect on attitude toward SSTs, but the effect of newness (0.047, 95% CI = −0.056 to 0.149) was nonsignificant. The explained variance in the three reflective attitude measures was 0.81, 0.90, and 0.81, respectively, and the five formative measures accounted for 51% of the variance in the formative construct. By conventional standards, these results indicate an acceptable measurement model.

It is possible to take into account measurement error in the five formative measures by specifying a first-order reflective measurement model for them. The resulting model fits the data somewhat less well, but the fit is still adequate: $\chi^2(120) = 260.953$, p = 0.000; SRMR = 0.037; RMSEA = 0.049 (with a 90% CI interval ranging from 0.041 to 0.057); CFI = 0.976; TLI = 0.969. Only perceived usefulness, perceived ease of use, and fun are significant determinants of attitude toward SSTs; reliability is borderline significant (z = 1.87). Together, the five formative measures account for 55% of the variance in the construct. The results for the two models are similar, but taking into account measurement error does change the findings somewhat. The major advantage of the second approach is that a more explicit measurement analysis of the independent variables or formative "indicators" is possible.

9.5 Extension 1: Relaxing the Assumption of Zero Non-target Loadings

9.5.1 Conceptual Development

The congeneric measurement model assumes that the loadings of indicators on factors other than the target factor are zero (which is sometimes referred to as an independent cluster confirmatory factor analysis). This is a strong assumption and even mild violations of the assumption of zero cross-loadings may decrease the overall fit of a model substantially, especially when the sample size is reasonably large. Furthermore, forcing zero cross-loadings when they are in fact non-zero may have other undesirable effects, such as inflated factor correlations and misleading evidence about (lack of) discriminant validity.

One approach to relaxing the assumption of zero cross-loadings is exploratory structural equation modeling (ESEM) (Marsh et al. 2014). ESEM basically replaces the confirmatory factor model used in traditional SEM with an exploratory factor model (or a combination of exploratory and confirmatory factor models), although the exploratory factor analysis is used in a more confirmatory fashion because the researcher usually posits a certain number of factors and expects a certain pattern of factor loadings. In early applications of this approach geomin rotation was used to find a simple structure solution following initial factor extraction, but it is also possible to use target rotation (Marsh et al. 2014). Furthermore, if a good reference indicator for each factor is available (which should have a strong loading on the target factor and zero or near-zero loadings on nontarget factors), the loadings of the reference indicators on the target factors can be fixed at one and the loadings on the non-target factors can be fixed at zero. This makes the solution rotationally determinate because J^2 restrictions are imposed on the factor space (where J is the number of factors). However, it should be kept in mind that a solution that is rotationally unique may not be identified in general (Millsap 2001).

Since the congeneric factor model is nested within the unconstrained factor model estimated within ESEM, it is possible to conduct a χ^2 difference test to compare the two specifications. If the congeneric factor model fits as well as the ESEM model, the more parsimonious model with zero cross-loadings is preferred. However, as the number of indicators and factors increases, it is likely that the ESEM model will fit better. The researcher should also check the similarity of the factor correlations between the two specifications. Particularly if the factor correlations are inflated when cross-loadings are assumed to be zero, the congeneric measurement model is probably not a useful representation of the data.

A second approach to modeling a more flexible factor pattern is based on Bayesian Structural Equation Modeling (BSEM) (Muthén and Asparouhov 2012). In this approach, informative priors with a small variance (e.g., a normal prior with a mean of zero and a variance of 0.01 for the standardized loadings, which implies a 95% confidence interval for the loadings ranging from −0.2 to +0.2) are specified for the cross-loadings. Although a model in which all loadings are freely estimated would not be identified in a frequentist approach to estimation, with the Bayesian approach identification is achieved by supplying strong priors. However, the priors should be neither too informative (in which case the result will be similar to a specification with zero cross-loadings, which can be thought of as a prior with zero mean and zero variance) nor too uninformative (which may lead to lack of model identification). As pointed out by Muthén and Asparouhov (2012), the Bayesian credibility intervals around the cross-loadings can be used as alternatives to modification indices to decide whether the assumption of zero cross-loadings is strongly inconsistent with the data (i.e., if the interval does not include zero, the assumption is violated).

9.5.2 Empirical Example

The approaches discussed in this section are quite recent and we are not aware of marketing applications, but we will present an illustrative application. In a survey dealing with subjective well-being (among other things), 1181 U.S. respondents completed the Satisfaction with Life Scale (Diener et al. 1985), which consists of five items rated on five-point agree-disagree scales (e.g., I am satisfied with my life), and they also rated their current level of general happiness based on how often they experienced five positive affective states (e.g., confident, enthusiastic) and five negative affective states (e.g., depressed, hopeless), based on five-point scales ranging from 1 = none of the time to 5 = all the time. These items are a subset of the items contained in the Affectometer 2 scale (Kammann and Flett 1983).

A congeneric three-factor measurement model with zero cross-loadings fits relatively poorly: $\chi^2(87) = 730.886$, p = 0.000; SRMR = 0.044; RMSEA = 0.079 (with a 90% CI ranging from 0.074 to 0.085; CFI = 0.922; TLI = 0.906. In an exploratory factor analysis in which the third life satisfaction, the second positive affect, and the fifth negative affect items are used as reference indicators (whose loading on the target factor are set to 1 and whose non-target loadings are set to zero, with the other loadings freely estimated), the fit of the model improves significantly based on the χ^2 statistic ($\chi^2(63) = 534.302$, p = 0.000) and some other fit statistics (SRMR = 0.030; CFI = 0.943), but the fit indices that take into account model parsimony actually deteriorate (RMSEA = 0.080, with a 90% CI ranging from 0.073 to 0.086; TLI = 0.905). Furthermore, in an absolute sense, the χ^2 statistic indicates a lack of fit. Although 12 of the 30 possible cross-loadings are significant, the largest is 0.38, with most below 0.1.

A Bayesian analysis with a variance prior of 0.01 yields similar results, but only four cross-loadings have 95% credibility intervals that do not include zero. Although the specification of non-zero non-target loadings leads to a somewhat better model fit, it does not appear that the assumption of zero cross-over loadings creates major problems in the analysis. This is also confirmed by the fact that the factor correlations are quite similar in the confirmatory and exploratory factor analyses.

9.6 Extension 2: Relaxing the Assumption of Uncorrelated Unique Factors

9.6.1 Conceptual Development

In the congeneric measurement model, shared variance between the indicators is solely due to shared substantive factors, either the same common factor or correlated common factors on which the items load. Other sources of covariation are not allowed; in particular, $\boldsymbol{\Theta}$ is specified to be diagonal. Below we will discuss a variety of reasons why this may not be the case and describe models that relax this assumption.

9.6.1.1 Sources of Shared Method Variance Among the Indicators

Many sources of systematic measurement error that may lead to covariation among the indicators have been identified in the methodological literature. Often, these are referred to as common method biases (Podsakoff et al. 2003). In general, factors responsible for shared variance other than shared substantive variance can be classified into those due to the respondent, the items (both the item itself and the response scale associated with the item), and more general characteristics of the survey (including properties of the survey instrument used and the conditions under which the survey is administered). Sometimes, these factors interact and are hard to separate, but the categorization serves a useful organizing function.

Many individual difference variables have been implicated in common method bias, including need for consistency, leniency, and positive or negative affectivity (see Podsakoff et al. 2003). However, some of these are not specific to response behavior in surveys, or they are restricted to particular contexts. Here we will focus on response styles and response sets (Baumgartner and Weijters 2015). Response styles refer to systematic response tendencies that are more or less independent of content, such as acquiescence and disacquiescence response style (ARS and DARS, i.e., a preference for the agreement or disagreement options or, more generally, the right or left of the response scale), extreme response style (ERS, i.e., a preference for the most extreme options on the response scale), and midpoint response style (MRS, i.e., a preference for the midpoint of the response scale). In contrast, response sets refer to systematic response tendencies in which the content of the items is taken into account, but the answers provided are inaccurate for various reasons. The best-known response set is social desirability, which leads to responses motivated by a desire to create a favorable impression (Steenkamp et al. 2010). Individual differences in response styles and response sets are sources of covariation among all the indicators or subsets of indicators that do not properly reflect shared substantive variation.

A multitude of item-related factors can also generate method effects. First, properties of the item stem (i.e., the statement to which participants are asked to respond) may induce similarities in how different items are comprehended, how relevant data are retrieved, how the information is integrated into a judgment, and how the response is edited, which have little or nothing to do with the substantive content of the items. For example, if items are difficult to understand or ambiguous, the respondent may be more likely to choose a noncommittal midpoint response. Probably the most well-studied item characteristic has been the keying direction of an item (i.e., whether or not the item is a reversed item). Numerous studies have shown that a common keying direction creates shared variation among items. Furthermore, when both regular and reversed items are included in an instrument, a substantively unidimensional scale may even appear to be multidimensional (Weijters and Baumgartner 2012; Weijters et al. 2013). Second, features of the response scale may generate method variance. For example, a common scale format for different items can lead to correlated measurement errors, although this problem is difficult to correct unless multiple scale formats are used in a survey. Even subtle aspects of the response scale such as the anchors used can have undesirable effects. As an illustration, Weijters et al. (2013) demonstrated that supposedly similar translations of response category labels (e.g., strongly agree vs. tout à fait d'accord) may differ in meaning across languages and encourage differential scale usage.

Finally, characteristics of the survey instrument or the survey setting may introduce method effects. As an example of the former, the positioning of items in a questionnaire can have a pronounced effect on how strongly the items are correlated. To illustrate, Weijters et al. (2009) show that the positive correlation between two items that measure the same construct decreases the farther apart the two items are in a questionnaire. If the two items were placed right next to each other, the correlation was 0.62, but it decreased to 0.35 when the items were 6 or more items apart. In another study, Weijters et al. (2014) demonstrated that a unidimensional scale could be turned into a multidimensional scale simply through item positioning. Specifically, they divided an established 8-item scale into multiple sets of two blocks of 4 items each, where each block of 4 items was shown on a separate screen and 32 filler items were placed between the two blocks. Depending on the item arrangement, the four items shown on the same page formed a distinct factor. As an example of the latter, Weijters et al. (2008) showed that mode of data collection (e.g., paper-and-pencil questionnaire, telephone interview, and online questionnaire) had a significant effect on stylistic responding (e.g., a telephone survey led to higher ARS and lower MRS), which in turn produced inflated factor loading and average variance extracted estimates in a measurement model for the construct of trust.

It is apparent that items can be correlated for many reasons other than substantive considerations. Models that incorporate this covariation are discussed next.

9.6.1.2 Models for Method Effects

In early applications of SEM, researchers often specified correlated errors in order to improve the fit of the model (Baumgartner and Homburg 1996). Nowadays, this practice is frowned up when it comes across as too data-driven. However, there are two principled ways in which method effects can be implemented. The first approach is generally referred to as the correlated uniqueness model (Marsh 1989). This method consists of allowing correlations among certain error terms, but instead of introducing the error correlations in an ad hoc fashion, they are motived by a priori hypotheses. For example, correlated uniquenesses might be specified for all items that share the same keying direction (i.e., the reversed items, the regular items, or both), which reflects the recognition that the polarity of item wording may introduce systematic covariation among items in addition to substantive covariation. Although the approach seemed initially promising (partly for methodological reasons, such as improved convergence during estimation) and has some benefits (e.g., it does not require unidimensionality of method effects), it also has important drawbacks, including the assumption that method effects are assumed to be uncorrelated (when multiple method effects are present) and the fact that method effects cannot be readily related to other variables of interest, including direct measures of the hypothesized method effects (Lance et al. 2002). Some care is also required with the use of correlated uniquenesses because in our experience, the estimated error correlations are sometimes garbage parameters that do not contribute to a better understanding of method effects (e.g., the correlations may have the wrong sign or the signs may differ across different pairs of variables).

The second approach involves specifying method factors for the hypothesized method effects. As in the case of correlated uniquenesses, we will assume that method effects are modeled at the individual-item level, rather than at the scale level, since we are interested in measurement analysis. Sometimes, a global method factor is posited to underlie all items in one's model, but this is only meaningful under special circumstances (e.g., when both regular and reversed items are available to measure a construct or several constructs; see Weijters et al. 2013), because otherwise method variance will be confounded with substantive variance. More likely, method factor will be specified for subsets of items that share a common method (e.g., reversed items). Of course, it is possible to model multiple method factors if several sources of method bias are thought to be present.

It is important to distinguish between inferred method effects and directly measured method effects, which leads to a distinction between implicit or explicit method factors. The difference can be illustrated in the context of acquiescence. Assume that a construct is measured with three regular and three reversed items that have not been recoded to make the keying direction consistent across all items. In this case the tendency to agree with items regardless of content (i.e., ignoring the keying direction) might be modeled by specifying a method factor with positive loadings on all items (in addition to a substantive factor that has positive loadings on the regular and negative loadings on the reversed items). This is an example of an implicit acquiescence factor because the individual differences captured by this factor may be due to mechanisms other than acquiescence (although acquiescence is a plausible explanation). An alternative would be to use a direct measure of acquiescence (e.g., measured by the extent of agreement to unrelated control items that are free of shared substantive content) and to specify this explicit acquiescence factor as a source of shared variation among the substantive items. In the latter case, it is even possible to model measurement error in the method factor by using several different acquiescence measures as multiple indicators of an underlying acquiescence factor. Figure 9.3 contains illustrative examples of measurement models that incorporate sources of shared covariation among the indicators other than substantive commonalities. The various models are described in more detail below.

9.6.2 Empirical Example

In the marketing research literature, some authors have used correlated uniqueness models to account for method effects. Van Auken et al. (2006) implement correlations between the unique factors associated with three different scaling formats (semantic differential, ratio scale and Likert scale formats) of a cognitive age measure that they validate across Japan and the US. Bagozzi and Yi (2012) present another example of such an approach, focusing on attitude, desire and intention factors measured by means of three scaling methods (Likert, semantic differential,

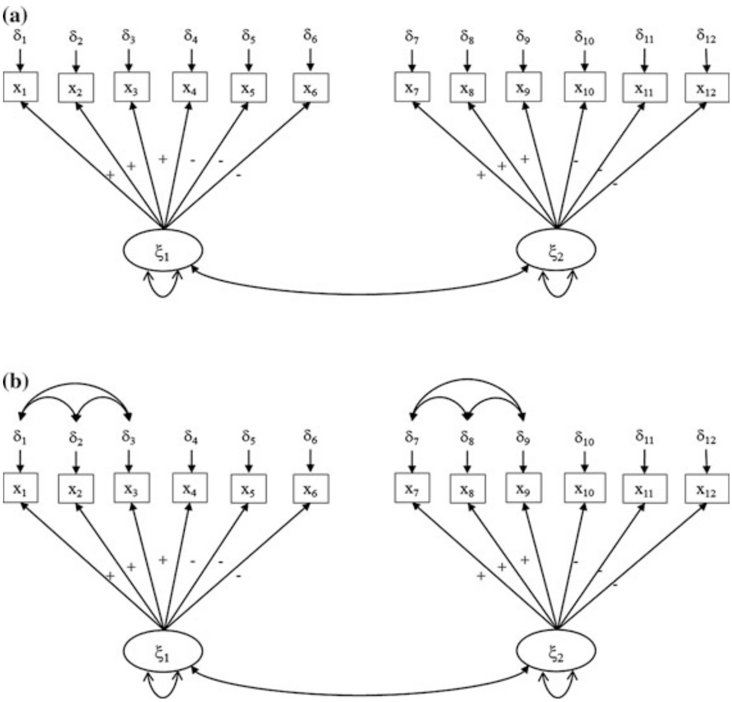


Fig. 9.3 Models to account for method effects in non-reversed and/or reversed items. **a** Congeneric measurement model (baseline). **b** Separate correlated uniquenesses for the regular (non-reversed) items in each scale. **c** Separate correlated uniquenesses for the reversed items in each scale. **d** Correlated uniquenesses for the regular (non-reversed) items in both scales. **e** Correlated uniquenesses for the reversed items in both scales. **f** Separate uncorrelated method ("acquiescence") factors with unit factor loadings for each scale. **g** Separate correlated method ("acquiescence") factors with unit factor loadings for each scale. **h** Overall method ("acquiescence") factor with unit factor loadings. Note for Fig. 9.3: In the illustrative example of Sect. 6.2, x_1–x_3 are regular and x_4–x_6 are reversed environmental concern items; x_7–x_9 are regular and x_{10}–x_{12} are reversed health concern items (see Table 9.2 and Sect. 6.2); ξ_1 refers to environmental concern, and ξ_2 refers to health concern

and peer evaluation). The specification of method factors is quite common in the literature reporting multi-trait multi-method models.

In our empirical example, we will compare correlated uniqueness models with method factor specifications. This empirical example uses the data from N = 740 Belgian consumers who responded to an environmental concern scale and a health concern scale, which we used earlier to illustrate the basic congeneric factor model. But this time, we do not only include the three regular items per construct analyzed previously, but also the three reversed items that we have disregarded so far. For environmental concern, the reversed items are 'I don't really worry about the

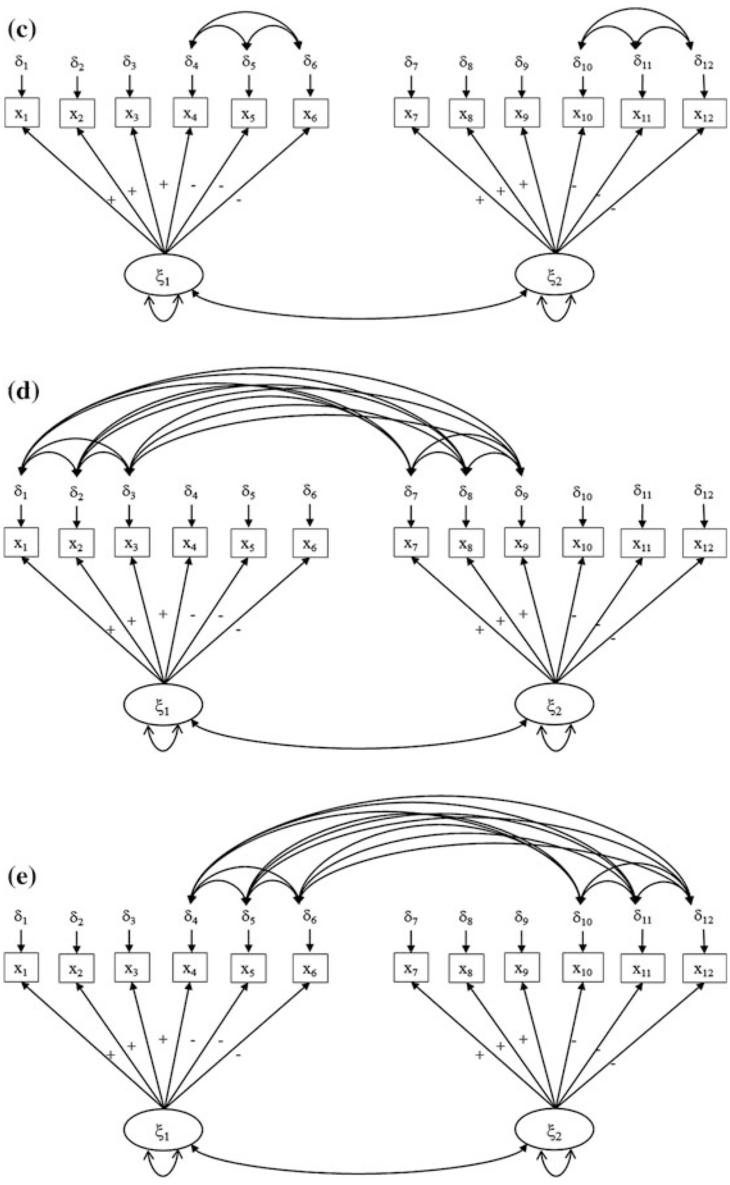


Fig. 9.3 (continued)

environment,' 'I don't want to put myself to the trouble to do things that are more environmentally friendly,' and 'It doesn't really matter whether products I use damage the environment or not'. For health concern, the reversed items are 'I have the impression that other people pay more attention to their health than I do,'

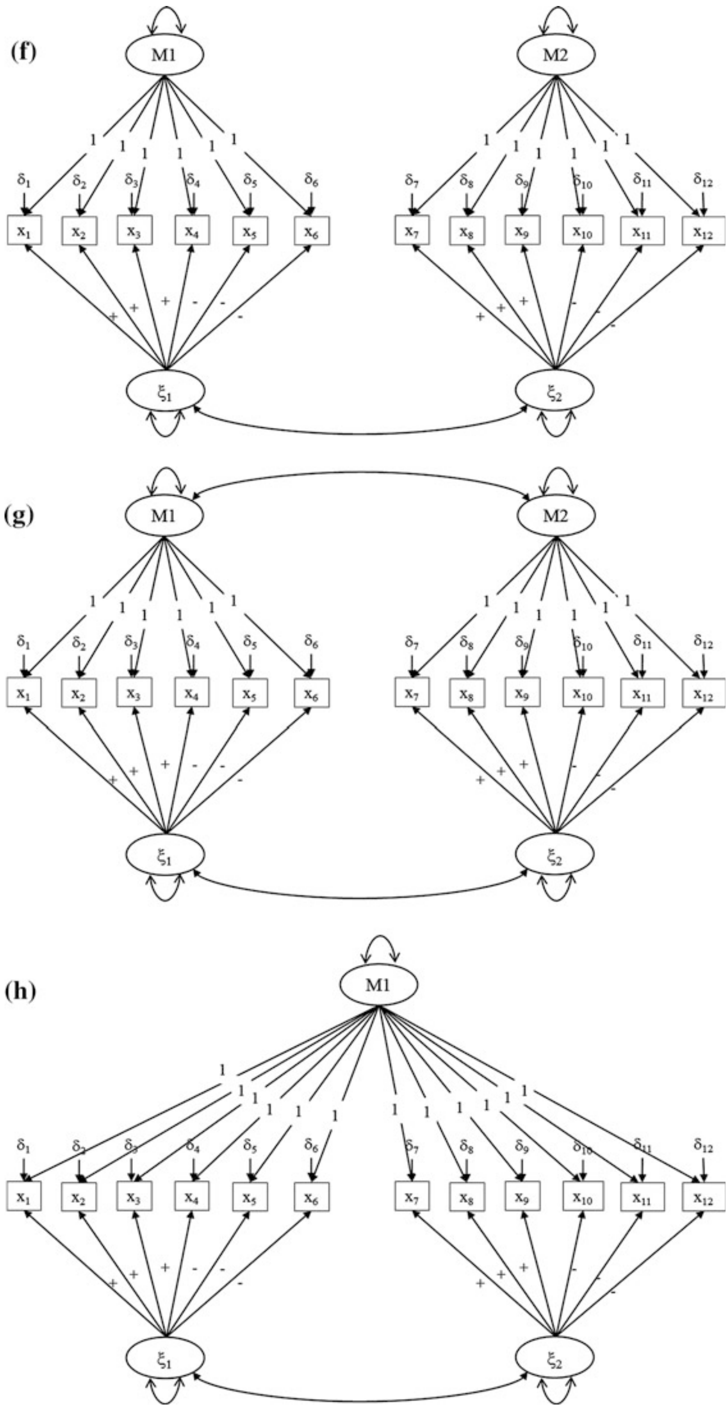


Fig. 9.3 (continued)

'I don't really think about whether everything I do is healthy,' and 'I refuse to ask myself all the time whether the things I eat are good for me' (see Table 9.2 for a listing of the non-reversed items).

We estimate a two-factor congeneric model where each factor has six reflective indicators, using the ML estimator in Mplus 7.3, and test all the models represented in Fig. 9.3. Note that the reversed items were not recoded, so the regular (reversed) items are expected to have positive (negative) loadings on the underlying substantive factor. Model specifications other than those shown in Fig. 9.3 are possible, but the models discussed below serve as a relevant illustration of the modeling choices one faces when trying to account for method effects. Table 9.4 reports model fit indices for all models, as well as the AVE for regular items and the AVE for reversed items (averaged over health and environmental concern) and the correlation between the two substantive factors. Model A, the basic congeneric measurement model, clearly is problematic in terms of fit, and the modification indices suggest the need to include residual correlations. Two decisions need to be made in this regard. First, one can freely correlate the residual terms of the regular items (see Fig. 9.3b and d) or the residual terms of the reversed items (see Fig. 9.3c and e). Even though researchers may often arbitrarily include residual correlations for reversed items (rather than for regular items), in the current data, the model fit indices favor correlating the residuals of the regular items. The results show that when the residuals of the reversed items are correlated, the AVE of the reversed items (averaged over health and environmental concern) decreases relative to the AVE of the non-reversed items, and vice versa. So researchers should be aware that accounting for method variance in a subset of items will likely diminish the contribution of these items to the measure of interest. It is also important to note that the identification of which keying direction is considered to be reversed is essentially arbitrary as it depends on how the construct is named (e.g., the constructs could have been labeled 'lack of environmental/health concern').

Second, correlated uniquenesses can be specified for each construct separately (as shown in Fig. 9.3b and c) or across both constructs (as shown in Fig. 9.3d, e). The within-factor correction leads to a substantial improvement in model fit already, but if the researcher is interested in controlling for the biasing effects of method effects on factor correlations, the cross-factor approach is usually preferable. The major downside of the latter is the prohibitively large number of additional parameters that need to be estimated, many of which turn out to be nonsignificant and hard to interpret. In the current data, there are six within-factor residual correlations, plus nine cross-factor residual correlations. Note that not including the cross-factor residual correlations implies that one assumes that common method variance does not carry over from one scale to another, which seems very unlikely.

In light of these issues with correlated residuals, the method factor approach seems more parsimonious and conceptually more meaningful. We do not test method factors that are specific to one type of item (regular or reversed) for several

Table 9.4 Fit indices for the health and environmental concern model including reversed items and method effects

	Model fit							AVE	AVE	Estimated
Model (see Fig. 9.3)	χ^2	df	RMSEA (90%CI)	CFI	TLI	SRMR	BIC	Regular items	Reversed items	Factor correlation
A Congeneric measurement model (baseline)	656.2	53	0.124 (0.116, 0.133)	0.833	0.793	0.068	20516.47	0.589	0.327	0.378
B Separate correlated uniquenesses for the regular (non-reversed) items in each scale	191.6	47	0.064 (0.055, 0.074)	0.960	0.944	0.052	20091.53	0.294	0.512	0.391
C Separate correlated uniquenesses for the reversed items in each scale	198.0	47	0.066 (0.057, 0.076)	0.958	0.941	0.048	20097.93	0.623	0.239	0.376
D Correlated uniquenesses for the regular (non-reversed) items in both scales	79.1	38	0.038 (0.026, 0.050)	0.989	0.980	0.027	20038.44	0.301	0.514	0.371
E Correlated uniquenesses for the reversed items in both scales	103.3	38	0.048 (0.037, 0.059)	0.982	0.969	0.030	20062.68	0.624	0.247	0.369
F Separate uncorrelated method ("acquiescence") factors with unit factor loadings for each scale	241.9	51	0.071 (0.062, 0.080)	0.947	0.932	0.062	20115.37	0.528	0.426	0.385
G Separate correlated method ("acquiescence") factors with unit factor loadings for each scale	149.9	50	0.052 (0.043, 0.062)	0.972	0.964	0.034	20029.93	0.528	0.427	0.362
H Overall method ("acquiescence") factor with unit factor loadings	211.5	52	0.064 (0.055, 0.074)	0.956	0.944	0.035	20078.37	0.539	0.409	0.365

Note See Table 9.1 for an explanation of these fit indices; AVE—average variance extracted reasons, including the issues mentioned in the context of correlated residuals for regular or reversed items (also see Weijters et al. 2013). Instead, we test three method factor models (see Fig. 9.3f, g and h) that vary in the extent to which they imply a scale-specific and/or a generic method effect. The model fit indices favor model G, which accounts for two scale-specific method effects that are correlated.

The model comparisons show several other things. First, among all the models tested both the overall chi-square test and most of the alternative fit indices suggest that model D is the preferred model. Unfortunately, this is not a very parsimonious model, and the assumptions of correlated uniquenesses for only the non-reversed items seems somewhat arbitrary. Second, a model with two correlated method factors (model G) is attractive for these data because the availability of both non-reversed and reversed items makes it possible to clearly separate substantive from stylistic variance. Furthermore, this model has few additional parameters compared to the baseline model, has the best fit in terms of BIC, and fits almost as well as model D based on the fit indices that take into account model parsimony. Third, even though the models that take into account method variance achieve a much better fit than the model that does not, the effect on the estimated factor correlation is relatively small.

9.7 Extension 3: Relaxing the Assumption of Continuous, Normally Distributed Observed Measures

9.7.1 Conceptual Development

Although observed measures are probably never literally continuous and normally distributed, simulation evidence suggests that this assumption may be relatively innocuous if the response scale has at least 5 to 7 distinct categories, particularly if the scale used is symmetric and the category labels were chosen carefully (e.g., a 5-point Likert scale with response options of strongly disagree, disagree, neither agree nor disagree, agree, and strongly agree). However, there are cases where the assumption of a continuous, normally distributed variable is clearly violated, such as when there are only two response options (e.g., yes or no). Even when there are more than two scale steps, it is not obvious that certain scales should be treated as interval scales. For example, in a frequency scale with response options of none of the time, rarely, sometimes, often, and all the time, numerical values of 1–5 may not properly reflect the spacing of the scale steps on the frequency scale. Another possible reason for non-normality of measurement variables is that nonlinear structural relations imply non-normality in the measures (Bauer and Curran 2004). Consequently, when estimating models with nonlinear structural relations (e.g., interactions, quadratic effects), researchers should consider explicitly accounting for non-normality in the measurement model (see also van der Lans et al. 2014).

Estimation procedures that do not require multivariate normality or correct for violations of multivariate normality exist (Andreassen et al. 2006; Chuang et al. 2015), and they are implemented in all of the commonly used programs for structural equation modeling. Here, we will focus on an approach that explicitly takes into account the discreteness of the data (and thus one potential violation of normality). Specifically, the conventional congeneric factor model can be extended to accommodate categorical (binary or ordinal) observed measures by assuming that the variables that are actually observed are discretized versions of underlying continuous response variables. Thus, a binary or ordinal factor model has to specify not only how the continuous response variable underlying the discretized observed variable is related to the latent construct that the researcher wants to measure (this is the usual factor model when the observed variable is assumed to be continuous), but also how the discretized variable that is actually observed is related to the underlying continuous response variable.

When the observed variables are binary, the model can be stated as follows:

$$\boldsymbol{x}^{*} = \boldsymbol{\tau} + \boldsymbol{\Lambda\xi} + \boldsymbol{\delta} \tag{9.9}$$

This is the same model as stated in Eq. (9.1), except that, as signaled by the asterisk, $\mathbf{x}$ is not directly observed. The $\mathbf{x}$ actually observed is a binary version of $\mathbf{x}^*$ such that $x_i = 1$ if $x_i^* \geq \nu_i$ and $x_i = 0$ otherwise. Therefore, the model contains not only intercept and slope parameters, but also threshold parameters ν_i. As explained by Kamata and Bauer (2008), for identification the intercept $\boldsymbol{\tau}$ is generally set to zero (since the intercept and the threshold cannot both be uniquely determined) and a scale for the latent variables in $\mathbf{x}^*$ and $\boldsymbol{\xi}$ has to be chosen. Different parameterizations can be employed, but one possibility is to (a) set the variance of δ_i to unity (called the conditional parameterization of the continuous response variable by Kamata and Bauer 2008) and (b) constrain the mean of ξ_j to zero and the variance of ξ_j to one (called the standardized parameterization of the latent construct by Kamata and Bauer 2008).

It turns out that this type of model is equivalent to the two-parameter model of item response theory (IRT). In the IRT model, the probability that a person will provide a response of 1 on item i, given ξ_j, is expressed as follows:

$$P\left(x_i = 1 | \xi_j\right) = F\left(\alpha_i \xi_j + \beta_i\right) = F\left(\alpha_i\left(\xi_j - \gamma_i\right)\right) \tag{9.10}$$

where F is either the normal or logistic cumulative distribution function. Equation (9.10) specifies a sigmoid relationship between the probability of a response of 1 to an item and the latent construct (referred to as an item characteristic curve); α_i is called the discrimination parameter (which shows the sensitivity of the item to discriminate between respondents having different ξ_j around the point of inflection of the sigmoid curve) and γ_i the difficulty parameter (i.e., the value of ξ_j at which the probability of a response of 1 is 0.5). The model is similar to logistic regression, except that the explanatory variable ξ_j is latent rather than observed (Wu and Zumbo 2007). The correspondence between the binary factor model in (9.9) and the two-parameter IRT model in (9.10) is given as follows:

$$\alpha_i = \lambda_{ij} \tag{9.11}$$

$$\beta_i = -\alpha_i \gamma_i = -\nu_i \tag{9.12}$$

for Var (δ_i) = 1 as in the conditional parameterization. In other words, the slopes in the two models are the same, but the threshold in the binary factor model is equal to the product of the slope and item difficulty in the IRT model.

The factor or IRT model for binary data can be extended to ordinal responses. For the ordinal factor model,

$$x_i = k \text{ if } \nu_{k-1} < x^* \leq \nu_k \tag{9.13}$$

where K is the number of response options (k = 1, 2, …, K) and $-\infty = \nu_0 < \nu_1 < \cdots < \nu_K = \infty$ (ν_1 to ν_{K-1} are thresholds to be estimated). In the corresponding IRT model, the so-called graded response model (Samejima 1969), the response behavior for the K ordinal categories is summarized by (K − 1) sigmoid curves (called operating characteristic curves or cumulative response curves) that express the probability of providing a response of k or higher, that is,

$$P\left(x_i \mathrm{k} | \xi_j\right) = F\left(\alpha_i \xi_j + \beta_{ik}\right) = F\left(\alpha_i\left(\xi_j - \gamma_{ik}\right)\right) \tag{9.14}$$

where k = 2, …, K. Obviously, $P(x_i \geq 1) = 1$. The interpretation of the α_i and γ_{ik} parameters is the same as in the binary IRT model, except that the γ_{ik} refer to thresholds of a response of k or higher. Note that the α_i are constant for the different response categories, which means that the slopes of the different sigmoid curves are parallel. The probability of a response in the kth interval is given by the difference of $P(x_i \geq k)$ and $P(x_i \geq k + 1)$, that is,

$$P\left(x_i = \mathrm{k} | \xi_j\right) = P\left(x_i \mathrm{k} | \xi_j\right) - P\left(x_i \mathrm{k} + 1 | \xi_j\right) \tag{9.15}$$

The curves describing the relationship between $P(x_i = k)$ and ξ_j are called category characteristic, category response or item characteristic curves. The points on the ξ_j axis where these curves intersect are the item difficulty parameters and the midpoint between the item difficulties for k and (k + 1) is the peak of the curve for the kth category.

Muraki (1990) developed a modified graded response model specifically designed for Likert-type items in which there is a separate γ_i parameter for each item, but the distances between adjacent thresholds are the same across items (i.e., $\gamma_{ik} = \gamma_i + c_k$).

The IRT literature is voluminous and we cannot do justice to important recent developments in this chapter. Of particular interest are extensions that not only consider the ordinal nature of observed responses, but also take into account scale usage heterogeneity across respondents (e.g., disproportionate use of the middle response option or the extremes of the rating scale). Illustrative examples include Javaras and Ripley (2007), Johnson (2003), and Rossi et al. (2001). De Jong et al. (2010)

describe an ordinal response model that implements the randomized response technique and enables researchers to get more truthful responses to sensitive questions. Finally, IRT models can be used to compare measurement instruments across different populations of respondents (e.g., in cross-cultural research, where the groups might be different countries), and the item parameters need not be modeled as fixed effects but can be specified as random effects. The interested reader is referred to de Jong et al. (2007).

9.7.2 *Empirical Example*

Given their availability in standard software packages for Structural Equation Modeling, estimation methods that account for non-normality are quite commonly used. Using satisfaction survey data, Andreassen et al. (2006) illustrate how alternative estimation methods that account for non-normality may lead to different results in terms of model fit, model selection, and parameter estimates and, as a consequence, managerial priorities in the marketing domain.

Table 9.5 Parameter estimates of the graded response model for five positive affect items

Item	Parameter	Estimate	Standard Error	p
pa1	Threshold 1	−3.252	0.182	<0.0001
	Threshold 2	−2.501	0.110	<0.0001
	Threshold 3	−0.973	0.057	<0.0001
	Threshold 4	1.240	0.062	<0.0001
	Slope	0.817	0.059	<0.0001
pa2	Threshold 1	−4.183	0.289	<0.0001
	Threshold 2	−2.786	0.147	<0.0001
	Threshold 3	−0.882	0.070	<0.0001
	Threshold 4	1.700	0.097	<0.0001
	Slope	1.239	0.092	<0.0001
pa3	Threshold 1	−4.207	0.240	<0.0001
	Threshold 2	−2.400	0.117	<0.0001
	Threshold 3	−0.378	0.058	<0.0001
	Threshold 4	2.176	0.109	<0.0001
	Slope	1.178	0.080	<0.0001
pa4	Threshold 1	−3.165	0.163	<0.0001
	Threshold 2	−1.608	0.072	<0.0001
	Threshold 3	−0.026	0.049	0.2957
	Threshold 4	1.766	0.078	<0.0001
	Slope	0.893	0.059	<0.0001
pa5	Threshold 1	−4.084	0.332	<0.0001
	Threshold 2	−2.957	0.141	<0.0001
	Threshold 3	−1.288	0.069	<0.0001
	Threshold 4	1.024	0.062	<0.0001
	Slope	0.973	0.067	<0.0001

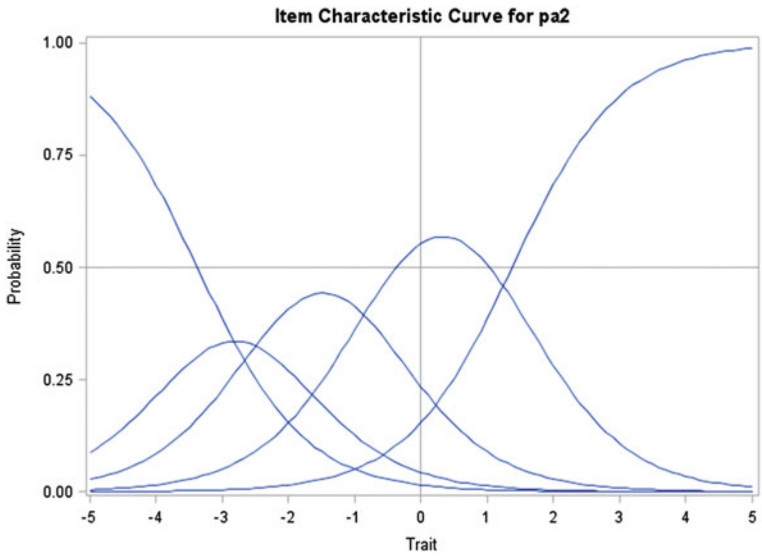


Fig. 9.4 Item characteristic curves for the second positive affect item

As an example of an ordinal factor or graded response model, we will use the data on the subjective well-being of 1181 U.S. respondents analyzed earlier. For simplicity, we will restrict our analysis to the experience of five positive affective states (e.g., confident, enthusiastic) rated on five-point scales ranging from 1 = none of the time to 5 = all the time. Table 9.5 provides the parameter estimates based on SAS 9.4 (Mplus produced identical results). The model was estimated with a probit link and maximum likelihood. The log likelihood of the model was −5,724.54, and the BIC was 11,625.94. As seen in Table 9.5, all five items are good measures of the experience of positive affect, but items 2 and 3 are the most discriminating. Figure 9.4 shows the item characteristic curves for the second item as an illustration. Restricting the slopes to be the same across items decreases the fit of the model (BIC = 11,634.39). However, a model in which the differences between adjacent thresholds are restricted to be the same across items (i.e., a modified graded response model) has a better fit than the baseline model (BIC = 11,608.59).

9.8 Conclusion

Measurement is an important aspect of empirical research in Marketing. Many constructs cannot be assessed directly and multiple, imperfect indicators of the intended construct are needed to approximate the theoretical entity of interest. Even if a variable seems relatively straightforward (e.g., sales or other measures of the financial performance of a firm), it is likely that the observed measures contain measurement error (both random and systematic) that should be taken into account. In this chapter, we discussed a wide variety of measurement models that researchers can use to evaluate the adequacy of the measures that are available to represent the phenomena investigated. It is hoped that the application of these models will further improve the correspondence between what the researcher hopes to capture and what is actually contained in the observed data.

References

Andreassen, T.W., B.G. Lorentzen, and U.H. Olsson. 2006. The impact of non-normality and estimation methods in SEM on satisfaction research in marketing. *Quality and Quantity* 40 (1): 39–58.

Asparouhov, T., and B. Muthén. 2014. Multiple-group factor analysis alignment. *Structural Equation Modeling* 21 (4): 495–508.

Bagozzi, R.P., and Y. Yi. 2012. Specification, evaluation, and interpretation of structural equation models. *Journal of the Academy of Marketing Science* 40 (1): 8–34.

Batra, R., A. Ahuvia, and R.P. Bagozzi. 2012. Brand love. *Journal of Marketing* 76 (2): 1–16.

Bauer, D.J., and P.J. Curran. 2004. The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods* 9 (1): 3–29.

Baumgartner, H., and C. Homburg. 1996. Applications of structural equation modeling in marketing and consumer research: A review. *International Journal of Research in Marketing* 13 (2): 139–161.

Baumgartner, H., and B. Weijters. 2015. Response biases in cross-cultural measurement. In *Handbook of culture and consumer psychology*, ed. S. Ng, and A.Y. Lee, 150–180. New York, NY: Oxford University Press.

Bergkvist, L., and J.R. Rossiter. 2007. The predictive validity of multiple-item versus single-item measures of the same constructs. *Journal of Marketing Research* 44 (2): 175–184.

Browne, M. W., and R. Cudeck.1992. Alternative ways of assessing model fit. *Sociological Methods & Research* 21 (November): 230–258.

Chen, M.-F. 2009. Attitude toward organic foods among Taiwanese as related to health consciousness, environmental attitudes, and the mediating effects of a healthy lifestyle. *British Food Journal* 111 (2–3): 165–178.

Chuang, J., V. Savalei, and C.F. Falk. 2015. Investigation of type I error rates of three versions of robust chi-square difference tests. *Structural Equation Modeling* 22 (4): 517–530.

Coltman, T., T.M. Devinney, D.F. Midgley, and S. Venaik. 2008. Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research* 61 (12): 1250–1262.

Dagger, T.S., J.C. Sweeney, and L.W. Johnson. 2007. A hierarchical model of health service quality scale development and investigation of an integrated model. *Journal of Service Research* 10 (2): 123–142.

De Jong, M.G., R. Pieters, and J.-P. Fox. 2010. Reducing social desirability bias through item randomized response: An application to measure underreported desires. *Journal of Marketing Research* 47 (February): 14–29.

De Jong, M.G., J.-B.E.M. Steenkamp, and J.-P. Fox. 2007. Relaxing measurement invariance in cross-national consumer research using a hierarchical IRT model. *Journal of Consumer Research* 34 (August): 260–278.

Diamantopoulos, A., P. Riefler, and K.P. Roth. 2008. Advancing formative measurement models. *Journal of Business Research* 61 (12): 1203–1218.

Diener, E., R.A. Emmons, R.J. Larsen, and S. Griffin. 1985. The satisfaction with life scale. *Journal of Personality Assessment* 49 (1): 71–75.

Edwards, J.R. 2011. The fallacy of formative measurement. *Organizational Research Methods* 14 (2): 370–388.

Fornell, C., and D.F. Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research* 18 (1): 39–50.

Groves, R.M., F.J.J. Fowler, M.P. Cooper, J.M. Lepkowski, E. Singer, and R. Tourangeau. 2004. *Survey Methodology*. Hoboken, NJ: Wiley.

Hu, L., and Bentler, P.M. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: A multidisciplinary journal* 6 (1): 1–55.

Jarvis, C.B., S.B. MacKenzie, and P.M. Podsakoff. 2003. A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research* 30 (2): 199–218.

Javaras, K.N., and B.D. Ripley. 2007. An 'unfolding' latent variable model for Likert attitude data: Drawing inferences adjusted for response style. *Journal of the American Statistical Association* 102 (June): 454–463.

Johnson, T.R. 2003. On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. *Psychometrika* 68: 563–583.

Kamata, A., and D.J. Bauer. 2008. A note on the relation between factor analytic and item response theory models. *Structural Equation Modeling* 15 (1): 136–153.

Kammann, R., and R. Flett. 1983. Affectometer 2: A scale to measure current level of general happiness. *Australian Journal of Psychology* 35 (2): 259–265.

Lance, C.E., C.L. Noble, and S.E. Scullen. 2002. A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods* 7 (2): 228–244.

Little, T.D. 2000. On the comparability of constructs in cross-cultural research: A critique of Cheung and Rensvold. *Journal of Cross-Cultural Psychology* 31 (2): 213–219.

MacCallum, R.C. 1986. Specification searches in covariance structure modeling. *Psychological Bulletin* 100 (1): 107–120.

MacCallum, R.C., and M.W. Browne. 1993. The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin* 114 (3): 533–541.

MacKenzie, S.B., P.M. Podsakoff, and C.B. Jarvis. 2005. The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions. *Journal of Applied Psychology* 90 (4): 710–730.

MacKenzie, S.B., P.M. Podsakoff, and N.P. Podsakoff. 2011. Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS Quarterly* 35 (2): 293–334.

Marsh, H.W. 1989. Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement* 13 (4): 335–361.

Marsh, H.W., A.J. Morin, P.D. Parker, and G. Kaur. 2014. Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. *Annual Review of Clinical Psychology* 10: 85–110.

McIntosh, C.N., J.R. Edwards, and J. Antonakis. 2014. Reflections on partial least squares path modeling. *Organizational Research Methods* 17 (2): 210–251.

Millsap, R.E. 2001. When trivial constraints are not trivial: The choice of uniqueness constraints in confirmatory factor analysis. *Structural Equation Modeling* 8 (1): 1–17.

Muraki, E. 1990. Fitting a polytomous item response model to Likert-type data. *Applied Psychological Measurement* 14 (1): 59–71.

Muthén, B., and T. Asparouhov. 2012. Bayesian structural equation modeling: A more flexible representation of substantive theory. *Psychological Methods* 17 (3): 313–335.

Parasuraman, A., V.A. Zeithaml, and L.L. Berry. 1988. Servqual. *Journal of Retailing* 64 (1): 12–40.

Podsakoff, P.M., S.B. MacKenzie, J.-Y. Lee, and N.P. Podsakoff. 2003. Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology* 88 (5): 879–903.

Rossi, P.E., Z. Gilula, and G.M. Allenby. 2001. Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *Journal of the American Statistical Association* 96 (March): 20–31.

Rossiter, J.R. 2002. The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing* 19 (4): 305–335.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometric Monograph No. 17,* Richmond, VA: Psychometric Society.

Schertzer, S.M., D. Laufer, D.H. Silvera, and J. Brad McBride. 2008. A cross-cultural validation of a gender role identity scale in marketing. *International Marketing Review* 25 (3): 312–323.

Singh, J.J., S.J. Vitell, J. Al-Khatib, and I. Clark III. 2007. The role of moral intensity and personal moral philosophies in the ethical decision making of marketers: A cross-cultural comparison of China and the United States. *Journal of International Marketing* 15 (2): 86–112.

Steenkamp, J.-B.E.M., and H. Baumgartner. 1998. Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research* 25 (June): 78–90.

Steenkamp, J.-B.E.M., M.G. De Jong, and H. Baumgartner. 2010. Socially desirable response tendencies in survey research. *Journal of Marketing Research* 47 (2): 199–214.

Strizhakova, Y., R.A. Coulter, and L.L. Price. 2008. The meanings of branded products: A cross-national scale development and meaning assessment. *International Journal of Research in Marketing* 25 (2): 82–93.

Traub, R.E. 1994. *Reliability for the social sciences: Theory and applications*. Thousand Oaks, CA: Sage.

Van Auken, S., T.E. Barry, and R.P. Bagozzi. 2006. A cross-country construct validation of cognitive age. *Journal of the Academy of Marketing Science* 34 (3): 439–455.

Van der Lans, R., B. van den Bergh, and E. Dieleman. 2014. Partner selection in brand alliances: An empirical investigation into the drivers of brand fit. *Marketing Science* 33 (4): 551–566.

Walsh, G., and S.E. Beatty. 2007. Customer-based corporate reputation of a service firm: Scale development and validation. *Journal of the Academy of Marketing Science* 35 (1): 127–143.

Weijters, B., and H. Baumgartner. 2012. Misresponse to reversed and negated items in surveys: A review. *Journal of Marketing Research* 49 (5): 737–747.

Weijters, B., H. Baumgartner, and N. Schillewaert. 2013a. Reversed item bias: An integrative model. *Psychological Methods* 18 (3): 320–334.

Weijters, B., A. De Beuckelaer, and H. Baumgartner. 2014. Discriminant validity where there should be none: Positioning same-scale items in separated blocks of a questionnaire. *Applied Psychological Measurement* 38 (6): 450–463.

Weijters, B., M. Geuens, and H. Baumgartner. 2013b. The effect of familiarity with the response category labels on item response to Likert scales. *Journal of Consumer Research* 40 (2): 368–381.

Weijters, B., M. Geuens, and N. Schillewaert. 2009. The proximity effect: The role of interitem distance on reverse-item bias. *International Journal of Research in Marketing* 26 (1): 2–12.

Weijters, B., N. Schillewaert, and M. Geuens. 2008. Assessing response styles across modes of data collection. *Journal of the Academy of Marketing Science* 36 (3): 409–422.

Wu, A.D., and B.D. Zumbo. 2007. Thinking about item response theory from a logistic regression perspective. In *Real Data Analysis*, ed. S.S. Sawilowsky, 241–269. Charlotte, NC: Information Age Publishing.

Yi, Y., and T. Gong. 2013. Customer value co-creation behavior: Scale development and validation. *Journal of Business Research* 66 (9): 1279–1284.

Chapter 10
Marketing Models for the Customer-Centric Firm

Eva Ascarza, Peter S. Fader and Bruce G. S. Hardie

10.1 Introduction

The past two decades have seen marketing academics and practitioners move from a product-centric, transaction-focused view of marketing towards one that is more customer-centric and relationship-oriented in nature (e.g., Fader 2012; Galbraith 2005; Hoekstra et al. 1999; Lamberti 2013; Ravi and Sun 2016; Seybold et al. 2001). With such a mindset, a firm's customers are viewed as (intangible) assets that generate cash flow not just this period but in future periods as well (Blattberg et al. 2001; Gupta and Lehmann 2005).

While all firms care about their customers, there are several factors that clearly distinguish those that are truly "customer-centric" from those that are merely "customer-oriented." A genuinely customer-centric firm (i) has the ability to track individual customers over time (and across channels) and seeks to calculate forward-looking metrics (e.g., customer lifetime value, hereafter CLV) at a granular level, (ii) seeks to identify the high CLV customers and sees them as a "growth engine" for the enterprise (in the same way that a product-centric firm views its best products in such a manner), and (iii) sees its product development efforts as a "means to an end,"

E. Ascarza
Columbia Business School, Uris Hall, 3022 Broadway, New York, NY
10027-6902, USA
e-mail: ascarza@gsb.columbia.edu

P.S. Fader
The Wharton School, University of Pennsylvania, 3730 Walnut Street,
Philadelphia, PA 19104-6340, USA
e-mail: faderp@wharton.upenn.edu

B.G.S. Hardie (✉)
London Business School, Regent's Park, London NW1 4SA, UK
e-mail: bhardie@london.edu

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_10

i.e., to elevate the value of its customers (and attract valuable new ones), instead of seeing it as "an end unto itself."

With this characterization in mind, a customer-centric firm takes the view that there are three key drivers of (organic) growth and overall profitability: Customer acquisition, customer retention, and customer development (i.e., increasing the value of each existing customer (per unit of time) while they remain a customer).[1] In order to make informed decisions in these three key areas, the firm must have access to rich customer-level data from both internal and external sources,[2] along with the capabilities to analyze these data. At the heart of this is a database (or collection of databases) that tracks customers' purchases and their interactions with the firm (Imhoff et al. 2001).

By the very nature of their operations, mail-order catalog companies, along with firms that have a contractual/subscription-based business model (such as many magazine publishers and financial services firms), have been in a position to build customer-level databases from the beginning of their operations. Historically, the challenge faced by all such firms was the cost of collecting, storing, and processing this customer data. (See, for example, Howard's (1978) description of operations at Sears, Roebuck and Company in the late 1950s.) Starting in the 1960s, the ever-increasing power and ever-decreasing cost of computing resources meant that more and more firms could collect customer data, with the more innovative firms developing analytical tools that would help them improve the performance of their marketing activities. (See Petrison et al. (1997) for an historical review of direct and database marketing.)

In this chapter we review the key data-based tools and methods that have been developed by marketing scientists (and researchers and practitioners in related fields such as operations research, statistics, and computer science) to assist firms in their customer acquisition, retention, and development activities. We start by reviewing the work on customer acquisition (Sect. 10.2).

While the literature focusing on customer acquisition is, understandably, quite distinct from that which focuses on creating/extracting value from existing customers (i.e., retention and development), it is much harder to cleanly separate papers that focus on retention from those primarily centered on development issues. The overlap (or, perhaps, lack of clarity) in our literature reflects a similar pattern in practice: while some businesses clearly distinguish between their customer retention and customer development-related marketing activities, these efforts are deeply (and inextricably) intertwined for most others. As a result, we do not offer separate coverage

[1] There are other ways of expressing this basic idea. For example, instead of talking about retention and development, Bolton et al. (2004) talk of the length, depth, and breadth of the relationship between a customer and a service provider, where "the depth of a relationship is reflected in the frequency of service usage over time [... and ...] in customers' decisions to upgrade and purchase premium (higher margin) products instead of low-cost variants [, ... and ...] the breadth of a relationship is reflected in cross-buying or 'add-on' buying; that is, the number of additional (different) products or services purchased" (p. 273).

[2] See Deighton and Johnson (2013) for an examination of the complex network of firms that collect and use data about individuals for marketing purposes.

of retention and development issues. Rather, we discuss the various models that have been developed to guide decisions concerning the overall management of acquired customers, encompassing the length, depth, and breadth of their relationship with the firm (Sect. 10.3).

We then briefly consider work on the coordination of acquisition and retention activities (Sect. 10.4), and conclude with a brief discussion of key areas that warrant the attention of researchers interested in developing marketing models for the customer-centric firm (Sect. 10.5).

10.2 Customer Acquisition

Despite its obvious importance, and with the obvious exception of work in the traditional direct mail and database marketing literature, "there is very little research on acquisition marketing. The traditional marketing literature does not separate the issue of acquiring customers from retaining customers. Positioning, segmentation, targeting is a generic concept. Research in advertising studies the general impact of communications but does not separate newly acquired customers from retained customers" (Blattberg et al. 2008). Many papers that at first glance appear to have a customer acquisition focus are actually "acquiring" customers for the *product*, which is not the same as acquiring customers for the *firm*.[3] For example, in Schwartz et al.'s (2016) work on optimizing the design of an online display advertising campaign using multi-armed bandit experiments, a customer is "acquired" if they open an account having clicked on the display advertisement; no distinction is made between those account openers who are first-time customers of the bank versus those who already have an account with the bank. The work on generating new product trial and the vast majority of the work on the adoption/diffusion of innovations is similarly product-centric. (This is not to say that these models are of no value to the customer-centric firm wanting to model customer acquisition (e.g., McCarthy et al. 2017); it is simply the case that their application has been product-centric in nature.)

Our review of the literature first considers the methods and models developed by those working with traditional direct marketers, and then explores the broader acquisition-related literature. In what follows, we take the view that a customer is "acquired" when they make their first purchase of the company's products or services (or make their first donation in a charity setting, etc.). The notion of "acquisition" is not so clear in a freemium business setting, where some parts of the organization may view someone as being acquired when they sign-up for the free service, while other parts may focus on the receipt of the first payment (thereby viewing the free service as an acquisition channel).

[3] A notable exception is the work of Natter et al. (2015).

10.2.1 Direct Approaches to Customer Acquisition

While firms do make use of direct response advertising (be it via print, radio, TV, or online), inserts, and other such media (Tapp et al. 2014), the classic direct marketing acquisition campaign sees the firm contacting a list of prospects, be it via mail, outbound telemarketing, or email.[4] "A prospect is someone you hope to be able to attract to become a customer, but he is not a customer until he has made a purchase" (Rosenwald 2004, p. 22).

In its simplest form, the firm sends the same message/offer to everyone on the list, and each prospect either responds or does not. This is effectively mass marketing by mail (Petrison et al. 1997) and email. There is a long tradition of experimenting with the message and offer before deciding on the *single* message to roll out to the whole list. This ranges from simple A/B tests to more complex methods such fractional factorial designs (Almquist and Wyner 2001) and Plackett-Burman designs (Bell et al. 2006). When the firm has the option of buying lists from different sources, it is standard practice to undertake a test in which mailings are sent to a sample of prospects from each list, and the choice of list(s) is made on the basis of the observed response rate(s).

When any experiment or test is undertaken, the test mailing response rate is used as a prediction of the response to the rollout mailing. It is not uncommon to find that the rollout response rate is actually lower than that observed in the test. Allenby and Blattberg (1987), Ehrman (1990), Ehrman and Miescke (1989), and Morwitz and Schmittlein (1998), amongst others, have proposed methods that adjust the test results to arrive at a more accurate prediction of rollout response.

The practices described above see all the prospects on the list receiving the offer, even though we expect them to vary in their propensity to respond. When the list contains data on each prospect (e.g., demographic, socioeconomic, geographic, psychographic variables) we can start to be selective. The simplest approach is to create *a priori* segments on the basis of some of the variables and conduct an experiment in which the offer is mailed to a sample from each segment. The offer is then rolled out to those segments whose test response rate (ideally adjusted, as noted above) is above a threshold.

A more sophisticated approach involves making the offer to a random sample of the list and recording each contacted prospect's response. Using logistic regression, discriminant analysis, CHAID, CART, or some more advanced method, the analyst builds a model that identifies those prospect characteristics that are predictive of

---

[4]While a sale may be the "direct response" to the advertisement, it is frequently a referral (i.e., the individual revealing that they are a prospect very interested in the product or service being advertised), which may or may not result in a sale. Calli et al. (2012) and Tellis et al. (2000) are examples of work that model the response to direct response advertising on radio and/or TV, both focusing on referrals and making no distinction between new referrals (i.e., prospects) and repeat customers.

response to the offer (e.g., Bult 1993).[5] This model is then used to score the rest of the prospects, with those above a threshold being contacted and the rest ignored. Such an approach can be extended to test different offers, with the goal of identifying which offer to send to which types of customers (Hansotia and Wang 1997).

The test response rate or probability of response rollout threshold is based on a breakeven calculation (i.e., roll out if expected profit > 0). While this could be the expected profit associated with the new customer's first transaction with the firm, it has long been recommended that it should be based on the expected lifetime value of a new customer—see Simon (1967) and Petrison et al.'s (1997) discussion of industry practices in the 1940–1960s. Simon (1993) suggests doing so using data from a sample of 300 customers—active and inactive—who first bought over 3 years ago. (More sophisticated approaches for calculating lifetime value are discussed in Sect. 10.3.1.)

Just as prospects are expected to vary in their propensity to respond, they can also differ in their value to the firm assuming they respond. As such, it may make sense to target those prospects with lower response probabilities but higher value given acquisition than ones with higher response probabilities but lower value given acquisition. In order to take such an approach, we must model expected customer value (given acquisition) as a function of the prospect covariates (Hansotia and Wang 1997). Ainslie and Pitt (1998) go one step further, modeling response to the mailing, profitability given acquisition, and riskiness; also see Liu et al. (2015) for a consideration of risk in the form of bad debt. When developing such models, it is important to control for sample selection bias (e.g., Vaidya and Cassidy 1999).

The work discussed above considers the questions of who to contact and (to a lesser extent) with what message. Another question is what to do with those prospects who do not respond to the mailing. Should the marketer send a second solicitation? A third? Buchanan and Morrison (1988) develop a model of consumer response to direct mail solicitations that can be used to determine the number of profitable solicitations for a customer acquisition campaign. Rao and Steckel (1995) extend Buchanan and Morrison's model to accommodate descriptor variables that characterize the individuals on the list of prospects. In turn, their model is extended by Ehrman and Funk (1997) and Pfeifer (1998) to account for non-readers of direct mail.

10.2.2 Beyond Classic Direct Marketing

While the classic direct marketing approach discussed above still holds for some firms, the reality is more complex for most. We only have to reflect on how we were

[5]Note that the vast majority of the response/predictive/classification models presented in the direct marketing related literature are not acquisition focused. Rather, they consider the response to mailings to existing customers (as opposed to prospects). Any model that includes past purchasing behavior as a covariate obviously falls in this category. We review this work in Sect. 10.3.3.

"acquired" by the numerous companies of which we are customers to realize that, regardless of whether the path to acquisition was long and winding or short and direct, our purchase decisions have been influenced by both actions of the firm (be they explicitly focusing on customer acquisition or not) and our interactions with its customer base. (Within the diffusion literature (e.g., Peres et al. 2010), these are labeled as external and internal influence.)

Since the early work of Katz and Lazarsfeld (1955) and Whyte, Jr. (1954), both academics and practitioners have been interested in the impact of word-of-mouth (WOM) on buyer behavior. Developments in electronic communications technologies over the past 15 years have further stimulated this interest. Rather than simply rely on organic WOM, firms are interested in actions that can stimulate WOM, such as seeding campaigns (e.g., Hinz et al. 2011; Libai et al. 2013) and the development of viral marketing campaigns (e.g., Van der Lans et al. 2010). One form of WOM marketing activity that has a particular customer acquisition focus is the referral program, in which existing customers are rewarded when they bring in new customers. See Kumar et al. (2010), Schmitt et al. (2011), and Van den Bulte et al. (2015) for analyses of the effectiveness of such programs.

More generally, several researchers have examined the relative value of customers acquired through different acquisition channels. For example, Steffes et al. (2011) compare internet, direct mail, direct sales, and telesales, Trusov et al. (2009) compare WOM referrals, traditional media, and promotional event activity, Verhoef and Donkers (2005) compare direct-response advertising in mass media, direct marketing, website, and WOM, and Chan et al. (2011) compare Google search advertising and other search engines. Less attention has been paid to the issue of the impact of various acquisition-related promotions on the value of acquired customers. Datta et al. (2015) consider the impact of offering a free trial, while Lewis (2006) considers the impact of introductory discounts.[6,7]

Reflecting on this body of work on customer acquisition, it is clear that Blattberg et al. (2008) are correct when they comment on the limited amount of research on customer acquisition. We are now in a multichannel world in which firms must decide how to allocate their marketing efforts across paid and owned media, as well as on attempts to influence "earned" media and WOM. Technology increasingly allows us to track an individual's online journey to their first purchase, yet the influence of offline activities is still hard to track. The relative importance of different media

[6]It is also important to consider the impact of acquisition campaigns on the behavior of existing customers. For example, offering new customers better deals than existing customers can potentially result in customer dissatisfaction—"I've been a loyal customer for many years and I'm getting a worse deal than new customers!" See Lhoest-Snoeck et al. (2014) for a discussion and examination of these issues.

[7]Other work looks at the long-term implications of customer behaviour at the time of acquisition. For example, Fader et al. (2007) compare the repeat-buying behaviour of two groups of customers that differ in terms of the size of their initial purchase, and find that those with a higher initial transaction value have higher repeat-buying rates and lower attrition rates. Padilla and Ascarza (2017) explore how differences in customer behavior at the time of acquisition explain differences in expected value and customers' sensitivity to marketing actions.

changes as prospects are developed (e.g., Carroll 2006) and it can be argued that the nature of the message in a given media channel should change as the prospect is developed (e.g., Lambrecht and Tucker 2013). Furthermore, it must be recognized that many non-acquisition-specific activities (e.g., brand advertising, PR) have a positive impact on customer acquisition, even if partialling out their effect is difficult. There is a lot of scope for researchers to develop models that help the manager answer acquisition-related questions such as "How much should we spend on our acquisition activities?"[8] "Whom do we target?" "Which messages do we use in which channels?" and so on.

10.3 Managing Acquired Customers

The notion of customer acquisition, retention, and development being the three key drivers of (organic) growth is widely accepted, and has even made its way into core marketing teaching materials (e.g., Gupta 2014). The logic of these three drivers is clear, and the notion of organizing a firm's activities around these drivers is attractive. However, in many business settings, the distinction between "retention" and "development" activities is not at all clear. For example, is getting the next transaction "retention" or "development"? This blurring of retention and development is also present in a lot of the modeling work by academic researchers. We therefore structure our review of the literature around the idea of managing acquired customers. We start by reviewing the literature on computing customer lifetime value and follow this with an examination of the literature that relates to the topic of churn management. We then review the literature on modeling the response to contacts by the firm, and conclude with a review of work on contact customization.

10.3.1 Computing Customer Lifetime Value

A fundamental marketing metric for any customer-centric firm is customer lifetime value (CLV), which can be defined as "the present value of the future cash flows attributed to the customer relationship" (Pfeifer et al. 2005, p. 17).[9] (The term "customer equity" (CE) denotes the sum of the lifetime values of a firm's customers, both current and future; see Kumar and Shah (2015) for a comprehensive guide to the literature on customer equity.)

As we look to the future, we do not know the customer's lifetime or the timing and nature of their purchasing while they are "alive" as a customer. These quantities

[8]This question is partially addressed by research on allocating marketing expenditures between acquisition and retention activities, which we consider in Sect. 10.4.

[9]This section draws on material presented in Fader and Hardie (2009, 2015). Readers are referred to these references for a deeper review of this literature.

must be considered as random variables and we therefore need to think of *expected* customer lifetime value, $E(CLV)$. Following Rosset et al. (2003), we can express this mathematically as

$$E(CLV) = \int_0^\infty E[V(t)]S(t)d(t)dt\,, \tag{10.1}$$

where, for $t > 0$ (with $t = 0$ representing the "birth" of the customer), $E[V(t)]$ is the expected net cash flow of the customer at time t (assuming they are alive at that time), $S(t)$ is the probability that the customer has remained alive to at least time t, and $d(t)$ is a discount factor that reflects the present value of money received at time t.

It is important to distinguish between the lifetime value of an as-yet-to-be-acquired customer, the lifetime value of a just-acquired customer, and the *residual* lifetime value (RLV) of an existing customer. (The difference between the first two quantities is simply the value of the first transaction that signals the start of the relationship.[10]) We can express the notion of RLV mathematically as

$$E(RLV) = \int_{t'}^\infty E[V(t)]S(t|t > t')d(t - t')dt\,, \tag{10.2}$$

where t' is the "age" of the customer at the point in time where their residual lifetime value is computed.

Reflecting on these definitional formulas, it is important to note that any calculation of CLV or RLV cannot terminate the calculation at, say, three years and call the resulting quantity *lifetime* value. Furthermore, we should not assume that the customer is "alive" (i.e., actively contemplating transactions) throughout the whole of this finite period (cf. Kumar et al. 2008; Rust et al. 2004; Venkatesan and Kumar 2004; Venkatesan et al. 2007). It also raises the fundamental problem of trying to incorporate the effects of time-varying covariates (e.g., marketing activities) in any true calculation of CLV or RLV. Any analyst wishing to do so will need to forecast the values of these covariates far into the future, which clearly introduces a lot of additional noise into the exercise. As a result, the stream of literature that has developed models for computing lifetime value has tended to ignore the effects of time-varying covariates and drawn on the well-established traditions of stochastic models of buyer behavior, which have been part of the marketing science literature from its very beginning (Fader et al. 2014).[11,12]

[10]We may also wish to include the acquisition cost in the calculation of the second quantity.

[11]A related stream of work uses homogeneous Markov chains to characterize customer behavior (e.g., Deming and Glasser 1968; Pfeifer and Carraway 2000; Soukup 1983). Such work does not account for heterogeneity in the underlying behavioral characteristics, which can lead to misleading interferences about the nature of buying behavior (e.g., Frank 1962). See Ching et al. (2004) for an example of how these simple Markov models of customer behavior can be embedded in broader marketing optimization models.

[12]Of course, if it is possible to characterize these time-varying covariates by a separate stochastic process, we could take the expectation of the covariate-dependent process over the distribution of

As we think about operationalizing (10.1) and (10.2), we must ask ourselves whether we are in a business setting where the loss (or "death") of an individual customer is actually observed by the firm (e.g., the customer terminates their contract or fails to renew their fixed-term subscription) or one where it is unobserved (Schmittlein et al. 1987). It is now standard to use the term contractual to characterize a relationship when the death of a customer is observed by the firm, and the term noncontractual to characterize a relationship where the death of a customer is unobserved by the firm.[13,14] The vast majority of businesses fit into this categorization. As we shall see, this categorization underpins most of the tools developed by marketing scientists to support businesses in the management of acquired customers, and we structure our review of literature on computing customer lifetime value around it.

Note that we are starting to see the emergence of some business settings in which the firm has a "hybrid" contractual/noncontractual relationship with its customers (i.e., we can expect observed and unobserved attrition in the same pool of customers). See Ascarza, Netzer and Hardie (2017) for an examination of such settings.

10.3.1.1 Contractual Settings

Since we observe the loss of a customer in contractual settings, it is a straightforward exercise to fit a survival model (also called a duration-time model or hazard-rate model) to the data (thereby giving us $S(t)$). This analysis strategy has been used by a number of researchers examining the correlates of the duration of a customer's relationship with the firm (e.g., Bolton 1998; Jamal and Bucklin 2006; Schweidel et al. 2008a). However, as noted above, including time-varying covariates in a model for $S(t)$ creates problems when we want to use it as the basis for computing CLV or RLV, as the analyst will need to forecast the values of these covariates far into the future.

Standard marketing textbook discussions of CLV use a discrete-time version of (10.1) and express the survivor function in terms of a constant retention rate (i.e., $S(t) = r^t$). While such "CLV formulas" have pedagogical value as a means of introducing the concept of lifetime value to students, they are of limited value in the "real world" (Fader and Hardie 2012, 2014c). If we consider a cohort of customers acquired at the same time, we typically observe that the cohort-level retention rates increase over time (e.g., Kumar and Reinartz 2012, Fig. 5.2), which challenges the textbook assumption of constant retention rates. (It also has implications for work that explores the linkage between the value of a firm's customer base and its stock

---

covariate paths. How the resulting estimates of $E(CLV)$ and $E(RLV)$ would differ from those based on models of customer behavior that do no consider time-varying covariates is an open question.

[13] David Shepard Associates (1999) use the labels contractual and implied; "an implied relationship is one in which there is no obligation on either party's part to do anything in the future" (p. 416).

[14] This is not the same as Jackson's (1985) lost-for-good versus alway-a-share classification. Following Fader and Hardie (2014a), we feel that the contractual versus noncontractual classification is a better way of thinking about the nature of a firm's relationship with its customers, as the notion of latent attrition is missing from the basic always-a-share "model."

market valuation, such as Gupta et al. (2004), Schulze et al. (2012); see McCarthy et al. (2017).) While it is tempting to tell a story of individual-level time dynamics (e.g., increasing loyalty as the customer gains more experience with the firm, and/or increasing switching costs with the passage of time), a far simpler story—and one consistent with the fundamental marketing concept of segmentation—is that of a sorting effect in a heterogeneous population.

A simple stochastic model for the duration of a customer's relationship with the firm that captures the phenomenon of increasing retention rates is the beta-geometric (BG) distribution (Potter and Parker 1964). Despite what may seem to be overly simplistic assumptions, the analyses presented in Fader and Hardie (2007a, 2014b) demonstrate that this two-parameter model generates very accurate forecasts of retention. This model for $S(t)$ can be substituted into (10.1) and (10.2) and used to compute CLV and RLV in contractual settings, something explored in (Fader and Hardie 2010). Note that this work simply focuses on *when* customers choose to terminate their relationship with the firm. Braun and Schweidel (2011) extend this to account for "competition" among the different reasons that ultimately lead to termination.

10.3.1.2 Noncontractual Settings

The challenge of noncontractual settings is that the loss of a customer is not observed and so we cannot estimate any model of $S(t)$ directly from the data. What we do observe are realizations of the product of $V(t)$ and $S(t)$, and the challenge facing the analyst to identify these two components of behavior from the observed behavior. In other words, how do we differentiate those customers with low purchase propensities who have ended their relationship with the firm (without informing it) from those who are simply in the midst of a long hiatus between transactions? While we can never know for sure which of these two states a customer is in, we can use statistical models to make probabilistic statements.

The seminal work in this area is Schmittlein et al. (1987), which introduced a latent-attrition framework in which a customer's relationship with a firm has two phases: they are "alive" for an unobserved period of time, then "dead." Ignoring the effect of random purchasing around their means, individual customers purchase the product at steady but different underlying rates. At different unobservable points in time they "die."[15] In their operationalization of this framework, Schmittlein et al. (1987) assume that (i) while "alive" the customer's purchasing is characterized by the NBD (negative binomial distribution) model (i.e., a gamma mixture of Poissons), and (ii) the unobserved customer lifetimes are treated as-if random and are characterized by the Pareto Type II distribution (i.e., a gamma mixture of exponen-

[15] What lies behind this death? It could be a change in customer tastes, financial circumstances, and/or geographical location, the outcome of bad customer service experiences, or even physical death, to name but a few possible causes. But given the modeling objectives, why this death occurs is of little interest to the analyst; the primary goal is to ensure that the phenomenon is captured by the model.

tials); the resulting model of buyer behavior in noncontractual settings is called the Pareto/NBD.

Empirical validations of the model are presented in Schmittlein and Peterson (1994) and Fader et al. (2005b), amongst others; its predictive performance is impressive. Applications of this model include the work of Reinartz and Kumar (2000, 2003) on customer profitability, Hopmann and Thede (2005) on "churn" prediction, and Huang (2012) and Wübben and v. Wangenheim (2008) on managerial heuristics for customer-base analysis.

The basic Pareto/NBD model has been modified and extended by a number of researchers. An important stream of work has focused on variants that are easy to implement, resulting in the BG/NBD (Fader et al. 2005a) and BG/BB (Fader et al. 2010) models, both of which can be implemented in a standard spreadsheet environment. Other work has relaxed the assumption of Poisson counts, exponential lifetimes and/or gamma heterogeneity (e.g., Abe 2009; Bemmaor and Glady 2012; Jerath et al. 2011; Platzer 2008; Singh et al. 2009), explored estimation issues (e.g., Jerath et al. 2016; Ma and Liu 2007), allowed for multi-category purchasing (Park et al. 2014), or relaxed the "buy"/"die" nature of customer behavior (e.g., Ma and Büschken 2011; Romero et al. 2013; Schwartz et al. 2014).

Several researchers have sought to incorporate the effects of covariates. While this is easy for the case of time-invariant covariates (e.g., Abe 2009; Fader and Hardie 2007b), the inclusion of time-varying covariates is less straightforward. Schweidel and Knox (2013), Schweidel et al. (2014) and Padilla and Ascarza (2017) build on the foundations of the BG/BB model, allowing covariates to influence the customer's behavior while alive and/or their likelihood of dying: Schweidel and Knox (2013) incorporate the effects of direct marketing activity, while Schweidel et al. (2014) incorporate the effects of past customer activity, and Padilla and Ascarza (2017) incorporate the effects of firm communications (both email and direct marketing) and new product introductions. Braun et al. (2015) and Knox and Van Oest (2014) both build on the foundation of the BG/NBD, allowing covariates to impact the latent attrition process: Braun et al. (2015) incorporate the effects of the customer's service experience, while Knox and Van Oest (2014) incorporate the effects of customer complaints.

The Pareto/NBD (and the variants discussed above) is a model for the flow of transactions. Models for spend per transaction were proposed by Colombo and Jiang (1999) and Schmittlein and Peterson (1994). Despite all the components being in place, Fader et al. (2005b) were the first to bring them together via (10.1) and (10.2) to come up with explicit formulas for CLV and RLV (conditional on the customer's observed behavior) in noncontractual settings. Drawing on the work of Colombo and Jiang (1999) and Schmittlein et al. (1987), their key result is that we only need to know three things about a customer's buying behavior in a given time period in order to compute their residual lifetime value: recency, frequency, and monetary value (i.e., RFM). Fader et al. (2005b) model spend per transaction and assume a constant margin. McCarthy et al. (2017) extend this by allowing for heterogeneity in margin per transaction.

10.3.2 Churn Management

The defining characteristic of contractual settings is that attrition is observed. For most firms operating in such settings, churn rate is an important KPI, and the management of churn is of great interest to decision makers. For many firms, the efforts to retain a customer are *reactive*; for example, a mobile phone operator offers some incentive to a customer who indicates that they wish to cancel their contract, etc. Increasingly, firms are becoming *proactive*, undertaking their retention marketing activities before the customer has the opportunity to churn. (See Passant (1995) for an early critique of retention marketing practices.) Both logic and limited budgets mean that a firm's retention efforts should be focused on a subset of those customers whose contracts are coming up for renewal. (Why, for example, spend money trying to retain a customer who has no intention of churning?) As a result, a key component of any proactive churn management exercise is a churn model that predicts a customer's likelihood of churning (voluntarily).[16]

When developing a standard churn model, the dependent variable of interest is binary—whether or not the customer churned in a given time interval. The predictor variables are measured over a specified time period that ends at or before the start of the churn interval. As discussed below, numerous researchers working in the areas of data mining/machine learning, marketing, and statistics have studied the problem of which predictor variables to use and what analysis methodology to use to identify the relationship between churn and the predictor variables.

The predictor variables are typically customer characteristics, measures of customer behavior (e.g., utilization of the service), and their interactions with the firm (e.g., calls to customer service). See Ballings and Van den Poel (2012, Table 1), Lemmens and Croux (2006, Table 1) and Zhang et al. (2012, Table 1) for illustrative lists of the variables commonly used in churn models. Note that these variables focus on the individual customer, ignoring the broader context in which (s)he operates. In recent years, a number of researchers have also considered variables that capture interpersonal influence (e.g., Dasgupta et al. 2008; Haenlein 2013; Verbeke et al. 2014; Zhang et al. 2012),[17] finding that a customer is more likely to churn if individuals with whom (s)he is connected have recently churned from the service provide. Ascarza, Ebbes et al. (2017) show that retention campaigns can have a positive impact (in term of usage and retention) on non-targeted customers who are connected to the targeted customers.

The classic statistical technique used to develop a churn model is logistic regression. Other methods include decision trees (e.g., C4.5, CART), neural networks, support vector machines, and ensemble methods (e.g., random forests, bagging, boost-

[16]Churners are typically categorized as voluntary or involuntary. Voluntary churn occurs when the customer decides to terminate their relationship with the firm, whereas involuntary churn occurs when the firm terminates the relationship (e.g., as a result of nonpayment or fraud). Involuntary churners are typically excluded when developing a churn model or modeling survival (cf. Braun and Schweidel 2011).

[17]Nitzan and Libai (2011) examine such effects in a duration time (i.e., survival) model.

ing). Illustrative marketing studies include Coussement and Van den Poel (2008), De Bock and Van den Poel (2011), Larivière and Van den Poel (2005), and Lemmens and Croux (2006). A number of studies have been undertaken, comparing and contrasting the various analysis methods, two recent examples being Verbeke et al. (2011) and Verbeke et al. (2012). Other issues that have received less attention include the length of the time period over which the predictor variables are measured (Ballings and Van den Poel 2012), the "staying power" of the model (Risselada et al. 2010) (i.e., for how long can the estimated model be used before its parameters need to be re-estimated or different variables added to the model), and the development of churn models in situations where privacy concerns limit the amount of data available to the analyst (Holtrop et al. 2017).

How does the model builder determine the best model specification? Since the standard objective of a churn model is to identify those customers with the highest risk of churning (with the view of targeting them with some proactive retention campaign), it is common to assess the performance of any given model specification on a validation sample in terms of its top-decile lift (which is the proportion of actual churners in the 10% of customers that the model predicts as having the highest risk of churning). Depending on the setting, this can be reduced to the top 5% (or smaller).

Building on Neslin et al. (2006), Blattberg et al. (2008) propose a framework for identifying the tradeoffs inherent in a (single) proactive churn management campaign. With reference to Fig. 10.1, the set of customers at risk of churning is split into those customers that are contacted/targeted with the retention campaign (i.e., those at most risk according to the churn model) and those that are not. Among those contacted (α) at cost c per customer with an incentive valued at δ, a proportion would have been churners (β), and among those, a proportion will be "rescued" (γ) given the company's intervention. (For a customer with lifetime value CLV,[18] the realized value is $CLV - c - \delta$.) The rest of those contacted $(1 - \beta)$ would not have been churners, yet some (ψ) might take the incentive with an expected increase in their lifetime value of Δ but at a cost of $c + \delta$. (Note that this ignores the possibility that contacting "not-would-be churners" can actually result in their churning (Ascarza et al. 2016).) As we consider the profitability of a proactive retention campaign, the tradeoff is between the upside effects of the campaign (coming from the lifetime value (CLV) obtained from the "rescued" customers, $\beta\gamma CLV$, and those would-be non-churners that take the incentive, $(1 - \beta)\psi\Delta CLV$ (assuming $\Delta > 0$)) and the downside effects of the campaign (coming from the costs of contacting the customer, c, and the expected cost of the incentive, $[\beta\gamma + (1 - \beta)\psi]\delta$).

This framework was initially used by Neslin et al. (2006) to evaluate a set of models developed in a churn modeling tournament. Verbeke et al. (2012) use this framework to develop a new model selection criterion. Rather than choosing the churn model specification that maximizes lift for some arbitrary fraction of the customer

[18]Neslin et al. (2006) and Blattberg et al. (2008) use the abbreviation LTV, which we have replaced with CLV. Strictly speaking, this should be $E(RLV)$, but the distinction raised in (10.1) and (10.2) is ignored in most of the literature, including the work of Blattberg et al. (2008) and Neslin et al. (2006).

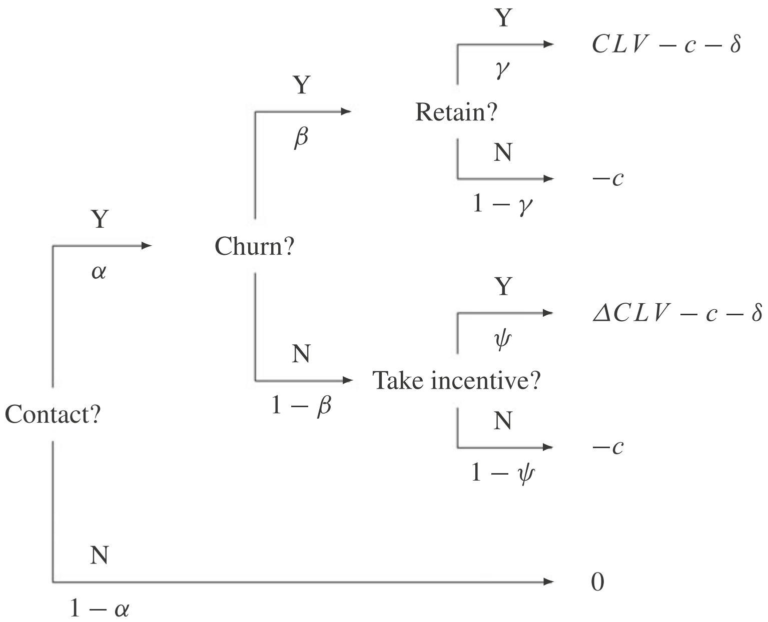


Fig. 10.1 A profitability framework for proactive churn management (after Blattberg et al. 2008, Fig. 24.6)

base (10% for the top-decile lift criterion), they propose the maximum profit (MP) criterion, which calculates the profit "generated by including the optimal fraction of customers with the highest predicted probabilities to attrite in a retention campaign" (p. 211). (See Lemmens and Gupta (2017) for a similar approach to model estimation and selection.) This is extended by Verbraken et al. (2013) to account for uncertainty in campaign costs and benefits.

Similar profit-based frameworks have been developed by Mozer et al. (2000) and Piatetsky-Shapiro and Masand (1999). Rosset et al. (2003) present a framework for campaign management based on the expected change in CLV resulting from the associated intervention; they explicitly recognize that an intervention could reduce the customer's long-term underlying propensity to churn, or simply lock them in for a fixed period of time without any change in their underlying propensity to churn.

Reflecting on the development of churn models, Hansen (2015) makes the following comment: "The business objective is never 'predict the churners', it is 'reduce the value and rate of churn'. For that, predicting the churner is simply a first step to taking action to dissuade the potential churner. That may sound like two distinct steps but do not waste time improving identification of churners, *focus on identifying those that can be dissuaded* [emphasis added]." At first glance it would appear that profit-based model selection criteria address this. But this is not the case. All this work targets those with the highest risk of churning, but ignores the fact that many of those with a high risk of churning are very dissatisfied and cannot be dissuaded (at least profitably) by the firm's intervention. Recognizing this, Ascarza (2016) proposes an alternative approach to proactive churn management in which the firm targets those customers with the highest sensitivity to the intervention.

Many firms would simply see "lost" customers as re-entering the prospect pool for future acquisition campaigns. Other firms have specific programs that attempt to reacquire/reactivate lost customers, a practice known as "customer winback" (Griffin and Lowenstein 2001; Stauss and Friege 1999). Gerpott and Ahmadi (2015), Kumar et al. (2015), and Thomas et al. (2004) develop models that can be used to guide such decisions; also see Pick et al. (2016).

10.3.3 Contact Response Models

By definition, attrition is unobserved (and unobservable) in noncontractual settings. As a result, there is no standard attrition-related KPI. Instead, the focus tends to be on purchasing by individuals in the firm's customer database, and this has driven most of the modeling efforts in this space.

In a classic direct marketing setting, mailings are sent out at regular intervals (e.g., quarterly for a charity, monthly for a mail-order company) and the customer may or may not respond (i.e., make a donation or purchase) to the contact. Over time, the firm grows its customer database in which it records the identity of those customers contacted on each mailing and their response. Given printing and mailing costs, budget constraints mean that it may not be possible to contact every customer; even in the absence of any explicit budget constraints, we can assume that it is not profitable to contact every customer. Therefore, for any given campaign, the key decision facing the firm is which customers it should contact.

Historically, the gold-standard approach to supporting this decision was to run an experiment in which the customer base was segmented using some *a priori* scheme and the mailing sent to a sample of customers from each segment. As with the use of experiments for customer acquisition, the firm would "roll out" to those segments where the response rate was above some threshold. A common segmentation scheme is the quintile-based RFM method championed by Hughes (1996). Customers are first ranked on the basis of how recently they made a purchase and are divided into quintiles, with 5 denoting the top quintile (i.e., most recent purchasers) and 1 the bottom quintile (i.e., least recent purchasers). This is repeated on the basis of how many times they made a purchase in a given time period, and then on the basis of their average spend per order in that time period. Thus each customer has a "recency" (R) coding, a "frequency" (F) coding, and a "monetary value" (M) coding, resulting in $5 \times 5 \times 5 = 125$ segments.

However, given the information in the customer database, it is not necessary to undertake such an experiment.[19] For those individuals contacted in the most recent mailing(s), we know their characteristics and whether or not they responded to the

[19]Of course, a firm will make use of experiments to determine the best mailing package (e.g., Bult et al. 1997).

mailing. We can therefore build a contact response model in which response to the mailing (Y/N) is modeled as a function of the customers' purchase histories (often summarized in terms of recency, frequency, and monetary value) and demographics.[20] This can be done using a basic logit or probit model, semiparametric methods such as the Cosslett estimator (Bult and Wansbeek 1995), or one of the other methods used in the development of acquisition scoring and churn models. The whole database can now be scored and customers ranked on the basis of their predicted probability of responding to a mailing. Those customers with a response probability above a certain threshold (often profit-based) are sent the mailing. Bult and Wansbeek (1995) propose a targeting method that seeks to maximize expected profit. Gönül et al. (2000) consider an alternative targeting rule for a catalog-based mail-order company: contact the customer only if the expected profit with the mailing a catalog exceeds the expected profit without the mailing a catalog.[21] Whatever decision rules used, they typically ignore estimation uncertainty in the parameter estimates and can therefore lead to suboptimal decisions. Muus et al. (2002) derive an optimal Bayes decision rule to address this problem.

These standard scoring models often suffer from the problems of selection bias and endogeneity, the first occurring because the firm's targeting rules mean that the model is estimated on a non-random sample of customers, the second typically occurring because of the use of variables that summarize the customer's purchase history (often summarized in terms of RFM variables). The use of a targeting rule that is correlated with the customer's past behavior can lead to biased estimates of the coefficients associated with the RFM variables. Solutions such as the use of instrumental variables, policy functions, and latent trait models have been explored by, amongst others, Cui et al. (2006), Donkers et al. (2006), Hruschka (2010), Rhee and McIntyre (2008, 2009), and Rhee and Russell (2009).

The standard scoring model considers whether or not the contacted customer responds to the mailing. In most situations, we are not just interested in whether or not someone responds but also in the nature of their response (e.g., how much they donate/buy). A number of researchers have proposed models of both phenomena (e.g., Donkers et al. 2006; Levin and Zahavi 1998; Otter et al. 2000; Schröder and Hruschka 2016; van Diepen et al. 2009), with the Type II Tobit being the most common model. Other researchers have gone further, considering the issue of returns (e.g., Baumgartner and Hruschka 2005; Koning et al. 2002).

The decision considered above is who to contact with a given single mailing. However, the reality is that a company is making multiple mailings over a given

[20]In addition to the frequency of response, a number of researchers have considered the impact of contact history (e.g., frequency of contact) on the customer's likelihood of responding to the current mailing, including possible irritation effects; see Schröder and Hruschka (2016) for a review. This is especially an issue in today's permission marketing world where, for example, too much contact could result in the customer opting-out of communications all together (e.g., Drèze and Bonfrer 2008).

[21]Whereas the work discussed above implicitly assumes that customers can only place an order (i.e., respond) in a given period if they receive a mailing, Gönül et al. (2000) recognize that customers can place orders from old catalogs, even though they did not receive a catalog in the current period.

period of time. Applying selection rules mailing by mailing can lead to suboptimal outcomes for the firm; see, for example, Elsner et al. (2003, Fig. 1). As Kestnbaum et al. (1998, p. 58) note, "If a customer is not selected because he or she falls a little below the cutoff point used for the decision criterion, the customer is not contacted. This may happen for every campaign, so the customer is *inadvertently abandoned*. Receiving no contacts for an extended period of time, he or she is not very likely to buy and the poor performance becomes worse. Rather than abandon a customer by default, wouldn't it be better to make a conscious decision to make one or two contacts per year or to stop contacting a customer based on the overall prognosis for that particular customer?" As a result, the decision problem changes from whether or not to contact each customer to one of how many mailings to send to each customer over a given time period. Proposed solutions to this problem include Bitran and Mondschein (1996), Elsner et al. (2003, 2004), Gönül and Ter Hofstede (2006), Jonker et al. (2006), Piersma and Jonker (2004), Neslin et al. (2013), and Simester et al. (2006).

Note that the work reviewed in this section models transactions *given* contact by the firm. Another stream of research models the flow of transactions (and the associated cash flows) in time, without any explicit conditioning on contact by the firm; see, for example, Kumar et al. (2008), Venkatesan and Kumar (2004) and Venkatesan et al. (2007). The more sophisticated models account for underlying dynamics in customer behavior, typically using hidden Markov models (HMMs) (e.g., Chang and Zhang 2016; Mark et al. 2013, 2014; Montoya et al. 2010; Netzer et al. 2008). Work in the tradition of Schmittlein et al. (1987)—as discussed in Sect. 10.3.1.2 above—can be viewed as a constrained form of such HMMs (Schwartz et al. 2014).

10.3.4 Contact Customization

The work discussed above focuses on increasing response rates by using scoring models to improve targeting given an offer. An alternative approach focuses on increasing response rates by improving the relevance of the offer to each customer via some form of customization (Malthouse and Elsner 2006). (Of course these two approaches are not mutually exclusive). The nature and scope of the customization obviously depends on the media used by the firm when contacting the customer; it is obviously far cheaper to customize emails than it is paper mailings.

When talking about customization, we typically think of customizing the offer given the decision to contact. A variant practiced by a number of catalog retailers that have both general and category-specific catalogs considers which subset of all catalogs that will be mailed over a given time period to send to each customer so as to maximze some profit-related objective function (e.g., Campbell et al. 2001; Elsner et al. 2004; George et al. 2013).[22]

[22]Note that most of this work has focused on which *products* to offer to the firm's customers. Khan et al. (2009) develop a model for determining which *promotions* to offer, if any, over a finite planning

Looking beyond catalog retailers, the idea of customization goes hand-in-hand with the concepts of cross-selling and up-selling (which lie the heart of any discussion of "customer development"). Cross-selling is where the firm tries to get the customer to buy products from the firm's product line that the customer does not currently own, and up-selling is where the firm tries to get the customer to buy more expensive variants of (or add-ons to) products they are buying (or currently own). In some digital settings, the goal is simply to increase usage (with no distinction being made between cross- and up-selling); see, for example, Ansari and Mela (2003) and Chung et al. (2016).

The simplest form of cross-selling model is the so-called "next product to buy" model, which models the purchase of a focal product as a function of current product ownership (e.g., Knott et al. 2002) and similarity to other customers (e.g., Moon and Russell 2008). Such a model can be used to identify who to target when next promoting that product. Such an approach to cross-selling is very campaign focused; Li et al. (2011, p. 684) argue that a more customer-centric approach to cross-selling asks "How do we introduce the right product to the right customer at the right time using the right communication channel to ensure long-term success?"

In settings where the decision is which one of several possible products the firm should feature as their recommended product, Bodapati (2008) argues that the firm should not automatically choose the product that has the high probability of purchase (given the customer's purchase history), but rather focus on the product whose purchase probability increases the most with recommendation. (Why recommend a product the customer was going to buy anyhow?)

A number of researchers have explored the idea that consumers acquire certain non-consumable products (e.g., financial products, durables) in the same order (e.g., Kamakura et al. 1991; Paas 1998). Building on such acquisition pattern analysis, a number of researchers have developed models that aim to predict which product will be acquired next by each customer (e.g., Li et al. 2005, Prinzie and Van den Poel (2006) and when (e.g., Prinzie and Van den Poel 2007), the output of which can be used to customize the next mailing sent to each customer.

Implicit in most of the work on cross-selling is the idea that the objective of the solicitation is to generate an immediate purchase. Li et al. (2011) suggest that cross-selling solicitations can have an educational and advertising effect in addition to the immediate promotional effect. Furthermore, customers differ in their preference for communication channels (e.g., mail versus email). They develop a model for making decisions about *when* to promote *which* product to *which* customer via *which* communication channel that takes into consideration the short- and long-term effects of any solicitation.

The issue of up-selling has received less direct attention. Working in a single category setting, Kim and Kim (1999) use a stochastic frontier model to estimate the inefficiency of the firm's customer-specific marketing activities, from which an estimate of each customer's upselling potential is calculated. Verhoef and Donkers

---

horizon. The promotions they consider are transaction, not product, specific (i.e., free shipping offers, discount coupons, and loyalty program rewards).

(2001) look at predicting customer potential value in a multicategory setting; this is, of course, identifying customers with both cross-selling and up-selling potential. In the same spirit is the work on estimating share of wallet (e.g., Chen and Steckel 2012; Du et al. 2007). Ballings and Van den Poel (2015) consider the task of identifying Facebook users who are expected to increase their usage frequency, with the view that those predicted not to increase their usage can be targeted with campaigns designed to increase their engagement with the social network. In all cases, the goal is identifying who to target, rather than what should be promoted as part of any up-selling campaign.

More generally, the question of what offer the customer should receive (be it for the purpose of cross-selling or up-selling) is a recommendation problem. Since the mid-1990s, a large number of researchers in both academia and industry (especially those with a computer science background) have focused on the problem of developing computer-based systems that generate recommendations, i.e., recommender systems. (Ansari et al. (2000) and Ying et al. (2006) are early examples of such work by marketing researchers. See Adomavicius and Tuzhilin (2005) for a survey of the early literature, and Ricci et al. (2015) for a comprehensive coverage of current methods and applications.) While such systems are used to generate a set of products to offer as part of a cross-selling or up-selling campaign (i.e., to customize communications from the firm), they are more widely used to customize the recommendations a customer sees when on the firm's website.[23]

Generally speaking, researchers in marketing have downplayed the optimization-related challenges associated with the design of large-scale cross-selling and up-selling campaigns. Various solutions have been proposed by operations researchers and computer scientists—see, for example, Cohen (2004), Delanote et al. (2013), Lu and Boutilier (2014), and Nobibon et al. (2011)—and these deserve consideration by marketing scientists (probably working with experts in optimization).

10.4 Coordinating Acquisition and Retention

While acquiring customers and managing acquired customers are important activities, they do not occur independently, isolated from one another. A fundamental question facing the manager of a customer-centric firm is how to allocate their marketing budget across various acquisition, retention, and development activities.

Several researchers have explored the trade-off between acquisition and retention spend using analytical models, with Fruchter and Zhang (2004) and Musalem and Joshi (2009) considering the case of a competitive environment, Lianos and Sloev (2013) considering the case of a monopolistically competitive industry, and Ovchinnikov et al. (2014) investigating the case of a firm facing capacity constraints.

[23] Once at the firm's website, a further form of customization matches the "look and feel" of a website to each customer (e.g., Hauser et al. 2009, 2014).

In their classic paper that introduced the idea of customer equity, Blattberg and Deighton (1996) present side-bar examples in which a simple decision calculus model is used to determine the optimal level of acquisition spending and another simple decision calculus model is used to determine the optimal level of retention spending. Berger and Bechwati (2001) build on this work to develop a model for optimizing the allocation of a promotion budget between spending on acquisition and retention. An real-world application of this model is presented in Berger and Bernstein (2002), and it is extended in Dong et al. (2007) to accommodate the notion of (acquisition) channel quality, which captures dependencies between acquisition and retention. Building on Dong et al., Swain et al. (2014) develop a model that allows the decision maker to explore the impact of margin-reducing incentives (such as discounts) that increase acquisition rates, but which may attract the "wrong" types of customers, on customer equity maximization.

This basic formulation is somewhat artificial in nature as it is a static/single-period formulation in which we have a fixed prospect pool, and the firm is looking at how much to spend this period per prospect. The assumption is that the firm will spend the same amount on retention in perpetuity, but the optimality of this is ignored given the static nature of the problem formulation. Fan and Berger (2001) consider the problem of how to allocate the fixed promotion budget over a finite time horizon with the objective of maximizing customer equity. Other research that builds on the primitives of Blattberg and Deighton (1996) includes Blattberg et al. (2008, Chap. 28), Calciu (2008), Pfeifer (2005), and Pfeifer and Ovchinnikov (2011).

All this work models each period's retention rate as a function of "retention spend" in that period (ignoring the phenomenon of cohort-level dynamics in retention rates); as such, it only applies to contractual settings. The analog formulation for noncontractual setting is not immediately obvious.[24] We also note that any customer development-related activities are excluded from the resource allocation exercise.

This work has a macro view, looking at high-level resource allocation. The more micro view considers the tradeoff at the level of the mailing (or, more generally, contact) decision. Bitran and Mondschien's (1996) model for the development of optimal mailing policies explicitly considers the trade-off between mailing to prospects on rented lists and mailing to existing customers. Mailing to prospects may not be profitable in the short term but is an investment for the future; mailing to existing customers allow the firm to reap the rewards of past such investments. In a fundraising setting, Stanford et al. (1996) present a linear programming model for determining the number and type of mailings to send to different groups of prospects and current

[24] At first glance, it may appear that the work that embeds Blattberg and Deighton's (1996) model in a brand-switching framework (e.g., Tsao et al. 2014; Williams and Williams 2015) would work in a noncontractual setting. However, this is not the case; the fact that someone purchases from a competitive firm between two purchases from the focal firm should not necessarily mean that they churned after the first purchase and were acquired (again) when they made their second purchase. The notions of acquisition and retention implicit in such models are quite different from those implict in most of the literature reviewed in this chapter. (See Fader and Hardie (2014a) for a further discussion of issues related to the treatment of competition in noncontractual settings.).

donors. The goal is to maximize funds raised given various constraints, including the size of the marketing budget and various prospecting goals.

In addition to the obvious dependency created by a budget constraint, we would intuitively expect there to be some additional relationship between acquisition and retention that should be taken into consideration when considering this resource allocation problem. For example, customer acquisition campaigns designed to acquire as many customers as possible may come back to haunt the firm when the retention team tries to retain what turns out to be the "wrong types" of customers. Research suggests these two processes are not independent. Thomas (2001) finds that the duration of a customer's relationship with the firm is correlated with their likelihood of being acquired in the first place; the same result is found by Reinartz et al. (2005). Schweidel et al. (2008b) find that a customer's relationship duration is correlated with the speed with which they were acquired (having entered the prospect pool). (In particular, customers who are acquired more quickly tend to have shorter relationships than those who took longer to start their "relationship" with the firm.) As discussed in Sect. 10.2.2, the value of customers can be a function of acquisition channels and type of acquisition-related promotion. These issues are largely ignored in the existing literature and deserve consideration in future work.

10.5 Discussion

We have reviewed the key data-based tools and methods that have been developed by researchers to assist a customer-centric firm in its customer acquisition, retention, and development activities. It is clear that while certain topics and/or industries have received a lot of attention, there are many opportunities for further research.

The whole topic of customer acquisition is under-addressed, with much of the published work coming from traditional direct mail settings. Given the recognition that customer retention and value varies across method and channel of acquisition, acquiring the "right" customers in the first place is of vital importance. Today's multichannel world poses a number of challenges. Which channels should be used? How should the message be tailored across channels and over time in recognition of the prospect's path to acquisition? How do we integrate online and offline activities? How do we integrate "direct" and "broadcast" activities? How do we account for the impact of non-acquisition-specific marketing activities on customer acquisition? How do we trade off the desire to acquire "quality" customers with the pressure for "quantity"?

As we reflect on the management of acquired customers, it is important to make the distinction between contractual and noncontractual settings. In contractual settings, the vast majority of research has focused on the problem of predicting churn. As we reflect on the management problem these models are supposed to address, we realize that developing a model that best-predicts churn could be missing the point. Managers need support in developing the best interventions to *reduce* churn, which may see them ignoring those customers with a high risk of churning (Ascarza 2016).

The whole issue of product and service usage while under contract has received little attention (cf. Ascarza and Hardie 2013). Similarly, the task of customer development has received little attention in contractual settings. A notable exception is the work of Thomas et al. (2015), which looks at the effects of two different types of campaigns (one focusing on retention, the other on customer development) in an opt-out setting. Customer development in contractual settings will typically involve signing up for a higher level of service and/or multiple services offered by the same firm. The associated modeling issues, including that of modeling switching between contracts of different lengths, have received little attention (cf. Heitz et al. 2011; Schweidel et al. 2011).

In noncontractual settings, much of the work has been campaign-oriented in nature, developing a model of customer response to a contact (typically some form of direct mail) and using that to decide on whom to target. While such a campaign orientation reflects the realities of marketing practices in most firms, this work has typically ignored the possible distinction between retention- and development-oriented activities. Furthermore, it is typically the case that the objective of the contact is to trigger an immediate purchase. We need to think more broadly, considering the advertising and educating roles of the firm's communications (e.g., Li et al. 2011), including simply keeping the firm in top-of-mind. In today's multichannel world, an added challenge is determining which channel and message to use to contact each customer (which could vary according to the objective of the communication). All this raises a number of optimization-related issues when implemented in real-world settings. Furthermore, while technological developments have created a data-rich world in online settings, the challenge is how to integrate online and offline activities, especially as omnichannel operations become more and more the norm. Finally, there is the challenge of controlling for selection bias and endogeneity when developing the response models that underpin such work.

An additional challenge in today's multichannel world is that of allocating "proportional credit to marketing communications and media activity across all channels, which ultimately leads to the desired customer action" (Moffett 2014, p. 3). How much credit should be given to the first- versus last-touch and the touchpoints in-between the two on the customer's path to purchase? This complex attribution problem is starting to receive the attention of researchers in both marketing and computer science (e.g., Abhishek et al. 2015; Li and Kannan 2014; Shao and Li 2011; Xu et al. 2014). Most of this work has focused on trying to attribute credit for a given transaction, failing to make the distinction between the first purchase that signals the start of the customer's relationship with the firm and subsequent transactions. A customer-centric firm will want attribution models that allow it to understand the impact of its activities (and the impact of other customers) on the acquisition, retention and development of its customers. (See Kannan et al. (2016) for an introduction to the topic.)

Reflecting on the methods developed to assist customer-centric firms in their targeting decisions, we want to bring attention to the use of incremental (or uplift) modeling. Rather than merely estimating the likelihood of a customer engaging in a certain behavior (e.g., churning), uplift models estimate the incremental impact of the firm's actions on such behavior (e.g., the difference in churn probability with and without the marketing intervention). Furthermore, they also recognize customer heterogeneity with respect to the incremental impact of the marketing intervention, identifying customers who will be most sensitive to specific marketing campaigns. While the literature has made some steps in this direction (e.g., Gönül et al. (2000) in the context of catalog mailing, Bodapati (2008) in the context of product recommendations, and Ascarza (2016) in the context of proactive churn management), an explicit focus on the incremental effect of targeted marketing campaigns has been more the exception than the norm. We encourage managers and researchers to re-orient their focus and seek to maximize the effectiveness of the marketing actions by comparing the expected behavior given the action to the counterfactual of what the behavior would have been in the absence of the intervention.

Finally, the issue of balancing the firm's acquisition and retention (let alone development) activities has received little attention, especially in noncontractual settings. There are questions of how to allocate a budget across various activities, as well as setting the size of the budget in the first place, at both the macro- and micro-level of the firm, while accounting for underlying dependencies between the activities.

References

Abe, M. 2009. "Counting your customers" one by one: A hierarchical Bayes extension to the Pareto/NBD model. *Marketing Science* 28 (May–June): 541–553.

Abhishek, V., P.S. Fader, and K. Hosanagar. 2015. Media exposure through the funnel: A model of multi-stage attribution. http://ssrn.com/abstract=2158421. Accessed 10 Aug 2016.

Adomavicius, G., and A. Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* 17 (June): 734–749.

Ainslie, A., and L. Pitt. 1998. UniBank and the analysis of the ExcursionCard customer database: A practical application of statistical techniques in database marketing. *Journal of Interactive Marketing* 12 (Summer): 57–66.

Allenby, G.M., and R.C. Blattberg. 1987. A new theory of direct market testing. *Journal of Direct Marketing* 1 (Autumn): 24–37.

Almquist, E., and G. Wyner. 2001. Boost your marketing ROI with experimental design. *Harvard Business Review* 79 (October): 135–141.

Ansari, A., S. Essegaier, and R. Kohli. 2000. Internet recommendation systems. *Journal of Marketing Research* 37 (August): 363–375.

Ansari, A., and C.F. Mela. 2003. E-customization. *Journal of Marketing Research* 40 (May): 131–145.

Ascarza, E. 2016. Retention futility: Targeting high risk customers might be ineffective. http://ssrn.com/abstract=2759170. Accessed 10 Aug 2016.

Ascarza, E., P. Ebbes, O. Netzer, and M. Danielson. 2017. Beyond the target customer: Social effects of CRM campaigns. *Journal of Marketing Research*, forthcoming.

Ascarza, E., and B.G.S. Hardie. 2013. A joint model of usage and churn in contractual settings. *Marketing Science*, 32 (July–August): 570–590.

Ascarza, E., R. Iyengar, and M. Schleicher. 2016. The perils of proactive churn prevention using plan recommendations: Evidence from a field experiment. *Journal of Marketing Research* 53 (February): 46–60.

Ascarza, E., O. Netzer, and B.G.S. Hardie. 2017. Some customers would rather leave without saying. Mark Sci, forthcoming.

Ballings, M., and D. Van den Poel. 2012. Customer event history for churn prediction: How long is long enough? *Expert Systems with Applications* 39 (18): 13517–13522.

Ballings, M., and D. Van den Poel. 2015. CRM in social media: Predicting increases in Facebook usage frequency. *European Journal of Operational Research* 244 (1): 248–260.

Baumgartner, B., and H. Hruschka. 2005. Allocation of catalogs to collective customers based on semiparametric response models. *European Journal of Operational Research* 162 (3): 839–849.

Bell, G.H., J. Ledolter, and A.J. Swersey. 2006. Experimental design on the front lines of marketing: Testing new ideas to increase direct mail sales. *International Journal of Research in Marketing* 23 (September): 309–319.

Bemmaor, A.C., and N. Glady. 2012. Modeling purchasing behavior with sudden "death": A flexible customer lifetime model. *Management Science* 58 (May): 1012–1021.

Berger, P.D., and N.N. Bechwati. 2001. The allocation of promotion budget to maximize customer equity. *Omega* 29 (February): 49–61.

Berger, P.D., and D. Bernstein. 2002. The optimal trade-off between acquisition and retention promotion—An application to the diagnostic self-testing market. *International Quarterly Journal of Marketing*, 2 (January–December): 47–54.

Bitran, G.R., and S.V. Mondschein. 1996. Mailing decisions in the catalog sales industry. *Management Science* 42 (September): 1364–1381.

Blattberg, R.C., and J. Deighton. 1996. Manage marketing by the customer equity test. *Harvard Business Review*, 74 (July–August): 136–144.

Blattberg, R.C., G. Getz, and J.S. Thomas. 2001. *Customer equity: Building and managing relationships as valuable assets*. Boston, MA: Harvard Business School Press.

Blattberg, R.C., B.-D. Kim, and S.A. Neslin. 2008. *Database marketing: Analyzing and managing customers*. New York, NY: Springer.

Bodapati, A.V. 2008. Recommendation systems with purchase data. *Journal of Marketing Research* 45 (February): 77–93.

Bolton, R.N. 1998. A dynamic model of the duration of the customer's relationship with a continuous service provider: The role of satisfaction. *Marketing Science* 17 (February): 45–65.

Bolton, R.N., K.N. Lemon, and P.C. Verhoef. 2004. The theoretical underpinnings of customer asset management: A framework and propositions for future research. *Journal of the Academy of Marketing Science* 32 (July): 271–292.

Braun, M., and D.A. Schweidel. 2011. Modeling customer lifetimes with multiple causes of churn. *Marketing Science*, 30 (September–October): 881–902.

Braun, M., D.A. Schweidel, and E.M. Stein. 2015. Transaction attributes and customer valuation. *Journal of Marketing Research* 52 (December): 848–864.

Buchanan, B., and D.G. Morrison. 1988. A stochastic model of list falloff with implications for repeat mailings. *Journal of Direct Marketing* 2 (Summer): 7–15.

Bult, J.R. 1993. Semiparametric versus parametric classification models: An application to direct marketing. *Journal of Marketing Research* 30 (August): 380–390.

Bult, J.R., H. van der Scheer, and T. Wansbeek. 1997. Interaction between target and mailing characteristics in direct marketing, with an application to health care fund raising. *International Journal of Research in Marketing* 14 (October): 301–308.

Bult, J.R., and T. Wansbeek. 1995. Optimal selection for direct mail. *Marketing Science* 14 (November): 378–394.

Calciu, M. 2008. Numeric decision support to find optimal balance between customer acquisition and retention spending. *Journal of Targeting, Measurement and Analysis for Marketing* 16 (3): 214–227.

Calli, M.K., M. Weverbergh, and P.H. Franses. 2012. The effectiveness of high-frequency direct-response commercials. *International Journal of Research in Marketing* 29 (March): 98–109.

Campbell, D., R. Erdahl, D. Johnson, E. Bibelnieks, M. Haydock, M. Bullock, and H. Crowder. 2001. Optimizing customer mail streams at Fingerhut. *Interfaces*, 31 (January–February): 77–90.

Carroll, B.J. 2006. *Lead generation for the complex sale*. New York, NY: McGraw-Hill.

Chan, T.Y., C. Wu, and Y. Xie. 2011. Measuring the lifetime value of customers acquired from Google search advertising. *Marketing Science* 30 (September–October): 837–850.

Chang, C.W., and J.Z. Zhang. 2016. The effects of channel experiences and direct marketing on customer retention in multichannel settings. *Journal of Interactive Marketing* 36 (November): 77–90.

Chen, Y., and J.H. Steckel. 2012. Modeling credit card share of wallet: Solving the incomplete information problem. *Journal of Marketing Research* 49 (October): 655–669.

Ching, W.K., M.K. Ng, K.-K. Wong, and E. Altman. 2004. Customer lifetime value: Stochastic optimization approach. *Journal of the Operational Research Society* 55 (August): 860–868.

Chung, T.S., M. Wedel, and R.T. Rust. 2016. Adaptive personalization using social networks. *Journal of the Academy of Marketing Science* 44 (January): 66–87.

Cohen, M.-D. 2004. Exploiting response models–Optimizing cross-sell and up-sell opportunities in banking. *Information Systems* 29 (June): 327–341.

Colombo, R., and W. Jiang. 1999. A stochastic RFM model. *Journal of Interactive Marketing* 13 (Summer): 2–12.

Coussement, K., and D. Van den Poel. 2008. Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques. *Expert Systems with Applications* 34 (1): 313–327.

Cui, G., M.L. Wong, and H.-K. Lui. 2006. Machine learning for direct marketing response models: Bayesian networks with evolutionary programming. *Management Science* 52 (April): 597–612.

Dasgupta, K., R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A.A. Nanavati, and A. Joshi. 2008. Social ties and their relevance to churn in mobile telecom networks. *Proceedings of the 11th international conference on extending database technology: Advances in database technology*, 668–677. New York, NY: Association for Computing Machinery.

Datta, H., B. Foubert, and H.J. van Heerde. 2015. The challenge of retaining customers acquired with free trials. *Journal of Marketing Research* 52 (April): 217–234.

David Shepard Associates. 1999. *The new direct marketing*, 3rd ed. New York: McGraw-Hill.

De Bock, K.W., and D. Van den Poel. 2011. An empirical evaluation of rotation-based ensemble classifiers for customer churn prediction. *Expert Systems with Applications* 38 (10): 12293–12301.

Deighton, J., & Johnson, P.A. 2013. *The value of data: Consequences for insight, innovation, and efficiency in the U.S. economy.* New York, NY: The Direct Marketing Association.

Delanote, S., R. Leus, and F.T. Nobibon. 2013. Optimization of the annual planning of targeted offers in direct marketing. *Journal of the Operational Research Society* 64 (December): 1770–1779.

Deming, W.E., & Glasser, G.J. 1968. A Markovian analysis of the life of newspaper subscriptions. *Management Science*, 14 (February): B-283–B-293.

Drèze, X., and A. Bonfrer. 2008. An empirical investigation of the impact of communication timing on customer equity. *Journal of Interactive Marketing* 22 (Winter): 36–50.

Dong, W., S.D. Swain, and P.D. Berger. 2007. The role of channel quality in customer equity management. *Journal of Business Research* 60 (December): 1243–1252.

Donkers, B., R. Paap, J.-J. Jonker, and P.H. Franses. 2006. Deriving target selection rules from endogenously selected samples. *Journal of Applied Econometrics* 21 (July/August): 549–562.

Du, R.Y., W.A. Kamakura, and C.F. Mela. 2007. Size and share of customer wallet. *Journal of Marketing* 71 (April): 94–113.

Ehrman, C.M. 1990. Correcting for “regression to the mean” in list selection decisions. *Journal of Direct Marketing* 4 (Spring): 21–30.

Ehrman, C.M., and G.M. Funk. 1997. Insights on "Selecting, evaluating, and updating prospects in direct mail marketing," by Vithala Rao and Joel Steckel. *Journal of Interactive Marketing* 11 (Summer): 8–13.

Ehrman, C.M., and K.J. Miescke. 1989. Structured decision rules for ranking and selecting mailing lists and creative packages for direct marketing. *Journal of Direct Marketing* 3 (Winter): 47–59.

Elsner, R., M. Krafft, and A. Huchzermeier. 2003. Optimizing Rhenania's mail-order business through dynamic multilevel modeling (DMLM). *Interfaces*, 33 (January–February): 50–66.

Elsner, R., M. Krafft, and A. Huchzermeier. 2004. Optimizing Rhenania's direct marketing business through dynamic multilevel modeling (DMLM) in a multicatalog-brand environment. *Marketing Science* 23 (Spring): 192–206.

Fader, P. 2012. *Customer centricity: Focus on the right customers for strategic advantage*, 2nd ed. Philadelphia, PA: Wharton Digital Press.

Fader, P.S., B.G.S. Hardie, and K. Jerath. 2007. Estimating CLV using aggregated data: The Tuscan Lifestyles case revisited. *Journal of Interactive Marketing* 21 (Summer): 55–71.

Fader, P.S., and B.G.S. Hardie. 2007a. How to project customer retention. *Journal of Interactive Marketing* 21 (Winter): 76–90.

Fader, P.S., and B.G.S. Hardie. 2007b. Incorporating time-invariant covariates into the Pareto/NBD and BG/NBD models. http://www.brucehardie.com/notes/019/. Accessed 10 Aug 2016.

Fader, P.S., and B.G.S. Hardie. 2009. Probability models for customer-base analysis. *Journal of Interactive Marketing* 23 (January): 61–69.

Fader, P.S., & Hardie, B.G.S. 2010. Customer-base valuation in a contractual setting: The perils of ignoring heterogeneity. *Marketing Science*, 29 (January–February), 85–93.

Fader, P.S., and B.G.S. Hardie. 2012. Reconciling and clarifying CLV formulas. http://www.brucehardie.com/notes/024/. Accessed 10 Aug 2016.

Fader, P.S., and B.G.S. Hardie. 2014a. The Pareto/NBD is not a lost-for-good model. http://www.brucehardie.com/notes/031/. Accessed 10 Aug 2016.

Fader, P.S., and B.G.S. Hardie. 2014b. A spreadsheet-literate non-statistician's guide to the beta-geometric model. http://www.brucehardie.com/notes/032/. Accessed 10 Aug 2016.

Fader, P.S., and B.G.S. Hardie. 2014c. What's wrong with this CLV formula? http://www.brucehardie.com/notes/033/. Accessed 10 Aug 2016.

Fader, P.S., and B.G.S. Hardie. 2015. Simple probability models for computing CLV and CE. In *The handbook of customer equity*, ed. V. Kumar and Denish Shah, 77–100. Cheltenham, UK: Edward Elgar Publishers.

Fader, P.S., B.G.S. Hardie, and K.L. Lee. 2005a. "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Science* 24 (Spring): 275–284.

Fader, P.S., B.G.S. Hardie, and K.L. Lee. 2005b. RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research* 42 (November): 415–430.

Fader, P.S., B.G.S. Hardie, and S. Sen. 2014. Stochastic models of buyer behavior. In *The history of marketing science*, ed. Russell S. Winer, and Scott A. Neslin, 165–205. Singapore: World Scientific Publishing.

Fader, P.S., B.G.S. Hardie, and J. Shang. 2010. Customer-base analysis in a discrete-time noncontractual setting. *Marketing Science*, 29 (November–December): 1086–1108.

Fan, S.S., and P.D. Berger. 2001. The optimal allocation between acquisition and retention spending over multiple time periods. *International Quarterly Journal of Marketing*, 1 (April–December): 199–210.

Frank, R.E. 1962. Brand choice as a probability process. *The Journal of Business* 35 (January): 43–56.

Fruchter, G.E., and Z.J. Zhang. 2004. Dynamic targeted promotions: A customer retention and acquisition perspective. *Journal of Service Research* 7 (August): 3–19.

Galbraith, J.R. 2005. *Designing the customer-centric organization*. San Francisco, CA: Jossey-Bass.

George, M., V. Kumar, and D. Grewal. 2013. Maximizing profits for a multi-category catalog retailer. *Journal of Retailing* 89 (December): 374–396.

Gerpott, T.J., and N. Ahmadi. 2015. Regaining drifting mobile communication customers: Predicting the odds of success of winback efforts with competing risks regression. *Expert Systems with Applications* 42 (21): 7917–7928.

Gönül, F.F., B.-D. Kim, and M. Shi. 2000. Mailing smarter to catalog customers. *Journal of Interactive Marketing* 14 (Spring): 2–16.

Gönül, F.F., and F. Ter Hofstede. 2006. How to compute optimal catalog mailing decisions. *Marketing Science*, 25 (January–February): 65–74.

Griffin, J., and M.W. Lowenstein. 2001. *Customer winback: How to recapture lost customers–And keep them loyal*. San Francisco, CA: Jossey-Bass.

Gupta, S. 2014. *Marketing reading: Customer management (Core curriculum readings series)*. Boston, MA: Harvard Business Publishing.

Gupta, S., and D.R. Lehmann. 2005. *Managing customers as investments*. Upper Saddle River, NJ: Wharton School Publishing.

Gupta, S., D.R. Lehmann, and J.A. Stuart. 2004. Valuing customers. *Journal of Marketing Research* 41 (February): 7–18.

Haenlein, M. 2013. Social interactions in customer churn decisions: The impact of relationship directionality. *International Journal of Research in Marketing* 30 (September): 236–248.

Hansen, K. 2015. Comment on the "Predictive modelling for churner/non-churner" conversation in the "Advanced Business Analytics, Data Mining and Predictive Modeling" LinkedIn group. https://www.linkedin.com/groups/35222/35222-5949032187344023556. Accessed 10 Aug 2016.

Hansotia, B.J., and P. Wang. 1997. Analytical challenges in customer acquisition. *Journal of Direct Marketing* 11 (Spring): 7–19.

Hauser, J.R., G. Liberali, and G.L. Urban. 2014. Website morphing 2.0: Switching costs, partial exposure, random exit, and when to morph. *Management Science* 60 (June): 1594–1616.

Hauser, J.R., G.L. Urban, G. Liberali, and M. Braun. 2009. Website morphing. *Marketing Science*, 28 (March–April): 202–223.

Heitz, C., M. Dettling, and A. Ruckstuhl. 2011. Modelling customer lifetime value in contractual settings. *International Journal of Services Technology and Management* 16 (2): 172–190.

Hinz, O., B. Skiera, C. Barrot, and J.U. Becker. 2011. Seeding strategies for viral marketing: An empirical comparison. *Journal of Marketing* 75 (November): 55–71.

Hoekstra, J.C., P.S.H. Leeflang, and D.R. Wittink. 1999. The customer concept: The basis for a new marketing paradigm. *Journal of Market Focused Management* 4 (1): 43–76.

Holtrop, N., J.E. Wieringa, M.J. Gijsenberg, and P.C. Verhoef. 2017. No future without the past? Predicting churn in the face of customer privacy. *International Journal of Research in Marketing*, 34 (March): 154–172.

Hopmann, J., and A. Thede. 2005. Applicability of customer churn forecasts in a non-contractual setting. In *Innovations in classification, data science, and information systems (Proceedings of the 27th annual conference of the gesellschaft fr klassifikation e.V., Brandenburg University of Technology, Cottbus, March 12–14, 2003)*, ed. Daniel Baier and Klaus-Dieter Wernecke, 330–337. Berlin: Springer.

Howard, R.A. 1978. Comments on the origin and application of Markov decision processes. In *Dynamic Programming and its Applications*, ed. M.L. Puterman, 201–205. New York: Academic Press.

Hruschka, H. 2010. Considering endogeneity for optimal catalog allocation in direct marketing. *European Journal of Operational Research* 206 (1): 239–247.

Huang, C.-Y. 2012. To model, or not to model: Forecasting for customer prioritization. *International Journal of Forecasting* 28 (2): 497–506.

Hughes, A.M. 1996. *The complete database marketer*, revised ed. Chicago, IL: Irwin.

Imhoff, C., L. Loftis, and J.G. Geiger. 2001. *Building the customer-centric enterprise*. New York, NY: Wiley.

Jackson, B.B. 1985. *Winning and keeping industrial customers*. New York, NY: Lexington Books.

Jamal, Z., and R.E. Bucklin. 2006. Improving the diagnosis and prediction of customer churn: A heterogeneous hazard modeling approach. *Journal of Interactive Marketing*, 20 (Summer/Autumn): 16–29.

Jerath, K., P.S. Fader, B.G.S. Hardie 2011. New perspectives on customer 'death' using a generalization of the Pareto/NBD model. *Marketing Science*, 30 (September–October): 866–880.

Jerath, K., P.S. Fader, and B.G.S. Hardie. 2016. Customer-base analysis using repeated cross-sectional summary (RCSS) data. *European Journal of Operational Research* 249 (1): 340–350.

Jonker, J.-J., N. Piersma, and R. Potharst. 2006. A decision support system for direct mailing decisions. *Decision Support Systems* 42 (November): 915–925.

Kamakura, W.A., S.N. Ramaswami, and R.K. Srivastava. 1991. Applying latent trait analysis in the evaluation of prospects for cross-selling of financial services. *International Journal of Research in Marketing* 8 (November): 329–349.

Kannan, P.K., W. Reinartz, and P.C. Verhoef. 2016. The path to purchase and attribution modeling: Introduction to special section. *International Journal of Research in Marketing* 33 (September): 449–456.

Katz, E., and P.F. Lazarsfeld. 1955. *Personal influence: The part played by people in the flow of mass communications*. Glencoe, IL: The Free Press.

Kestnbaum, R.D., K.T. Kestnbaum, and P.W. Ames. 1998. Building a longitudinal contact strategy^{TM}. *Journal of Interactive Marketing* 12 (Winter): 56–62.

Khan, R., M. Lewis, and V. Singh. 2009. Dynamic customer management and the value of one-to-one marketing. *Marketing Science* 28 (November–December): 1063–1079.

Kim, B.-D., and S.-O. Kim. 1999. Measuring upselling potential of life insurance customers: Application of a stochastic frontier model. *Journal of Interactive Marketing* 13 (Autumn): 2–9.

Koning, R., P. Spring, T. Wansbeek. 2002. Joint modeling of primary and secondary action in database marketing. Working paper, Department of Economics, University of Groningen.

Knott, A., A. Hayes, and S.A. Neslin. 2002. Next-product-to-buy models for cross-selling applications. *Journal of Interactive Marketing* 16 (Summer): 59–75.

Knox, G., and R. Van Oest. 2014. Customer complaints and recovery effectiveness: A customer base approach. *Journal of Marketing* 78 (September): 42–57.

Kumar, V., Y. Bhagwat, and X. Zhang. 2015. Regaining "lost" customers: The predictive power of first-lifetime behavior, the reason for defection, and the nature of the win-back offer. *Journal of Marketing* 79 (July): 34–55.

Kumar, V., J.A. Petersen, and R.P. Leone. 2010. Driving profitability by encouraging customer referrals: Who, when, and how. *Journal of Marketing* 74 (September): 1–17.

Kumar, V., and W. Reinartz. 2012. *Customer relationship management*, 2nd ed. Heidelberg: Springer.

Kumar, V., and D. Shah. 2015. *The handbook of customer equity*. Cheltenham, UK: Edward Elgar Publishers.

Kumar, V., R. Venkatesan, T. Bohling, and D. Beckmann. 2008. The power of CLV: Managing customer lifetime value at IBM. *Marketing Science* 27 (July–August): 585–599.

Lamberti, L. 2013. Customer centricity: The construct and the operational antecedents. *Journal of Strategic Marketing* 21 (7): 588–612.

Lambrecht, A., and C. Tucker. 2013. When does retargeting work? Information specificity in online advertising. *Journal of Marketing Research* 50 (October): 561–576.

Larivière, B., and D. Van den Poel. 2005. Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications* 29 (2): 472–484.

Lemmens, A., and C. Croux. 2006. Bagging and boosting classification trees to predict churn. *Journal of Marketing Research* 43 (May): 276–286.

Lemmens, A., and S. Gupta. 2013. Managing churn to maximize profits. http://ssrn.com/abstract=2964906. Accessed 10 May 2017.

Levin, N., and J. Zahavi. 1998. Continuous predictive modeling–A comparative analysis. *Journal of Interactive Marketing* 12 (Spring): 5–22.

Lewis, M. 2006. Customer acquisition promotions and customer asset value. *Journal of Marketing Research* 43 (May): 195–203.

Lhoest-Snoeck, S., E. van Nierop, and P.C. Verhoef. 2014. For new customers only: A study on the effect of acquisition campaigns on a service company's existing customers' CLV. *Journal of Interactive Marketing* 28 (August): 210–224.

Li, H., and P.K. Kannan. 2014. Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research* 51 (February): 40–56.

Li, S., B. Sun, and A.L. Montgomery. 2011. Cross-selling the right product to the right customer at the right time. *Journal of Marketing Research* 48 (August): 683–700.

Li, S., B. Sun, and R.T. Wilcox. 2005. Cross-selling sequentially ordered products: An application to consumer banking services. *Journal of Marketing Research* 42 (May): 233–239.

Lianos, G., & Sloev, I. 2013. Customer acquisition and customer retention in a monopolistically competitive industry. http://ssrn.com/abstract=2386586. Accessed 10 Aug 2016.

Libai, B., E. Muller, and R. Peres. 2013. Decomposing the value of word-of-mouth seeding programs: Acceleration versus expansion. *Journal of Marketing Research* 50 (April): 161–176.

Liu, H., J. Pancras, and M. Houtz. 2015. Managing customer acquisition risk using co-operative databases. *Journal of Interactive Marketing* 29 (February): 39–56.

Lu, T.L., and C. Boutilier. 2014. Dynamic segmentation for large-scale marketing optimization. *ICML-2014 workshop on customer life-time value optimization in digital marketing*, 31st international conference on machine learning (ICML 2014), Beijing, June 21–26.

Ma, S., and J. Büschken. 2011. Counting your customers from an "always a share" perspective. *Marketing Letters* 22 (3): 243–257.

Ma, S., & Liu, J.-L. 2007. The MCMC approach for solving the Pareto/NBD model and possible extensions. *Third international conference on natural computation (ICNC 2007)*, 505–512.

Malthouse, E.C., and R. Elsner. 2006. Customisation with crossed-basis sub-segmentation. *Journal of Database Marketing & Customer Strategy Management* 14 (1): 40–50.

Mark, T., K.N. Lemon, M. Vandenbosch, J. Bulla, and A. Maruotti. 2013. Capturing the evolution of customer-firm relationships: How customers become more (or less) valuable over time. *Journal of Retailing* 89 (September): 231–245.

Mark, T., K.N. Lemon, and M. Vandenbosch. 2014. Customer migration patterns: Evidence from a North American retailer. *Journal of Marketing Theory and Practice* 22 (3): 251–269.

McCarthy, D., P.S. Fader, and B.G.S. Hardie. 2017. Valuing subscription-based businesses using publicly disclosed customer data. *Journal of Marketing* 81 (January): 17–35.

Moffett, T. 2014. *Use cross-channel attribution to understand marketing effectiveness*. Cambridge, MA: Forrester Research Inc.

Montoya, R., O. Netzer, and K. Jedidi. 2010. Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Marketing Science* 29 (September–October): 909–924.

Moon, S., and G.J. Russell. 2008. Predicting product purchase from inferred customer similarity: An autologistic model approach. *Management Science* 54 (January): 71–82.

Morwitz, V.G., and D.C. Schmittlein. 1998. Testing new direct marketing offerings: The interplay of management judgment and statistical models. *Management Science* 44 (May): 610–628.

Mozer, M.C., R. Wolniewicz, D.B. Grimes, E. Johnson, and H. Kaushansky. 2000. Predicting subscriber dissatisfaction and improving retention in the wireless telecommunications industry. *IEEE Transactions on Neural Networks* 11 (3): 690–696.

Musalem, A., and Y.V. Joshi. 2009. How much should you invest in each customer relationship? A competitive strategic approach. *Marketing Science*, 28 (May–June): 555–565.

Muus, L., H. van der Scheer, and T. Wansbeek. 2002. A decision theoretic framework for profit maximization in direct marketing. In *Econometric models in marketing advances in econometrics*, vol. 16, ed. P.H. Franses, and A.L. Montgomery, 119–140. Oxford: Elsevier Science.

Natter, M., A.-M. Ozimec, and J.-Y. Kim. 2015. ECO: Entega's profitable new customer acquisition on online price comparison sites. *Marketing Science*, 34 (November–December): 789–803.

Neslin, S.A., S. Gupta, W. Kamakura, J. Lu, and C.H. Mason. 2006. Defection detection: Measuring and understanding the predictive accuracy of customer churn models. *Journal of Marketing Research* 43 (May): 204–211.

Neslin, S.A., G.A. Taylor, K.D. Grantham, and K.R. McNeil. 2013. Overcoming the "recency trap" in customer relationship management. *Journal of the Academy of Marketing Science* 41 (May): 320–337.

Netzer, O., J.M. Lattin, V. Srinivasan. 2008. A hidden Markov model of customer relationship dynamics. *Marketing Science*, 27 (March–April): 185–204.

Nitzan, I., and B. Libai. 2011. Social effects on customer retention. *Journal of Marketing* 75 (November): 24–38.

Nobibon, F.T., R. Leus, and F.C.R. Spieksma. 2011. Optimization models for targeted offers in direct marketing: Exact and heuristic algorithms. *European Journal of Operational Research* 210 (3): 670–683.

Otter, P.W., H. van der Scheer, and T.J. Wansbeek. 2000. Optimal selection of households for direct marketing by joint modeling of the probability and quantity of response. CCSO Working paper 2006/06, CCSO Centre for Economic Research, University of Groningen.

Ovchinnikov, A., B. Boulu-Reshef, and P.E. Pfeifer. 2014. Balancing acquisition and retention spending for firms with limited capacity. *Management Science* 60 (August): 2002–2019.

Paas, L.J. 1998. Mokken scaling characteristic sets and acquisition patterns of durable and financial products. *Journal of Economic Psychology* 19 (June): 353–376.

Padilla, N., and E. Ascarza. 2017. First impressions count: Leveraging acquisition data for customer management. http://ssrn.com/abstract=2933291. Accessed 5 May 2017.

Park, C.H., Y.-H. Park, and D.A. Schweidel. 2014. A multi-category customer base analysis. *International Journal of Research in Marketing* 31 (September): 266–279.

Passant, P. 1995. From the practitioners: Retention marketing needs a new vision. *Journal of Direct Marketing* 9 (Spring): 2–4.

Peres, R., E. Muller, and V. Mahajan. 2010. Innovation diffusion and new product growth models: A critical review and research directions. *International Journal of Research in Marketing* 27 (June): 91–106.

Petrison, L.A., R.C. Blattberg, and P. Wang. 1997. Database marketing: Past, present, and future. *Journal of Direct Marketing* 11 (Fall): 109–125.

Pfeifer, P.E. 1998. On using the beta-logistic model to update response probabilities given nonresponse. *Journal of Interactive Marketing* 12 (Spring): 23–32.

Pfeifer, P.E. 2005. The optimal ratio of acquisition and retention costs. *Journal of Targeting, Measurement and Analysis for Marketing* 13 (2): 179–188.

Pfeifer, P.E., and R.L. Carraway. 2000. Modeling customer relationships as Markov chains. *Journal of Interactive Marketing* 14 (Spring): 43–55.

Pfeifer, P.E., M.E. Haskins, and R.M. Conroy. 2005. Customer lifetime value, customer profitability, and the treatment of acquisition spending. *Journal of Managerial Issues* 17 (Spring): 11–25.

Pfeifer, P.E., and A. Ovchinnikov. 2011. A note on willingness to spend and customer lifetime value for firms with limited capacity. *Journal of Interactive Marketing* 25 (August): 178–189.

Piatetsky-Shapiro, G., and B. Masand. 1999. Estimating campaign benefits and modeling lift. *Proceedings of the fifth ACM SIGKDD international conference on knowledge discovery and data mining*, 185–193. Association for Computing Machinery: New York, NY.

Pick, D., J.S. Thomas, S. Tillmanns, and M. Krafft. 2016. Customer win-back: The role of attributions and perceptions in customers' willingness to return. *Journal of the Academy of Marketing Science* 44 (March): 218–240.

Piersma, N., and J.-J. Jonker. 2004. Determining the optimal direct mailing frequency. *European Journal of Operational Research* 158 (1): 173–182.

Platzer, M. (2008). *Stochastic models of noncontractual consumer relationships*. Unpublished Master's thesis, Vienna University of Economics and Business Administration.

Potter, R.G., and M.P. Parker. 1964. Predicting the time required to conceive. *Population Studies* 18 (1): 99–116.

Prinzie, A., and D. Van den Poel. 2006. Investigating purchasing-sequence patterns for financial services using Markov, MTD and MTDg models. *European Journal of Operational Research* 170 (3): 710–734.

Prinzie, A., and D. Van den Poel. 2007. Predicting home-appliance acquisition sequences: Markov/Markov for discrimination and survival analysis for modeling sequential information in NPTB models. *Decision Support Systems* 44 (November): 28–45.

Rao, V.R., and J.H. Steckel. 1995. Selecting, evaluating, and updating prospects in direct mail marketing. *Journal of Direct Marketing* 9 (Spring): 20–31.

Ravi, R., and B. Sun. 2016. *Customer-centric marketing: A pragmatic framework*. Cambridge, MA: The MIT Press.

Reinartz, W., and V. Kumar. 2000. On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *Journal of Marketing* 64 (October): 17–35.

Reinartz, W., and V. Kumar. 2003. The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing* 67 (January): 77–99.

Reinartz, W., J.S. Thomas, and V. Kumar. 2005. Balancing acquisition and retention resources to maximize customer profitability. *Journal of Marketing* 69 (January): 63–79.

Rhee, S., and S. McIntyre. 2008. Including the effects of prior and recent contact effort in a customer scoring model for database marketing. *Journal of the Academy of Marketing Science* 36 (December): 538–551.

Rhee, E., and S. McIntyre. 2009. How current targeting can hinder targeting in the future and what to do about it. *Journal of Database Marketing & Customer Strategy Management* 16 (1): 15–28.

Rhee, E., and G.J. Russell. 2009. Forecasting household response in database marketing: A latent trait approach. In *Advances in business and management forecasting*, vol. 6, ed. K.D. Lawrence, and R.K. Klimberg, 109–131. Bingley, UK: JAI Press/Emerald Group Publishing.

Ricci, F., L. Rokach, and B. Shapira. 2015. *Recommender systems handbook*, 2nd ed. New York, NY: Springer.

Risselada, H., P.C. Verhoef, and T.H.A. Bijmolt. 2010. Staying power of churn prediction models. *Journal of Interactive Marketing* 24 (August): 198–208.

Romero, J., R. Van der Lans, and B. Wierenga. 2013. A partially hidden Markov model of customer dynamics for CLV measurement. *Journal of Interactive Marketing* 27 (August): 185–208.

Rosenwald, P.J. 2004. *Accountable marketing: The economics of data-driven marketing*. New York, NY: Texere.

Rosset, S., E. Neumann, U. Eick, and N. Vatnik. 2003. Customer lifetime value models for decision support. *Data Mining and Knowledge Discovery* 7 (July): 321–339.

Rust, R.T., K.N. Lemon, and V.A. Zeithaml. 2004. Return on marketing: Using customer equity to focus marketing strategy. *Journal of Marketing* 68 (January): 109–127.

Schmitt, P., B. Skiera, and C. Van den Bulte. 2011. Referral programs and customer value. *Journal of Marketing* 75 (January): 46–59.

Schmittlein, D.C., D.G. Morrison, and R. Colombo. 1987. Counting your customers: Who they are and what will they do next? *Management Science* 33 (January): 1–24.

Schmittlein, D.C., and R.A. Peterson. 1994. Customer base analysis: An industrial purchase process application. *Marketing Science* 13 (Winter): 41–67.

Schröder, N., and H. Hruschka. 2016. Investigating the effects of mailing variables and endogeneity on mailing decisions. *European Journal of Operational Research* 250 (2): 579–589.

Schulze, C., B. Skiera, and T. Wiesel. 2012. Linking customer and financial metrics to shareholder value: The leverage effect in customer-based valuation. *Journal of Marketing* 76 (March): 17–32.

Schwartz, E.M., E.T. Bradlow, P.S. Fader. 2014. Model selection using database characteristics: Developing a classification tree for longitudinal incidence data. *Marketing Science*, 33 (March–April): 188–205.

Schwartz, E.M., E. Bradlow, P. Fader. 2016. Customer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*, forthcoming.

Schweidel, D.A., E.T. Bradlow, and P.S. Fader. 2011. Portfolio dynamics for customers of a multi-service provider. *Management Science* 57 (March): 471–486.
Schweidel, D.A., P.S. Fader, and E.T. Bradlow. 2008a. Understanding service retention within and across cohorts using limited information. *Journal of Marketing* 72 (January): 82–94.
Schweidel, D.A., P.S. Fader, E.T. Bradlow. 2008b. A bivariate timing model of customer acquisition and retention. *Marketing Science*, 27 (September–October): 829–843.
Schweidel, D.A., and G. Knox. 2013. Incorporating direct marketing activity into latent attrition models. *Marketing Science*, 32 (May–June): 471–487.
Schweidel, D.A., Y.-H. Park, Z. Jamal. 2014. A multiactivity latent attrition model for customer base analysis. *Marketing Science*, 33 (March–April): 273–286.
Seybold, P.B., R.T. Marshak, and J.M. Lewis. 2001. *The customer revolution*. New York, NY: Random House.
Shao, X., and L. Li. 2011. Data-driven multi-touch attribution models. *Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining*, 258–264. Association for Computing Machinery: New York, NY.
Simester, D.I., P. Sun, and J.N. Tsitsiklis. 2006. Dynamic catalog mailing policies. *Management Science* 52 (May): 683–696.
Simon, J.L. 1967. Expenditure policy for mail-order advertisers. *Journal of Marketing Research* 4 (February): 59–61.
Simon, J.L. 1993. *How to start and operate a mail-order business*, 5th ed. New York, NY: McGraw-Hill.
Singh, S.S., S. Borle, and D.C. Jain. 2009. A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. *Quantitative Marketing and Economics* 7 (2): 181–205.
Soukup, D.J. 1983. A Markov analysis of fund-raising alternatives. *Journal of Marketing Research* 20 (August): 314–319.
Stanford, R.E., W.S. Martin, and G.C. Myers. 1996. Fundraising vs. contributor prospecting trade-offs in direct mail response rate management: A linear programming analysis. *Journal of Direct* 10 (Autumn): 8–18.
Stauss, B., and C. Friege. 1999. Regaining service customers: Costs and benefits of regain management. *Journal of Service Research* 1 (4): 347–361.
Steffes, E.M., B.P.S. Murthi, and R.C. Rao. 2011. Why are some modes of acquisition more profitable? A study of the credit card industry. *Journal of Financial Services Marketing* 16 (2): 90–100.
Swain, S.D., P.D. Berger, and B.D. Weinberg. 2014. The customer equity implications of using incentives in acquisition channels: A nonprofit application. *Journal of Marketing Analytics* 2 (1): 1–17.
Tapp, A., I. Whitten, and M. Housden. 2014. *Principles of direct, database, and digital marketing*, 5th ed. Harlow, UK: Pearson Education Ltd.
Tellis, G.J., R.K. Chandy, and P. Thaivanich. 2000. Which ad works, when, where, and how often? Modeling the effects of direct television advertising. *Journal of Marketing Research* 37 (February): 32–46.
Thomas, J.S. 2001. A methodology for linking customer acquisition to customer retention. *Journal of Marketing Research* 38 (May): 262–268.
Thomas, J.S., R.C. Blattberg, and E.J. Fox. 2004. Recapturing lost customers. *Journal of Marketing Research* 41 (February): 31–45.
Thomas, S.A., S. Feng, and T.V. Krishnan. 2015. To retain? To upgrade? The effects of direct mail on regular donation behavior. *International Journal of Research in Marketing* 32 (March): 48–63.
Trusov, M., R.E. Bucklin, and K. Pauwels. 2009. Effects of word-of-mouth versus traditional marketing: Findings from an internet social networking site. *Journal of Marketing* 73 (September): 90–102.

Tsao, H.Y., C. Campbell, J. Ma, and L. Pitt. 2014. Budget allocation to grow market share and maximize customer equity: The effect of inertial segment size. *Journal of Marketing Analytics* 2 (4): 205–217.

Vaidya, R., and N. Cassidy. 1999. Prioritizing leads using response probabilities and expected purchase amount. In *Advanced Research Techniques Forum*, Santa Fe, NM, June 13–16.

van Diepen, M., B. Donkers, and P.H. Franses. 2009. Dynamic and competitive effects of direct mailings: A charitable giving application. *Journal of Marketing Research* 46 (February): 120–133.

Van den Bulte, C., E. Bayer, B. Skiera, P. Schmitt. 2015. How customer referral programs turn social capital into economic capital. Marketing Science Institute Working Paper Series, Report No. 15-102.

Van der Lans, R., G. van Bruggen, J. Eliashberg, B. Wierenga. 2010. A viral branching model for predicting the spread of electronic word of mouth. *Marketing Science*, 29 (March–April): 348–365.

Venkatesan, R., and V. Kumar. 2004. A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing* 68 (October): 106–125.

Venkatesan, R., V. Kumar, and T. Bohling. 2007. Optimal customer relationship management using Bayesian decision theory: An application for customer selection. *Journal of Marketing Research* 44 (November): 579–594.

Verbeke, W., K. Dejaeger, D. Martens, J. Hur, and B. Baesens. 2012. New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research* 218 (1): 211–229.

Verbeke, W., D. Martens, B. Baesens. 2014. Social network analysis for customer churn prediction. *Applied Soft Computing* 14 (Part C): 431–446.

Verbeke, W., D. Martens, C. Mues, and B. Baesens. 2011. Building comprehensible customer churn prediction models with advanced rule induction techniques. *Expert Systems with Applications* 38 (3): 2354–2364.

Verbraken, T., W. Verbeke, and B. Baesens. 2013. A novel profit maximizing metric for measuring classification performance of customer churn prediction models. *IEEE Transactions on Knowledge and Data Engineering* 25 (May): 961–973.

Verhoef, P.C., and B. Donkers. 2001. Predicting customer potential value an application in the insurance industry. *Decision Support Systems* 32 (December): 189–199.

Verhoef, P.C., and B. Donkers. 2005. The effect of acquisition channels on customer loyalty and cross-buying. *Journal of Interactive Marketing* 19 (Spring): 31–43.

Whyte, W.H., Jr. (1954). The web of word of mouth. *Fortune*, 50 (November): 140–143, 204, 206, 208, 210, 212.

Williams, C., and R. Williams. 2015. Optimizing acquisition and retention spending to maximize market share. *Journal of Marketing Analytics* 3 (3): 159–170.

Wübben, M., and F. von Wangenheim. 2008. Instant customer base analysis: Managerial heuristics often ‘get it right’. *Journal of Marketing* 72 (May): 82–93.

Xu, L., J.A. Duan, and A. Whinston. 2014. Path to purchase: A mutually exciting point process model for online advertising and conversion. *Management Science* 60 (June): 1392–1412.

Ying, Y., F. Feinberg, and M. Wedel. 2006. Leveraging missing ratings to improve online recommendation systems. *Journal of Marketing Research* 43 (August): 355–365.

Zhang, X., J. Zhu, S. Xu, and Y. Wan. 2012. Predicting customer churn through interpersonal influence. *Knowledge-Based Systems* 28 (April): 97–104.

Chapter 11
Eye Movements During Search and Choice

Ralf van der Lans and Michel Wedel

11.1 Introduction

The popular proverb: "The eye is the window to the soul" has been attributed to intellectuals such as Leonardo Da Vinci or William Shakespeare. It suggests that people move their eyes to objects that they are interested in and that the human eye is, thus, an important indicator of a person's cognitive and affective processes. Not surprisingly, there is a long history of research in marketing that has analyzed eye movements of consumers. This research, which started in the early 1900s, initially focused on how consumers process advertisements in magazines and newspapers (Karslake 1940; Nixon 1924; Poffenberger 1925; Tiffin and Winick 1954). After a period of relative paucity in research activity, researchers in the 1970s started using eye movements to develop and test theories of consumer search and choice (Russo and Leclerc 1994; Russo and Rosen 1975; van Raaij 1977). The further development of this pioneering research was initially hampered by laborious data collection and analysis, as well as relatively coarse and inaccurate measures of consumers' eye movements. For instance, Nixon (1924) hid himself behind a curtain to observe the eye movements of consumers while they were paging through a magazine. Karslake (1940) developed an eye-camera that photographed the direction of the eyes up to three times per second when consumers in a lab were reading a newspaper. He used multiple judges to determine whether consumers viewed specific ads in the newspaper, and found that the inter-rater reliability was high. Likewise,

R. van der Lans (✉)
School of Business and Management, Hong Kong University
of Science and Technology, Hong Kong, China
e-mail: rlans@ust.hk

M. Wedel
Robert H. Smith School of Business, University of Maryland
College Park, College Park, MD, USA
e-mail: mwedel@rhsmith.umd.edu

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_11

Van Raaij (1977) used a camera behind a one-way mirror to identify the choice processes that consumers used on different presentation formats. In spite of the fact that the very first eye tracking equipment was available already around 1900, i.e. Dodge's "Falling Plate Camera" which was later manufactured by Spindler, Hoyer, and Göttingen (Wade 2010), marketing researchers in the early to mid 1900s still primarily used direct observation of eye movements using photography and cameras. While companies such as Perception Research Services already used eye trackers in the early 1970s, academic marketing researchers in the early 1990s still relied on human observers to collect eye-movement data (Russo and Leclerc 1994). The reason is that eye trackers for consumer research need to be easy to use, efficient and accurate. Moreover, the need to enable the collection of eye movements in natural settings, for example when consumers are standing in front of a shelf or leafing through a magazine renders many of the earlier eye tracking devices, in which the viewer's head needed to be fixated, ineffective for marketing research. Instead, consumers should be allowed to freely move their heads during eye-tracking experiments, and consumer research requires the collection of relatively large samples of participants. It was not until the 1990s that eye trackers were developed that enabled the unobtrusive collection of eye movement data in natural settings, at larger scales and at low cost. The last two decades saw rapid developments in new user friendly eye trackers, such as those by Tobii,[1] ASL,[2] SMI[3] and EyeLink.[4]

The development of new eye-tracking technology that meets the demands mentioned has stimulated interest in eye-tracking research in academia as well as in practice. Companies such as Google, Yahoo, IBM, Microsoft, Unilever, P&G, Kraft Foods, Kimberly Clark, Heinz, Pepsico, and Pfizer use eye-tracking technology to design ads, web pages, packaging and shelf layouts. During the last 15 years, in the academic marketing literature an increasing number of articles involve eye-tracking data (see Fig. 11.1). Similar to the early eye-tracking studies by Nixon (1924) and Poffenberger (1925), researchers initially focused mostly on advertising (e.g., Pieters et al. 2002; Pieters and Wedel 2004; Wedel and Pieters 2000). The last decade, however, has seen an increasing interest from academic researchers in marketing to use modern eye-tracking equipment for developing and testing theories of search and choice (e.g., Stüttgen et al. 2012; van der Lans et al. 2008a). However, this is an emerging field of research, and there are still many opportunities to improve our understanding of consumer search and choice using eye tracking data. With the emergence of online shopping, eye-tracking experiments conducted in front of computer screens have high external validity and are becoming increasingly important to understand online search and decision making.

[1]http://www.tobii.com.

[2]http://www.asleyetracking.com.

[3]http://www.smivision.com.

[4]http://www.sr-research.com.

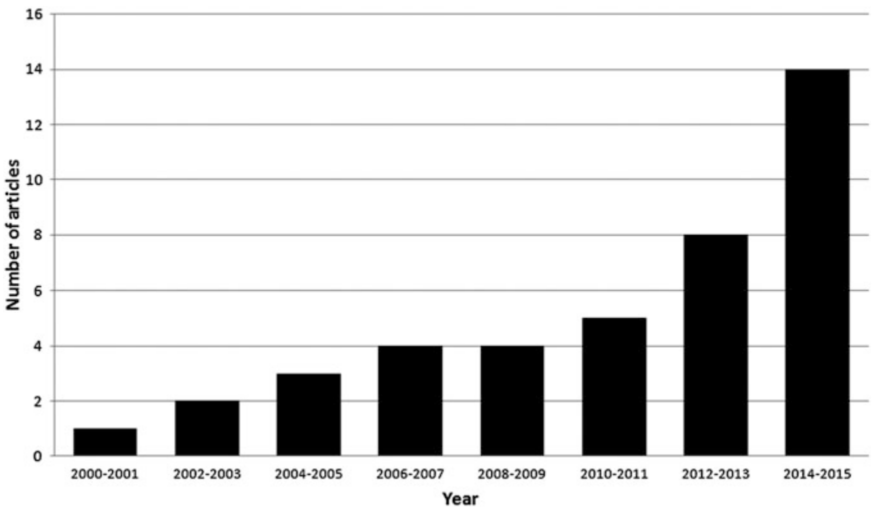


Fig. 11.1 Number of eye-tracking papers in marketing journals. *Note* We only counted papers that used eye-tracking technology from the following marketing journals: *International Journal of Research in Marketing, Journal of Advertising, Journal of Consumer Psychology, Journal of Consumer Research, Journal of Interactive Marketing, Journal of Marketing, Journal of Marketing Management, Journal of Marketing Research, Journal of Retailing, Journal of the Academy of Marketing Science, Marketing Letters, Marketing Science, and Psychology and Marketing*. For 2015, we only included papers that were published before September in that year

This chapter has two goals. First, we aim to provide an overview of the current eye-tracking literature in marketing on search and choice processes. Second, we provide a framework that may help researchers to collect and process eye tracking data and incorporate that data into their models, which will hopefully stimulate them to develop new models to uncover the underlying processes of search and choice. To attain these goals, we first discuss how eye movements relate to underlying visual processes in the next section. Section 11.3 details how to set up an eye-tracking study and which eye-movement measures are collected by modern eye trackers. Section 11.4 introduces eye-tracking metrics. Subsequently, Sects. 11.5 and 11.6 provide an overview of recent integrated models of search and choice in marketing, respectively. Section 11.7 highlights important directions for future research.

11.2 The Eye and Visual Processes

Our visual system is complex, and highly efficient. Although a detailed review of how the eye works and interacts with the visual brain is beyond the purpose of this chapter, a basic understanding of key aspects is important to incorporate eye-movement data into marketing decision models. We identify four key features of eye-movements and visual processes that are crucial for including them in the development of marketing decision models: (1) eye movements during decision making consist of fixations and saccades, (2) the perceptual field from which information is extracted is small, (3) eye movements are indicators of visual attention, and (4) eye movements are guided by bottom-up and top-down processes. We discuss each of these features next.

11.2.1 Fixations and Saccades

Introspection suggests that we move our eyes smoothly across visual scenes to process information. Except during smooth pursuits when our eyes are following moving objects, this impression is incorrect. At any point in time, we see only 1% of what is in our field of vision with high acuity. Therefore we need to move our eyes across the visual field: about 3–6 times per second, our eyes abruptly jump from one location to another. These fast jumps (up to 500° per second) are called saccades and last for about 20–40 ms. In between those jumps, the location of the eyes is relatively stable for about 100–400 ms. These periods are called fixations, and we are only able to extract visual information during those short moments (Rayner 1998). A sequence of fixations and saccades is called a scan path (see Fig. 11.2 for a scan path of the eyes on a shopping website). In addition to fixations

Fig. 11.2 Example of a scan-path of eye movements. *Note* Eye-tracking scan path of a consumer on a shopping website for cameras. The scan path consists of 53 fixations (numbered from 1 to 53) and 52 saccades, with the first and last fixations indicated by larger numbers and a box around the number

and saccades, we blink about 10–15 times per minute to moisten the eye, each blink taking about 100–150 ms (Burr 2005). As a consequence of saccades and blinks, we are functionally blind for up to 15% of the time.

The fact that our perception of visual input is stable despite frequent eye movements is one of the fascinating capabilities of the visual brain (Higgins and Rayner 2015). An explanation for the seemingly coherent percept that results from multiple fixations on different visual stimuli is that it is due to short-term or trans-saccadic memory. The visual brain actively fills in detail from memory, and thus what we perceive visually is not entirely what is processed by our eyes at that time. Change blindness is a phenomenon that provides evidence for this hypothesis. It refers to the fact that we do not observe large changes to objects or scenes that are produced experimentally while a saccade occurs (Simons and Rensink 2005).

11.2.2 The Perceptual Field Around a Fixation Is Small

The region from which the eye extracts detailed information is small: about 2° of visual angle (Anstis 1974). This corresponds to the size of a thumbnail at arm's length. Evidence for this comes for instance from forced-choice recognition tasks, in which participants were found to perform at chance levels when stimuli were presented more than 2° of visual angle away from the point of fixation (Nelson and Loftus 1980), and from the study by van der Lans et al. (2008b). They represented the perceptual field by a normal distribution and found that the best fitting model had a standard deviation that corresponded to 2° of visual angle.

The illusion that we are able to perceive much more visual information than what is obtained from the perceptual field at any moment in time is due to the fact that the brain rapidly and unconsciously moves the eyes to the regions or objects in which we are interested, and integrates the information thus obtained. Moreover, while our eyes are only able to perceive detailed information from a small region, we are able to process coarse basic features, such as colors, edges, luminance and movements from a much larger region, from "the corner of our eyes". These basic features help us with moving our eyes towards conspicuous or interesting areas for more detailed processing. In addition, it also allows us to identify the gist of visual objects and scenes. For instance, Pieters and Wedel (2012) show that consumers are able to discriminate ads from editorial content within one eye fixation, and that color plays a crucial role in identifying the gist of the ad in coarse exposure conditions (Wedel and Pieters 2015).

11.2.3 Eye Movements Are Indicators of Visual Attention

Similar to the measurement of latent constructs in surveys (see Chap. 9 of this handbook), eye fixations can be considered (probabilistic) indicators of visual attention. Whether eye fixations are a reliable and valid measure of visual attention is an important question. While it is possible to covertly attend to objects without fixating them, previous research has demonstrated a strong relationship between fixations and visual attention (Findlay 2005). Our typical visual environment contains many different objects and it would require large memory capacity to store all this information simultaneously for subsequent processing. People, therefore, tend to move their eyes to the parts of the scene they want to process at that moment, instead of relying on working memory (Droll and Hayhoe 2007). As a consequence, in almost all natural situations, eye movements are reliable and valid indicators of visual attention, and in particular reflect ongoing information uptake during search and choice tasks. Eye-movement measures have been shown to have much higher validity compared to self reported (memory-based) measures of visual attention (Aribarg et al. 2010). People are inaccurate in remembering which parts of a visual scene they attended to just before, and it is difficult to consciously move the eyes to regions one is not interested in, or suppress movements to regions that attract attention.

11.2.4 Bottom-up and Top-Down Process Guide Eye Movements

What attracts our attention and how do we determine to move our eyes to a specific location instead of another? This question has received much interest and researchers have found that eye movements are guided by both bottom-up and top-down processes (Folk, Remington, and Johnston 1992). Bottom-up processes are strictly driven by properties of the stimulus, including its colors, luminance, and objects embedded in it; top-down processes originate from cognitive processes such as memory, emotions, and goals. As discussed above, the perceptual field from which the eye extracts detailed information is small, but in the periphery our eyes are able to process coarse blobs of visual features, such as colors, luminance, size, motion and edges (Wolfe and Horowitz 2004). These basic visual features attract attention bottom-up if they are distinctive and pop-out from their surroundings (Duncan and Humphreys 1989). In addition to attending bottom-up to distinct features, people are also able to attend top-down to specific basic features depending on their goals. Indeed, in search and decision making, much of what people see is a result of top-down predictions rather than an accurate bottom-up representation of the visual field (Mathews et al. 2015). These top-down predictions bias bottom-up information acquisition (Desimone and Duncan 1995; Pieters and Wedel 2007). This combination of bottom-up and top-down acquisition of basic perceptual features in the visual field leads to a saliency map (Itti and Koch 2001; Koch and Ullman 1985). The salience map is a retinotopic map that represents the visual conspicuousness of locations in the visual field. That is, the salience map contains a spatial representation of the extent to which each location in the visual field stands out from its surroundings based on features such as colors, edges and contrast. Several studies have supported its physical location in the human brain (Thompson 2005; Treue 2003). Figure 11.5 provides an example.

Although the saliency map plays a crucial role in the guidance of eye-movements, the layout of the visual scene or display is also important for the guidance of eye-movements. For instance, people may use a reading strategy to systematically navigate visual stimuli (Monk 1984; Ponsoda et al. 1995). Such systematic strategies become apparent for example on retail shelves (van der Lans et al. 2008b) and on shopping and comparison websites (Shi et al. 2013).

In sum, our eyes quickly jump from fixation to fixation location and it is only during these short periods that our eye extracts detailed information from the visual display. The region, or perceptual field, from which the eye extracts visual information, is small. Moreover, people are able to strategically move their attention to locations of interest through top-down processes that may facilitate systematic scan patterns, next to the influence of the bottom-up organization of the visual display. These features have important implications for the analysis of eye-movement data, and marketing researchers need to develop models that describe fixations and saccades as function of features of the visual display (bottom-up factors and organization of the display) and individual characteristics (top-down factors).

11.3 Eye-Movement Recording

There are several techniques to track our eye movements. The most popular technique in marketing research is currently video-based tracking of infrared reflections on the eye (Duchowski 2003). Infrared eye-tracking cameras can be built into computer screens, goggles, or miniature stand-alone devices. Goggles are wearable devices that look like glasses, which have miniature cameras that record what the respondent is looking at, as well as the eye movements. They allow the collection of eye-tracking data in realistic environments, for instance when consumers are walking through a store or mall during shopping tasks. In addition to capturing the eye movements of participants, these goggles videotape what the participant is attending to, resulting in large amounts of data (i.e., fixation location and video data) that may be difficult to analyze, especially for larger samples of participants. Goggles are less useful in lab experiments in which the researcher is able to control external factors. As a consequence, computer screens with built-in eye-trackers and stand-alone eye-tracking devices have been predominantly used in academic marketing research. These eye-trackers yield high external validity in tasks such as offline shelf search and choice, and online browsing and shopping (Wedel and Pieters 2008).

Video-based eye-tracking of infrared reflections is a user-friendly methodology that allows participants to move their head within reasonable boundaries. It is based on the idea that (infrared) light reflected by the cornea (the front part of the eye that covers the iris, pupil and anterior chamber) generates a luminous spot, called a Purkinje reflection from the location of which, together with that of the pupil, it is possible to compute the gaze direction of the eye. After a calibration task, video image processing software is able to compute the location from which the eye extracts information. The calibration task is thus an important determinant of the accuracy of eye-tracking data (Tobii White Paper 2011). Commercial infrared eye-trackers usually sample the position of both eyes at a frequency of 50 or 60 Hz, with an accuracy of 0.5° or higher, which is sufficient for most commercial and academic applications in marketing.

Eye-tracking experiments can be time-consuming, because participants need to be recruited one-by-one and during the experiment the presence of a researcher is necessary to explain the procedures, to do the calibration and to monitor the data collection. It is therefore important to carefully prepare and pretest the experiment and the stimuli. In general, setting up an eye-tracking experiment consists of three steps: (1) preparation of the experiment, (2) calibration, and (3) data collection. When preparing the stimuli for the experiment, one needs to consider not only the presentation of the stimuli at realistic magnitudes, but also the resolution of the images, websites or videos and the commands that participants need to use to move from one stimulus to the next. The resolution of the stimulus needs to be aligned with the resolution of the eye-tracker. The eye-tracker measures the focus of the eye in pixel coordinates that are determined by this resolution. If the resolution of an image is too high or too low, the image is not presented full screen to participants. The commands that participants are instructed to use, for instance to move from one image to the next, is another important consideration, because they may interfere with or affect the eye-movements unintentionally. If participants are instructed to use the keyboard during the experiment, their eyes may alternate between the screen and the keyboard resulting in a loss of eye-tracking data. Alternatively, if participants are instructed to use the mouse, their eyes may follow the cursor which again may affect the eye-tracking data. When moving between stimuli, it is advisable to instruct participants to use a specific key (for instance, the space bar) on which they can place their fingers throughout the experiment, in order to minimize the need to look away from the screen.

The collection of eye-tracking data requires careful preparation and training of the researcher or a research assistant. First, the researcher needs to assess the exclusion criteria for the study. Contact lenses and glasses do generally not form a major impediment, but long eyelashes, droopy eyelids, corneal surgery, and some types of (bifocal) glasses, can obstruct or distort the recording. The researcher seats participants comfortably in front of the eye-tracker. The researcher needs to assure that the eye-tracker is able to track the eyes by adjusting the height of the screen as well as adjusting the seating. It is advisable to use a chair without wheels, such that the optimal distance is retained during the experiment. During calibration, participants are asked to look at one or more dots that move to different locations on the screen. The number of dots (usually two, six or nine) in the calibration can be set in the eye-tracking software. We recommend using a high number of calibration points for eye-movement analysis whenever possible, especially if relatively high accuracy is required. Depending on the outcome of the calibration, the researcher decides whether to readjust the eye-tracker and/or seating of the participant after which recalibration is necessary. If calibration is successful, the researcher can start the eye-tracking experiment. It is advisable that s/he monitors the eye trackers during the data collection process, to minimize the loss of data due to failures to track.

11.4 Analyzing Eye-Movement Data

Eye trackers generate large amounts of data. Depending on the sampling frequency (50 or 60 Hz), every second the eye-tracker collects 50 or 60 times the x and y coordinates of the gaze location, the distance of the eye to the screen, and the diameter of the pupil for one or both eyes. To illustrate, for an eye-tracking experiment lasting only 10 s, this could result into 500 data points per individual on four eye-tracking variables for each eye. If ten stimuli are used, which is not uncommon, the number of data points would be about 5,000 per respondent, for a total of 500,000 data points in a study with 100 participants. Eye-tracking data, therefore, results in a rich panel data structure that may provide a wealth of information. However, a significant percentage of the data may involve outliers or may be missing due to blinks and tracking problems. In addition, the raw eye-tracking data are seldom used for detailed analysis. Before analyzing the eye-tracking data, the raw gaze data needs to be aggregated to fixation-level data in which missing data and outliers are removed. During this process, researchers also need to determine whether the quality of the eye-tracking data for each participant is sufficient for data analysis. In this section, we illustrate how to determine the quality of eye-tracking data and how to aggregate it into fixations. After that, we review several eye-tracking measures that are useful for follow-up analysis in search and choice models.

11.4.1 Fixation Algorithms and the 80% Rule

There exist many algorithms to aggregate raw eye-tracking data into fixations for follow-up analysis (Duchowski 2003; Salvucci and Goldberg 2000). These algorithms usually use the velocity or distance of the eye between samples to determine whether an observation belongs to a fixation or saccade. Missing data and outliers are usually interpolated or smoothed. A disadvantage of most algorithms is that researchers need to set parameter values, for instance the threshold to distinguish fixations from saccades. In many situations, it is not clear how to set these values and optimal values may differ across tasks and individuals. As a consequence, many researchers use the default settings in available software packages, which may work well in some, but not in other cases. To alleviate these problems, van der Lans et al. (2011) developed the velocity-based BIT algorithm to classify raw eye-tracking data into fixations and saccades. Based on individual differences in the raw-eye-movement data of one or both eyes, the algorithm automatically determines individual specific thresholds to determine which raw gaze locations constitute a fixation. The algorithm also automatically removes blinks and other anomalies in the data. Moreover, in addition to the *x* and *y* coordinates of individual fixations, the algorithm computes for each individual the percentage of data that is classified as fixation. This percentage is useful to determine the validity of the eye-tracking data for each individual. As mentioned above, individuals are functionally blind for about 15% of the time, suggesting that on average up to 85% of the data can be classified into fixations. In some situations, the percentage of fixations is much lower, indicating poor quality eye-tracking data that need to be removed before follow-up analysis. Figure 11.3 illustrates for a sample dataset the number of participant/task combinations (out of 1,170) from which the percentage of eye-tracking data that were classified as fixations exceeded a specific threshold (ranging from 0 to 100% as indicated on the *x*-axis). For most participant/task combinations (995 or 85%), at least 80% of the eye-tracking data are classified as fixations, after which a sharp drop in the number of participants with valid data is observed due to blinks, saccades and outliers. Based on our experience of analyzing large numbers of eye-tracking samples, we recommend retaining participants for which the BIT algorithm classifies at least 80% of the eye-tracking data into fixations, and to remove the remaining participant/task combinations for further analysis. If a large percentage of the eye tracking data cannot be classified as fixations, this usually indicates measurement or calibration problems.

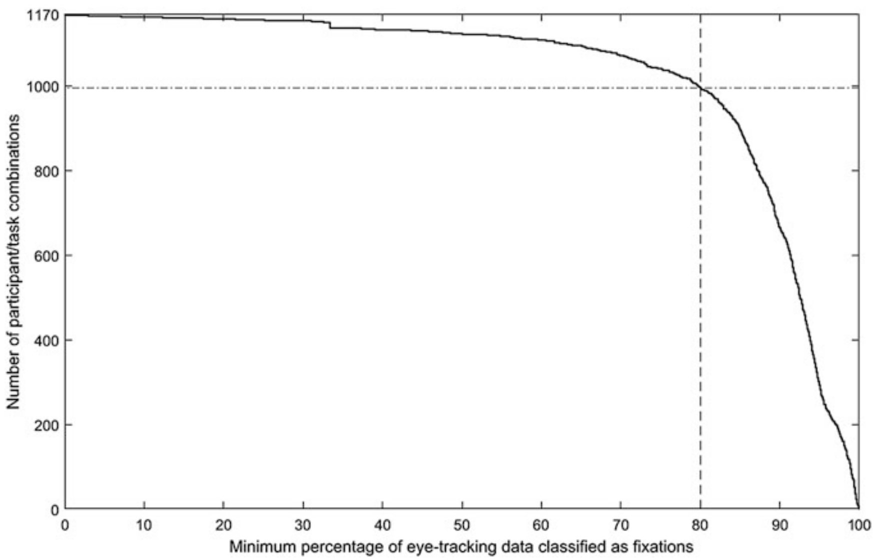


Fig. 11.3 Percentage of valid eye-tracking data in a sample eye-tracking study. *Note* In this particular eye tracking study, there is a sharp decline in the number of participant/task combinations for which the eye-tracking data contained more than 80% of fixations, as computed by the BIT algorithm of van der Lans et al. (2011)

11.4.2 AOIs and Eye-Movement Metrics

After aggregating the raw eye-tracking data into fixations and saccades, the following information is available for each eye fixation: (1) a unique identifier, usually represented by individual, task and sequence number; (2) the *x*-*y* coordinates in pixels; (3) the duration in milliseconds; (4) the distance of the eye to the stimulus; and (5) the diameter of the pupil. Most eye-tracking research in marketing uses the first two or three sources of information for follow-up analysis. Although it is possible to model the specific *x*-*y* coordinates of fixations (van der Lans et al. 2008a, b), most researchers spatially aggregate fixation locations to specific areas of interest (AOIs) . AOIs are spatial regions on the stimulus (packages on a shelf, banner ads on a website, the brand logo in an advertisement). AOIs are most often based on specific hypotheses that the researcher has formulated, and can be based on spatial grids or meaningful objects/regions. Spatial grids are useful if researchers are interested in the approximate distance between fixations or their spatial distribution. For instance, Liechty et al. (2003) divided images into 48 segments by overlaying a (8 × 6) grid over the image. Spatial grids are non-overlapping and exhaustively partition the stimulus of interest. Defining AOIs based on meaningful objects/regions is more common, however. For instance, in advertising research, AOIs of interest often represent the brand, pictorial, text and headlines (Jiang et al. 2014; Rosbergen et al. 1997). In choice and search tasks, brand and SKU facings on shelves are frequently used as AOIs (Chandon et al. 2009; van der Lans et al. 2008a). Even more specific, AOIs can represent package information, such as brand name, pictorial, and ingredient information (Pieters and Warlop 1999), or attribute information, which is often the case in conjoint tasks (Toubia et al. 2012; Yang et al. 2015). In segmenting text, AOIs can constitute the entire text, sentences, or words. In most cases, the AOIs are non-overlapping, but they can be multimodal (e.g. images and text, as can be the case for symbols, images and text referring to a brand), and do not need to exhaustively partition the stimulus so that whitespace may be left. Next to allowing for specific hypotheses to be tested, aggregating eye movements on a small set of AOIs instead of using the raw *x*-*y* coordinates offers the advantage of substantial data reduction. In some cases, image processing algorithms can be used to define AOIs (for example for faces), or the eye-tracking data itself can be used to identify AOIs using clustering algorithms. However, in most situations the researcher identifies the AOIs based on prior hypotheses using specialized software. Placement and size of the AOI is important, and needs to account for the size of the perceptual field. Consequently, AOIs should not be smaller than 2° of visual angle, and it is advisable to add a margin of at least 1° to AOIs to account for the perceptual field. Sometimes, especially if interest focuses on small AOIs, post hoc sensitivity analysis to the placement and size of the AOI is advisable. Dynamic stimuli such as TV commercials or video ads require dynamic AOIs that may change in position, size and shape as time progresses. For example, Teixeira et al. (2010) used the brand as a dynamic AOI in TV commercials.

After the spatial aggregation of eye movements on AOIs, they can be summarized through a variety of metrics. The choice of metrics depends on the goals of the study and specific hypotheses formulated. Eye-movement metrics can be classified into time, space, and time-space based metrics (see Table 11.1; see Holmqvist et al. 2011 for an extensive overview).

Table 11.1 Overview of essential eye-movement metrics

Eye-movement metric	Description
Time:	
Viewing time	Total viewing time on an stimulus
Number of fixations	Number of fixations on the stimulus
Average fixation duration	Average fixation duration of fixations on the stimulus
Scan-path length	Number of fixated AOIs
Global/local ratio	The ratio of global versus local saccades
Space:	
Location of first fixation	The (x, y)-coordinates or AOI of the first fixation
Proportion of AOIs fixated	The percentage of AOIs in an stimulus fixated by a participant
Proportion of AOIs skipped	The percentage of AOIs in an stimulus that have not been fixated by a participant
Noting	Whether or not the AOI has been fixated
Reexamination	Whether the AOI has been fixated more than once consecutively
Returning	Whether the AOI is re-fixated after another AOI was fixated
Dispersion of saccades	Variability in the length of saccades
Time and Space:	
Gaze duration	Sum of fixation durations on a specific AOI
Fixation frequency	Number of fixations on a specific AOI
Average fixation duration	Average fixation duration of fixations on a specific AOI
First pass dwell time	Sum of fixation durations during the first visit on a specific AOI
Time until first fixation	The total time before the first fixation on a specific AOI
Time after first fixation	The total time on a stimulus after the first fixation on a specific AOI
Number of AOIs fixated before	The number of AOIs fixated before the first fixation on a specific AOI
Number of AOIs fixated after	The number of AOIs fixated after the first fixation on a specific AOI
Scan-path	Entire string of AOIs visited
Sub-scan of order N	String of subset of N AOIs visited consecutively
(Markov) switching matrix	A matrix indicating the switching frequencies or probabilities between AOIs
Other metrics:	
Pupil diameter	The size of the pupil
Blink rate	Number of blinks in a specific time interval
Distance eye to screen	The distance of the eye to the screen.
Facial expressions	Facial expressions as recorded by a webcam

Popular time-based metrics include *viewing time* and the *number of fixations* on a stimulus as a whole. These two metrics are correlated indicators of overall efficiency and effort needed to complete the task. *Average fixation duration* can be used to reflect task difficulty, as more difficult tasks tend to generate longer fixation durations (Vlaskamp and Hooge 2006), or time pressure, as reflected by shorter fixation durations (Pieters and Warlop 1999). The *length of the scan-path* in terms of the number of AOIs visited can be an indication of the complexity of the design of stimuli, or the effort required to process them. Another metric is the *ratio of global to local saccades*, where local saccades can be defined as occurring on contiguous, and global saccades on noncontiguous spatial AOIs (Liechty et al. 2003). The latter two metrics can also be computed directly from the fixation locations in pixel coordinates (Krischer and Zangemeister 2007).

Location of the first fixation and proportion of AOIs fixated (or skipped) are popular spatial metrics. The *first fixation* is important to understand the gist of the image (Pieters and Wedel 2012), which can be used to plan future fixations towards informative regions. Atalay et al. (2012) found that consumers in choice tasks tended to direct their first fixation towards the center of the image and, as a consequence, are more likely to choose the option in the center. The *proportion of AOIs fixated* (or skipped) is a useful metric to understand the amount of information that is processed in the image. Yang et al. (2015) used this information to summarize the attributes processed during conjoint choice tasks and found that the amount of extracted information decreases over time. These metrics can be especially useful when characterizing non-compensatory processing. *Noting* and *reexamination* metrics indicate, respectively, whether an AOI is fixated at least once or more than once (Chandon et al. 2008). *Returning* metrics indicate whether or not (or how often) a participant returns to an AOI after having inspected another. These metrics are called “regressions” in eye movement research for reading. Especially when the AOIs constitute a spatial grid, measures such as the length and direction of saccades between AOIs can be useful metrics. Typically, longer horizontal saccades are most common, especially on stimuli with a row-column layout, such as shelves and comparison sites (van der Lans et al. 2008a; Shi et al. 2013). Finally, *dispersion* metrics, such as the variance or range of fixation points in AOI or *x*-*y* coordinates, can be used as measures of attentional focus or concentration (Teixeira et al. 2010).

The eyes move in time and space. Therefore, most eye-tracking metrics focus on these two dimensions jointly. First, *gaze durations* and *fixation frequencies*, respectively the time and the number of fixations on a specific AOI, are probably the most popular eye-movement metrics in marketing. They are also highly correlated. For instance, Aribarg et al. (2010) simultaneously model gaze duration and fixation frequencies on different ad elements to explain ad recognition and recall. *Average fixation durations* can also be computed for specific AOIs. Here, they indicate depth of processing, because more complex stimuli tend to generate longer fixations (Holmqvist et al. 2011). The *first pass dwell time* reflects the duration of the first visit of the eye to an AOI until it leaves to a different AOI. This metric has been used to measure comprehension in reading tasks, and informativeness of objects in scene perception. For instance, Stewart et al. (2004) found that first passage times on brand names increased if brands were presented in the context of less plausible brand extensions, reflecting processing difficulties. *Time until the first fixation* and the *time after the first fixation* on an AOI are important indicators for visual search efficiency. The first indicates how quickly an object attracts attention, while the latter indicates how much time is needed to make a decision after observing the AOI. Van der Lans et al. (2015) used these metrics to show that during search, online advertising reduces the number of fixated brands after fixating the target brand, while it did not affect the number of AOIs fixated before the target brand. *Sub-scans of order N* are the most frequent or modal sequences of *N* AOIs visited consecutively, where *N* usually equals 2 or 3. A *Markov switching matrix* summarizes the probabilities of the eyes moving from one AOI to another AOI. Pieters and Warlop (1999) used this information in choice tasks to model switching between choice alternatives and attributes of products. Various metrics, including the entropy and the stationary distribution, can be computed to summarize the transition matrix.

Finally, in addition to the *x-y* coordinates of the position of the eye, often eye-trackers also record the pupil diameter, blinks, the distance of the eye to the screen, and facial expressions. The *size of the pupil* depends on the brightness of the image, as well as the cognitive load and arousal (Brisson et al. 2013). If participants are aroused or need to use more cognitive resources to process visual information, the pupil tends to dilate. However, because the measurement of the size of the pupil may depend on many factors, including change of lightning conditions, and the direction of the eyes, pupil dilation has not been used frequently and demonstrating effects may be difficult. Pieters and Wedel (2007), for example, did not find differences in pupil diameter between different experimentally induced goals during advertising viewing, but found that the pupil was wider when looking at the pictorial than when looking at the text. This metric could be an interesting measure to include in models of eye movements during search and choice in future research. The same holds for the *blink rate*, which, although under the influence of various factors, may be a metric that indicates workload (Van Orden et al. 2001). The *distance of the eye to the screen* is another potentially useful metric that has not yet been used in eye-tracking research in marketing. People tend to approach stimuli they like and move away from stimuli they dislike (Chen and Bargh 1999) and hence, future research could look into the use of the distance to the screen as an indicator of preference. Finally, eye-tracking equipment often contains a webcam that records the *facial expressions* of participants. This information is useful to infer emotions of consumers while processing visual information. For instance, Teixeira et al. (2012) automatically inferred joy and surprise from facial expressions and combined this with eye-tracking data to understand avoidance of online video ads.

11.5 Eye Movements During Search

Search is ubiquitous in our daily lives and one of the most common tasks that consumers engage in when they do grocery shopping, browse websites, or play video games. Because of its importance in day to day life, search has been studied extensively in psychology (Wolfe 1998) and other fields where the outcomes of search are critical, such as airport security screening (McCarley et al. 2004), aircraft inspection (Gramopadhye et al. 1997), and radiology (Robinson 1997). In marketing, van der Lans et al. (2008a) developed a brand search model that describes how consumers search for products on a shelf (or shopping websites). Figure 11.4 summarizes the brand search process and the main components of this model, which we discuss in more detail next.

11.5.1 Brand Search Theory

As illustrated in Fig. 11.4, brand search involves two key states: (1) localization and (2) identification (van der Lans et al. 2008b). During the localization state, consumers search for and with their eyes fixate a candidate brand that potentially is the target. In the subsequent identification state, the candidate brand is processed in more detail to determine its identity. If the fixated candidate is determined to be the target that a consumer is looking for, search terminates. If the candidate turns out to be a distracter brand, the process returns to the localization state and a new candidate is searched for.

In the localization state, eye movements are guided by a salience map that is represented in the visual brain (Itti and Koch 2001; Niebur and Koch 1998; van der Lans et al. 2008a, b). The salience map is formed from perceptual features such as color, luminance and contrast of the brands and packages on the (virtual) shelf (Wolfe and Horowitz 2004). The salience at each location in the display is represented as a weighted average of the perceptual features at that location. Visual salience guides eye-movements, such that more salient areas are visited before less salient ones, ceteris paribus. In addition to the salience map, consumers may also use a systematic strategy in which the eye scans the visual display in a predetermined order, for example using a reading strategy (Monk 1984; Ponsoda et al. 1995).

In the identification state, a candidate brand on the display is matched to a search template that is made temporarily available in visual working memory. This template (see Fig. 11.4 middle-right) is the consumer's reconstruction of what the target brand and/or SKU looks like (Carlisle et al. 2011; Desimone and Duncan 1995). It may consist of a representation of the brand at various levels of specificity, ranging from a collection of perceptual features to an exact image of the brand logo or pack (Vickery et al. 2005). If the template in working memory is diagnostic and a highly specific representation of the target, identification can be fast within one

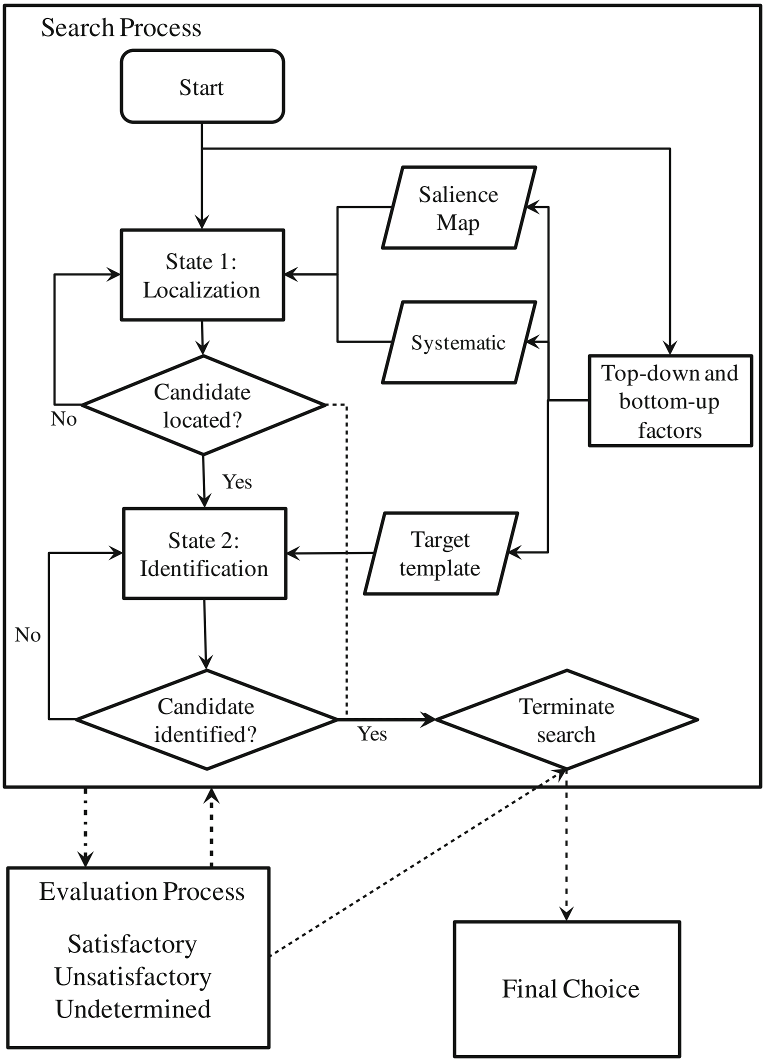


Fig. 11.4 Attention process during brand search and choice. *Notes—Dashed arrows and the boxes* "evaluation process" and "final choice" are part of the choice model proposed by Stuttgen et al. (2012)

eye-fixation. Otherwise, identification can be slow and consumers may need multiple fixations to determine whether a fixated product is the target or a distracter.

Both attention states are influenced by a continuous interaction between bottom-up perceptual features of the display and top-down consumer-specific cognitive factors (Atalay et al. 2012; van der Lans et al. 2008a), as shown in the right top in Fig. 11.4. Bottom-up factors are determined by product design and shelf layout, and represent the visual information that consumers may extract from the shelf. In the localization state, product designs that are different from other, competitive, products on the shelf may receive more bottom-up weight in the salience map. In the identification state, package designs with clear, larger and contrasting logos and text are easier to identify. Top-down factors depend on involvement, memory, and goals (Pieters and Wedel 2007; Yantis and Egeth 1999), and are affected by advertising, experience with the product or the shopping task. In the localization state, consumers searching for a brand may give more top-down weight in the salience map to diagnostic features, such that the target product becomes more salient. In the identification state, top-down factors may influence the accuracy and diagnosticity of the memory template, leading to a more efficient identification process.

11.5.2 Brand Search Model

The brand search process described above, has been formalized as a Bayesian spatial point process model in which the latent localization and identification states are represented by a hidden Markov model component (van der Lans et al. 2008a, b). This model describes for each customer c and eye fixation i the exact location $y_{ci} \in (u, v)$ in pixel coordinates, with $(u, v) \in (D_1, D_2)$ the size of the image in pixels. Given an image resolution of $D_1 = 1,024 \times D_2 = 1,280$, this results in a large choice model with 1,310,720 choice alternatives, each representing a pixel. The explanatory variables $x_{ci}(u, v)$ in this model are the characteristics of the pixels, such as their colors, brightness, edges, and to which AOI they belong. Colors, brightness and edges can be automatically computed using RGB or CIELAB values (i.e., standard color spaces that are used by computers to represent colors) and image processing software, which is available in software packages such as Matlab. Because the eye extracts information from a region of about 2° of visual angle and most eye-tracking devices have a spatial accuracy of 0.5°, pixel information is spatially smoothed with a normal kernel of 2° of visual angle. The spatial smoothing assures that each pixel contains information from its neighbors, with more distant neighbors receiving less weight. The individual-specific effects β_{cs} of each explanatory variable $x_{ci}(u, v)$ on the probability of fixating pixel (u, v) for consumer c depends on the latent attention state $s \in \{1, 2\}$, representing respectively localization or identification. The latent attention states are modelled using a hidden Markov model component with a variable $z_{ci} \in \{1, 2\}$ that indicates the latent state that fixation i of consumer c is generated in. Using this formulation, the likelihood of eye-fixation locations y_c of consumer c can be represented as follows:

$$P(y_c|\beta_c, z_c, x_c) = \prod_{s=1}^{2} p(\beta_{cs}|\mu_s, \Sigma_s) \prod_{i:\, z_{ci}=s} p_s(y_{ci} = (u, v)|\beta_{cs}, x_{ci}(u, v)). \tag{11.1}$$

In (11.1), μ_s and Σ_s are respectively the mean and covariance matrix of the multivariate normal distribution of the individual specific parameter β_{cs}. In (11.1), the constant β_{cs0} is set to 1 for identification.

The explanatory variables $x_{ci}(u, v)$ can be divided into three categories: (1) features $x_{ciF}(u, v)$ that represent the saliency map, (2) systematic strategies $x_{ciS}(u, v)$, and repetition variables $x_{ciR}(u, v)$ representing identification. In the localization state, colors, brightness and edges are used as explanatory variables $x_{ciF}(u, v)$ to represent the saliency map van der Lans et al. (2008b). Using these variables, the saliency map in location (u, v), $S(u, v)$, can be computed as follows:

$$S(u, v) = \frac{1 + \sum_{k \in F} x_{cik}(u, v) \cdot \beta_{c1k}}{\sum_{a \in D_1} \sum_{b \in D_2} \left(1 + \sum_{k \in F} x_{cik}(a, b) \cdot \beta_{c1k}\right)}. \tag{11.2}$$

In (11.2), the saliency in location (u, v) is normalized, such that the total saliency of the visual display equals 1. To capture systematic strategies in $x_{ciS}(u, v)$, each brand $b \in \{1, \ldots, B\}$ on the shelf is represented by an AOI$_b$, which is a component of a possible systematic zigzag or reading strategy. If consumer c follows a reading strategy, and fixation $i-1$ is on AOI$_b$, fixation i should land on AOI$_{b'}$, with b' the brand adjacent to brand b according to the specific strategy. In the identification state, two explanatory variables $x_{ciR}(u, v)$ correspond to AOIs of brands and text on packages. These AOIs reflected repeated fixations on the same AOI.

Using MCMC estimation, van der Lans et al. (2008b) found that consumers switched frequently between localization and identification states. As expected, in the localization state, eye movements were guided by both the saliency map and systematic search strategies, while in the identification state consumers repeatedly fixated on text and package information to determine whether a fixated brand was the target or a distracter (see Fig. 11.5 for an example of an estimated saliency map and attention switching between localization and identification). Interestingly, in a post hoc analysis, the authors found that saliency of brands strongly predicted search performance (speed and accuracy), while the time spent in the identification state had a negative effect on accuracy.

This research was extended in various ways. First, using an experimental procedure in which search goals were varied across experimental conditions, van der Lans et al. (2008a) decomposed brand saliency into a bottom-up contribution that is common across the experimental conditions, and a top-down contribution specific to the search goal induced in each condition. To do so, they extended their brand search model such that the individual level parameters $\beta_{c,s=1}$ corresponding to the saliency map consisted of a common bottom-up and condition-specific top-down component. They modelled this by using the following hierarchical structure:

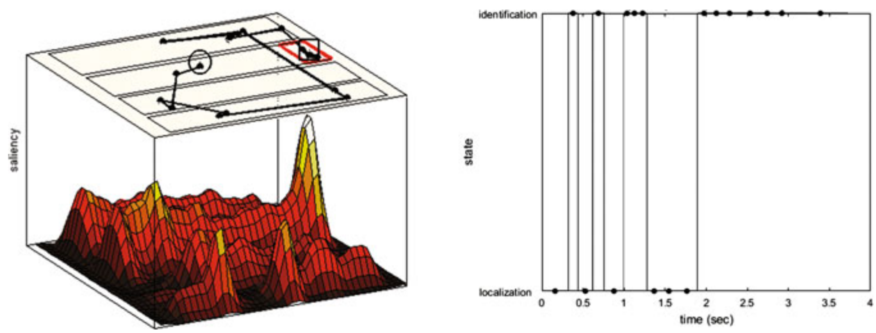


Fig. 11.5 Examples of estimated saliency map and attention switching. *Notes*—The *left* panel illustrates the estimated saliency map of a participant searching on a shopping website. On *top* of this panel is the actual scan path of eye movements, with the final fixations on the area of the target SKU (*red block*). The target SKU is most salient, indicating that the participant is able to weight the features on the shopping website efficiently. The *right* panel indicates switching between the two latent attention states (localization and identification). Note the frequent switching between localization and identification during the search task

$$\beta_{c,s=1} \sim N\left(\mu + \tau_g, \Sigma\right), \tag{11.3}$$

where μ represents the bottom-up part of the saliency map and τ_g the top-down component. The top-down part τ_g depends on the search goal g that was experimentally manipulated. Using this model, they uncovered how brands competed for attention and which features attracted attention. They found that on average two-thirds of a brand's saliency is determined by bottom-up factors that can be influenced by in-store marketing. The remaining one-third of brand salience depended on top-down factors that can be influenced by out-of-store marketing activities such as advertising.

Second, more recently the brand search model has been extended to understand how online display advertising may affect online shopping (van der Lans et al. 2015). Here the Hidden Markov model was extended to three states, where two identification states capture identification of brands and SKUs of a particular brand, respectively. Using different online display ads, this research found that the pack-shot is crucial to increase shopping efficiency as measured by shorter search times. This effect was fully mediated by more efficient switching from identification to localization, indicating that online advertising improves the identification process (template matching) but not the localization process. This is in line with findings from eye tracking research on driving behavior (Huestegge et al. 2010).

11.6 Eye Movements During Choice

Eye movements provide powerful metrics to understand consumer decision making. Economic theories of consumer choice assume that consumers assign utilities to choice alternatives and maximize utility of the choice outcome under budget constraints. However, as discussed in Chap. 7 of this Handbook, maximizing utility takes mental effort and consumers may use screening rules, form consideration sets, and use non-compensatory decision strategies in order to reduce the effort. The strategies that consumers use to search and scan the choice alternatives therefore have important implications of the final choice outcome (see Orquin and Loose 2013 for an extensive review). As a case in point, Drèze et al. (1994) show that simply relocating the shelf position of fast moving consumer goods (FMCGs), such as cereals, juices, and tissues, may double sales of these products. Eye-movement metrics, as summarized in Table 11.1, may provide input to choice models to infer the screening rules and the information that consumers use during decision making. Alternatively, researchers may integrate models that simultaneously explain visual attention patterns and choice outcomes. We will review the first stream of research in Sect. 11.6.1, while we review the latter in Sect. 11.6.2, focusing on the approach by Stüttgen et al. (2012).

11.6.1 Choice Models with Eye-Tracking Metrics as Explanatory Variables

In this stream of literature, researchers aim to understand which information is processed during choice and the order it has been attended. To obtain these insights, eye-movement data is summarized using specific metrics, such as the ones reported in Table 11.1. Russo and Rosen (1975) were among the first to investigate eye-movement patterns of consumers during choice among six alternatives. They were interested in whether consumers are more likely to use a paired comparison choice heuristic (i.e., the multi-alternative choice task is simplified by a sequence of paired comparisons) or the elimination by aspects heuristic proposed by Tversky (1972). To investigate this, each of the six alternatives represented an AOI, and Markov switching between brands was investigated. More specifically, scan-path sequences reflecting binary comparisons between two brands *X* and *Y* (i.e., sub-scans such as *X-Y-X;* see Table 11.1) were coded. Using this procedure, Russo and Rosen found that the majority of refixations were part of a binary comparison, supporting the paired comparison heuristic. In follow-up research in choices for nondurables, Russo and Leclerc (1994) found that choice consisted of three stages: orientation, evaluation, and verification. The paired-comparison heuristic was much more prevalent in the evaluation stage, while the orientation and verification stages consisted of fixations on different products. Reutskaja et al. (2011) used a similar approach to uncover the stopping rule that consumers use in choice under time pressure and under the assumption that search is random. They compared the stopping rule of the optimal search model (i.e., a model that assumes that consumers stop after fixating all alternatives or after running out of time) and of the satisficing model that assumes that search stops after consumers fixate an alternative which values exceed a reservation value. To compare these stopping rules, they analyzed eye-tracking measures such as the number of fixations, average fixation durations and number of AOIs fixated, where each AOI is an alternative. Using these metrics, it was concluded that consumers use a hybrid choice model that is a combination of optimal search and satisficing. Meißner and Decker (2010) used a choice-based conjoint analysis to study how consumers change information acquisition strategies over time. Brand-attribute combinations were presented in matrix format and were coded as AOIs. To capture information acquisition strategies, they compared information acquisition by choice alternative versus attribute across alternatives. They found that respondents switch from attribute-based strategies (for instance, comparing prices across alternatives) to a more holistic strategy (processing multiple attributes within an alternative). In follow-up research using choice-based conjoint, Meißner et al. (2016) related estimated importance of attributes and other related variables such as progression in the choice task and experience to the number of fixations on each attribute using a Poisson count model. Across three choice-based conjoint studies they found that respondents tended to focus on attractive alternatives and that incidental fixations did not bias choices, suggesting that conjoint studies are relatively free from distorting effects from task layouts and random exposures.

Related research used similar approaches, but often related the eye-tracking metrics to choice outcomes. For instance, Pieters and Warlop (1999) recorded the switching patterns during choice among three alternatives. In addition to coding brand packages as AOIs, they also divided the package of the choice alternatives into AOIs representing different attributes (brand name, pictorial and ingredient information). They recorded the following eye-tracking metrics: (1) average fixation duration on an AOI, (2) proportion of attribute AOIs skipped, (3) switching between brands, and (4) switching between attributes within a brand. They found that switching patterns, and to a lesser extend average fixation durations, were strong predictors of choice. Switching patterns were operationalized as the number of changes within a specific brand (intra-brand saccades) and between brands (inter-brand saccades), which both positively related to choice. Average fixation durations on a specific brand also related positively to the choice probability of this brand. Moreover, the fixation patterns differed under time pressure and motivation, which were experimentally manipulated. Using this approach, eye-movement metrics can be used as mediators to understand the effects of visual marketing elements, such as package design, position on the display, and shelf or website layout on choice. Chandon et al. (2009) used this approach to test the effects of in-store (number of facings, shelf position and price) and out-of-store (consumer and brand characteristics) factors on choice. Using noting and reexamination (see Table 11.1) as attention metrics, they found that these metrics mediated the effects of in-store factors. The effect of number of facings of a brand on choice was fully mediated by their attention metrics, while position on the shelf was partly mediated. Similarly, Atalay et al. (2012) demonstrated that gaze durations and fixation frequencies on brands mediated the effect of shelf location on choice. This explained why consumers tend to choose the middle option, as the centre position tends to receive relatively more attention. Deng et al. (2016) demonstrated that horizontally versus vertically organized assortments influenced processing fluency under time pressure. Processing fluency was measured by the number of fixated alternatives per second (scan-path length by second, see Table 11.1). They found that horizontal displays were easier to process, which subsequently led to higher choice quantities. Toubia et al. (2012) demonstrated using eye-tracking data that by making conjoint choice tasks incentive compatible, participants utilized up to 20% more of the information presented to them. Finally, using a Bayesian mediation analysis, Zhang et al. (2009) linked eye tracking data on feature ads to national sales data for the featured brand, collected in the same weeks as when the ads had appeared in the newspapers. They used gaze durations on AOIs representing each feature ad on a display page as eye movement metrics. This study did not model individual level choice but aggregate level sales. It revealed that the size of feature ads positively affected sales, an effect that was fully mediated by attention to the ads. This study thus revealed a critical downstream impact of attention on sales, generalized across thousands of ads and consumers. Shi et al. (2013) investigated the layout of comparison websites on search and choice strategies. They manipulated website design in which product-attribute combinations were either presented in a vertical format (products are presented in columns) or a horizontal format (products are presented in rows). Eye-movements were modeled through two information search processes: attribute-based versus product-based acquisition, in which information is either sampled within attributes across products or within products across attributes. Using a multi-layer hidden Markov model, consumers were allowed to switch between these information acquisition patterns. Overall, consumers tended to start and end in the product-based information search state, while in the middle search tended to be attribute-based. Interestingly, switching between attribute- and product-based patterns depended on the layout of the comparison website. The horizontal format led to more product-based information acquisition, which resulted in significantly different choice outcomes. In addition to (shelf) location and number of facings, researchers also have investigated the effects of visual saliency of packaging (Milosavljevic et al. 2012) and of nutrition information (Bialkova and van Trijp 2011) on visual attention and subsequent choice. Confirming earlier expectations, in both studies visual saliency attracted attention and subsequently influenced choice. Finally, Townsend and Kahn (2013) investigated the effects of assortment size and visual versus verbal stimuli on eye-movement patterns during choice. They found that larger assortment size resulted in more systematic processing, especially for verbal stimuli.

11.6.2 Integrated Models of Choice and Visual Attention

The previous stream of literature demonstrates that eye-movement metrics, reflecting information acquisition during choice, are important predictors of the final choice outcome. A limitation of this literature is that it treats the visual attention process as exogenous, while in reality search and choice are dynamically linked (Shimojo et al. 2003). Thus, in addition to explaining the final choice outcome as a function of eye-movement metrics, it is important to model the allocation of eye-fixations in the component search process to fully understand how consumers make choices and how this is affected by aspects of the marketing mix. Integrating models of choice and visual attention is a relatively new area of research and there have been only few papers in marketing. Here, we focus on Stüttgen et al. (2012), who developed one of the first formal models that integrated search and choice, which they applied to eye-tracking data in a conjoint choice experiment. Their model incorporates a satisficing choice rule in which consumers choose a product if it's value exceeds a certain threshold. In contrast to the model by Reutskaja et al. (2011) as discussed in the previous Section, they explicitly modeled the eye-movement search pattern based on the brand search model developed by van der Lans et al. (2008a, see Fig. 11.4). However, they extended this search model significantly to capture choice through a satisficing stopping rule. The assumed process is as follows. First, a consumer assigns a status: undetermined, unsatisfactory or satisfactory to each product on the shelf. If a product has not been fixated, its status is undetermined. To determine whether a product is satisfactory, a consumer needs to fixate all attributes a of that product, and the value of each attribute needs to exceed a specific level, which was assumed to follow a Bernoulli distribution. Second, in addition to saliency and systematic strategies in the localization state and repetitions in the identification state, the location of the next fixation also was assumed to depend on the status of the product, as consumers are expected to fixate undetermined products more. Third, in addition to global and local latent attention states, they included a third termination state in which the consumer makes a choice and terminates the process. Because these variables make consumers more likely to switch to the termination state, switching probabilities depend on the status of products and the number of fixations, and are proportional to:

$$p(z_{ci}=3) \propto \exp(\lambda_{30}+\lambda_{31}I(z_{ci-1}=1)+\lambda_{32}i^{*}+\lambda_{33}\max(\wp)). \tag{11.4}$$

In (11.4), λ_{30} captures the baseline switching probability to the termination state ($z_{ic} = 3$), λ_{31} is a parameter that captures the dependence of that switching probability on the previous state z_{ci-1} as in van der Lans et al. (2008a). Moreover, switching also depends on the (adjusted) number of fixations i^{*} through the parameter λ_{32}, and whether any of the choice alternatives is currently determined satisfactory (i.e., $\wp_x=1$ if product x is satisfactory), through the parameter λ_{33}.

Once a consumer is in the termination state 3, s/he chooses a product x with the following probabilities:

$$\Pr(x|x \text{ is unsatisfactory}) = 0, \tag{11.5}$$

$$\Pr(x|x \text{ is satisfactory}) = \frac{1}{\aleph}, \tag{11.6}$$

$$\Pr(x|x \text{ is undertermined}) = \frac{\tau}{\aleph}, \text{ with } 0 \leq \tau \leq 1. \tag{11.7}$$

In (11.6) and (11.7), $\aleph$ is a normalizing constant such that the probabilities sum to 1, and τ is expected to be close to zero, reflecting that consumers most likely choose satisfactory products.

Stüttgen et al. (2012) found the satisficing choice stopping rule was supported for consumers searching for instant noodles. Moreover, in a holdout sample, the model demonstrated superior predictive power compared to standard choice models. The results by Stüttgen et al. (2012) suggest that integrating search and choice may provide a detailed understanding of how product design and shelf layout may influence choice, via consumers visual search for the information that they intend to use in making decisions.

In another effort along these lines, using a vertical format to present product-attribute combinations in a choice-based conjoint experiment, Yang et al. (2015) developed a dynamic discrete choice model that incorporates the costs of searching for information. Their model assumes bounded rationality. Consumers choose the alternative that maximizes utility, using only information that has been acquired. Moreover, they assume that acquiring additional information is costly and that consumers are forward looking and tradeoff these costs with the expected benefits of finding better choice alternatives. In addition to the cost-benefit tradeoffs, their model incorporates fatigue and imperfect memory as well as proximity effects, such that consecutive fixations are more likely to be used to inspect contiguous cells (Shi et al. 2013). Their results show that modeling eye-movements as a tradeoff between search and choice, instead of modeling only choice in conjoint experiments, significantly improves parameter estimates, the fundamental understanding of the choice process, as well as out-of-sample predictions.

11.7 Conclusion

While early eye-tracking research in marketing was hampered by labor intensive and relatively inaccurate measures of visual attention, today's eye-tracking devices are unobtrusive, user friendly and accurate. These developments have greatly increased interest in eye-movement research and provided new opportunities for marketing academics and practitioners to model the key processes of search and choice. This research consistently shows that eye-movement data provides critical information to better predict choices and highlights the factors that influence search and choice success.

Marketing research industry is increasingly embracing eye tracking as a powerful tool to glean insight into the effectiveness of visual marketing effort, and several major product and service companies even have their own eye tracking labs. While this is an exciting development and the investments have most often paid off, the vast majority of eye-movement research in practice, unfortunately, still stops short of presenting more than the results in heat-maps and other graphical representations of eye-movement patterns. While initially appealing, these visual displays lack deep and actionable insights, and their appeal may therefore wane to the detriment of the use of eye-tracking research in practice. On the contrary, the statistical analysis of metrics, and model based approaches to the prediction of search and choice from eye-movements, such as the ones discussed in this chapter, hold the promise of enabling optimization of marketing effort and predictive analytics. In particular, the explanation and prediction of search and choice results is important and holds great promise. This holds not only for consumer search and choice, but importantly these methods could also be used to understand, improve and support manager's decision making. For instance, how do managers read policy documents, or scan tables and figures and how can information be optimally presented. For instance, Duclos (2015) uses eye-tracking patterns to understand how investors process graphical information and how visual attention could explain potential biases in decision making. Nowadays, eye-tracking solutions based on front-facing cameras are incorporated in many electronic devices, such as laptops, smart-TVs, mobile phones, and tablets (Wilson 2015). Such devices will in the near future likely produce large quantities of eye-tracking data that provide deep insights into the underlying search and choice processes of consumers on these devices, and enable predictions of the outcomes of these processes before they are finalized. Model-based approaches will be critical in that regard. An important challenge for future research is to develop realistic eye tracking models of search and choice that are scalable to these large scale datasets. As the number of stimuli for which eye tracking data is available increases, it will also become important to develop image processing tools that automatically extract the most important information from stimuli that can be used as input to search and choice models. We hope that this overview stimulates researchers in practice and academia to play an active role in those future developments and capitalize on the opportunities to use eye-tracking data to provide deeper insights and better forecasts of consumer search and choice.

Acknowledgements The first author gratefully acknowledges financial support from the Hong Kong Research Grant Council (GRF645511).

References

Anstis, S.M. 1974. A Chart demonstrating variations in acuity with retinal position. *Vision Research* 14: 589–592.

Aribarg, A., R. Pieters, and M. Wedel. 2010. Raising the bar: Bias adjustment of recognition tests in advertising. *Journal of Marketing Research* 47 (3): 387–400.

Atalay, A.S., H.O. Bodur, and D. Rasolofoarison. 2012. Shining in the center: central gaze cascade effect on product choice. *Journal of Consumer Research.* 39 (4): 848–866.

Bialkova, S., and H.C.M. van Trijp. 2011. An efficient methodology for assessing attention to and effect of nutrition information displayed front-of-pack. *Food Quality and Preference* 22: 592–601.

Brisson, J., M. Mainville, D. Mailloux, C. Beaulieu, J. Serres, and S. Sirois. 2013. Pupil diameter measurement errors as a function of gaze direction in corneal reflection eyetrackers. *Behavior Research Methods* 45: 1322–1331.

Burr, D. 2005. Vision: In the Blink of an Eye. *Current Biology* 15 (14): 554–556.

Carlisle, N.B., J.T. Arita, D. Pardo, and G.F. Woodman. 2011. Attentional templates in visual working memory. *The Journal of Neuroscience.* 21 (25): 9315–9322.

Chandon, P., J.W. Hutchinson, E.T. Bradlow, and S.H. Young. 2008. Measuring the value of point-of-purchase marketing with commercial eye-tracking data. In *Visual marketing: From attention to action*, ed. M. Wedel, and R. Pieters, 225–258. New York: Lawrence Erlbaum Associates.

Chandon, P., J.W. Hutchinson, E.T. Bradlow, and S.H. Young. 2009. Does in-store marketing work? Effects of the number and position of shelf facings on brand attention and evaluation at the point of purchase. *Journal of Marketing.* 73 (November): 1–17.

Chen, M., and J.A. Bargh. 1999. Consequences of automatic evaluation: immediate behavioral predispositions to approach or avoid the stimulus. *Personality and Social Psychology Bulletin* 25 (2): 215–224.

Deng, X., B.E. Kahn, H.R. Unnava, and H. Lee. 2016. A “Wide” variety: Effects of horizontal versus vertical display on assortment processing, perceived variety, and choice. *Journal of Marketing Research* 53 (5): 682–698.

Desimone, R., and J. Duncan. 1995. Neural mechanisms of selective visual attention. *Annual Review of Neuroscience* 18: 193–222.

Drèze, X., S.J. Hoch, and M.E. Purk. 1994. Shelf management and space allocation. *Journal of Retailing* 70 (4): 301–326.

Droll, J.A., and M.M. Hayhoe. 2007. Trade-offs betrween gaze and working memory use. *Journal of Experimental Psychology: Human Perception and Performance* 33 (6): 1352–1365.

Duchowski, A.T. 2003. *Eye tracking methodology: Theory and practice*. London: Springer-Verlag.

Duclos, R. 2015. The psychology of investment behavior: (De)Biasing financial decision-making one graph at a time. *Journal of Consumer Psychology.* 25 (2): 317–325.

Duncan, J., and G.W. Humphreys. 1989. Visual search and stimulus similarity. *Psychological Review* 96 (3): 433–458.

Findlay, J.M. 2005. Covert attention and saccadic eye movements. In *Neurobiology of attention*, ed. L. Itti, G. Rees, and J.K. Tsotsos, 114–116. Amsterdam: Elsevier Academic Press.

Folk, C.L., R.W. Remington, and J.C. Johnston. 1992. Involuntary covert orienting is contingent on attentional control settings. *Journal of Experimental Psychology: Human Perception and Performance* 18 (4): 1030–1044.

Gramopadhye, A.K., C.G. Drury, and J. Sharit. 1997. Feedback strategies for visual search in airframe structural inspection. *International Journal of Industrial Ergonomics* 19: 333–344.

Higgins, E., and K. Rayner. 2015. Transsaccadic processing: Stability, integration, and the potential role of remapping. *Attention, Perception, & Psychophysics* 77 (1): 3–27.

Holmqvist, K., M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. van de Weijer. 2011. *Eye tracking: A comprehensive guide to methods and measures*. Oxford, UK: Oxford University Press.

Huestegge, L., E.-M. Skottke, S. Anders, J. Müsseler, and G. Debus. 2010. The development of hazard perception: Dissociation of visal orientation and hazard processing. *Transportation Research Part F* 13: 1–8.

Itti, L., and C. Koch. 2001. Computational modelling of visual attention. *Nature Reviews Neuroscience* 2 (3): 194–203.

Jiang, Y., R. Adaval, Y. Steinhart, and R. S. Wyer Jr. 2014. Imagining yourself in the scene: The interactive effects of goal-driven self-imagery and visual perspectives on consumer behavior. *Journal of Consumer Research* 41 (Aug): 418–35.

Karslake, J.S. 1940. The purdue eye-camera: A practical apparatus for studying the attention value of advertisements. *Journal of Applied Psychology* 24 (4): 417–440.

Koch, C., and S. Ullman. 1985. Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology.* 4: 219–227.

Krischer, C.C., and W.H. Zangemeister. 2007. Scanpaths in reading and picture viewing: Computer-assisted optimization of display conditions. *Computers in Biology and Medicine* 37 (7): 947–956.

Liechty, J., R. Pieters, and M. Wedel. 2003. Global and local covert visual attention: Evidence from a Bayesian hidden Markov model. *Psychometrika* 68 (4): 519–541.

Mathews, Z., R. Cetnarski, and P.F.M.J. Verschure. 2015. Visual anticipation biases conscious decision making but not bottom-up visual processing. *Frontiers in Psychology.* 5 (1443): 1–14.

McCarley, J.S., A.F. Kramer, C.D. Wickens, E.D. Vidoni, and W.R. Boot. 2004. Visual skills in airport-security screening. *Psychological Science* 15 (5): 302–306.

Meißner, M., and R. Decker. 2010. Eye-tracking information processing in choice-based conjoint analysis. *International Journal of Market Research.* 52 (5): 591–610.

Meißner, M., A. Musalem, and J. Huber. 2016. Eye tracking reveals processes that enable conjoint choices to become increasingly efficient with practice. *Journal of Marketing Research* 53 (1): 1–17.

Milosavljevic, M., V. Navalpakkam, C. Koch, and A. Rangel. 2012. Relative visual saliency differences induce sizable bias in consumer choice. *Journal of Consumer Psychology.* 22: 67–74.

Monk, T.H. 1984. Search. In *Sustained attention in human performance*, ed. J.S. Warm, 293–321. New York: Wiley.

Nelson, W.W., and G.R. Loftus. 1980. The functional visual field during picture viewing. *Journal of Experimental Psychology: Human Learning and Memory.* 6 (4): 391–399.

Niebur, E., and C. Koch. 1998. Computational architectures for attention. In *The attentive brain*, ed. R. Parasuraman, 163–186. Cambridge, MA: The MIT Press.

Nixon, H.K. 1924. Attention and interest in advertising. *Archives of Psychology.* 72: 5–67.

Orquin, J.L., and S.M. Loose. 2013. Attention and choice: A review on eye movements in decision making. *Acta Psychologica* 144 (1): 190–205.

Pieters, R., and L. Warlop. 1999. Visual attention during brand choice: The impact of time pressure and task motivation. *International Journal of Research in Marketing* 16: 1–16.

Pieters, R., L. Warlop, and M. Wedel. 2002. Breaking through the clutter: Benefits of advertisement originality and familiarity for brand attention and memory. *Management Science* 48 (6): 765–781.

Pieters, R., and M. Wedel. 2004. Attention capture and transfer in advertising: Brand, pictorial and text size effects. *Journal of Marketing.* 68 (2): 36–50.

Pieters, R., and M. Wedel. 2007. Goal control of visual attention to advertising: The yarbus implication. *Journal of Consumer Research.* 34 (2): 224–233.

Pieters, R., and M. Wedel. 2012. Ad gist: Ad communication in a single eye fixation. *Marketing Science.* 31 (1): 59–73.

Poffenberger, A.T. 1925. *Psychology in advertising*. Chicago, Ill: A.W. Shaw Company.

Ponsoda, V., D. Scott, and J.M. Findlay. 1995. A Probability vector and transition matrix analysis of eye movements during visual search. *Acta Psychologica* 88: 167–185.

Rayner, K. 1998. Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin* 124 (3): 372–422.

Reutskaja, E., R. Nagel, C.F. Camerer, and A. Rangel. 2011. Search dynamics in consumer choice under time pressure: An eye-tracking study. *American Economic Review* 101 (April): 900–926.

Robinson, P.J.A. 1997. Radiology's Achilles' heel: Error and variation in the interpretation of the Röntgen image. *The British Journal of Radiology* 70: 1085–1098.

Rosbergen, E., R. Pieters, and M. Wedel. 1997. Visual attention to advertising: a segment-level analysis. *Journal of Consumer Research.* 24 (3): 305–314.

Russo, J.E., and F. Leclerc. 1994. An eye-fixation analysis of choice processes for consumer nondurables. *Journal of Consumer Research.* 21: 274–290.

Russo, J.E., and L.D. Rosen. 1975. An eye fixation analysis of multialternative choice. *Memory and Cognition* 3 (3): 267–276.

Salvucci, D.D., and J.H. Goldberg. 2000. Identifying fixations and saccades in eye-tracking protocols. In *Proceedings of the eye tracking research and applications (ETRA) symposium*, 71–78.

Shi, S.W., M. Wedel, and R. Pieters. 2013. Information acquisition during online decision making: A model-based exploration using eye-tracking data. *Management Science* 59 (5): 1009–1026.

Shimojo, S., C. Simion, E. Shimojo, and C. Scheier. 2003. Gaze bias both reflects and influences preference. *Nature Neuroscience* 6 (12): 1317–1322.

Simons, D.J., and R.A. Rensink. 2005. Change blindness: Past, present, and future. *TRENDS in Cognitive Sciences.* 9 (1): 16–20.

Stewart, A.J., M.J. Pickering, and P. Sturt. 2004. Using eye movements during reading as an implicit measure of the acceptability of brand extensions. *Applied Cognitive Psychology.* 18: 679–709.

Stüttgen, P., P. Boatwright, and R.T. Monroe. 2012. A Satisficing choice model. *Marketing Science.* 31 (6): 878–99.

Teixeira, T., M. Wedel, and R. Pieters. 2012. Emotion-induced engagement in internet video advertisements. *Journal of Marketing Research* 49 (April): 144–159.

Teixeira, T.S., M. Wedel, and R. Pieters. 2010. Moment-to-moment optimal branding in tv commercials: preventing avoidance by pulsing. *Marketing Science.* 29 (5): 783–804.

Thompson, K.G. 2005. Dissociation of selection from saccade programming. In *Neurobiology of attention*, ed. L. Itti, G. Rees, and J.K. Tsotsos, 124–129. San Diego, CA: Elsevier.

Tiffin, J., and D.M. Winick. 1954. A comparison of two methods of measuring the attention-drawing power of magazine advertisements. *The Journal of Applied Psychology.* 38 (4): 272–275.

Tobii. 2011. Accuracy and precision test method for remote eye trackers.

Toubia, O., M.G. De Jong, D. Stieger, and J. Füller. 2012. Measuring consumer preferences using conjoint poker. *Marketing Science.* 31 (1): 138–56.

Townsend, C., and B.E. Kahn. 2013. The "visual preference heuristic": The influence of visual versus verbal depiction on assortment processing, perceived variety, and choice overload. *Journal of Consumer Research.* 40 (Feb): 993–1015.

Treue, S. 2003. Visual attention: The where, what, how and why of saliency. *Current Opinion in Neurobiology* 13: 428–432.

Tversky, A. 1972. Elimination by aspects: A theory of choice. *Psychological Review* 79 (4): 281–299.

van der Lans, R., R. Pieters, and M. Wedel. 2008a. Competitive brand salience. *Marketing Science.* 27 (5): 452–461.

van der Lans, R., R. Pieters, and M. Wedel. 2008b. Eye-movement analysis of search effectiveness. *Journal of the American Statistical Association.* 103 (482): 452–461.

van der Lans, R., M. Wedel, and R. Pieters. 2011. Defining eye-fixation sequences across individuals and tasks: The binocular-individual threshold (Bit) algorithm. *Behavior Research Methods* 43 (1): 239–257.

van der Lans, R., M. Wedel, and R. Pieters. 2015. How online advertising can help the implementation of brand choices: A process analysis using eye tracking. *Working Paper*.

Van Orden, K.F., W. Limbert, S. Makeig, and T.-P. Jung. 2001. Eye activity correlates of workload during a visuospatial memory task. *Human Factors* 43 (1): 111–121.

van Raaij, F. W. 1977. Consumer information processing for different information structures and formats. In *Advances in consumer research*, vol. 4, ed. J. W.D. Perreault. Association of Consumer Research.

Vickery, T.J., L.-W. King, and Y. Jiang. 2005. Setting up the target template in visual search. *Journal of Vision* 5 (1): 81–92.

Vlaskamp, B.N.S., and I.T.C. Hooge. 2006. Crowding degrades saccadic search performance. *Vision Research* 46 (3): 417–425.

Wade, N.J. 2010. Pioneers of eye movement research. *i-Perception.* 1: 33–68.

Wedel, M., and R. Pieters. 2000. Eye fixations on advertisements and memory for brands: A model and findings. *Marketing Science.* 19: 297–312.

Wedel, M. and R. Pieters. 2008. Eye-tracking for visual marketing. *Foundations and Trends (R) in Marketing.* 1 (4): 231–320.

Wedel, M., and R. Pieters. 2015. The buffer effect: The role of color when advertising exposures are brief and blurred. *Marketing Science* 34 (1): 134–143.

Wilson, M. (2015). *Tobii Eye-Tracking Tech Is Making Its Way into More Devices.* http://www.Kitguru.Net/Channel/Generaltech/Matthew-Wilson/Tobii-Eye-Tracking-Tech-Is-Making-Its-Way-in-to-More-Devices/.

Wolfe, J.M. 1998. Visual search. In *Attention*, ed. H. Pashler, 13–73. Hove, East Sussex, UK: University College London Press Ltd.

Wolfe, J.M., and T.S. Horowitz. 2004. What attributes guide the deployment of visual attention and how do they do it? *Nature Reviews Neuroscience* 5 (June): 1–7.

Yang, L.C., O. Toubia, and M.G. De Jong. 2015. A bounded rationality model of information search and choice in preference measurement. *Journal of Marketing Research.* 52 (April): 166–83.

Yantis, S. and H.E. Egeth. 1999. On the distinction between visual salience and stimulus-driven attentional capture. *Journal of Experimental Psychology: Human Perception and Performance.* 25 (3): 661–76.

Zhang, J., M. Wedel, and R. Pieters. 2009. Sales effects of attention to feature advertisements: A Bayesian mediation analysis. *Journal of Marketing Research.* 46 (Oct): 669–81.

Chapter 12
Business-Cycle Research in Marketing

Barbara Deleersnyder and Marnik G. Dekimpe

12.1 Introduction

The recent Global Financial crisis has reminded companies that macro-economic developments, and especially business-cycle fluctuations, can be among the most influential determinants of a firm's activities and performance. According to researchers of the U.S. National Bureau of Economic Research (NBER) , business cycles consist of expansions occurring at about the same time in many economic activities, followed by similarly general contractions, and revivals which merge into the expansion phase of the next cycle (see Burns and Mitchell 1946, p. 3, or Stock and Watson 1999, p. 5). These cycles are recurrent but not strictly periodic, and are normally visible in aggregate economic series such as real Gross Domestic Product (GDP), real income, or employment, among others (Stock and Watson 1999). Often, aggregate information on the state of the national economy is considered. However, cyclical expansions and contractions need not have an equal impact on every industry, nor on all firms and brands within that industry. This has spurred a new stream of research in marketing that considers business-cycle relationships with (marketing and performance) variables at a more disaggregate level. Existing marketing research in this area can be broadly divided into three research streams. First, several studies have evaluated the impact of business-cycle fluctuations on brand, firm or industry *performance*. Second, a number of studies have focused on

B. Deleersnyder (✉) · M.G. Dekimpe
Tilburg University, Tilburg, The Netherlands
e-mail: b.deleersnyder@tilburguniversity.edu; B.Deleersnyder@uvt.nl

M.G. Dekimpe
e-mail: m.g.dekimpe@tilburguniversity.edu

M.G. Dekimpe
KU Leuven, Leuven, Belgium

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_12

the relation between the business cycle and *marketing conduct*. Finally, a third set of studies has evaluated how *marketing effectiveness* changes over the business cycle.

12.2 Main Insights from Business-Cycle Research in Marketing

Similar to a regression framework that specifies brand, firm or industry Performance (P) in terms of its Marketing input (M) to capture marketing effectiveness (β) in the form: $P = \alpha + \beta \cdot M$, business-cycle research in marketing can be broadly categorized into three research streams: research on the impact of business cycles on performance, research on the impact of business cycles on marketing conduct, and research on the changes in marketing effectiveness over the business cycle. Importantly, each of the three factors (P, M and β) has been found to vary with the general state of the Economy (E). We briefly elaborate on the general conclusions and insights from each of these research streams, and then introduce the methodological approaches to study these relations.

Within the first research stream, marketing studies have looked at a variety of *performance* (P) metrics. At the category level, the impact of business cycles has been evaluated on durable sales by Deleersnyder et al. (2004), on international tourism by Dekimpe et al. (2016), and on healthcare expenditures by Cleeren et al. (2016). Studies by Lamey et al. (2007, 2012) and Deleersnyder et al. (2009) have linked business-cycle swings to private-label performance, while Lamey (2014) extended this to discounter shares. Others, such as van Heerde et al. (2013), have examined the impact on brand sales. Overall, these marketing studies show that disaggregate performance P at the category, firm and brand level can vary strongly in relation to the business cycle. Most performance series move in the same direction as the general state of the economy, even though the swings in the economy are sometimes amplified in these series. Interestingly, discounter share and private-label performance move in the opposite direction as the business cycle, as they seem to flourish each time the economy winds down (Lamey 2014; Lamey et al. 2007, 2012).

Second, several studies have determined how *marketing conduct* (M) is adjusted in relation to business-cycle swings. Most studies in this research stream focused on advertising spending decisions that are found to be cut dramatically during economic downturns (see, e.g., Deleersnyder et al. 2009; Özturan et al. 2014; Kashmiri and Mahajan 2014). But also innovation activity (Kashmiri and Mahajan 2014) and prices (Deleersnyder et al. 2004) have been shown to move closely with the business cycle. Lamey et al. (2012) evaluated management adjustments across a broad range of seven marketing instruments (major and minor innovations, advertising spending, regular price, display activity, feature promotions, and temporary price cuts) in relation to the business cycle in the U.S. Studies in this research stream typically show that reducing marketing support in a contraction may actually amplify the impact of economic fluctuations on brand or firm performance.

Finally, a third research stream has focused on how the *effectiveness* (β) of various marketing-mix actions changes over the business cycle. This has, for example, been documented for advertising (see, e.g., Srinivasan et al. 2011; Steenkamp and Fang 2011; van Heerde et al. 2013), innovations (e.g., Srinivasan et al. 2011; Steenkamp and Fang 2011), and price promotions (e.g., Gordon et al. 2013; Sethuraman et al. 2011; van Heerde et al. 2013). Research on advertising and innovation have typically represented fairly general (aggregate) numbers (Srinivasan et al. 2011; Steenkamp and Fang 2011; van Heerde et al. 2013), but more research should explore differences between, for instance, online and offline advertising media, or various kinds of innovations, and this across different industries. Furthermore, Srinivasan et al. (2011) show that most firms overspend on advertising in a recession, and van Heerde et al. (2013) show that advertising elasticities are lower in a recession, suggesting that advertising spending should be reduced in economic downturns. In contrast, evidence of a higher advertising effectiveness during the contraction is reported in Steenkamp and Fang (2011) and Frankenberger and Graham (2003).

A variety of methods has been used to make these business-cycle-related inferences. The objective of this article is to present an overview of the most commonly used techniques. We catalogue them along two dimensions: (i) how the business cycle is inferred (GDP-based, through a discrete categorization, or based on filtering techniques); and (ii) how the inferred business cycle is subsequently linked to the marketing variables of interest.

12.3 How Is the Business Cycle Inferred?

To evaluate the impact of the business-cycle, researchers need to first assess or measure the state of the economy in E. Marketing research has used three quite different approaches to infer (or identify) the overall state of the economy. A first set of studies, reviewed in Sect. 12.3.1, relied on official economic indicators such as national GDP. Other studies, reviewed in Sect. 12.3.2, have used the official recession dates as published by governmental institutions. A final set of studies, discussed in Sect. 12.3.3, inferred business-cycle fluctuations through business-cycle filtering procedures applied to macro-economic measures (mostly GDP).[1] These three approaches to infer business-cycle fluctuations from aggregate economic indicators differ according to the type of indicator underlying the

[1]While the majority of the studies in this area rely on objective or "hard" economic data to assess the state of the economy, occasionally, marketing studies have also used surveys to evaluate consumers' and/or managers' perception about the 'severity of the recession' affecting their industry (e.g., Srinivasan et al. 2005).

business cycle (GDP or other economic aggregate), the number of economic indicators used (single vs. multiple indicators), the amount of information contained in the resulting business-cycle metric (discrete vs. continuous), and the treatment of the general growth pattern underlying the economic series E (retained or removed in the business cycle).

All three approaches have been used to infer business-cycle fluctuations from aggregate economic indicators E. Note, however, that filtering techniques introduced in Sect. 12.3.3 are equally suitable and can be applied to separate directly business cycle related ups and downs from *any* series of interest to marketers such as (disaggregate) performance (P) or marketing conduct (M).

12.3.1 Proxied by Official Economic Indicator(S)

A country's GDP (or GNP) is an official indicator most often used to determine the general state of an economy. It represents the total output produced in a country or region during a certain period. GDP figures are published periodically (e.g., quarterly and/or yearly) by official (national or international) institutions such as the International Monetary Fund or the Worldbank, and are publicly available for most countries. GDP is often expressed on a per-capita basis to take into account population growth/differences. Fluctuations in aggregate output have been found to be at the core of the business cycle, making it a good proxy for the country's economic activity as a whole (Stock and Watson 1999, p. 15).

However, marketing studies have also relied on (or supplemented GDP data with) other economic indicators such as household income (Gordon et al. 2013), unemployment (Cha et al. 2015; Kumar et al. 2014), inflation (Özturan et al. 2014), currency depreciation (Dutt and Padmanabhan 2011), or consumer confidence (Lemmens et al. 2007), among others. In Ma et al. (2011), the authors focus on the price of gasoline as a relevant macro-economic factor that significantly influences consumers' weekly shopping behavior.

Macro-economic indicators have been used to infer both the absolute level of an economy, and to study changes in that level. In business-cycle studies, researchers are mainly interested in period-to-period variation in the economic activity. Therefore, they often include the growth rate (in percentages), rather than the absolute levels of the respective series in their analyses (see, e.g., Gordon et al. 2013).

Apart from working with a continuous economic variable (level and/or growth rate), some researchers such as Sethuraman et al. (2011) or Kamakura and Du (2012) have also used these metrics to delineate discrete periods of economic adversity. Quite often, researchers refer to the concept of a 'recession' when economic activity significantly weakens for an extended period of time. The financial and business press commonly defines an economic 'recession' as two (or more) consecutive quarters of negative growth in GDP (or an alternative economic indicator) (Shiskin 1974). Some marketing researchers (e.g., Sethuraman et al. 2011) have followed this definition, while others already classify *any* period with negative GDP growth as a recessionary period (e.g., Gordon et al. 2013).

Such a binary classification is clearly less informative than a continuous metric on features such as the strength of the downturn and/or the speed of decline/recovery. Moreover, it has been recognized (Baxter and King 1999; Hodrick and Prescott 1997) that the business cycle is intrinsically a continuous phenomenon, rather than a discrete concept, and that it is hard even for specialized agencies to pinpoint the exact starting and ending dates of a recession period. In some countries, like the U.S., the government has appointed an official committee to agree on when the national economy is turning into a recession.

12.3.2 Discrete Categorization by Official Economic Instances

In the U.S., the National Bureau of Economic Research (NBER) established the 'Business Cycle Dating Committee' which identifies and maintains a chronology of the U.S. business cycle (see http://www.nber.org/cycles/recessions.html). The Committee identifies months of peaks and troughs in economic activity, and a recession/contraction is defined as the period between a peak and a trough, while an expansion is determined as the period between the trough and the next peak. Details on the monthly U.S. turning points as published by the NBER are presented in Appendix A. Based on these dates, several marketing studies examining yearly U.S. data including Frankenberger and Graham (2003), Graham and Frankenberger (2011), Srinivasan et al. (2011), Kamakura and Du (2012) and Kashmiri and Mahajan (2014) have classified a year as a recession year if the majority of the year (more than 6 months) occurred during an NBER-classified recessionary trough (cf. Srinivasan et al. 2011, p. 55).

The official NBER U.S. business cycle dates are well-accepted. On the positive side, it does not require that the researcher observes a full (or multiple) business cycle(s); as such, no long time series are required. Unfortunately, the approach also suffers from a number of drawbacks. First, the Committee does not have a fixed definition or approach. It uses a judgement-based procedure that has been criticized for a lack of statistical foundation and for its rigid focus on absolute declines (as opposed to growth slowdowns) in various output and other economic indicators (Stock and Watson 1999; Lamey et al. 2012).

Second, the NBER Business Cycle Dating Committee is unique to the U.S. While there are several attempts to identify cycles by national statistical agencies and central banks in other countries, there is little agreement on what economic indicators are (most) informative, nor on what patterns of (absolute or relative) growth or decline in these economic aggregates are indicative of a recession. Some more coordinated efforts to identify official business cycle-turning points for larger economic regions like the Eurozone are undertaken by the Center for Economic

Policy Research (CEPR). Alternatively, the Economic Cycle Research Institute (ECRI) publishes dates of business-cycle peaks and troughs for a select number of 21 strong economic entities around the world (such as the U.K., China, Japan or South Africa). Still, even though important international interdependencies across economic markets exist and some shocks can hit the economic activity globally (Baxter and Kouparitsas 2005), evidence accrues that economic contractions do not always affect all countries, and countries' business cycles do not necessarily coincide with the U.S. or even with a neighboring country's business cycle. Accordingly, when examining business-cycle phenomena in other Western or emerging countries, researchers should be careful when relying on (or "adopting") such official recession dates (even from geographically close or apparently fairly similar countries).

Finally, official business cycle dates ignore the strength and uniqueness of each individual contraction period within a given market. Zarnowitz (1985), for example, showed that there is significant variation in duration and phase across successive business-cycle swings. Some contractions are more dramatic than others, and also within a single contraction period, different phases can be distinguished where the impact and behavior may differ. The same holds for economic booms that can be more or less extreme. Note that with temporally aggregated data, the recession classification is only a rough approximation. With yearly data, for instance, this approach is unable to distinguish between a year in which all months are classified as a recession, as opposed to a year where only 7 months are part of a recession.

12.3.3 *Infer Cyclical Information Using Business-Cycle Filters*

To overcome these limitations, the economic literature has developed filtering techniques that allow researchers to self-extract the relevant business-cycle information from a series. Also an increasing number of marketing researchers (see, e.g., Deleersnyder et al. 2004, 2009; Lamey et al. 2007, 2012) have come to rely on them. Contraction periods identified by applying business-cycle filters to U.S. real per-capita GDP data have been shown by Christiano et al. (2011, pp. 323–324 and Fig. 4) to closely mimic the periods marked as recessions by the official NBER Business Cycle Dating Committee. We conduct a similar comparison in this section to illustrate how business-cycle filters work (see Fig. 12.1, panel C, to be discussed later). However, as in Christiano et al. (2011), we also point out some differences, and emphasize the richness and additional insights gained from implementing filtering techniques over the NBER or other discrete classification methods.

Not all over-time variation in a series can be attributed to business-cycle swings. Business-cycle filters are designed to separate fluctuations related to the business cycle from other sources of variation in the series of interest (such as seasonality or a long-term trend). Business-cycle fluctuations are recurrent ups and downs of

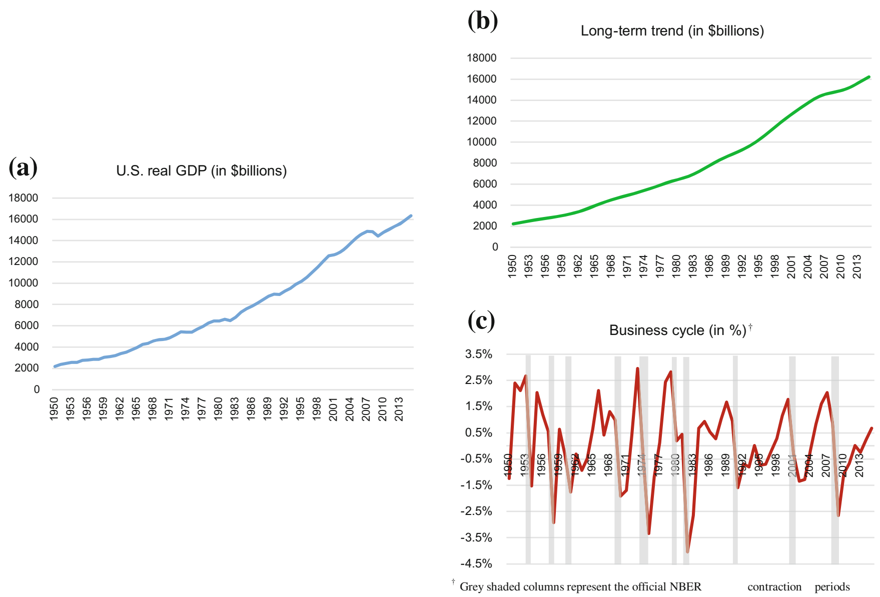


Fig. 12.1 Decomposing yearly U.S. real GDP (1950–2015) using the HP business-cycle filter (λ = 10)*. **Data source* http://www.bea.gov/national/

varying length that are not strictly periodic (for instance, they do not recur exactly every 5 years). Based on the observation from several NBER researchers (see, e.g., Burns and Mitchell 1946; Christiano and Fitzgerald 1998) that business cycles typically last between 1.5 and 8 years, the underlying idea of business-cycle filters is to pass through all components of the time series with periodic fluctuations between, say, 6 and 32 quarters. These filters, with proper adaptation, can be used on data series with different levels of temporal aggregation. Interestingly, even though the filters have been designed and applied in the economics literature to detect business cycles from various aggregate economic series (E) (see, e.g., Stock and Watson 1999), these techniques can also be implemented directly to the performance (P) or marketing conduct (M) series of interest in order to extract the variation that occurs at (and is related to) business-cycle periodicities.

While the use of business-cycle filters is well accepted in economics, little agreement exists on which filter should be preferred. We refer to Canova (1998) or Baxter and King (1999) for an in-depth review and evaluation of alternative filtering methods. These studies conclude that while different filters need not produce exactly the same cyclical components (as the filters tend to extract slightly different frequencies from the data), the more popular ones (as the Hodrick-Prescott filter and the Baxter and King filter) tend to extract quite similar cyclical information from a series. These two filters are also the most often used filters in recent marketing studies. Below, we elaborate on the specification and implementation of both filters.

12.3.3.1 The Hodrick-Prescott (HP) Filter

The HP filter decomposes a time series y_t into a trend component, y_t^{trend}, which varies smoothly over time, and a cyclical component, y_t^c, by fitting a smooth curve through the series. To enhance the comparability across series, y_t is typically analyzed in logarithms, so that the units of y_t^c, when multiplied by 100, represent percentage deviations from the series' growth path (Stock and Watson 1999). To identify both components, we need to minimize the variance of the cyclical component subject to a penalty for variation in the second difference of the trend component. The cyclical component, which fluctuates around that trend, is then obtained by subtracting the long-term trend from y_t, that is, $y_t^c = y_t - y_t^{trend}$. As such, the low frequencies are eliminated from the original series. More formally, the HP filter obtains y_t^{trend} by minimizing:

$$\sum_{t=1}^{T} (y_t - y_t^{trend})^2 + \lambda \sum_{t=2}^{T-1} ((y_{t+1}^{trend} - y_t^{trend}) - (y_t^{trend} - y_{t-1}^{trend}))^2 \tag{12.1}$$

where λ is a penalty parameter that determines the degree of smoothing; the larger its value, the smoother the resulting growth component. As business cycles exhibit cycles of varying length that tend to last no longer than 8 years in duration (Christiano and Fitzgerald 1998), the smoothing constant is chosen to generate a trend accounting for all fluctuations longer than 8 years. Baxter and King (1999) recommend a value of λ equal to 10 for annual series, and λ equal to 1,600 for quarterly series. These values produce a good correspondence between the HP filter and an ideal BP filter that passes through cycles between 1.5 and 8 years. To evaluate the robustness of the findings, we recommend researchers to apply different business-cycle filters and/or adjust the smoothing values to assess if the results can be reproduced. Lamey et al. (2007), for instance, relied on the HP filter with λ = 10 for yearly data, but elaborately evaluated the sensitivity of the main findings by adopting the alternative BP filter. Also, all results could be replicated when changing the recommended smoothing parameter 6 times from a values as low as 4 up to a value of 30 for λ (see Lamey et al. 2007, Table 4 for details).

12.3.3.2 Band-Pass (BP) Filters

Deleersnyder et al. (2004), and Lamey et al. (2007), among others, have applied the BP filter developed by Baxter and King (1999) as an alternative way to isolate the business-cycle fluctuations in a time series. Even though this filter originated in the spectral domain, it can be undertaken entirely in the time domain. We refer to the original study of Baxter and King (1999) for a detailed discussion on both the design of the filter in the frequency domain, and its translation back into the time domain in the form of an easy-to-use symmetric (in terms of leads and lags) moving average. In this section, we limit our discussion to the intuition underlying this filter, and its implementation in marketing studies.

An 'ideal' or optimal band-pass filter would isolate only those components in the series that lie within the specified (business-cycle) periodicity range, while it eliminates both very slow moving components (e.g., the underlying trend) and high frequency components (e.g., irregular and/or seasonal fluctuations). Such an ideal filter, however, would require an infinite-order moving average in the time domain, so that in practice an approximation is needed. The approximation proposed by Baxter and King (1999) is based on a symmetric 3-year centered moving-average transformation, where the weights are chosen to approximate as close as possible the optimal filter. For annual data, this approximate filter can be shown to equal:

$$y_t^c = 0.7741 y_t - 0.2010(y_{t-1} + y_{t+1}) - 0.1351(y_{t-2} + y_{t+2}) - 0.0510(y_{t-3} + y_{t+3}), \tag{12.2}$$

where y_t is the logarithm of the original time series in year t, and y_t^c the cyclical component to be used in further analyses. For quarterly data, an equivalent symmetric moving-average filter based on 13 quarterly filter weights can be constructed. These filter weights are provided in the second and third column of Table 4 in the study by Baxter and King (1999, p. 591). This results in a flexible and easy to implement filter that can be adapted easily to the series' aggregation level, and to the desired periodicities a researcher wants to capture.

Because of the leads and lags in Eq. 12.2, 6 years of data are lost in the derivation of the cyclical component (i.e., 6 observations due to 3 leads and lags in the yearly filter, or 24 observations due to 12 leads and lags in a quarterly filter). To avoid this, van Heerde et al. (2013) worked with the Christiano-Fitzgerald (CF) random walk filter (Christiano and Fitzgerald 2003). The CF filter is a band-pass filter built on the same principles as the BP filter by Baxter and King (1999). However, the CF filter is designed to use the entire time series for the calculation of each filtered data point and can be expressed as:

$$\begin{aligned} y_t^c = {} & B_0 y_t + B_1 y_{t+1} + \cdots + B_{T-1-t} y_{T-1} + \tilde{B}_{T-t} y_T \\ & + B_1 y_{t-1} + \cdots + B_{t-2} y_2 + \tilde{B}_{t-1} y_1, \end{aligned} \tag{12.3}$$

for t = 3, 4, …, T − 2, with:

$$\begin{aligned} & B_j = \tfrac{\sin(jb) - \sin(ja)}{\pi j}, j \geq 1 \\ & B_0 = \tfrac{b-a}{\pi}, a = \tfrac{2\pi}{p_u}, b = \tfrac{2\pi}{p_l}, \\ & \tilde{B}_k = -\tfrac{1}{2} B_0 - \sum_{j=1}^{k-1} B_j \end{aligned} . \tag{12.4}$$

The parameters p_l and p_u are the cut-off cycle length. If the data are monthly, recommended values for p_l = 18 and p_u = 96 are used to pass on fluctuations in y_t

with business-cycle periodicities between 1.5 and 8 years. Cycles longer than p_l and shorter than p_u are preserved in the cyclical term y_t^c. We refer to Christiano and Fitzgerald (2003) or Nilsson and Gyomai (2011) for a detailed discussion on the CF filter and a comparison with some other frequently-used filters.

Even though the HP and BP filters are both popular and have been used frequently in business-cycle research in marketing, when working with time series at a more disaggregate level than the yearly level (e.g., quarterly or monthly data), a BP filter is more appropriate and should be preferred over the HP filter. Researchers can design BP filters to pass through all components of the time series with a business cycle periodicity between 6 and 32 quarters, while both lower and higher frequency variation will be removed from the series. The HP filter, in contrast, suppresses only the low-frequency fluctuations, implying that what is passed through consists of both the cyclical fluctuations (of main interest) and higher-frequency noise. This is especially relevant when working with more disaggregate data.

To illustrate how business-cycle filters work, we implemented the HP filter on yearly postwar U.S. real GDP (source: http://www.bea.gov/national/), and show in Fig. 12.1 panel A the original GDP in billions of (chained 2009) dollars from 1950 to 2015 in blue, the green line in panel B represents the long-run trend extracted from the series, while the difference between the original series (blue) and the HP trend component (green) represents the cyclical fluctuations in U.S. real GDP in panel C (red line). The extracted cyclical component after filtering reflects percentage deviations from the long-term growth rate, and fluctuates around 0. The grey shaded columns in Fig. 12.1 panel C correspond to the official contraction periods as identified by the U.S. Business Cycle Dating Committee of the NBER (see Appendix A for the corresponding dates). During the postwar period, so far, 10 official economic contractions occurred.

As is evident in Fig. 12.1 panel C, the cyclical fluctuations obtained after filtering coincide to a large extent with the official NBER recession years. Note, however, two important differences between these NBER contractions and the information contained in the filtered business cycle series: First, the NBER does not identify the strength of a recession (or alternatively, expansion) and as a result, it will pick up especially the more severe economic contractions, while mild economic slowdowns remain unnoticed (as in e.g., 1967 or 1986). Second, the NBER contraction peaks slightly lag the peaks identified by the business-cycle filter.

After the cyclical component from the economic activity (E) has been extracted using one of the business-cycle filters, marketing researchers have used it in three ways: (i) they directly work with the cyclical series (e.g., Deleersnyder et al. 2004, 2009; or Lamey et al. 2012, Eqs. 4–5), (ii) they redefine it into a classification variable (e.g., Lamey et al. 2012, Eq. 6), or (iii) they converted it into the drop (rise) relative to the previous peak (trough) (e.g., Lamey et al. 2007; Steenkamp and Fang 2011; or van Heerde et al. 2013).

12.4 Methods and Metrics to Link the Inferred Business Cycle to Marketing Variables

In marketing, we aim to evaluate the impact of the business cycle on marketing conduct (M), marketing effectiveness (β) and/or performance (P) series. There are different approaches to link the business cycle to these variables, and the choice of method and/or metric may differ across the three methods used to infer the business cycle (see Sect. 12.3).

12.4.1 Methods Based on a Discrete Recession Classification

When the business cycle is discretized and periods of economic recession/expansion are identified, marketing researchers often create a 'recession dummy' variable to summarize the cyclical fluctuations in E, and use that dummy variable in their subsequent time series analysis. Such a recession dummy has been related to the marketing-conduct (M) (see, e.g., Kashmiri and Mahajan 2014) and/or firm-performance (P) variables (e.g., Srinivasan et al. 2011) as an independent variable. This way, the recession dummy can capture the main effect of a recession on the respective dependent variable. Moreover, the recession dummy has also been included in an interaction effect with the marketing conduct variables to test for differences in marketing effectiveness across expansion and recession phases (Srinivasan et al. 2011; or Kashmiri and Mahajan 2014) as in:

$$P_t = \beta_0 + \beta_1 M_t + \beta_2 Dcontr_t + \beta_3 [M_t \times Dcontr_t], \tag{12.5}$$

where Dcontr = 1 during a recession/contraction year (0 otherwise), and all other factors as defined before. In Eq. 12.5, β_1 captures the effect of M on P during an expansion period, while (β_1 + β_3) represents the effect of M on Y during a contraction period. Hence, β_3 captures the differential effect in marketing effectiveness during a contraction phase.

12.4.2 Methods Based on Business-Cycle Filtered Data

A business-cycle filter when applied to an economic indicator E, or directly to a performance (P) or marketing conduct (M) series will produce a new cyclical time series for further analysis. The resulting cyclical time series, y_t^c, forms the basis to determine the *extent* and *nature* of the series' sensitivity to business cycles. Both univariate and multivariate approaches have been used to summarize different business-cycle properties in a variety of disaggregate marketing series.

12.4.2.1 Univariate Approaches to Capture Business-Cycle Phenomena

Univariate business-cycle statistics focus on describing the nature of the cyclical ups and downs in a given time series: (i) the extent (size) of the cyclical fluctuations, and (ii) potential asymmetries between expansion and contraction periods.

Cyclical Volatility

Cyclical variability or 'volatility' measures the extent of cyclical swings in a series, and is quantified through the standard deviation, $\sigma(y_t^c)$, of the filtered cyclical component of a series. These standard deviations are comparable across series when y_t^c (the business-cycle component) has been determined on the log-transformed original series y_t. When multiplied by 100, they represent percentage deviations from the series' growth path (Stock and Watson 1999, p. 29).

Cyclical-volatility values are typically used to compare the cyclical sensitivity across different series, either in absolute terms, or relative to the cyclical volatility in the aggregate economy E (captured in the filtered GDP series) over the same period. Deleersnyder et al. (2004), for example, evaluated the cyclical sensitivity in performance (P) for a wide range of consumer durable products, and found durable sales to be, on average, more than 4 times more cyclically volatile than the U.S. economy as a whole.

Cyclical Asymmetries

Fluctuations associated with business cycles, in performance (P) and/or in the marketing-mix (M) series, may not only be described in terms of the extent to which they go up or down. We can also assess whether they evolve asymmetrically, in that their behavior in a contraction differs from their (opposite) behavior in an economic expansion.

Sichel (1993) distinguishes in this respect between two types of cyclical asymmetry: *deepness* asymmetry and *steepness* asymmetry. These are graphically illustrated in Fig. 12.2, and pertain to differences in the extent (=deepness asymmetry in panel A) or speed (=steepness symmetry in panel B) of adjustments in the series across contraction and expansion phases.

Negative deepness asymmetry is defined as the characteristic that troughs are deeper, i.e. further below mean or trend, than peaks are tall (and the opposite in case of positive deepness asymmetry). In the case of negative deepness asymmetry, we expect that the negative deviations from the mean or trend during contraction periods are larger, in absolute value, than the positive deviations during expansion periods. Steepness asymmetry, in turn, pertains to the speed or rate of change with which a series P or M (or the economy as a whole, E) falls into a contraction as

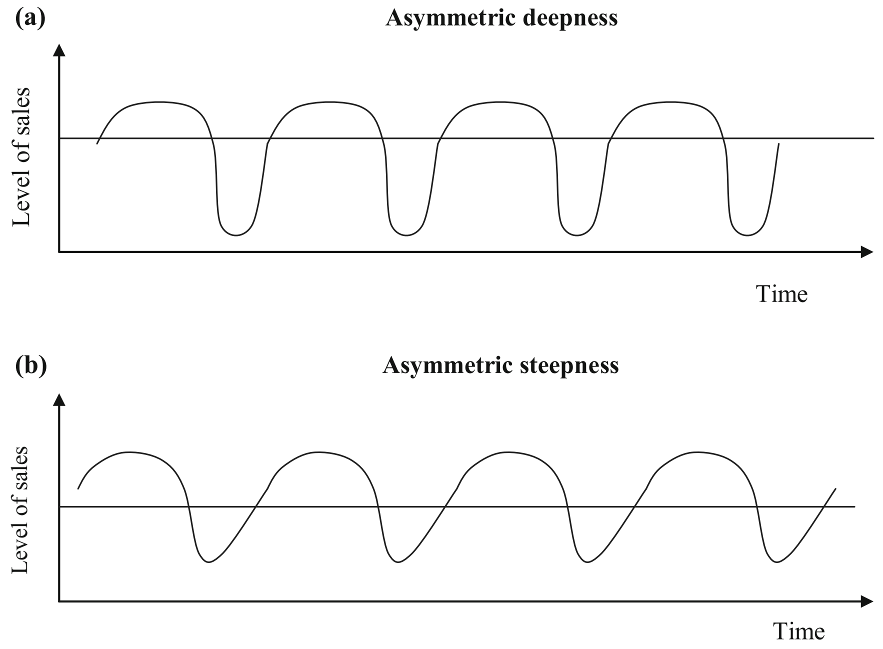


Fig. 12.2 Steepness and deepness asymmetry

compared to its speed of recovery. Following the pioneering work of Sichel (1993), cyclical (a)symmetries are tested through the third-order moment, i.e. the skewness statistic, of the filtered series y_t^c. To construct a formal test for deepness asymmetry, the following coefficient of skewness is computed:

$$D(y_t^c) = \frac{\left[T^{-1} \sum_{t=1}^{T} (y_t^c - \overline{y^c})^3\right]}{\sigma(y^c)^3}, \tag{12.6}$$

where $\overline{y^c}$ is the mean of the cyclical component y_t^c (after filtering, this mean should be close to zero), $\sigma(y^c)$ its standard deviation, and T the sample size (Sichel 1993). If a time series exhibits steepness asymmetry, its first difference, representing the slope or rate of change, should exhibit skewness. We refer to Fig. 12.2 panel B for a graphical illustration of such behavior. The formal test statistic for steepness asymmetry is based on the coefficient of skewness for Δy_t^c, the first difference of the cyclical component in the series of interest:

$$ST(\Delta y_t^c) = \frac{\left[T^{-1} \sum_{t=1}^{T} (\Delta y_t^c - \overline{\Delta y^c})^3\right]}{\sigma(\Delta y^c)^3}, \tag{12.7}$$

where $\overline{\Delta y^c}$ and $\sigma(\Delta y^c)$ are, respectively, the mean and standard deviation of Δy_t^c (Sichel 1993). To determine the significance of both test statistics, asymptotic standard errors are derived as follows. For deepness asymmetry, we regress $z_t = (y_t^c - \overline{y^c})^3 / \sigma(y^c)^3$ on a constant, the significance of which corresponds to the significance of $D(y_t^c)$. Indeed, the coefficient estimate associated with the constant equals the deepness statistic, and the corresponding standard error measures its statistical reliability. Since the observations on y_t^c are serially correlated, the correction suggested by Newey and West (1987) is implemented in the derivation of the standard errors. Asymptotic, Newey-West corrected, standard errors for the steepness statistic $ST(y_t^c)$ can be calculated using a similar procedure, but with $z_t = (\Delta y_t^c - \overline{\Delta y^c})^3 / \sigma(\Delta y^c)^3$.

While frequently used and intuitively appealing, the parametric skewness test proposed by Sichel (1993) has been criticized for having low power to reject the null hypothesis of symmetry, and for being sensitive to outliers (Verbrugge 1997; Razzak 2001). A non-parametric triples test, first developed by Randles et al. (1980), and introduced in the economics literature by Verbrugge (1997), has been suggested as an alternative, more powerful test to derive cyclical asymmetry (Verbrugge 1997; Razzak 2001).

In the marketing literature, Deleersnyder et al. (2004) and Lamey et al. (2007) found evidence of deepness and steepness asymmetry in, respectively, the cyclical fluctuations in the sales of multiple U.S. durable categories, and in the cyclical fluctuations of private-label share (which they studied for Belgium, the U.K., the U.S., and West Germany). Dekimpe et al. (2016), in turn, found no evidence for an asymmetric speed of adjustment or steepness asymmetry in the international-tourism industry from and to New Zealand. Thus far, these asymmetries have been assessed in various performance series P, but (to the best of our knowledge) not yet in marketing conduct (M). We refer to Sect. 12.4.1 for details how asymmetries in β have been studied in marketing.

12.4.2.2 Multivariate Approaches to Capture Business-Cycle Phenomena

Univariate business-cycle statistics focus on understanding the extent of the cyclical ups and downs *within* a given series P or M, but are not concerned with explicitly linking these cyclical patterns with fluctuations in the aggregate economy E (Christiano and Fitzgerald 1998). This is captured in a number of multivariate approaches in which temporal fluctuations in a variable of interest y (which can represent a P or M variable) are explicitly related to (cyclical) fluctuations in the general state of the economy E.

Two conceptually different types of multivariate approaches have been advanced and used in marketing research. *Cyclical comovement models* capture the temporary effect, and measure how cyclical fluctuations in the economy translate into cyclical fluctuations in a marketing series, with the up- and downward movements in the latter always returning to the underlying mean or trend (hence the measurement of a temporary effect). Thus, if consumers change their behavior during an economic contraction, but restore their original behavior in the subsequent expansion, this can be captured by means of a cyclical comovement model (discussed in section "The Temporary Effect: Cyclical Comovement Models"). However, what if the contraction behavior is not restored, and consumers fully or partially stick to the new contraction-induced behavior due to, for instance, consumer learning or inertia? If researchers suspect that an economic contraction has a long-lasting or permanent effect, in which cyclical shocks affect the marketing or performance series' underlying growth pattern, *business-cycle growth models* are called for. These models are introduced and discussed in section "The Permanent Effect: Business-Cycle Growth Models".

The Temporary Effect: Cyclical Comovement Models

To quantify the extent to which cyclical ups and downs in firms' marketing (M) or performance (P) series move together with the economy as a whole (E), a cyclical comovement elasticity can be derived. This is done by regressing the cyclical component extracted from the time series of interest, y_t^c, on the cyclical component filtered from that market's GDP (i.e., gdp_t^c) (or an equivalent aggregate economic indicator) over the corresponding period:

$$y_t^c = \gamma \, gdp_t^c + \varepsilon_t, \tag{12.8}$$

As both cyclical components are expressed in percentage deviations, the resulting parameter γ is an elasticity estimate.[2] The sign and significance of γ indicates the direction of the impact. A series evolves *pro-cyclically* ($\gamma > 0$) when changes occur in the same direction as the aggregate economy, *counter-cyclically* ($\gamma < 0$) when movements are in the opposite direction as the economy, or *a-cyclically* ($\gamma = 0$) when the cyclical fluctuations in the series are unrelated to the general economic activity. The magnitude of the comovement estimate, on the other hand, reflects the extent to which fluctuations in the general economy get *attenuated* ($|\gamma| < 1$) or *amplified* ($|\gamma| > 1$) in the marketing or performance series.

This elasticity only looks at the instantaneous or coincident effect, while potential dynamic influences are ignored. But marketing investments and/or firm performance may, however, respond to cyclical turns in advance and/or with some delay. To account for such dynamics, researchers can derive a dynamic comovement elasticity by adding lead and/or lag terms to Eq. 12.8:

$$y_t^c = \sum_{k=-K}^{L} \gamma_k gdp_{t-k}^c + \varepsilon_t, \tag{12.9}$$

[2]It is not necessary to include an intercept in Eq. 12.8, as both series are zero-reverting after filtering.

with 'K' ('L') the number of lead (lagged) dynamics accounted for. Based on the estimates from Eq. 12.9, a dynamic comovement elasticity can be derived as $\gamma^{K+L+1} = \sum_{-K}^{L} \gamma_k$.

A substantial number of marketing studies have derived a cyclical comovement elasticity, making it a fairly standard approach to describe business-cycle sensitivity. Indeed, the representation of the cyclical fluctuations into a unit-free comovement elasticity facilitates the interpretation and comparison across brands, firms, industries, marketing tools as well as countries. Example studies that reported such cyclical comovement statistics include among others Cleeren et al. (2016), Dekimpe et al. (2016), Deleersnyder et al. (2004, 2009) and Lamey et al. (2007, 2012). In Deleersnyder et al. (2009), for instance, the average comovement elasticity in advertising spending across a broad set of 37 countries in 4 different (traditional) media was 1.4, implying that when the national economic activity winds down with 1%, a reduction of cyclical advertising spending can be expected of 1.4%.

Although both the univariate cyclical volatility and the multivariate comovement elasticity describe the extent of business-cycle sensitivity in a series, they approach it from a distinct, yet complementary, perspective. Some studies report both (Deleersnyder et al. 2004; Dekimpe et al. 2016), and show how part of the information in the cyclical volatility measure is also reflected in the co-movement elasticity. However, the cyclical volatility is always positive, and therefore cannot distinguish between pro- and counter-cyclical behavior. But as shown by Lamey (2014) or Lamey et al. (2007, 2012), some marketing series such as discounter share or private-label performance can move in the opposite direction as the state of the economy. Accordingly, the cyclical co-movement offers a more informative and easy-to-understand measure to describe cyclical patterns in a series, and is therefore often preferred by marketing researchers.

The Permanent Effect: Business-Cycle Growth Models

Business-cycle statistics based on filtered time series describe the behavior of the cyclical fluctuations around the series' overall trend or mean level, after the underlying trend or mean has been removed. As such, statistics based on filtered series cannot address the question whether this cyclical behavior also contributes to long-run changes in the series (e.g., by altering the series' growth rate), or whether they just represent temporary deviations from the underlying trend or mean that eventually cancel out (Beaudry and Koop 1993). For example, if all customers who are found to switch to private labels during the contraction would eventually return to buying national brands once the economy recovers, no long-run growth would be observed that could be attributed to the earlier contraction. As such, deepness and steepness asymmetries can exist "*with or without there being asymmetry in persistence*" (Beaudry and Koop 1993, p. 151). Inspired by the idea that business-cycle changes may not always fully rebound, business-cycle studies in marketing have moved beyond focusing on only the filtered or extracted business-cycle information in the series, and started to link these cyclical fluctuations to the series' underlying long-term growth rate. Lamey et al. (2007, 2012), for instance, show that economic expansions and contractions affect private-label share evolution to a different degree, and thus, alter the series' underlying growth pattern. Contractions, for example, cause a substantial positive impact on private-label growth that is not fully offset in the subsequent expansion, because some consumers who switch to private labels in the contraction keep buying them even when bad economic times are long over. This leaves permanent 'scars' on national brands' performance.

Model to capture the growth implications of alternative business-cycle phases. To formally assess whether cyclical shocks affect a series' underlying growth, and to check whether this effect differs across contraction and expansion periods, an asymmetric growth model can be specified with a performance (P) or marketing conduct (M) series as the dependent variable. In line with Beaudry and Koop (1993) and Thoma (1994), two new variables can be defined reflecting the general state of the economy (E) at a certain point in time t:

$$\begin{array}{ll} exp_t & \begin{cases} = gdp_t^c - (prior\,trough\,in\,gdp^c) & if\,\Delta gdp_t^c > 0 \\ = 0 & if\,\Delta gdp_t^c \leq 0 \end{cases} \\ contr_t & \begin{cases} = 0 & if\,\Delta gdp_t^c > 0 \\ = (prior\,peak\,in\,gdp^c) - gdp_t^c & if\,\Delta gdp_t^c \leq 0. \end{cases} \end{array} \tag{12.10}$$

Decreases (increases) in the cyclical component of GDP correspond to contractions (expansions). The variable exp_t measures the magnitude of the expansion by calculating how much the business cycle, reflected in the filtered GDP series, has increased relative to its previous trough. Similarly, the variable $contr_t$ measures the magnitude of a contraction by calculating how much the business cycle has dropped compared to its previous peak when the economy is winding down. Through this operationalization, all values of exp_t and $contr_t$ will be nonnegative, making the interpretation of their corresponding coefficients in Eq. 12.11 (see next) more straightforward.

The series underlying growth rate, Δy_t, is subsequently linked to current and lagged values of the exp_t and $contr_t$ variables. By assessing whether they have additional explanatory power over lagged growth terms $\Delta y_{t-j}(\mathrm{j} = 1 \ldots \mathrm{J})$, we test whether the business cycle Granger causes the series' underlying growth:

$$\Delta y_t = \alpha + \sum_{j=1}^{J} \omega_j \Delta y_{t-j} + \sum_{k=0}^{K} \varphi_k^- \,\mathrm{contr}_{\mathrm{t-k}} + \sum_{l=0}^{L} \varphi_l^+ \,\mathrm{exp}_{\mathrm{t-1}} + \varepsilon_t. \tag{12.11}$$

The lag lengths J, K and L in Eq. 12.11 can be determined on the basis of information criteria (Judge et al. 1988). Note that this asymmetric growth model is specified in differences. Preliminary unit-root tests can be used to empirically determine whether the series is indeed non-stationary (as implicitly assumed in Eq. 12.11).

By splitting the business cycle in two phases, we allow contractions and expansions to affect the series' underlying growth rate differently. Thus, the effect of a contraction is not necessarily cancelled out in a subsequent expansion. Unlike the cyclical comovement elasticity derived in Eqs. 12.8 and 12.9, we no longer assess the linkage between the cyclical movements in, respectively, the series of interest and the economy as a whole. Instead, we now test whether changes in the latter contribute to the focal series' growth (and hence, their long-run level). The sum of the parameters φ_k^- (φ_l^+) associated with contraction (expansion) periods, i.e. $\sum_{k=0}^{K} \varphi_k^- = \varphi_{ST}^- \left(\sum_{l=0}^{L} \varphi_l^+ = \varphi_{ST}^+ \right)$, gives the combined short-run impact on the series underlying growth. As Eq. 12.11 contains lags of the dependent variable, it can be shown that the impact on the series' long-run or steady-state growth rate becomes, respectively, $\frac{\sum_{k=0}^{K} \varphi_k^-}{1 - \sum_{j=1}^{J} \varpi_j} (= \varphi_{LT}^-)$ and $\frac{\sum_{l=0}^{L} \varphi_l^+}{1 - \sum_{j=1}^{J} \varpi_j} (= \varphi_{LT}^+)$ (Franses 2005). Standard errors of these ratios can be derived using the well-known delta-method. Under the assumption that a contraction period inhibits (stimulates) the series' growth, we expect its impact to be negative (positive), thus, $\sum_{k=0}^{K} \varphi_k^- < 0 (> 0)$. When the impact of an expansion on the same series' growth rate turns out to have an impact in the opposite direction (derived as $\sum_{l=0}^{L} \varphi_l^+$), the size of the expansions will determine to what extent the growth-inhibiting (or stimulating) effect of earlier contractions will be off-set.

Some marketing studies have used a simplified version of Eq. 12.11 and directly link the economic contractions to the series long-run growth rate acquired after business-cycle filtering, Δy_t^{trend} (see, e.g., Lamey et al. 2012; or Lamey 2014). Here, an asymmetric growth model is specified where only a recession dummy (instead of the size of the expansion and size of the contraction) influence the series underlying (filtered) trend (instead of the yearly growth rate):

$$\Delta y_t^{trend} = \delta + \phi \, Dcont_t + \mu_t \tag{12.12}$$

In Eq. 12.12, a contraction dummy, $Dcont_t$, is set to one when the economy is contracting, and zero when the economy is expanding. Existing research by Lamey et al. (2012) and Lamey (2014) relied on changes in filtered GDP, with a contraction (expansion) determined as $\Delta gdp_t^c \leq 0 (\Delta gdp_t^c > 0)$. Alternatively, also other recession classification techniques discussed in Sect. 12.3 could be used to construct the contraction dummy in Eq. 12.12. With this dummy coding, the parameter δ reflects the average annual long-term growth in the series when the economy is booming, whereas (δ + ϕ) measures the average long-term growth in the series when the economy is contracting. The parameter ϕ, in turn, quantifies the *incremental* long-term growth in the series in a contraction year that is not cancelled out in future expansion periods.

Cyclically-induced growth model. Apart from linking the aggregate economic contraction variables to the underlying long-term growth in the series (P or M) over time, also the cyclical information in the series itself has been linked to its (average) underlying growth for a cross-sectional sample of time series on P or M. In particular, to capture the long-run consequences of cyclical fluctuations in the series, one (or more) measure(s) of cyclical sensitivity, y_i^{CYC}, have been linked to the average long-term growth rate in the respective series, $\overline{\Delta y_i^{trend}}$, in a cross sectional regression, while controlling for other factors driving long-term growth in y_t such as, for instance, population growth (captured by matrix $X_{i,j}$):

$$\overline{\Delta y_i^{trend}} = \eta_1 + \eta_2 y_i^{CYC} + \eta_j X_{i,j} + \varepsilon_i, \tag{12.13}$$

with $\overline{\Delta y_i^{trend}}$ determined as the average (over time) of the first difference of the trend component in y_i. This average is calculated by regressing the first difference of the trend component of each underlying time series (derived through a business-cycle filter) on a constant. The cyclical properties in the series captured by y_i^{CYC} can relate to the univariate cyclical volatility or asymmetry measures derived in Sect. 12.4.2.1, the cyclical comovement elasticity derived in section "The Temporary Effect: Cyclical Comovement Models", or a combination of these measures (see, e.g., Dekimpe et al. 2016). To address the estimated nature of the dependent variable and to obtain efficient estimates, weighted least square (WLS) is used with the inverse of the dependent variable's standard error as weight. Likewise, (some of) the independent variables are often estimated quantities as well, such as the cyclical comovement elasticity derived from Eq. 12.8 or 12.9. To obtain unbiased estimates for the parameters' standard errors, a bootstrap algorithm is typically used (we refer to the Technical Appendix in the study by Lamey et al. 2012 for details on this). Marketing studies linking the growth in the series of interest to the same series' cyclical characteristics include, among others, Cleeren et al. (2016) and Dekimpe et al. (2016) in the context of performance series (P), and Deleersnyder et al. (2009) for advertising spending (M).

12.5 Conclusion

Recently, an increasing number of marketing studies have focused on business-cycle phenomena, in which a diverse set research techniques has been used. This chapter reviewed the techniques most commonly used in marketing research to study business cycles. In so doing, we distinguished between different methods to assess the general state of the economy (E), and methods or metrics to describe/summarize various business-cycle patterns in marketing conduct (M) and/or performance (P) variables. Both discrete and continuous metrics to quantify the state of the economy were discussed, while we also reviewed measures to describe the nature of the cyclicality in a time series. In this respect, we distinguished between univariate and multivariate approaches, and between techniques aimed at quantifying, respectively, temporary or long-run cyclical implications.

Given that the impact of the business cycle has been shown to differ between industries and countries, more research is needed to fine tune the emerging empirical generalizations, and to develop a solid contingency framework. Also, more research is needed to allow for different cyclical effects (both in conduct and effectiveness) between different forms of a given marketing-mix instrument (for example, between online and offline advertising). Finally, more detailed normative recommendations are needed. Some studies base their recommendations on the cyclical sensitivity of performance P (as in Deleersnyder et al. 2009 and Lamey et al. 2007), while others focus on the cyclical sensitivity in marketing effectiveness β (as in Steenkamp and Fang 2011 or van Heerde et al. 2013). Also, little is known to what extent deviations from an “optimal” pro- or counter-cyclical adjustment matter, or whether the flat-maximum principle (suggesting little profit implications; see Tull et al. 1986) applies also in this context.

We hope that this review will add to a better understanding of the advantages and limitations of the various techniques that can be used to tackle these research question, and therefore offer an impetus to future research on the marketing implications of macro-economic fluctuations. Indeed, as the recent Global Financial Crisis brought to the fore, these implications can be very profound, requiring a better managerial understanding.

Appendix A

See Table 12.1.

Table 12.1 NBER U.S. business cycle expansions and contractions*

Business cycle reference dates		Duration in months			
Peak	Trough	Contraction	Expansion	Cycle	
Quarterly dates are in parentheses		Peak to trough	Previous trough to this peak	Trough from previous trough	Peak from previous peak
	December 1854 (IV)	_	_	_	_
June 1857 (II)	December 1858 (IV)	18	30	48	–
October 1860 (III)	June 1861 (III)	8	22	30	40
April 1865 (I)	December 1867 (I)	32	46	78	54
June 1869 (II)	December 1870 (IV)	18	18	36	50
October 1873 (III)	March 1879 (I)	65	34	99	52
March 1882 (I)	May 1885 (II)	38	36	74	101
March 1887 (II)	April 1888 (I)	13	22	35	60
July 1890 (III)	May 1891 (II)	10	27	37	40
January 1893 (I)	June 1894 (II)	17	20	37	30
December 1895 (IV)	June 1897 (II)	18	18	36	35
June 1899 (III)	December 1900 (IV)	18	24	42	42
September 1902 (IV)	August 1904 (III)	23	21	44	39
May 1907 (II)	June 1908 (II)	13	33	46	56
January 1910 (I)	January 1912 (IV)	24	19	43	32
January 1913 (I)	December 1914 (IV)	23	12	35	36
August 1918 (III)	March 1919 (I)	7	44	51	67
January 1920 (I)	July 1921 (III)	18	10	28	17
May 1923 (II)	July 1924 (III)	14	22	36	40
October 1926 (III)	November 1927 (IV)	13	27	40	41
August 1929 (III)	March 1933 (I)	43	21	64	34
May 1937 (II)	June 1938 (II)	13	50	63	93
February 1945 (I)	October 1945 (IV)	8	80	88	93
November 1948 (IV)	October 1949 (IV)	11	37	48	45
July 1953 (II)	May 1954 (II)	10	45	55	56
August 1957 (III)	April 1958 (II)	8	39	47	49
April 1960 (II)	February 1961 (I)	10	24	34	32
December 1969 (IV)	November 1970 (IV)	11	106	117	116
November 1973 (IV)	March 1975 (I)	16	36	52	47
January 1980 (I)	July 1980 (III)	6	58	64	74
July 1981 (III)	November 1982 (IV)	16	12	28	18
July 1990 (III)	March 1991(I)	8	92	100	108
March 2001 (I)	November 2001 (IV)	8	120	128	128
December 2007 (IV)	June 2009 (II)	18	73	91	81

(continued)

Table 12.1 (continued)

Business cycle reference dates		Duration in months			
Peak	Trough	Contraction	Expansion	Cycle	
Quarterly dates are in parentheses		Peak to trough	Previous trough to this peak	Trough from previous trough	Peak from previous peak
Average, all cycles:					
1854–2009 (33 cycles)		17.5	38.7	56.2	56.4
1854–1919 (16 cycles)		21.6	26.6	48.2	48.9
1919–1945 (6 cycles)		18.2	35.0	53.2	53.0
1945–2009 (11 cycles)		11.1	58.4	69.5	68.5

Source http://www.nber.org/cycles/cyclesmain.html

References

Baxter, Marianne, and Robert G. King. 1999. Measuring business cycles: Approximate band-pass filters for economic time series. *The Review of Economics and Statistics* 81 (4): 575–593.

Baxter, Marianne, and Michael A. Kouparitsas. 2005. Determinants of business cycle comovement: A robust analysis. *Journal of Monetary Economics* 52 (1): 113–157.

Beaudry, Paul, and Gary Koop. 1993. Do recessions permanently change output? *Journal of Monetary Economics* 31 (2): 149–163.

Burns, Arthur F., and Wesley C. Mitchell. 1946. *Measuring business cycles*. New York, NY: National Bureau of Economic Research.

Canova, Fabio. 1998. Detrending and business cycle facts. *Journal of Monetary Economics* 41 (3): 475–512.

Cha, William, Pradeep Chinatagunta, and Sanjay Dhar. 2015. Food Purchases During the Great Recession, *Working paper*, Booth School of Business, the University of Chicago.

Christiano, Lawrence J., and Terry J. Fitzgerald. 1998. The business cycle: It's still a puzzle. *Federal Reserve Bank of Chicago Economic Perspectives* 22 (4): 56–83.

Christiano, Lawrence J., and Terry J. Fitzgerald. 2003. The band pass filter. *International Economic Review* 44 (2): 435–465.

Christiano, Lawrence J., Mathias Trabandt, and Karl Walentin. 2011. DSGE models for monetary policy analysis. In *Handbook of monetary economics*, vol. 3A, ed. Benjamin M. Friedman, and Michael Woodford, 285–367. The Netherlands: North-Holland.

Cleeren, Kathleen, Lien Lamey, Jan-Hinrich Meyer, and Ko De Ruyter. 2016. How business cycles affect the healthcare sector: A cross-country investigation. *Health Economics* 28: 14–14.

Dekimpe, Marnik G., Yuri Peers, and Harald J. van Heerde. 2016. The impact of the business cycle on service providers: Insights from international tourism. *Journal of Service Research* 19 (1): 22–38.

Deleersnyder, Barbara, Marnik G. Dekimpe, Miklos Sarvary, and Philip M. Parker. 2004. Weathering tight economic times: The sales evolution of consumer durables over the business cycle. *Quantitative Marketing and Economics* 2 (4): 347–383.

Deleersnyder, Barbara, Marnik G. Dekimpe, Miklos Sarvary, Philip M. Parker, Jan-Benedict E.M. Steenkamp, and Peter S.H. Leeflang. 2009. The role of national culture in advertising's sensitivity to business cycles: An investigation across continents. *Journal of Marketing Research* 46 (5): 623–636.

Dutt, Pushan, and V. Padmanabhan. 2011. Crisis and consumption smoothing. *Marketing Science* 30 (3): 491–512.

Frankenberger, Kristina D., and Roger C. Graham. 2003. Should Firms Increase Advertising Expenditures During Recessions? MSI Report No. 03-115.

Franses, Philip Hans. 2005. On the use of econometric models for policy simulation in marketing. *Journal of Marketing Research* 42 (1): 4–14.

Gordon, Brett R., Avi Goldfarb, and Yang Li. 2013. Does price elasticity vary with economic growth? A cross-category analysis. *Journal of Marketing Research* 50 (1): 4–23.

Graham, Roger C., and Kristina D. Frankenberger. 2011. The earnings effects of marketing communication expenditures during recessions. *Journal of Advertising* 40 (2): 5–24.

Hodrick, Robert J., and Edward C. Prescott. 1997. Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking* 29 (1): 1–16.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lütkepohl, and Tsoung-Chao Lee. 1988. *Introduction to the theory and practice of econometrics*. New York: Wiley.

Kamakura, Wagner A., and Du RexY. 2012. How economic contractions and expansions affect expenditure patterns. *Journal of Consumer Research* 39 (2): 229–247.

Kashmiri, Saim, and Vijay Mahajan. 2014. Beating the recession blues: Exploring the link between family ownership, strategic marketing behavior and firm performance during recessions. *International Journal of Research in Marketing* 31 (1): 78–93.

Kumar, V., Nita Umashankar, Kihyun H. Kim, and Yashoda Bhagwat. 2014. Assessing the influence of economic and customer experience factors on service purchase behaviors. *Marketing Science* 33 (5): 673–692.

Lamey, Lien. 2014. Hard economic times: A dream for discounters. *European Journal of Marketing* 48 (3/4): 641–656.

Lamey, Lien, Barbara Deleersnyder, Marnik G. Dekimpe, and Jan-Benedict E.M. Steenkamp. 2007. How business cycles contribute to private-label success: Evidence from the United States and Europe. *Journal of Marketing* 71 (1): 1–15.

Lamey, Lien, Barbara Deleersnyder, Jan-Benedict E.M. Steenkamp, and Marnik G. Dekimpe. 2012. The effect of business cycle fluctuations on private-label share: What has marketing conduct got to do with it? *Journal of Marketing* 76 (1): 1–19.

Lemmens, Aurélie, Christophe Croux, and Marnik G. Dekimpe. 2007. Consumer confidence in Europe: United in diversity? *International Journal of Research in Marketing* 24 (2): 113–127.

Ma, Yu., Kusum L. Ailawadi, Dinesh K. Gauri, and Dhruv Grewal. 2011. An empirical investigation of the impact of gasoline prices on grocery shopping behavior. *Journal of Marketing* 75 (2): 18–35.

Newey, Whitney K., and Kenneth D. West. 1987. A simple positive semi-definite heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55 (3): 703–708.

Nilsson, Ronny and Gyorgy Gyomai. 2011. Cycle Extraction: A Comparison of the Phase-Average Trend Method, the Hodrick-Prescott and Christiano-Fitzgerald Filters. *Organisation for Economic Co-operation and Development* (OECD). https://www.oecd.org/std/leading-indicators/41520591.pdf. Accessed 25 Apr 2016.

Özturan, Peren, Ayşegül Özsomer, and Rik Pieters. 2014. The role of market orientation in advertising spending during economic collapse: The case of Turkey in 2001. *Journal of Marketing Research* 51 (2): 139–152.

Randles, Ronald H., Michael A. Fligner, George E. Policello II, and Douglas A. Wolfe. 1980. An asymptotically distribution-free test for symmetry versus asymmetry. *Journal of the American Statistical Association* 75 (369): 168–172.

Razzak, Weshah A. 2001. Business cycle asymmetries: International evidence. *Review of Economic Dynamics* 4 (1): 230–243.

Sethuraman, Raj, Gerard J. Tellis, and Richard A. Briesch. 2011. How well does advertising work? Generalizations from meta-analysis of brand advertising elasticities. *Journal of Marketing Research* 48 (3): 457–471.

Shiskin, Julius. 1974. The Changing Business Cycle, 222. *New York Times*, 1 December 1974.

Sichel, Daniel E. 1993. Business cycle asymmetry: A deeper look. *Economic Inquiry* 31 (2): 224–236.

Srinivasan, Raji, Gary L. Lilien, and Shrihari Sridhar. 2011. Should firms spend more on research and development and advertising during recessions. *Journal of Marketing* 75 (3): 49–65.

Srinivasan, Raji, Arvind Rangaswamy, and Gary L. Lilien. 2005. Turning adversity into advantage: Does proactive marketing during a recession pay off? *International Journal of Research in Marketing* 22 (2): 109–125.

Steenkamp, Jan-Benedict E.M., and Eric (Er) Fang. 2011. The impact of economic contractions on the effectiveness of R&D and advertising: Evidence from U.S. companies spanning three decades. *Marketing Science* 30 (4): 628–645.

Stock, James H., and Mark W. Watson. 1999. Business cycle fluctuations in U.S. macroeconomic time series. In *Handbook of macroeconomics*, ed. John Taylor, and Michael Woodford, 3–64. Amsterdam: Elsevier Science.

Thoma, Mark A. 1994. Subsample instability and asymmetries in money-income causality. *Journal of Econometrics* 64 (1/2): 279–306.

Tull, Donald S., Van R. Wood, Dale Duhan, Tom Gillpatrick, Kim R. Robertson, and James G. Helgeson. 1986. “Leveraged” decision making in advertising: The flat maximum principle and its implications. *Journal of Marketing Research* 23 (1): 25–32.

van Heerde, Harald J., Maarten Gijsenberg, Marnik G. Dekimpe, and Jan-Benedict E.M. Steenkamp. 2013. Price and advertising effectiveness over the business cycle. *Journal of Marketing Research* 50 (2): 177–193.

Verbrugge, Randal J. 1997. Investigating cyclical asymmetries. *Studies in Nonlinear Dynamics and Econometrics* 2 (1): 15–22.

Zarnowitz, Victor. 1985. Recent work on business cycles in historical perspective: Review of theories and evidence. *Journal of Economic Literature* 23 (2): 523–580.

Chapter 13
Marketing Models for the Life Sciences Industry

Vardan Avagyan, Vardit Landsman and Stefan Stremersch

13.1 Introduction

The life sciences industry forms the innovative producer side of therapies in the healthcare industry. The life sciences industry has several unique features (Stremersch and Van Dyck 2009; Stremersch 2008). Companies that produce therapies are significantly more strongly linked to science compared to any other industry. The therapies that are launched in the healthcare market alter the trajectory of many debilitating diseases and have substantial impact on people's well-being. Because of its great importance, the life sciences industry is strictly regulated (Stremersch and Lemmens 2009; Verniers et al. 2011). Moreover, its value chain is unique: care providers, like physicians, (co-)decide on therapy choices for patients, and payers, such as insurance companies or government institutions, typically pay for (part of) the treatment.

Figure 13.1 represents the healthcare value chain (Stremersch and Van Dyck 2009). The healthcare value chain is comprised of a healthcare payment flow and a healthcare delivery flow. Payments for healthcare products flow from left to right, from payers to healthcare providers. Employers, patients, and/or government pay

V. Avagyan (✉) · V. Landsman · S. Stremersch
Erasmus School of Economics, Erasmus University Rotterdam, Rotterdam, The Netherlands
e-mail: avagyan@ese.eur.nl

V. Landsman
e-mail: landsman@ese.eur.nl

S. Stremersch
e-mail: stremersch@ese.eur.nl

V. Landsman
Leon Recanati Graduate School of Business, Tel Aviv University, Tel Aviv, Israel

S. Stremersch
IESE Business School, University of Navarra, Pamplona, Spain

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_13

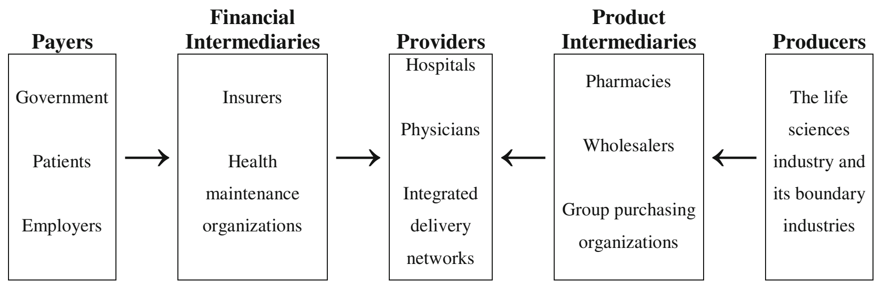


Fig. 13.1 The healthcare value chain (Stremersch and Van Dyck 2009)

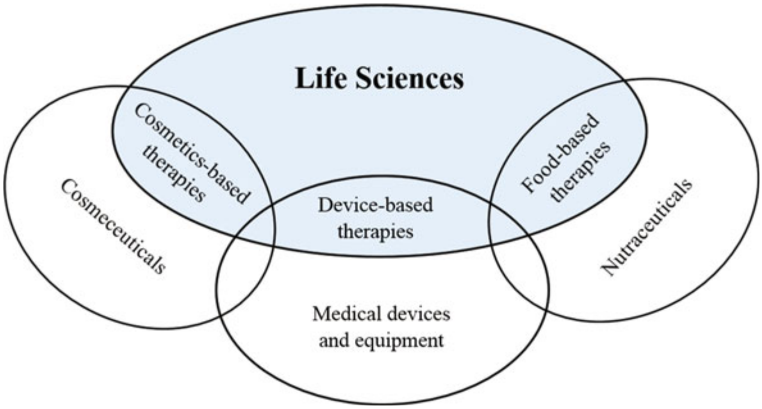


Fig. 13.2 The life sciences industry and its boundaries (Stremersch and Van Dyck 2009)

for healthcare products. Often, the payments flow through financial intermediaries such as health maintenance organizations (HMOs) and insurance companies. Healthcare delivery starts from the life sciences industry (Stremersch and Van Dyck 2009). The industries that constitute the life sciences industry are the pharmaceutical, biotechnology, and (therapeutic) medical devices industries. Industries that are adjacent to life sciences are the food (e.g., nutraceuticals), high-tech (e.g., medical imaging), and cosmetics industries (e.g., cosmeceuticals) (See Fig. 13.2).

The life sciences industry is an important part of the economy. For instance, for the US, global annual sales in the life sciences industry approached $286 billion representing 1.6% of the GDP (PhRMA 2015a). The world pharmaceutical market, one of the largest components of the life sciences industry, was worth an estimated $865 billion in 2014 representing 1.1% of world GDP (EFPIA 2015). It takes more than 10 years to develop a drug and costs on average $800 million up to $2.6

billion (PhRMA 2015b), and only 1 in 50,000 drug candidates are eventually commercialized (Grewal et al. 2008).

The life sciences industry gives rise to interesting research questions, as well as enables new model development to support managerial decision making. For instance, the ultimate decision maker (i.e., the provider) is not the consumer (i.e., the patient), while the latter has a certain degree of influence over the former (Camacho et al. 2010). The payment is typically covered by a third party who puts substantial influence on the decisions of providers. The nuanced relationship among life sciences companies, providers, patients and payers takes place within an environment controlled by regulators. Regulators interfere in life sciences companies' decisions regarding the launch time and prices and influence diffusion pattern of therapies. Regulators may also influence the marketing efforts directed to patients or physicians. They do so to control costs and to secure quality access to healthcare for the public at large.

The academic marketing literature has produced a sizeable array of decision-support tools for the life science marketers. In this chapter, we present to researchers and managers in the life sciences industry a broad overview of these analytical tools, categorized according to subject areas, and the key managerial insights that have been derived from them.

This chapter should also be of interest to managers from other industries for two reasons. First, these decision-support models are highly sophisticated, state-of-the-art, and are applicable to decision problems outside of the life sciences. For instance, physician-level detailing response models can be easily generalized to consumer-level marketing-mix response models, or, key-opinion-leader models in pharmaceuticals can be generalized to other social influence settings. Second, other industries may benefit from some of the insights of pharmaceutical marketing studies. For instance, insights on the decision making process of the physicians and patients and how such process affects therapy compliance of patients can be generalized to other expert-advice decision-making settings, e.g., client-counsel decision making and the effect thereof on client's adherence to counsel.

We structure the chapter as follows. Section 2 presents the typical models employed in the following modeling traditions: choice model, count model, learning model, modeling key opinion leaders, diffusion model, sales growth model, and launch model (Table 13.1 presents an overview of such models discussed in the section). Section 3 presents the findings on the role of marketing categorized according to the following decision areas: direct-to-physician promotion, direct-to-consumer advertising (DTCA), pricing, and product usage adherence (Table 13.2 presents a summary of key findings in each decision area). We conclude with a number of areas that we think need more research.

Table 13.1 Overview of models employed

Main types of models	Data and sources	Necessary considerations	References
Physician choice models			
Multinomial logit with random effects	Physician panel data that are available from specialized data providers, pharmacies, insurers and syndicated surveys such as NAMCS	Unobserved heterogeneity in physicians' choice Endogeneity of marketing variables	Gonzalez et al. (2008), Janakiraman et al. (2008), Wosinska (2002)
Probit model with random effects			Hellerstein (1998), Lundin (2000)
Latent class multinomial logit			Gönül et al. (2001)
Prescription Count Models			
Poisson models	Prescription information from pharmacies, insurance agencies, specialized providers (e.g., ImpactRx, IMS Health), individual pharmaceutical firms	Dispersion of prescriptions Extension to multiple product categories Endogeneity of marketing variables	Datta and Dave (2013), Dong et al. (2011), Manchanda and Chintagunta (2004)
Negative binomial models			Manchanda et al. (2004), Stremersch et al. (2013), Venkataraman and Stremersch (2007)
Physician learning models			
Bayesian learning models	Disaggregate-level data at the physician and at the patient level that are available from specialized providers, third-party providers such as the Italian National Institute of Health or medical schools (e.g., the IPCI database) Aggregate-level data that are available from specialized providers	Source of information signals Direct versus indirect role of information Learning of the quality of treatment versus characteristics of treatment Myopic or forward-looking physicians Risk-aversion of physicians	Chan et al. (2013), Ching and Ishihara (2010), Chintagunta et al. (2009, 2012), Coscelli and Shum (2004), Janakiraman et al. (2009), Kalra et al. (2011), Narayanan et al. (2005), Narayanan and Manchanda (2009)
Quasi-Bayesian learning models		Differential versus equal weights on different types of information	Camacho et al. (2011)
Modeling key opinion leaders			
Linear response model	Survey of physicians about their network ties and self-reported opinion leadership Distance between physician office locations Anonymous patient-movement data	Endogenous group formation Correlated unobservables Simultaneity of decisions of opinion leaders and other physicians	Bhatia and Wang (2011), Nair et al. (2010)
Hazard model			Iyengar et al. (2011, 2015)

(continued)

Table 13.1 (continued)

Main types of models	Data and sources	Necessary considerations	References
Diffusion Models			
Bass model	Observed behavior in physician panels (e.g., IMS Health) Surveys and interviews (e.g., Medical Innovation dataset)	Trial and repeat prescriptions Effect of marketing actions Dynamics of such effects Heterogeneity among adopters	Hahn et al. (1994), Kolsarici and Vakratsas (2010), Rao and Yamada (1988), Shankar et al. (1998), Vakratsas and Kolsarici (2008)
Logistic model			Desiraju et al. (2004), Van den Bulte (2000)
Discrete-time hazard			Iyengar et al. (2011, 2015), Van den Bulte and Lilien (2001)
Sales models			
Brand-level models	Wholesalers Pharmacies Specialized providers	Incorporation of lagged sales Time-varying or constant parameters Lagged effect of marketing actions	Parsons and Vanden Abeele (1981), Shankar et al. (1999), Stremersch and Lemmens (2009)
Market-share models			Montgomery and Silk (1972)
Category-level models			Chintagunta and Desariju (2005)
Launch models			
Cox propotional hazard	*Data on Launch time* IMS Health, PJB Publications, World Health Organization, regulatory approvals, firm announcements, local media reports, etc. *Data on Regulations* URCH Publishing, OECD reports, GenericsWeb (IMS Health)	Right censoring problem Endogeneity of launch or launch price decisions Endogeneity of launch timing and regulatory regimes	Danzon et al. (2005)
Discrete-time hazard			Kyle (2006), (2007), Kyle and Qian (2014)
Tobit-I			Verniers et al. (2011)

Table 13.2 Summary of findings on the role of marketing

Key findings	References
The role of direct-to-physician promotion	
Detailing	
Some studies find that detailing has a strong and positive impact on prescriptions written, others find that it has, at best, only a modest impact on prescriptions.	Azoulay (2002), Chintagunta and Desiraju (2005), Gönül et al. (2001), Manchanda and Chintagunta (2004), Mizik and Jacobson (2004), Parsons and Vanden Abeele (1981), Stremersch et al. (2013)
Detailing is more effective for drugs that are highly effective or have many side effects.	Venkataraman and Stremersch (2007)
In the early stages of the treatment life cycle, detailing has primarily an indirect impact by helping the physicians reduce uncertainty about the treatment. In the late stages, the indirect effect of detailing becomes increasingly smaller as its effect becomes more direct.	Narayanan et al. (2005)
Journal advertising	
Journal advertising influences physicians' adoption.	Van den Bulte and Lilien (2001)
Journal advertising is less effective for promoting market size expansion in expensive categories but is more effective for chronic care compared to acute care categories.	Fischer and Albers (2010)
Sampling	
Sampling has a positive influence on physician prescription decisions.	Manchanda et al. (2004), Mizik and Jacobson (2004)
Sampling has diminishing returns to scale.	Gönül et al. (2001)
Empirical generalizations	
Detailing elasticity is positive.	Albers et al. (2010), Kremer et al. (2008), Sridhar et al. (2014)
Detailing elasticity is lower in the pharmaceutical industry than the mean of personal selling elasticity across all industries.	Albers et al. (2010)
Detailing elasticity is the highest, followed by direct-to-physician advertising, DTCA and other direct-to-physician instruments.	Kremer et al. (2008)

(continued)

Table 13.2 (continued)

Key findings	References
Detailing elasticity is higher for new drugs than for mature drugs.	Sridhar et al. (2014)
Detailing elasticity is higher in Europe than in the US.	Sridhar et al. (2014)
The Role of DTCA	
The effect on physician prescriptions Some studies suggest that the effect of DTCA on prescriptions is positive and large, others suggest that DTCA has a very limited effect or no effect. Some even report a negative effect of DTCA in certain disease categories.	Calfee et al. (2002), Iizuka and Jin (2005), (2007), Kremer et al. (2008), Law et al. (2008, 2002), Narayanan et al. (2005), Stremersch et al. (2013)
The effect on patient requests The mean effect of DTCA on patient requests is negative but small. The effect is heterogeneous across socio-demographic characteristics and brands.	Liu and Gupta (2011), Stremersch et al. (2013)
Informational role of DTCA DTCA increases physician visits, but has a small or no effect at all on patient adherence.	Brekke and Kuhn (2006), Bowman et al. (2004), Calfee et al. (2002), Liu and Gupta (2011), Wosinska (2005)
The effect on investors DTCA increases stock returns, lowers systematic risk, and raises idiosyncratic risk.	Osinga et al. (2011)
The role of pricing	
Physicians' sensitivity to prices A significant segment of price-insensitive physicians exists. Studies report mixed results on the effect of direct-to-physician promotional efforts on physicians' price sensitivity.	Gönül et al. (2001), Narayanan et al. (2004), Rizzo (1999), Windmeijer et al. (2006)
Launch prices Treatments that offer larger therapeutic advancement or are indicated for acute conditions have higher launch prices than those that offer smaller therapeutic advancement or are indicated for chronic conditions. Launch time has an inverted U-shape effect on the launch price.	Ekelund and Persson (2003), Lu and Comanor (1998), Verniers et al. (2011)
Price dynamics over the life cycle	

(continued)

Table 13.2 (continued)

Key findings	References
In non-regulated markets, treatments that offer significant therapeutic advancement are introduced with a price-skimming strategy, whereas treatments that provide only a small therapeutic gain are launched with a penetration strategy. In regulated markets, all classes of therapeutic innovation are launched with price-skimming strategies.	Bhattacharya and Vogt (2003), Ekelund and Persson (2003), Lu and Comanor (1998)
Price reactions to the entry of generics Studies report mixed results on the effect of generic entry on the prices of brand-name drugs.	Bhattacharya and Vogt (2003), Caves et al. (1991), Frank and Salkever (1997)
Price reactions to parallel imports Parallel imports lower the prices of brand-name drugs, but have no effect on the prices of generic drugs. Parallel imports weaken the negative effect of price cap regulations on prices.	Brekke et al. (2015), Duso et al. (2014), Ganslandt and Maskus (2004)
Product usage adherence	
Patient empowerment Patient-initiated informational empowerment improves adherence. Physician-initiated informational empowerment leads to higher nonadherence. Decisional empowerment increases nonadherence.	Camacho et al. (2014), Prigge et al. (2015)
Patient-physician relationship related factors The attitudinal homophily, the time that has elapsed from the medical intervention, and the duration of the relationship influence adherence.	Bowman et al. (2004), Dellande et al. (2004), Schwartz et al. (2011)
Therapy-related factors The consequences of the treatment, experienced side effects, the complexity of medical regimen such as the dosage frequency, salience of symptoms, and treatment progress influence adherence.	Bowman et al. (2004), Kahn and Luce (2003), Lamiraud and Geoffard (2007), Lien et al. (2010)
Firm controlled factors Drug prices, detailing, and DTCA have a small or no effect on patient adherence. Warning labels improve adherence.	Bowman et al. (2004), Donohue et al. (2004), Ferguson et al. (1987), Wosinska (2005)

13.2 Typical Models

13.2.1 Physician Choice Models

Lawmakers and public administrators empower healthcare providers as expert gatekeepers of patient care. Legislators and administrators believe that without strong regulatory system patients may fall prey to malpractices. Therefore, marketing modelers have shown a strong interest in the therapy choices of healthcare providers like physicians. Such choices can have far-reaching consequences, for example, the survival or death of patients or the commercial success or failure of therapies and their corresponding suppliers. Therapy choice is at the core of the medical profession. The *primum non nocere* ("first do no harm") precept of the Hippocratic Oath that all physicians take, also includes not treating certain conditions if the potential harm resulting from the treatment outweighs the potential benefits.

To examine physicians' prescription choices, scholars typically use physician panel data from specialized data providers. These data include information on physicians' prescription choices at each prescription occasion combined with some information about the prescribed drug (e.g., dosage), the physician (e.g., specialty), or the patient (e.g., gender). Examples of databases are Scott-Levin (now part of IMS Health), Jigsaw (a database of Synovate Healthcare, now part of Ipsos), and IMS Health. Alternatively, researchers may inventory data on physician prescription choices from pharmacies or insurers who record the identities of prescribing physicians, patient, and the date when the drug was prescribed (e.g., Lundin 2000; Wosinska 2002). In addition, researchers take the opportunity to use data from syndicated surveys such as the National Ambulatory Medical Care Survey (NAMCS). In this particular survey, nonfederal office-based physicians complete a one-page questionnaire for each patient visit during a one-week reporting period. The survey data include physician characteristics, patient demographics, and visit characteristics (e.g., patients' symptoms, physician's diagnoses, visit disposition, time spent with physician, etc.).

Researchers routinely use probit or logit models to study physician choice decisions. For instance, choice models have been used to investigate the responsiveness of physicians' choice to the promotional activities of firms (Gonzalez et al. 2008; Gönül et al. 2001; Janakiraman et al. 2008; Wosinska 2002), to price and co-pay (Gonzalez et al. 2008; Lundin 2000), to physicians' habit persistence (e.g., Janakiraman et al. 2008; Lundin 2000), to patient age, gender, and insurance status (Gönül et al. 2001; Scott and Shiell 1997), and to the type of physician remuneration from the payer, i.e., whether the physician is remunerated based on either the consultation length or its content (Scott and Shiell 1997).

However, physicians' prescription choice decisions might be strongly influenced by other explanatory factors that are unobservable to a researcher. Take for example the quality of the relationship between doctor and patient. While this can be proxied by homophily (in gender, age, race, etc.) it cannot really be measured in a large sample of prescription occurrences. Another example is the drug profile as perceived by the individual physician as it evolves over time, informed by scientific studies as well as by feedback from the physician's patients.

To account for such unobserved heterogeneity in physicians' choices, researchers employ several methodologies. Marketing scientists often opt for multinomial logit with random effects (e.g., Gonzalez et al. 2008; Janakiraman et al. 2008; Wosinska 2002). This model accounts for unobserved heterogeneity in physicians' choices via a random coefficient formulation. Specifically, physician i's valuation of drug j at patient visit occasion t is defined as:

$$U_{ijt} = X_{ijt}^{'}\beta_{ij} + \varepsilon_{ijt}, \text{ for } i = 1, \ldots N, j = 1, \ldots, J, \text{ and } t = 1, \ldots, T_i, \tag{13.1}$$

where X_{ijt} is a vector of explanatory variables, β_{ij} is the corresponding vector of physician-specific parameters, J is the number of alternative drugs, N is the number of physicians, T_i is the number of physician i's prescription occasions, and ε_{ijt} is an unobserved random term that is independent and identically distributed extreme value, independent of observed variables and coefficients. Physician i's probability of prescribing drug j on occasion t is then defined as:

$$P_{ijt} = \frac{exp\left(X_{ijt}^{'}\beta_{ij}\right)}{\sum_{k=1}^{J} exp\left(X_{ikt}^{'}\beta_{ik}\right)} \tag{13.2}$$

This specification assumes that physician-specific parameters are normally distributed; i.e., $\beta_{ij} \propto MVN(\beta, \Sigma)$ where β is a vector of population-level means and Σ is the corresponding variance-covariance matrix. The stochastic deviation around the means accounts for unobserved factors and also eliminates from standard logit models the restrictive assumption of independence from irrelevant alternatives.

Other scholars have employed random effects probit (e.g., Hellersten 1998; Lundin 2000) or latent class multinomial logit model that allows for semiparametric distribution of heterogeneity (e.g., Gönül et al. 2001). In contrast to multinomial logit with random effects, latent class multinomial logit is slightly less flexible because it approximates the underlying continuous distribution of physician-specific parameters with a discrete one. However, this model does not require the researcher to make specific assumptions about the distributions of parameters across physicians.

13.2.2 Prescription Count Models

Beyond the choice models discussed in the previous section, the most commonly used individual-level models include count models. The reason is that the most commonly used data provide the monthly number of prescriptions by a physician.

Moreover, typical of such data is that the count includes a relatively large number of zeros and a small number of frequently occurring outcomes.

A couple of sources can be used to obtain prescription count data. One can acquire the prescription information from pharmacies that inventory the scripts as part of their compliance with government or insurance regulation. Such data include the prescriber's name, which, in some countries, the data provider is not legally prohibited from sharing with the pharmaceutical firm. One can also inventory data from insurance agencies that obtain prescription information, including prescriber identity, dosage, etc. for reimbursement purposes. Alternatively, one can secure prescription data from specialized data providers who ask a sample of physicians to keep diaries of the number of prescriptions written, of firm promotion efforts (e.g., the detailing and samples that they receive from sales representatives), of the drug and of its dosage prescribed. Data providers such as ImpactRx, IMS Health, and individual pharmaceutical firms can make such information available to researchers. Scholars often integrate these kinds of data with other data sets that cover, for example, DTCA expenditures (from KantarMedia or Nielsen) or census data on demographic characteristics of different location (see, e.g., Stremersch et al. 2013).

The count models widely used in pharmaceutical marketing are Poisson (Datta and Dave 2013; Dong et al. 2011; Manchanda and Chintagunta 2004) and negative binomial regression models (e.g., Manchanda et al. 2004; Stremersch et al. 2013; Venkataraman and Stremersch 2007). In the Poisson regression model, the number of prescriptions written RX_{it} by each physician i in period t is given by:

$$Pr(RX_{it} = k|\lambda_{it}) = \frac{\lambda_{it}^{k} exp(-\lambda_{it})}{k!} \tag{13.3}$$

where λ_{it} is the physician-specific mean prescription rate. Manchanda and Chintagunta (2004) specify λ_{it} to be a function of detailing, where the effect of detailing is allowed to be physician-specific and to be a function of detailing characteristics, observed physician characteristics, and unobserved factors.

The Poisson regression model has been extended to multiple product categories. Multivariate count models represent a natural way of accommodating prescription data from multiple drug categories. Marketing scientists have proposed a multi-category count data model that allows for the Poisson parameter in one equation to be a function of the Poisson parameters in other equations (Dong et al. 2011).

In Poisson regression models, the conditional mean and variance of prescriptions are specified to be equal. If prescriptions are over-dispersed (i.e., variance larger than the mean), negative binomial models are commonly used. The negative binomial distribution model with mean λ_{it} and over-dispersion parameter α is represented by:

$$Pr(RX_{it} = k|\lambda_{it}) = \frac{\Gamma(\alpha + k)}{\Gamma(\alpha)\Gamma(k+1)} \left(\frac{\alpha}{\alpha + \lambda_{it}}\right)^{\alpha} \left(\frac{\lambda_{it}}{\alpha + \lambda_{it}}\right)^{k} \tag{13.4}$$

The most common specification for the conditional mean of the number of prescriptions is a log-link function that specifies the log of the mean of the conditional distribution as linear in parameters. The mean number of prescriptions is then a function of, for example, detailing D_{it}. To incorporate time dynamics in the prescription process, the mean number of prescriptions can be specified as a function of the number of time periods since the introduction of the drug:

$$ln(\lambda_{it}) = \beta_0 + \beta_{1i}t + \beta_{2i}D_{it} + \gamma_i X_{it} + \varepsilon_{it} \tag{13.5}$$

where X_{it} includes a set of time-varying physician-specific covariates.

This flexible count model has been adopted in many studies examining physicians' prescription, request accommodation, and sample-dispensing behaviors. Stremersch et al. (2013) develop a system of four hierarchical negative binomial models to specify physician prescriptions, patient requests, detailing, and DTCA to be estimated simultaneously. The authors include in their model the physicians' responsiveness to requests as a determinant of the number of requests; the responsiveness of prescriptions to detailing as a determinant of detailing; and the responsiveness of prescriptions to requests, the responsiveness of prescriptions to DTCA, and the responsiveness of requests to DTCA as determinants of DTCA spending. Moreover, the authors specify correlated error terms across four negative binomial models. These links between the four equations enables one to overcome possible biases from endogeneity of requests, detailing, and DTCA budget allocations. Venkataraman and Stremersch (2007) use a negative binomial distribution model to examine both the prescription- and sample-dispensing behaviors of physicians.

13.2.3 Physician Learning Models

Learning models build on the premise that, over time, decision makers receive information signals that enable them to learn about a treatment, from its launch to its mature use in clinical practice. In the context of the healthcare industry, there is considerable uncertainty about the characteristics and benefits of a new pharmaceutical treatment (e.g., efficacy, side effects, drug interactions, etc.) Such uncertainty decreases over time as physicians gain experience from prescribing the treatment, and as they gain information from firms and independent researchers and institutions. How fast such uncertainty decreases and the treatment's perceived level of quality are of prime interest to firms and to the other players in the healthcare value chain, i.e., care providers, payers, regulators, etc. Care providers will only administer to patients treatments they feel have more benefits than risks, taking into account the uncertainty surrounding both. For these reasons, marketing scholars have built physician learning models. These models are typically developed within the (quasi-) Bayesian updating tradition.

Research uses either disaggregate- or aggregate-level data to estimate learning models. Disaggregate-level data are typically at the physician level over time and possibly at the patient level as well. Scholars usually obtain such data from providers like ImpactRx and IMS Health. Other scholars secure physician panel from third-party providers such as the Italian National Institute of Health (Coscelli and Shum 2004) or medical schools (e.g., the IPCI (Integrated Primary Care Information) database, which is a panel of general physicians that records the full prescription history of each patient, including all refills) (Camacho et al. 2011). At the patient level, one may use Anonymous Patient-Level Data provided by IMS Health, which contains individual physicians' prescription choices over time at the patient level. For each individual prescription, one may observe information about the prescribing physician (e.g., specialty and location), the patient characteristics (e.g., age, gender, insurance coverage), and the prescription information (e.g., the drug name, units of the medication dispensed, drug strength, etc.) Such data can be combined with physician-level promotional information, also provided by IMS Health. Aggregate-level data are typically from data providers, such as IMS Health, that provide information on the aggregate demand of a brand at a monthly level and the promotional expenditures for each brand. Note that physician learning models are individual-level models. Therefore, scholars have developed methods that enable estimating learning models with aggregate data (see e.g., Ching and Ishihara 2010; Narayanan et al. 2005).

Existing work that models physician learning typically focuses on the uncertainty and learning of *overall* treatment quality. The mean quality of treatment j for physician i (defined as Q_{ij}) may be heterogeneous across patients. Therefore, the quality of treatment j for patient p visiting physician i can be defined as the sum of the true mean of quality of treatment across patients and a patient-specific deviation from this mean, that is,

$$Q_{ipj} = Q_{ij} + q_{ipj}, \text{ where } q_{ij} \sim N\left(0, \sigma^2_{q,ipj}\right). \tag{13.6}$$

Before receiving any information, physicians already have a prior belief about the quality of treatment that is normally distributed. Physicians receive (truthful) information signals that help them reduce their uncertainty about the quality of the treatment. The signals that physicians receive are commonly specified to be normally distributed around the true mean quality.

The assumptions of normally distributed prior beliefs and signals guarantee that the physician's posterior beliefs are also normally distributed. Physicians integrate the information signals they receive with their prior beliefs to update their beliefs about the treatment's mean quality and the patient-specific deviation from this mean. Most scholars who model physician learning assume that physicians learn in a Bayesian fashion. Bayesian learning models assume that physicians optimally weigh past (or prior) and new information according to the Bayesian updating rule.

At each encounter with a patient, a physician chooses the treatment that maximizes the expected utility of the patient. The physician's utility is modeled as a function of the physician’s updated beliefs about the quality of the treatment, the degree of the physician’s risk-aversion, and other factors that directly influence the utility (e.g., firm’s promotions).

Existing physician learning models critically hinge upon the assumptions and specifications that modelers make. The first assumption with respect to the learning process pertains to the information source that reduces physician uncertainty in the Bayesian updating process. Several studies have specified physician learning models when the only source of information signals is the physicians’ direct experience, i.e., feedback retrieved from patients (e.g., Camacho et al. 2011; Chintagunta et al. 2009; Coscelli and Shum 2004). In other models, physicians learn from a range of sources based on direct and indirect experience, i.e., pharmaceutical firms’ marketing efforts, attendance at medical conferences, etc. (e.g., Ching and Ishihara 2010; Chan et al. 2013; Chintagunta et al. 2012; Narayanan et al. 2005).

The second assumption concerns the role of information in the physicians’ treatment decisions (Narayanan et al. 2005). The information physicians obtain can be specified to have either an indirect or a direct role. The former refers to the indirect influence of information on physicians’ utility by enabling them to update their prior beliefs and thus learn about the true quality of the treatment (e.g., by raising awareness about potential side effects of the treatment). The direct role of information refers to all effects that are not indirect (e.g., reminders about the treatment) that influence physicians’ preferences through goodwill accumulation and thus directly shift physicians’ utility. These definitions of indirect and direct effects loosely correspond to the “informative” (indirect) and the “persuasive” (direct) roles of advertising. Informative advertising informs consumers about the characteristics of the product, while persuasive advertising has the potential to shift consumer tastes (Narayanan et al. 2005).

One of the fundamental assumptions of Bayesian learning models is that all information is weighted equally. However, psychology literature has documented several behavioral effects that may cause physicians to place systematically different weight on one type of information over other types of information in the learning process. Examples of such behavioral effects are forgetting and effects related to information salience, recency and extremity, to name a few. In recent years, several scholars have developed quasi-Bayesian models to allow for such effects. For example, Camacho et al. (2011) propose a quasi-Bayesian learning model in which physicians deviate from pure Bayesian updating in a predictable and systematic manner. Specifically, when learning about a new drug, physicians place more weight than Bayesian updating would predict on the feedback from patients who switched away from the drug relative to the feedback of patients who continued therapy and refilled the drug. The authors argue that the negative feedback from switching patients is more salient in the physician’s memory than the (positive) feedback from satisfied patients. Camacho et al. (2011) find evidence that negative patient feedback receives 7–10 times more weight than positive feedback in physician learning. The authors show that this effect significantly diminishes the speed of diffusion of the new drug.

The next assumption of physician learning models refers to the characteristics of the treatment about which the physician is learning. The majority of prior work focuses on uncertainty and learning on *overall quality* of treatments. However, physicians' uncertainty can manifest itself differentially on treatment characteristics such as side effects, efficacy, etc. Chan et al. (2013) propose a learning model wherein physicians learn both about a treatment's effectiveness and its side effects, rather than about its overall quality.

One of the distinctive characteristics of the physician learning process is the level at which a physician is learning. When a patient reports her treatment experience to the physician, the reported information may reflect the treatment's average quality across patients or the patient's idiosyncratic match with the treatment. Chintagunta et al. (2009) develop a physician learning model that distinguishes physician learning across-patients about the efficacy of the new treatment from physician learning (heterogeneous across patients) about match between the treatment and each patient.

An important aspect of learning models is whether physicians are specified to be forward-looking or myopic, that is, whether physicians engage in strategic trials. Typically, physician learning models assume myopic physicians. One exception is the study by Chintagunta et al. (2012) that specifies forward-looking physicians who engage in strategic trials of treatments.

The risk aversion of physicians is another key distinctive feature of designing physician learning models. The medical profession has several institutional characteristics that may trigger risk-averse behaviors among physicians such as the *primum non nocere* precept, the threat of medical malpractice liability suits (especially in the US), and the emotions involved in making decisions that will determine whether a patient lives or dies. Several learning models allow for physicians' risk aversion, whereas others assume risk-neutral physicians. Importantly, as Coscelli and Shum (2004) point out, risk aversion is hard to identify when the physicians' prior means are estimated separately from the treatment's true mean utilities because resistance to treatment adoption may come from either risk aversion or low prior means.

13.2.4 Modeling Key Opinion Leaders Among Physicians

Life sciences firms often stimulate the adoption of their treatments through key opinion leaders who exert strong influence on the attitudes and behavior of their peers (Stremersch and Van Dyck 2009). The role of key opinion leaders can be significant especially when there is great uncertainty about the treatment. They typically participate in premarket product testing and help the firm to reduce clinical uncertainty. Key opinion leaders can also ensure market access by helping firms receive more favorable formulary status (a higher price tier in the list of drugs that the payers pay). Given the significant economic benefits of key opinion leaders, their identification is crucial for life sciences firms to be able to target their marketing activities towards such opinion leaders.

There are several approaches for obtaining data on key opinion leaders. The first is a survey of physicians to gather information on physicians' network ties and self-reported opinion leadership (e.g., Iyengar et al. 2011, 2015; Nair et al. 2010). Alternatively, one can also use distance between physician office locations (Manchanda et al. 2008). Another common approach is to use information on patient movements between physicians, which enables the identification of key opinion leaders (e.g., Bhatia and Wang 2011). Examples of companies that provide such anonymous patient-movement data are Surveillance Data Incorporated (later acquired by IMS Health), Wolters Kluwer, and Dendrite.

The models adopted by researchers to study key opinion leaders can be categorized into two main types: linear response models and hazard models. Nair et al. (2010) employ a linear specification for a physician i's prescription at time t, denoted as y_{it}, as a function of her nominated opinion leader $j(i)$'s prescriptions $RX_{j(i),t}$, detailing D_{it}, control variable $X_{-i,t}$, which are computed as the mean prescription of all other physicians in physician i's zip code, and physician- and time-specific fixed effects, α_i and γ_t.

$$y_{it} = \alpha_i + \gamma_t + \beta D_{it} + \delta RX_{j(i),t} + \gamma X_{-it} + \varepsilon_{it}, \text{for } i = 1, \ldots, N; t = 1, \ldots, T. \quad (13.7)$$

The authors further specify the prescriptions of physician i's opinion leader as:

$$RX_{j(i),t} = \alpha_{j(i),t} + \tau_t + \omega D_{j(i),t} + \xi y_{it} + \zeta X_{-j(i),t} + \varepsilon_{j(i),t}, t = 1, \ldots, T. \quad (13.8)$$

Nair et al. (2010) use a two-stage fixed-effects panel data instrumental variables regression to estimate their model. Nair et al. (2010) find that compared to their peers, key opinion leaders are more sensitive to the promotional activities of firms. Bhatia and Wang (2011) use a similar linear specification to model physicians' prescriptions. They use a simultaneous equations regression to estimate their model. Bhatia and Wang (2011) find that primary care physicians (PCP) negatively influence other PCPs, whereas specialists have a significantly positive effect on the PCPs but not vice versa.

Other scholars employ a discrete-time hazard model (Iyengar et al. 2011, 2015). Iyengar et al. (2011) study the adoption of a treatment, whereas Iyengar et al. (2015) study physicians' adoption of (i.e., the trial) and repeat prescription of a treatment. The trial of a drug can be expressed as:

$$P\left(Y_{it}^a = 1 | Y_{it-1}^a = 0\right) = \Phi\left(\beta_{0i}^a + X_{it}^a \beta_1^a\right), \quad (13.9)$$

and the probability of the repeat prescription, conditional on having adopted earlier as:

$$P\left(Y_{it}^r = 1 | Y_{it-1}^a = 1\right) = \Phi\left(\beta_{0i}^r + X_{it}^r \beta_1^r\right). \quad (13.10)$$

Here, Y_{it}^a is an indicator variable that equals 0 before adoption and equals 1 at the time of adoption and later. Y_{it}^r is an indicator variable that takes a value of 1 if physician i prescribes at a time t and is 0 otherwise. Φ is the normal cumulative distribution function. β^a_{0i} and β^r_{0i} are a physician-specific baseline trial and repeat utility, respectively, which are normally distributed and control for unobserved characteristics. X^a_{it} and X^r_{it} contain covariates before and after adoption, and β^a_1 and β^r_1 are the corresponding vectors of parameters. The covariates of the hazard models are specified to include self-reported assessments of leadership, the number of discussions and referral nominations received from other physicians (in-degree centrality), social contagion measures, and control variables.

Iyengar et al. (2015) find that who is the most influential varies across physicians' trial and subsequently, across physician's repeat prescriptions of a treatment. In the case of treatment trials, physicians who are central in the physician network and heavy prescribers, as well as immediate colleagues drive the contagion. However, central physicians and heavy prescribers do not drive contagion at the repeat stage. Instead, for the repeat prescription decisions only immediate colleagues are influential. Moreover, who is the most influenceable also varies across physicians' trial and repeat prescription stages (Iyengar et al. 2015). For the trial stage, the most influenceable are the physicians who do not see themselves as opinion leaders. For the repeat stage, the physicians in the middle of the status distribution, as measured by network centrality, are the most susceptible.

There are three significant modeling challenges in identifying the effects of opinion leaders on other physicians: endogenous group formation, correlated unobservables, and simultaneity (see also Nair et al. 2010). The endogenous group formation—often referred to as "homophily"—occurs because physicians may choose as their opinion leaders doctors with similar "tastes" for prescriptions. For instance, physicians may meet and form relationships with experts at conferences about specific therapeutic treatment options. If physicians choose these experts as opinion leaders, subsequent correlation in their (physicians') prescription behavior may reflect these similar tastes and not a causal effect of the opinion leader's behavior on the physician. The availability of panel data may solve the problem of endogenous group formation. With panel data, the econometrician can control for the endogeneity of group formation through physician fixed effects.

The second concern arises when correlated unobservables drive the behavior of both the physician and the opinion leader. For instance, there might be common spatial- and temporal-specific factors or direct-to-physician marketing that similarly influence the physicians in the group. These factors may lead to correlation in physician prescriptions that are not due to opinion-leadership effects but are simply caused by these unobservable factors. Including time and physician fixed effects may solve the problem of correlated unobservables. In addition, one can adopt a difference-in-difference approach by using the behavior of other physicians not in the focal physician's reference group to control for common unobservable factors (see for example Nair et al. 2010 for such an approach).

The third challenge is simultaneity, which arises due to the potentially simultaneous nature of the decisions of opinion leaders and other physicians. Due to peer effects, opinion leaders affect other physicians, while the physicians simultaneously affect the opinion leaders. This will create an upward bias of key opinion leaders' effect. The problem of simultaneity can be solved via instrumental variables (e.g., detailing visits to opinion leaders) that influence the prescriptions of opinion leaders but are excluded in the prescription equation of other physicians in the reference group.

13.2.5 Diffusion Models

Diffusion models represent the dynamic process of new treatment adoption. Such models support in predicting the dynamic process of new treatment adoption and gauge their commercial potential. Scholars commonly estimate diffusion models on two types of data. The first is the observed behavior of physicians across time aggregated at the brand or category level. Such data are typically obtained from IMS Health or directly from the company. The second is the stated prescription behavior of physicians gathered from surveys. An early example of such dataset widely used in academia is the *Medical Innovation* dataset (Coleman et al. 1966). The dataset contains information from a physician survey in four Illinois communities on the adoption of a new pharmaceutical drug (tetracycline). The physicians were also asked to name three doctors who acted as a friend, as an adviser, and as discussion partner. This dataset was later used by other studies investigating diffusion processes (e.g., Strang and Tuma 1993; Valente 1996; Van den Bulte and Lilien 2001).

One way in which marketing scholars have modeled the diffusion pattern of a new treatment is the Bass model (Bass 1969). The standard Bass model is represented by a differential equation where the change in the number of adopters n_t is increasing in the total number of past adopters N_t:

$$n_t = \frac{\mathrm{d}N_t}{\mathrm{d}t} = p(M - N_t) + \frac{q}{M} N_t (M - N_t), \tag{13.11}$$

where M is the total number of eventual adopters or the market potential. The parameters q and p represent the internal and external influence, respectively. The internal influence parameter captures the interaction of adopters (i.e., physicians who have already adopted the treatment) and potential adopters (i.e., physicians who have not yet adopted at present and are expected to adopt the treatment in the future). The internal influence may reflect the social contagion as well as the adoption of common treatment standards among physicians. The external influence parameter captures all influence outside the social system of the physicians, for instance, the promotion of a treatment by firms (i.e., detailing, advertising, sampling, etc.).

The original Bass model has subsequently been extended to accommodate several market complexities. In the context of the life sciences industry, marketing scholars have extended the original Bass model to incorporate both trial and repeat prescriptions in a treatment diffusion model (Hahn et al. 1994; Kolsarici and Vakratsas 2010; Rao and Yamada 1988). The common approach is to model the repeat prescriptions as a constant (Shankar et al. 1998) or as a dynamic (Kolsarici and Vakratsas 2010) percentage of cumulative trials of the treatment up to that point. Other extensions of the Bass model include the marketing actions of life sciences firms. In particular, such models extend the original Bass model to capture the influence of detailing and journal advertising on new drug trials (Hahn et al. 1994; Kolsarici and Vakratsas 2010; Rao and Yamada 1988). Kolsarici and Vakratsas (2010) consider also the influence of free samples and DTCA. Such extensions include both own and competitive marketing actions. The effects of these marketing actions are commonly specified to have a delayed effect on the diffusion of drugs (e.g., Hahn et al. 1994) and diminishing returns to scale (e.g., Kolsarici and Vakratsas 2010). The marketing actions of life sciences firms may have also *dynamic* effect on prescriptions. For instance, Kolsarici and Vakratsas (2010) employ an augmented Kalman filter approach in a Bass diffusion model to estimate the dynamic effects of marketing actions on prescriptions.

The Bass model has also been extended to incorporate heterogeneity among new treatment adopters. For instance, Vakratsas and Kolsarici (2008) distinguish between early market and main market adopters of a new pharmaceutical drug. An early market is created due to persistent and severe symptoms suffered by a class of patients. This market may be formed even before the introduction of the drug due to well-defined diagnosed needs of patients forming this market. The main market corresponds to patients with milder conditions whose adoption may have been triggered by the launch itself. This distinction between two segments of adopters is similar to the dual-market notion suggested for technological markets (e.g., Goldenberg et al. 2002).

Another approach to characterizing market penetration of new treatments is a logistic model presented by Van den Bulte (2000). Desiraju et al. (2004) use this model to study the effect of market characteristics on the maximum penetration potential and diffusion speed of a new category of prescription drugs in both developing and developed countries. The logistic model specifies the growth rate of a new treatment in country i at time t as a function of cumulative adopters at the start of period t, denoted as N_{it}, and the population of country i, M_{it}:

$$\frac{\mathrm{d}N_{it}}{\mathrm{d}t} = \frac{\theta_i}{\alpha_i M_{it}} N_{it}(\alpha_i M_{it} - X_{it}), \tag{13.12}$$

where α_i is the maximum penetration fraction of country i, and θ_i represents the diffusion speed. As Van den Bulte (2000) notes, the time to go from a penetration level P_1 to penetration P_2 equals $\theta_i^{-1} ln\{(1-P_1)P_2/(1-P_2)P_1\}$ and is inversely related to the diffusion speed.

Models that use physician-level data commonly characterize the diffusion of treatments via a discrete-time hazard model (Iyengar et al. 2011, 2015; Van den Bulte and Lilien 2001). In such a model, the physician adoption of a treatment is expressed as in Eq. 13.9, where the covariates include factors that influence physician's decision to adopt a treatment such as a physician's socio-demographic characteristics, type of physician (e.g., primary care or a specialist), contagion measures, firm promotional activities, seasonal and location factors, and the like.

13.2.6 Sales Growth Models

Sales (growth) models represent the amount of a drug's biologically active ingredient or the number of units of the drug sold in a given market or region. Sales differ from adoption in that the former captures repeat purchases as well as (possibly) volume sold. Therefore, the distinction between diffusion models and sales growth models is important. The estimation of the Bass diffusion model is shown to create an estimation bias when sales data are used instead of adoption data (e.g., Van den Bulte and Stremersch 2004). While repurchase decisions are relatively rare in durable markets, the repurchase rate is very high in pharmaceutical markets. Hence, marketing scholars have developed models to capture sales rather than adoption. Life sciences companies can use sales models to make sales forecasts once the product is available in the market. Alternatively, the pattern of sales growth of a similar molecule can be used to predict the sales pattern of a new molecule that the firm intends to launch.

Sales data are typically obtained from wholesalers. Such data can be broken down to the region in which the wholesaler delivers to pharmacies (e.g., states), but they do not break down to the level of the prescribing physician (thus, it is aggregate-level data). Alternatively, sales data can also be acquired from the pharmacies. A different source of data is the life sciences firm itself, which possesses data on the past sales of its own therapies. However, in such cases, the firm sometimes does not possess sales data of competitor brands. Finally, sales data can be obtained from providers, such as IMS Health and Wolters Kluwer. For the own brands, such data can be cross-validated with the firm's own data on its shipments to wholesalers.

Models that capture aggregate sales of an individual brand are typically represented in the following form:

$$sales_{it} = f(sales_{it-1}, g_1(D_{it}), g_2(JA_{it}), g_3(Samples_{it}), g_4(Mail_{it}), g_5(DTCA_{it}), X_{it}), \tag{13.13}$$

where D_{it}, JA_{it}, $Samples_{it}$, $Mail_{it}$, and $DTCA_{it}$ capture the firm's expenditures on detailing, journal advertising, samples, direct mail, and DTCA, respectively.

Commonly, $f(\cdot)$ is specified in a linear or logarithmic form, whereas $g(\cdot)$ takes on a logarithmic or square-root form. When $f(\cdot)$ and $g(\cdot)$ are specified in logarithmic form, an underlying multiplicative demand model is implied. This enables the researcher to interpret the coefficients of firm's direct-to-physician promotions and DTCA as elasticities. Furthermore, lagged sales are included because they influence a drug's current sales in three ways. First, once a physician starts prescribing a new drug, she may trigger adoption by her peers due to contagion effects. Second, physicians who currently prescribe a certain drug are likely to maintain similar behavior, also referred to as habit persistence (Janakiraman et al. 2008). Third, in the case of chronic diseases, patients take the same drug for a long time, receiving refill prescriptions repeatedly.

Parsons and Vanden Abeele (1981) use the model presented in Eq. 13.13 with $f(\cdot)$ and $g(\cdot)$ specified in logarithmic form to examine the effectiveness of promotional tools on the sales of a single pharmaceutical manufacturer. Other scholars have extended the simple sales model as specified in Eq. 13.13 to competitive settings. In such a model, the sales of an individual brand are specified to be a function of both its own and the competitors' marketing mix, the cumulative sales of competitors, the perceived quality of the brand, as well as of the order of entry in the category (e.g., Shankar et al. 1999). To capture heterogeneity in parameters for different brands, one can specify the parameters to vary depending on whether the brand is a pioneer or whether it entered at the growth or mature stage of its life cycle (Shankar et al. 1999).

Researchers modeling treatment sales also tackle the underlying assumptions with regard the trial-repeat purchase process of treatments. One option is via semi-parametric methods, which do not entail any assumption on the purchase process. For instance, Stremersch and Lemmens (2009) adopt regression splines to model the role of regulatory regimes on the sales of new drugs across the globe. In contrast to other specifications, splines do not impose any assumption (linear, quadratic, or cubic) about the interactions of explanatory variables with time. Stremersch and Lemmens (2009) specify sales of drug i belonging to category c sold in country j at time t as follows:

$$sales_{icjt} = \beta_{rt} REG_{rjt} + \beta_{pt} X_{picjt} + \varepsilon_{icjt}, \tag{13.14}$$

where REG_{rjt} represents r country regulations in country j, and X_{picjt} represents p other variables such as other country, drug and category characteristics. The general idea behind splines is that any smoothly varying function can be represented as a linear combination of basis functions. These functions are usually polynomial functions of low degree (e.g., linear in case of linear splines, cubic in case of cubic splines). Hence, the time-varying coefficients of any explanatory variables can be represented as:

$$\beta_t = \beta_0 + \beta_1 t + \sum_{k=1}^{K} u_k^{\beta} (t - \kappa_k)_+ \tag{13.15}$$

where K is the number of linear spline basis functions, and κ_k is the truncation points where the broken lines are tied together.

Researchers have used other suitable methods to model time-varying coefficients. For instance, Osinga et al. (2010) employ state space models to capture persistent (i.e., enduring influence) and transient (i.e., short-lived influence) effects of the marketing actions of life sciences firms on sales. Such models provide managers with clear guidelines to allocate their marketing expenditures over time.

Other works focus on market shares rather than on sales as dependent variable. Montgomery and Silk (1972) model the market share of a therapeutic drug, which can be calculated with reference to a particular therapeutic class of drugs. The authors adopt a Koyck distributed lag model for market share that is specified as:

$$\begin{aligned}\ln(MS_t) = \alpha_0 + \textstyle\sum_{j=0}^{J} a_j \ln JA_{t-j} + \sum_{k=0}^{K} b_k Samples_{t-k} + \sum_{i=0}^{I} c_i Mail_{t-i} \\ + \gamma \textstyle\sum_{l=0}^{\infty} \lambda^l \ln JA_{t-J-l} + \delta \sum_{l=0}^{\infty} \lambda^l Samples_{t-K-l} + \varphi \sum_{l=0}^{\infty} \lambda^l Mail_{t-I-l} + \varepsilon_t,\end{aligned} \tag{13.16}$$

where $0 < \lambda < 1$. In this model, *I, J,* and *K* are the number of lags for each promotion tool of the firm, i.e., direct mail (denoted as *Mail*), samples (*Samples*), and journal advertising (*JA*), respectively. After the lag periods *I, J,* and *K*, the effects of promotional tools are specified to decline geometrically with the same decay rate $(1 - \lambda)$. Provided that $\gamma \neq \delta \neq \varphi$, the decay rate and not the magnitudes of the effects are specified to be identical. In addition, the specific lags allow each promotional tool to exhibit an individual decay rate up to the period in which the geometric decay sets in. This makes the assumption of identical decay rate considerably less restrictive because one may expect significant differences across promotional tools to show up in the first few periods.

A different approach is to decompose the sales of each brand in a given region and at a given time period as the product of category sales and the share of that brand. For instance, Chintagunta and Desariju (2005) specify category sales $(Csales_t)$ as a function of category-level marketing activities:

$$\ln(Csales_t) = \omega + \nu \ln(CP_t) + \rho \ln(CD_t) + \varphi \ln(COME_t) + \textstyle\sum_{s=1}^{3} \kappa_s SD_{st} + \tau_1 t + \tau_2 t^2 + e_t, \tag{13.17}$$

where CP_t, CD_t, and $COME_t$ are share-weighted price of the individual brands in the category, category detailing, and other marketing expenditures. SD_t captures the seasonality effects, whereas linear and nonlinear time trends (t and t^2) intend to capture possible diffusion effects resulting from the introduction of new brands and category growth. The conditional share of each brand in period t is specified by a mixed logit formulation, where the parameters of price, detailing, other marketing expenditures, and seasonality effects are not fixed but are drawn from some distribution. There are three advantages of adopting a mixed logit formulation. First, with such a model one can account for cross-sectional variations resulting from aggregation across heterogeneous decision makers. Second, as noted above, the mixed logit model does not suffer from the Independence of Irrelevant Alternative (IIA) property. Third, the predicted shares from mixed logit can be shown to be arbitrarily close to those obtained from more complex models.

13.2.7 Launch Models

Life sciences companies need to make optimal decisions on when and where to launch a new therapy. Global launch-time decisions are among the most critical that life sciences firms need to make. A new drug is never launched in all countries at once (often dubbed as a "sprinkler launch"); rather it is launched sequentially ("waterfall strategy".)

Scholars studying international launch decisions in the life sciences industry typically obtain data on international launch timing from providers such as IMS Health (e.g., IMS Lifecycle New Product Focus database) and PJB Publications (e.g., Pharmaprojects database). The Pharmaprojects database includes a drug's chemical and brand names, the name and nationality of the firm that developed it, the drug's status (in clinical trials, registered, or launched) in several countries and the year of launch, if applicable. Alternatively, one can obtain information on treatment launches from the World Health Organization (WHO) or, as IMS does, from various sources such as regulatory approvals, announcements by life sciences firms, local media reports, etc. Because of highly regulated markets, one can also consider the launch time as the first period in which positive sales occur.

Launch decisions are subject to strict regulations by regulatory agencies such as the FDA (Food and Drug Administration) in the US or the EMA (European Medicines Agency) in Europe. There are several types of regulations that are salient to marketers, including ex-manufacturer price regulation, cross-country and therapeutic reference pricing, regulations on marketing efforts to physicians, regulation on DTCA, patent protection and post patent-expiry regulations. Other important regulations include profit and pharmaco-economic evidence regulations, regulations on physician prescription budgets and on patient copayment. Such regulations have substantial influence on firms because these local regulatory controls are incorporated into a firm's global launch plans. Data on various regulations across countries are inventoried from multiple sources. One of the main sources is URCH Publishing, an independent information provider for the life sciences industry. Another source is the OECD, which publishes reports on health regulations among its member states (e.g., Jacobzone 2000). A third source includes data on patent protection, which one can obtain from the Pipeline Scope database available from GenericsWeb (recently acquired by IMS Health).

Researchers studying the global launch of therapies typically focus on either the probability of the therapy's launch in a given country or on the launch window. The latter is defined as the difference in time between the first worldwide launch and the subsequent launch in a specific country (Verniers et al. 2011). The launch probability or the launch window is typically specified to depend on time-varying and time-invariant covariates such as therapy characteristics (e.g., therapy's age, importance, etc.), firm characteristics (e.g., the firm's total number of drugs, international experience, the firm's number of drugs in the country, etc.), competition (e.g., the number of similar drugs in the market), launch price, home country (i.e., whether the firm has its headquarters in the country), and country characteristics such as income, income distribution, size of population, national culture, regulatory regimes, etc.

There are several significant methodological challenges for researchers studying international launch decisions of life sciences firms. One of the main difficulties is the right censoring problem. Censoring occurs for the therapy-country combinations for which a launch is not observed at the end of the observation window. To correct for right censoring, studies that focus on the probability of a therapy launch typically use a semi-parametric Cox proportional hazard model (Danzon et al. 2005). Others account for right censoring by using discrete-time hazard models. For instance, Kyle (2006, 2007) uses a logit transformation of probability of a therapy's launch $P(t)$:

$$log\left(\frac{P(t)}{1-P(t)}\right)=X_{ij}(t). \tag{13.18}$$

The main advantages of discrete-time hazard models are that they lead to a simple model formulation, account for right censoring, and they can allow for time-varying covariates. Scholars who focus on international launch windows have adopted a Tobit I model to account for right censoring (Verniers et al. 2011).

Another challenge is rooted in the firms' simultaneous determination of different launch decisions. Both launch timing and launch prices are subject to negotiations between the firm and the regulator. Both the firm and the regulator may use the launch time as an instrument to affect the launch price level and use it as a tool to affect the launch time. To account for the endogeneity of these firms' launch decisions, research on the international launch window and pricing decisions of pharmaceutical drugs simultaneously estimate launch window and price equations by adopting the three-stage least squares procedure (Verniers et al. 2011).

Models studying the global launch of therapies typically account for differences in regulatory regimes across countries. For instance, research shows that several forms of regulations, such as ex-manufacturer price controls, profit caps, therapeutic reference pricing, and pharmaco-economic evidence, delay the launch time of new drugs (Cockburn et al. 2014; Danzon et al. 2005; Kyle 2007; Lanjouw 2005; Verniers et al. 2011). The story is more nuanced with respect to the influence of patent protection: stronger patent protection accelerates, rather than delays, the launch timing of a new drug because it protects drug companies from competition (Cockburn et al. 2014; Kyle and Qian 2014; Verniers et al. 2011). However, this raises another source of endogeneity. For instance, patent and price control regimes are outcomes of a political process, which engenders a concern about endogeneity (Cockburn et al. 2014; Kyle and Qian 2014). The most likely source of endogeneity is the unobserved heterogeneity in political institutions across countries, which affects both the choice of policy regime and the timing of new drug launches. For example, firms are more likely to lobby for strong patent protection where entry is more profitable; this may cause an over-estimation of the effect of patent protection on the timing of drug launches. To overcome such concerns, researchers typically use an instrumental-variable approach. The instruments need to be correlated with the regulatory policy choice but should not directly influence the firms' launch time decisions. For instance, Cockburn et al. (2014) use instrumental variables based on political, legal, and demographic characteristics of a country such as the degree to which voting rights within the political structure constrains policy change, ethno-linguistic diversity, legal origin, and the number of bilateral trade agreements signed by a country.

13.3 Select Findings on the Role of Marketing

13.3.1 Select Findings on the Role of Direct-to-Physician Promotion

Given the physician's key role in therapy choice as a prescriber of such therapy, marketing modelers have devoted a lot of attention to one specific driver of physician's prescription behavior, namely the promotion of drugs to the physician through detailing (i.e., sales visits), sampling, medical journal advertising or conferences/meetings.

Life sciences firms typically spend the lion's share of their promotional budget on promotional efforts to doctors (Fig. 13.3). This is especially true outside the US and New Zealand where DTCA is forbidden by law (e.g., EU) or severely curtailed (e.g., Canada). Therefore, the effect of detailing on doctors' prescription behavior has received most attention. The data on detailing can come from at least two sources: the firm's own CRM system, in which case the firm lacks information on competitive detailing; or the firm secures data from an external data provider such as IMS and ImpactRx. Given that external data providers have physician panels in many countries, these data can be either at the aggregate brand level or at the individual physician level. In such panels, doctors are asked to record the detailing visits they received, the drugs that were discussed, and the order in which they were

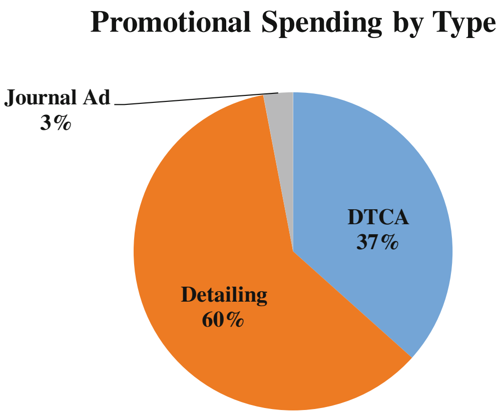


Fig. 13.3 Promotional spending share of detailing, DTCA, and journal advertising for prescription drugs in US, 2011

discussed. Depending on the respective panel, additional information is also sometimes inventoried such as the drug attributes that were discussed in a sales conversation or the duration of the sales conversation. Physician promotion data can be then cross-linked to demand data at either an aggregate level or at an individual physician level.

Existing research has found quite a bit of variation on the impact of detailing on physician prescriptions. Some studies find that detailing has a strong and positive impact on prescriptions written (e.g., Azoulay 2002; Berndt et al. 1995; Chintagunta and Desiraju 2005; Fischer and Albers 2010; Gönül et al. 2001; Liu et al. 2015; Manchanda and Chintagunta 2004). Other studies show that detailing has, at best, only a modest impact on prescriptions (e.g., Datta and Dave 2013; Mizik and Jacobson 2004; Parsons and Vanden Abeele 1981; Rosenthal et al. 2003; Stremersch et al. 2013).

Venkataraman and Stremersch (2007) show that detailing and meetings are more effective for drugs that are highly effective or have many side effects. Because there is sound scientific evidence to back up the detailing visit, when the firm promotes a more effective instead of a less effective drug, the firm is more able to lower physician uncertainty about the drug and increase physicians' affect toward it. With respect to the number of side effects, a drug with many side effects engenders a high level of physician uncertainty that a firm can effectively reduce through its marketing efforts. A drug with fewer side effects creates a low level of physician uncertainty, thus reducing the need for uncertainty reduction through the firms' marketing efforts.

Another reason for the variation in findings on detailing effectiveness can be the differential role of detailing at different stages of the treatment life cycle (Narayanan et al. 2005). In the early phases, marketing communications (i.e., detailing) have primarily an indirect impact by helping the physicians reduce uncertainty about the treatment. However, as time goes by, physicians learn more about the treatment from patient feedbacks and promotional efforts of firms. Therefore, the indirect effect of detailing becomes increasingly smaller as its effect becomes more direct (i.e., the impact on preferences through goodwill accumulation dominates).

In the presence of learning, physicians may also strategically substitute sources of information. Physicians typically do not choose to be addressed by promotional activities of firms, but they can choose whether or not to experiment with a new treatment to learn from patient feedback. However, such experimentation is costly compared to information obtained through firms' promotional activities (e.g., detailing). Forward-looking physicians may then be less likely to experiment if they anticipate receiving free information through detailing (Chintagunta et al. 2012). Therefore, detailing engenders two opposing forces. First, treatment adoption accelerates as physicians become informed. Second, if physicians expect more future detailing, treatment adoption slows down as the physicians reduce experimentation to learn from patient feedback and instead wait to obtain information from future detailing. To minimize the latter force, Chintagunta et al. (2012) suggest that firms should avoid announcing increases in their detailing activities.

The enormous detailing budgets of firms have prompted wide-ranging legislative proposals in different countries to regulate and curtail detailing activities. For instance, the United Kingdom limits the sales promotion expenditures of pharmaceutical firms to a certain proportion of their overall profits. Liu et al. (2015) find that regulatory policies that limit the amount of detailing may reduce the detailing of all firms but have differential effects on the firms' market shares and profits. Such restrictions benefit firms that have the weakest detailing effectiveness, but have a reverse effect on firms with the strongest detailing effectiveness.

With respect to medical journal advertising, Van den Bulte and Lilien (2001) show that several studies examining the diffusion of tetracycline may have confounded social contagion with such marketing effects. In particular, they show that when controlling for medical journal advertising in diffusion models, contagion may disappear. This underscores the importance of controlling for such potential confounds when studying social contagion in new treatment diffusion.

Scholars have derived important contingency factors that determine the effectiveness of medical journal advertising on primary demand, i.e., the market size of the category (Fischer and Albers 2010). In particular, journal advertising is less effective for promoting drugs in expensive categories but is more effective for chronic care compared to acute care categories.

Compared to detailing and medical journal advertising, sampling is an understudied topic in marketing. Prior work suggests that sampling has a positive influence on physician prescription decisions (e.g., Manchanda et al. 2004; Mizik and Jacobson 2004) but with diminishing returns (Gönül et al. 2001). Montoya et al. (2010) find that detailing is most effective as a physician acquisition tool, whereas sampling is most effective as a retention tool.

Across different promotion tools directed toward physicians, existing work suggests that journal advertising has a more pronounced effect for the advertised drug than do positive scientific information published in medical journals (Azoulay 2002), samples, and direct mail (Montgomery and Silk 1972). The effectiveness of medical journal advertising is typically lower than that of detailing (Berndt et al. 1995; Fischer and Albers 2010; Kremer et al. 2008; Kolsarici and Vakratsas 2010). In contrast, medical journal advertising is more effective than DTCA for most disease categories (Berndt et al. 1995; Fischer and Albers 2010; Kremer et al. 2008).

Marketing scholars have also conducted meta-analyses to formulate empirical generalizations (Albers et al. 2010; Kremer et al. 2008; Sridhar et al. 2014). The idea behind a meta-analytical work is to draw out a generalized quantitative estimate of the effectiveness of a promotion tool from the body of prior individual empirical analyses. Because it is unit-free and easily interpretable, such meta-analytical work often favors an elasticity measure that refers to the ratio of the percentage change in output (i.e., sales or prescriptions) to the corresponding change in promotion efforts. The key decisions of meta-analytical work pertain to the scope of the database, coding, and estimation method. Scope of database refers to the inclusion criteria (e.g., year and journal of publication, measurements to be reported unambiguously or derivable from the estimated coefficients, etc.) based on which past studies are selected into the database. Coding refers to the coding scheme and procedure for how all observations in the database are coded on selected variables (e.g., type of promotional instrument, treatment life-cycle stage, order of entry, etc.). Estimation method refers to the estimation methods used in the studies that are being meta-analyzed.

Some of the empirical generalizations resulting from the meta-analyses are:

- Detailing elasticity is positive, albeit its magnitude varies across studies: 0.21 reported in Sridhar et al. (2014), 0.245 in Albers et al. (2010), to 0.326 in Kremer et al. (2008).
- Detailing elasticity is lower in the pharmaceutical industry (0.245) than the mean of personal selling elasticity across all industries (0.34) (Albers et al. 2010).
- Detailing elasticity is the highest (0.326), followed by direct-to-physician advertising (0.123), DTCA (0.073) and other direct-to-physician instruments (0.062) (Kremer et al. 2008).
- Detailing elasticity is higher for new drugs (0.41 and 0.23 in Europe and the US, respectively) than for mature drugs (0.17 and 0.14, respectively) (Sridhar et al. 2014).
- Detailing elasticity is higher in Europe than in the US (Sridhar et al. 2014).

13.3.2 Select Findings on the Role of Direct-to-Consumer Advertising

DTCA is one of the most contentious topics in the healthcare industry (Wilkes et al. 2000). DTCA by life sciences firms is strictly regulated worldwide. It is legally allowed in New Zealand and the US, and to some extent, in Canada (see Table 13.3 for an overview of worldwide DTCA regulations). Proponents of DTCA argue that DTCA informs the society about the available treatments and thus it triggers the appropriate use of drugs. Critics argue that the information in advertising might be biased or misleading. Without sufficient information about the treatments, DTCA may only persuade rather than inform. A crucial concern about DTCA is that it may increase the demand for more expensive treatments. In addition, the huge costs associated with it may contribute to higher prices of treatments that payers and/or patients will eventually have to bear.

There are several data sources on DTCA. For instance, Mintzes et al. (2003) survey patients and physicians and compare their responses across two cities—Sacramento and Vancouver—that have differing DTCA regulations. Studies on DTCA that inventory data in controlled experiments either in the laboratory or on the field are rare (for an exception, see Amaldoss and He 2009, who conduct an experiment with business school students). More commonly, in observational designs, DTCA data are secured from providers such as Kantar Media (formerly TNS Media), AC Nielsen, Ipsos or Media Vest Global. Kantar tracks advertising

Table 13.3 DTCA Regulations (adapted from Vakratsas and Kolsarici 2014)

		Help-seeking advertisements	Reminder advertisements	Product claim advertisements
Regulatory content	Therapeutic category information	√	X	√
	Symptoms	√	X	√
	Health claims	√	X	√
	Risk information	NR	NR	R
	Price information	X	√	√
	Brand name	X	√	√
	Dosage information	X	√	R
	Direct to a physician	√	NR	R
Regulatory context	US and New Zealand	√	√	√
	Canada	√	√	X
	Australia and EU	X	X	X

NR Not Required, *R* Required, √: allowed, *X* forbidden
Help-seeking or *disease-oriented* advertisements do not mention a specific brand but discuss symptoms, conditions, and suggest that patients ask their doctors about the treatment. *Reminder advertisements* mention the drug name but refrain from health claims or statements about the use of the product. *Product claim advertisements* discuss both drug and disease information with a fair representation of benefits and risks of the drug

expenditures in national media (for example, network TV and national newspapers) as well as in local ones (for example, spot TV and local newspapers) at the Designated Media Area (DMA) level (a DMA is a region where the population receives similar TV and radio station offerings). One can also gather DTCA information from news archives, such as the Vanderbilt Television news archives, to ascertain when a particular treatment was advertised. Alternatively, scholars obtain DTCA data directly from life sciences firm (e.g., Kolsarici and Vakratsas 2010).

Academic research on the effect of DTCA on physician prescriptions shows mixed findings. Some studies suggest that the effect of DTCA on prescriptions is positive and large (e.g., Iizuka and Jin 2005; Ling et al. 2002). Other studies suggest that DTCA has a very limited effect or no effect at all on brand-level prescriptions (e.g., Calfee et al. 2002; Iizuka and Jin 2007; Law et al. 2008; Narayanan et al. 2005; Stremersch et al. 2013). Some even suggest that DTCA may have a negative effect in certain disease categories (Kremer et al. 2008). Other studies find that DTCA of prescriptions drugs has a positive spillover effect on same-brand over-the-counter (OTC) version of the drugs (Ling et al. 2002). For OTC drugs, recent work suggests that comparative advertising damages the targeted competitor's profits more than benefits the advertiser (Anderson et al. 2016).

In line with the trend of increasing patient involvement and empowerment in medical decision making (Camacho et al. 2010), a hotly contested topic is the extent to which DTCA triggers patient drug requests by brand name, which in turn may lead to more prescriptions. Patient requests and specifically, requests for a drug by brand name are pervasive in patient-physician relationships. Existing evidence shows that patients request for a specific medication in about 10% of all office visits (Paterniti et al. 2010). About a third of all patients in France, Germany, UK and the US report to have made a request for a drug by brand name at a certain point (Calabro 2003).

In turn, patient requests are commonly believed to positively influence the number of prescriptions for the requested drug. Physician accommodation of patient requests leads to more positive patient evaluation of care (Kravitz et al. 2002), whereas request denials lower patient's satisfaction with, trust in, and commitment toward the physician (Shah et al. 2006). In a recent survey, approximately four in ten of nearly 1,900 physicians across specialties report that they sometimes or often accommodate patient requests of brand-name drugs even when a generic one is available (Campbell et al. 2013). However, according to the FDA, the daily costs of drugs can fall on average by 14–16% if patients use a generic drug instead of its brand-name version.

Recent work by marketing scholars examining the full chain linking DTCA to requests and the latter to prescriptions (Stremersch et al. 2013) shows that although the effect of requests on prescriptions is significantly positive, the mean effect of DTCA on patient requests might be negative. Stremersch et al. (2013) point to the long list of side effects enumerated at the end of ads enforced by FDA regulations, which may trigger higher patient salience about such side effects and may cause the ads to backfire. The authors also find that large heterogeneity exists across these two effects. They show that specialists receive a larger number of requests than primary care physicians but that the former accommodate such requests to a lesser extent. In addition, they find that the socio-demographic characteristics of a physician's area of practice moderate the effects of DTCA on requests and of requests on prescriptions. For example, physicians practicing in areas with a higher proportion of minorities (i.e., blacks and Hispanics) receive more requests that are less triggered by DTCA and less frequently accommodated than physicians practicing in areas with a lower proportion of minorities.

Other research suggests that the effect of DTCA on requests may also vary across brands (Liu and Gupta 2011). With respect to the effect of requests on physicians' prescriptions, prior work shows also that drug characteristics moderate the effect of requests on brand prescriptions (Venkataraman and Stremersch 2007). Patient requests have a more positive effect on prescriptions for relatively more effective drugs compared to prescriptions for less effective drugs. Patient requests have a more negative effect on prescriptions for drugs with more side effects than on prescriptions for drugs with fewer side effects.

As already discussed, proponents of DTCA emphasize its informational role in educating potential patients. More informed patients may be able to understand their health conditions and are more likely to seek medical consultation with their physicians. Prior work indeed suggests that DTCA increases physician *visits* (e.g., Brekke and Kuhn 2006; Calfee et al. 2002; Liu and Gupta 2011). Furthermore, proponents of DTCA argue that DTCA may improve patient *adherence*. However, existing work in marketing suggests that DTCA has a small or no effect at all on patient adherence (e.g., Bowman et al. 2004; Wosinska 2005). Furthermore, Amaldoss and He (2009) suggest that DTCA for branded drugs can enable pharmaceutical companies to build goodwill among their consumers, which decreases consumer price sensitivity and increases profits. Osinga et al. (2011) suggest that the target audience of DTCA may not only be patients or physicians, but also shareholders. They find that shareholders value DTCA positively because it increases stock returns and lowers systematic risk. They also find that DTCA raises idiosyncratic risk that can be substantially mitigated or eliminated by using well-diversified portfolios.

13.3.3 Select Findings on the Role of Pricing

Some critics accuse life sciences firms of exploiting consumers through excessively high drug prices. However, R&D is approximately 17% of sales for the US-based pharmaceutical industry, compared to only 4% for other US industries (Danzon et al. 2015). Given the high R&D costs and the high risk involved in the development of a new drug, proponents of the current level of drug prices justify such prices by pointing at the desire to fuel the development of new drugs.

Typically, studies focus on ex-manufacturer drug prices, which is the price charged to wholesalers. Such data are ordinarily acquired from similar sources as sales data (including those of wholesalers and pharmacies): either directly from life sciences firms' data on its own therapies or from information providers such as IMS Health or Wolters Kluwer.

Physicians prescribe lower-cost drugs because they may want their patients to pay less or they may be incentivized to do so. For instance in the UK, all general practitioners receive additional funds if they maintain their prescription and other treatment costs below a certain threshold. Moreover, the UK Prescription Pricing Authority makes all general practitioners aware of their total drug spending with a comparison against the expenditures of similar physicians. However, Gönül et al. (2001) find that there could be at least a significant segment of physicians for whom price does not matter because their prescription behavior is driven by the severity of a patient's condition or by possible interactions with other types of medication taken by the patient, thus crowding out less critical price concerns. Gonzalez et al. (2008) examine physicians' reaction to the introduction of a new, cheaper generic subcategory of antidepressants. They show that generic entry leads physicians to decrease the prescription of the branded drug bioequivalent to the generics, but benefits other non-bioequivalent branded drugs as physicians sensitive to the detailing activities of competitors increase their prescriptions of these other branded alternatives. However, price-sensitive physicians increase their prescriptions of the new generics to the detriment of all branded alternatives. The overall effect of the generic entry is a slight decrease in the prescriptions of the patent-losing molecule.

Existing research suggests that physicians' price sensitivity depends on direct-to-physician promotions of life sciences firms. However, the studies find mixed results. Some studies find that firms' promotional efforts play a persuasive role and decrease physicians' price sensitivity (e.g., Rizzo 1999; Windmeijer et al. 2006). Others suggest that firms' promotional efforts have predominantly an informational role and raise physicians' price sensitivity (e.g., Narayanan et al. 2004).

Prior research has found that treatments that offer larger therapeutic advancement generally have higher launch prices than those that offer smaller therapeutic advancement (Ekelund and Persson 2003; Lu and Comanor 1998). Further, the number of branded substitutes has a negative effect on the launch price level of a new treatment in the category (Lu and Comanor 1998). However, competition seems to matter less in regulated countries. For instance, in Sweden, scholars find no effect of the presence of branded substitutes on launch prices (Ekelund and Persson 2003). Treatments indicated for acute conditions have higher launch prices than treatments indicated for chronic conditions (Lu and Comanor 1998). The reasons for setting lower launch prices for treatments for chronic illnesses are: (1) chronic conditions are more pervasive among the elderly who are more price-sensitive, and (2) repeat purchases of treatments for chronic ailments are more likely than ones for acute conditions.

Launch price is often an outcome of negotiations between the life sciences firm and the government or the insurance firm. In such negotiations, both parties may use the launch time to affect the agreed-on price. Prior work finds that launch time has an inverted U-shape effect on the launch price of treatments such that the launch price is highest at moderate launch times (Verniers et al. 2011). At very early launch times, the firm is likely to more easily accept a lower launch price because the treatment will enjoy a longer time under patent protection. At very late launch times, the health regulator's willingness to pay for the treatment may decrease because more information about the treatment spreads and the treatment loses its novelty. As its treatment nears patent expiration globally, the firm itself may prepare for generic competition. Therefore, at very late launch times both parties align more easily on a relatively low launch price.

Prior research has investigated treatment-price evolution over the product's life cycle (e.g., Bhattacharya and Vogt 2003; Lu and Comanor 1998). On the one hand, treatments that offer significant therapeutic advancement are generally introduced with a price-skimming strategy where prices start at a high introductory level and then decline moderately over time (Lu and Comanor 1998). On the other hand, treatments that provide only a small therapeutic gain (i.e., imitative treatments) are typically launched with a penetration strategy, where prices start at a low introductory level and then increase over time. In contrast to the Lu and Comanor (1998) results from the US market, Ekelund and Persson (2003) find in the Swedish market higher relative introduction prices and price-skimming strategies for all classes of therapeutic innovation. These differences in price dynamics between the US and Sweden may be rooted in the different regulatory regimes in these two countries. Unlike the US, Sweden has a price-cap regulation under which price increases are generally ruled out. Therefore, the pattern of high relative launch prices and declining prices is likely to occur because the regulator compensates the life sciences firms by allowing a relatively high introductory price.

Generic drugs enter the market after patent expiration. They are required to have the same active ingredients, strength, safety, quality and dosage as the original drug, while a small deviation in efficacy is allowed. Of course, the generic firm does not need to reproduce the branded firm's full assortment. Prior research suggests two possible price reaction strategies by brand-name firms to the entry of generics. On the one hand, firms of branded drugs may lower the prices to defend their market share. Consistently, Caves et al. (1991) find that prices of branded drugs decrease after generic entry. On the other hand, firms of brand-name drugs have significant customer loyalty advantage vis-à-vis their generic entrants (e.g., Grabowski and Vernon 1992). Therefore, brand-loyal customers may be willing to pay more for branded drugs. In line with this rationale, other scholars find that prices of branded drugs may actually increase after generic entry (e.g., Bhattacharya and Vogt 2003; Frank and Salkever 1997).

Drug prices are also subject to competition from parallel imports. Parallel imports are imports of a patented or trademarked product from a country where it is already launched (WHO). Parallel imports often take place when there are price differences of the same product, either a brand-name or a generic drug, in different markets. Existing research suggests that parallel imports lower the prices of brand-name drugs (Duso et al. 2014; Ganslandt and Maskus 2004). In contrast, parallel imports do not seem to influence the prices of generic drugs (Duso et al. 2014). Furthermore, parallel imports may affect the impact of price cap regulation by altering the bargaining power between the therapy producer and the distributor. For instance, Brekke et al. (2015) find that stricter price cap regulation lowers drug prices, but the effect is weaker for drugs with parallel imports.

13.3.4 Select Findings on Product Usage Adherence

Adherence to prescriptions not only protects the patient from unnecessary disease complications, but it also reduces the risk of hospitalization and even death (PhRMA 2015a). For life sciences firms, nonadherence to medical prescriptions leads to lost sales. A recent study estimated that the global pharmaceutical industry loses around hundreds of billions of US dollars in revenue annually due to medication nonadherence (Forissier and Firlik 2012). In addition, drugs may perform worse than expected when patients do not adhere to the drug regimen.

A common approach to obtain data on patient adherence is by conducting a primary survey asking patients to report their adherence behavior (e.g., Camacho et al. 2014; Dellande et al. 2004; Prigge et al. 2015; Williams et al. 1998). The other common approach is to use secondary data. Studies typically use prescription claims that disentangle new purchases from refills to infer patient adherence (e.g., Bowman et al. 2004; Wosinska 2005). These prescription claims are from patient-level panels that are owned by either market research companies, such as

Ipsos and Truven Health, or insurance companies. Alternatively, researchers secure patient-adherence data from clinical trials (see e.g., Lamiraud et al. 2007) or from health organizations that initiate treatment programs such as weight-control clinics or alcohol-abuse outpatient treatment programs (see e.g., Dellande et al. 2004; Lien et al. 2010). Less common is the use of an experimental approach to obtain information on patient adherence (see e.g., Kahn and Luce 2003).

An interesting issue arising from the increasing influence or *empowerment* of patients in treatment decisions (see Camacho et al. 2010) is the effect of patient empowerment on adherence. In a study of 11,735 patients in 17 countries, Camacho et al. (2014) find that patient-initiated informational empowerment, which occurs when the patient initiates a treatment-relevant dialogue, improves adherence. However, contrary to the beneficial attributions to patient empowerment in prior literature, Camacho et al. (2014) show that physician-initiated informational empowerment, which occurs when the doctor proactively exchanges treatment-relevant information with the patient, improves the quality of communication but may result in cognitive and emotional overload and impair informational processing on the patient's part, leading to higher nonadherence. The study also finds that decisional empowerment, which occurs when the physician leaves the final decision to the patient, increases nonadherence.

Several scholars find other drivers of patient adherence. Specifically with respect to patient-physician relationship related factors, the attitudinal homophily (Dellande et al. 2004), the time that has elapsed from the medical intervention (last prescription) (Bowman et al. 2004), and the duration of the relationship between patient and provider (Schwartz et al. 2011) all affect patient adherence. With respect to therapy-related factors, research shows that the consequences of the treatment (Kahn and Luce 2003), experienced side effects (Lamiraud and Geoffard 2007), the complexity of medical regimen such as the dosage frequency (Bowman et al. 2004; Lamiraud and Geoffard 2007), salience of symptoms (Bowman et al. 2004), and treatment progress (Lien et al. 2010) all influence patient adherence. With respect to factors that are more under the direct control of life sciences firms, there is only scant research on the role of life sciences firms in patient adherence. Drug prices (Bowman et al. 2004), detailing (Donohue et al. 2004), and DTCA (Bowman et al. 2004; Donohue et al. 2004; Wosinska 2005) have small or no effect on patient adherence. An exception is the study by Ferguson et al. (1987) that finds that warning labels with information on the results of poor adherence substantially improve adherence.

13.4 Future Research Directions

13.4.1 Physician Networks and Medical Crowdsourcing

In today's digital world, physician-dedicated platforms, such as Sermo in the US with more than 382,000 members and China's DXY with more than 3.2 million members, are becoming increasingly popular. Such platforms offer access to ready-made communities of doctors and can be used to deliver treatment-specific messages. Physician-dedicated platforms potentially can provide access to hard-to-reach physicians (e.g., specialists) and also to online statistics that give insight into physicians' web habits and interests. A potential area of future research is an investigation of online promotional activities directed at such physician communities.

In these physician-dedicated platforms, physicians share ideas ranging from drug information to clinical practice. For instance, Sermo's recent survey of physicians from the UK, the US, France, Spain, Italy and Germany reveals that the majority of physicians (up to 87%) indicated that at least 20% of their patient cases were in the medical "grey zone", where physicians would benefit from feedback from their peers. These platforms provide crowdsourcing forums of medical knowledge for physicians who need the input of colleagues with broader experience to address the "grey zone". There are several cases of medical crowdsourcing activities resulting in life-saving diagnoses. However, despite the high relevance of medical crowdsourcing, there is little research on it, and models on physician learning commonly treat physicians as independent agents. Future research may yield important insights by revealing the drivers of physician engagement in such medical platforms. Another interesting avenue for future research might be to use data from crowdsourcing platforms to incorporate this kind of social learning in physician learning models.

13.4.2 Digital Opinion Leaders

The new digital landscape may also make individuals, who might not have been considered as "opinion leaders" in the past, to become highly active influencers online. Take for example Dr Mehmet Oz, a US-based cardiac surgeon, who also frequently appears in talk shows. Dr Oz has a large group of followers in social media, with approximately 4 million on Twitter alone (@droz, see Fig. 13.4a). Other examples include Dr Kevin Pho and Kelly Young. Dr Kevin Pho has relatively fewer followers on Twitter than Dr Mehmet Oz (@kevinmd, see Fig. 13.4b) but is far more engaging with his approximately 129,000 followers (32,700 tweets) than the latter. Kelly Young, a patient suffering a chronic disease, created a blog RAWarrior.com that has now become a hub of resources on rheumatoid arthritis and is just one example of such a large community influencing hundreds of thousands of people. Her online influence through RAWarrior includes thousands of Facebook, Youtube, and Twitter followers. Such *digital opinion leaders* exist also in social networks of patients such as PatientsLikeMe, DiabeticConnect, etc.

The emergence of digital opinion leaders poses for the life sciences firms the considerable challenge of identifying, engaging with, and activating such digital opinion leaders among physicians and patients. This task is especially complex with regard to engagement with digital opinion leaders among patients. In some countries, online communication with patients is not allowed, while in others it is highly

Fig. 13.4 **a** Physician Engagement in Social Media: Dr. Oz (September, 2015) **b** Physician engagement in social media: Dr. Pho (September, 2015)

regulated. In the US, the FDA requires a comparable balance of risks and benefits, regardless of character-space constraints. The benefit should be accurate, factual, and not misleading; the risk should be part of each communication, and the primary link should be provided for more complete information about risk. Furthermore, life sciences firms may voluntarily correct misinformation about their own products that was created or disseminated by an independent third party. A firm may choose to provide appropriate corrective information that meets FDA guidance, or they may provide a reputable source from which to obtain the correct information such as the contact information of the firm's medical affairs department.

Future research on digital opinion leaders is potentially highly impactful as many firms in the life sciences industry aim at identifying and partnering with digital opinion leaders, which is imperative in our digital era (Ghinn 2012). Future research may show how we can identify and target digital opinion leaders. In addition, little is known about the role of digital opinion leaders in the diffusion of new treatments, how this role varies depending on contingency factors (e.g., treatments' effectiveness, side effects, or life cycle stages), and how it compares to the role of traditional key opinion leaders. Future research may also provide a fine-grained picture of the mechanisms that digital opinion leaders use to influence the decisions of physicians and patients. For instance, on the physician side, one can examine outcome variables such as the number of prescriptions, generic prescription, adoption of new medical devices or of treatment procedures, or sample dispensing. On the patient side, one can focus on the influence of digital opinion leaders on the number of requests or on patient adherence to medical treatments.

13.4.3 eDetailing

Many life sciences companies are piloting virtual detailing (i.e., eDetailing). eDetailing can take various forms ranging from a remote live discussion with a sales representative to a purely scripted interaction with a series of interactive screens or via an interactive voice-response phone line. With the advent of eDetailing, firms face several challenges such as how much salesforce to allocate to real-life detailing versus virtual detailing, whether there is a need for sales representatives to meet the physicians face-to-face after virtual detailing or whether online detailing is enough. Despite the increasing shift to eDetailing by life sciences companies and despite the increasing calls for more scholarly investigations on how firms can leverage digital technology (e.g., see research priorities of Marketing Science Institute for 2014–2016), current literature still lacks systematic studies on eDetailing. We know little about the differential effectiveness of the various forms of eDetailing with respect to their impact on physicians and how they compare to a real-life detailing visit. Another interesting question to examine is the cross-promotion tool tradeoffs of eDetailing with real-life detailing, with other direct-to-physician promotion tools, and with DTCA. Future work may also aim to provide guidance for life sciences firms on how the effectiveness of eDetailing varies depending on different physician types, on the drug's life cycle stage or its characteristics such as effectiveness and side effects. Another important area of future investigation is the content of eDetailing and how it needs to be aligned with the content of real-life detailing (see Kappe and Stremersch 2016 for a recent work on the content of real-life detailing). The content of detailing, i.e., which drug attributes are discussed, is a strategic choice of pharmaceutical firms. An interesting avenue of future research is to apply learning models and determine which (e) detailing content characteristics are a primary source of physician learning.

13.4.4 eDTCA

A related development to eDetailing is the emergent shift to online DTCA (i.e., eDTCA) by several major firms. This form of communication with patients occurs via Facebook pages, Twitter feeds, blogs, RSS feeds, dedicated YouTube channels, etc. It is important to study the differences between the short- and long-term effectiveness of eDTCA and of DTCA through more traditional media channels, including potential synergies between eDTCA and DTCA. Insights on which media channel is most suited may support life sciences firms in better media allocations of DTCA spending. Another relevant issue is the heterogeneity among patients in responsiveness to eDTCA. In light of prior work, such heterogeneity is likely to exist: Liu and Gupta (2011) have found substantial heterogeneity in responsiveness to DTCA among patients from different insurance groups, and Stremersch and Van Dyck (2009) suggest gender as a source of heterogeneity in patient responsiveness to DTCA. Another possible source of heterogeneity might be the severity of the medical condition.

13.4.5 Regulations and Generic Use

One of the serious challenges that life sciences firms are currently facing is the increasing pressure from governments and insurers to reduce healthcare costs. Because generic drugs are considerably cheaper than branded drugs, several countries have increased the pressure on the healthcare system to increase the transition from prescribing, dispensing and administering branded drugs to generic ones. Governments and insurers have implemented policies such as promoting or enforcing generic prescriptions by physicians, capping physicians' prescription budgets, promoting or enforcing generic substitutions by pharmacists, and promoting activities to increase patient awareness about the equivalence of generic and branded drugs. Consequently, the number of firms that manufacture generic drugs has increased. The increased competition from generics firms creates several challenges for branded firms, which deserve the attention of future research. First, branded firms may need to consider whether they should also supply generics along with their branded drugs, and if so, how much focus they need to devote to generics versus branded drugs. Second, firms need to reconsider their pricing strategy upon the entry of generics. Future research may shed light on conditions under which increase or decrease of prices is more likely to occur. Third, future research may examine how and when firms should alter their direct-to-physician promotions and DTCA upon the introduction of generics to retain their brand-loyal physicians and patients. Fourth, little is known about the international variation in the effect of generic entry on the prices and promotion strategies of branded firms.

The rising pressure on healthcare budgets also engenders discussions on increasing price regulations such as direct caps of ex-manufacturer price or cross-country reference pricing. For instance, with the recent announcement of a price increase for Daraprim, a drug that treats a parasitic infection, politicians have called for actions against the "outrageous" pricing. This call was based on the public outcry against the price increase (The Economist 2015). Research shows that regulations do not influence launch price (Verniers et al. 2011), but there is little empirical research on the influence of regulations on the price dynamics of treatments across life cycles, although such influence is often implied (Ekelund and Persson 2003). Cross-country reference-price enforcement may create cross-country spillovers of price, which may influence the global launch plans of life sciences firms. Future research may investigate the global entry and pricing decisions of life sciences firms, with particular focus on the interplay between pricing decisions and launch sequence across countries. Furthermore, global pricing decisions of life sciences firms are subject to parallel imports. In one market, parallel imports may create increased competition for the branded-drug producers but also take some market share from generics because patients who would consume generics in the absence of parallel imports switch to imported drugs (Duso et al. 2014). In another market, the branded-drug producers may increase profits by selling their drugs to parallel importers. Future research may develop rigorous models explicitly accounting for parallel imports to optimize the managers' decisions on global pricing of life sciences firms,

13.4.6 Corporate Image of the Pharmaceutical Industry

The pharmaceutical industry used to enjoy the status of being one of the most respected industries. However, the corporate image of the pharmaceutical industry in general and of some leading pharmaceutical companies in particular, has significantly deteriorated in recent years. The results of the PatientView survey,[1] conducted on 1,150 patient groups in 58 countries from November 2014 to January 2015, indicate that in 2014 the global pharmaceutical industry ranked sixth in terms of corporate reputation from among eight healthcare industry sectors. Those surveyed pointed to several reasons for the negative reputation of the pharmaceutical industry, including unfair pricing policies that make drugs unaffordable for many patients, drugs with only short-term health benefits, not serving the needs of neglected patient groups, inappropriate marketing activities, nontransparent corporate activities, adverse news about products, and not having a patient-centered strategy. Several pharmaceutical firms have also appeared at the center of public scandals and lawsuits for bribing physicians to prescribe specific drugs, bribing generic companies not to supply generic substitutes, and for suing generic companies over patent rights.

---

[1]Available in http://www.patient-view.com/uploads/6/5/7/9/6579846/patientview_11-2-2015_press_release_corp_rep_global.pdf (accessed October 2015).

The pharmaceutical industry is in desperate need to restore its damaged corporate image. However, much of the scholarly work focuses on the effects of the firms' actions, such as promotion, advertising, selection of key opinion leaders, on the brand- or category-level sales. We believe that more scholarly attention on the corporate image of pharmaceutical firms is needed. For instance, what is the influence of a firm's DTP promotion or of the DTCA of one brand on the firm's corporate image and subsequently, on the firm's other brands? What are the consequences of the withdrawal of one brand on the overall corporate image of the company? What are the effects of the corporate social responsibility activities of pharmaceutical firms? In December 2013, in line with the public outcry against obscure practices and with the increasing transparency requirements for financial relationships between pharmaceutical and medical device manufacturers and healthcare providers imposed by the US Physician Payment Sunshine Act, GSK (GlaxoSmithKline) decided to stop hiring external key opinion leaders. Instead, GSK decided to hire an internal team of physicians tasked with educating peers about GSK products. We know little about the effects of developing such an internal team or about their influences compared to those of hiring external speakers on the firm's corporate image.

13.4.7 Extending the Focus on Other Members of the Healthcare Value Chain

The above overview illustrates that academic research seems to be focused on the pharmaceutical industry rather than on the biotechnology and medical device industries. While the biotech industry received quite some attention from an innovation generation angle (Fang et al. 2015; Paruchuri et al. 2006; Prabhu et al. 2005; Wuyts et al. 2004), this is much less the case with the marketing angle. Still, the biotech industry faces quite some complexity upon the introduction of new therapies because for instance, smaller firms are often involved. Most of the very expensive therapies are also biologicals, which often leads to unique market rollout strategies. Equally, the medical devices industry has seen limited interest (the works by Kahn and Luce (2003) on mammography or Thirumalai and Sinha (2011) on recalls of medical devices are notable exceptions). This is unfortunate because the medical devices industry is at the intersection of the ICT, engineering, and the pharmaceutical world. Finally, cosmeceuticals and nutraceuticals have seen equally little interest from academic marketing research. Nonetheless, with claims on deep human needs such as anti-ageing and beauty, these areas present excellent opportunities for marketing researchers.

The scope concerning stakeholders has been quite limited as well. Most of the attention has been focused on the impact of firm actions on physician and patient behaviors. The literature on payer behavior and its impact on patients and physicians is practically nonexistent. Still payers' reimbursement policies and regulations have a profound impact on healthcare provision. Further, despite the increasing importance of the pharmacist in therapy advice, we have seen little research from a pharmacy perspective. Nonetheless, the influence of the pharmacy environment (e.g., in-store promotion, pharmacy margins) on patient behavior is increasingly becoming prominent.

In sum, there is a wealth of future research opportunities available for researchers with an interest in decision-support models in the life sciences industry. In the section above, we highlight some of such opportunities, while recognizing that the above list by no means is an exhaustive list of issues that warrant more research. The industry presents an area that does not only allow for the development of sophisticated tools on state-of-the-art data, but it also supports the decision making of highly professional firms in an area with a huge impact on society and on everyone's life. Such combination is relatively rare in marketing research.

References

Albers, S., M. Mantrala, and S. Sridhar. 2010. Personal selling elasticities: A meta-analysis. *Journal of Marketing Research* 47 (5): 840–853.

Amaldoss, W., and C. He. 2009. Direct-to-consumer advertising of prescription drugs: A strategic analysis. *Marketing Science* 28 (3): 472–487.

Anderson, S., F. Ciliberto, J. Liaukonyte, and R. Renault. 2016. Push-me, pull-you: Comparative advertising in the OTC analgesics industry. Working paper.

Azoulay, P. 2002. Do pharmaceutical sales respond to scientific evidence? *Journal of Economics and Management Strategy* 11 (4): 551–594.

Bass, F. 1969. A new-product growth model for consumer durables. *Management Science* 15 (5): 215–227.

Berndt, E., L. Bui, D. Reiley, and G. Urban. 1995. Information, marketing, and pricing in the U.S. antiulcer drug market. *American Economic Review* 85 (2): 100–105.

Bhatia, T., and L. Wang. 2011. Identifying physician peer-to-peer effects using patient movement data. *International Journal of Marketing Research* 28 (1): 51–61.

Bhattacharya, J., and W. Vogt. 2003. A simple model of pharmaceutical price dynamics. *Journal of Law and Economics* 23 (2): 599–626.

Bowman, D., C. Heilman, and P. Seetharaman. 2004. Determinants of product-use compliance behavior. *Journal of Marketing Research* 41 (3): 324–338.

Brekke, K., and M. Kuhn. 2006. Direct to consumer advertising in pharmaceutical markets. *Journal of Health Economics* 25 (1): 102–130.

Brekke, K., T. Holmas, and O. Straume. 2015. Price regulation and parallel imports of pharmaceuticals. *Journal of Public Economics* 129: 92–105.

Calabro, S. 2003. In DTC-free Europe, just as many patients ask for brand drugs. *Medical Marketing and Media* 38 (11): 30.

Calfee, J., C. Winston, and R. Stempski. 2002. Direct-to-consumer advertising and the demand for cholesterol-reducing drugs. *Journal of Law Economics* 45 (S2): 673–690.

Camacho, N., M. de Jong, and S. Stremersch. 2014. The effect of customer empowerment on adherence to expert advice. *International Journal of Research in Marketing* 31 (3): 293–308.

Camacho, N., V. Landsman, and S. Stremersch. 2010. The Connected Patient, in The Connected Customer: The Changing Nature of Consumer and Business Markets, ed. Stefan Wuyts, Marnik G. Dekimpe, Els Gijsbrechts, and F.G.M. Pieters, Hampshire, UK: Routledge Academic.

Camacho, N., B. Donkers, and S. Stremersch. 2011. Predictably non-Bayesian: Quantifying salience effects in physician learning about drug quality. *Marketing Science* 30 (2): 305–320.
Campbell, E., G. Pham-Kanter, C. Vogeli, and L. Iezzoni. 2013. Physician acquiescence to patient demands for brand-name drugs: Results of national survey of physicians. *JAMA Internal Medicine* 173 (3): 237–239.
Caves, R., M. Whinston, M. Hurwitz, A. Pakes, and P. Temin. 1991. Patent expiration, entry, and competition in the U.S. pharmaceutical industry: Comments and discussion. *Brookings Papers* Economic *Activity* 12 (1): 1–66.
Chan, T., C. Narasimhan, and Y. Xie. 2013. Treatment effectiveness and side effects: A model of physician learning. *Management Science* 59 (6): 1309–1325.
Ching, A., and M. Ishihara. 2010. The effects of detailing on prescribing decisions under quality uncertainty. *Quantitative Marketing and Economics* 8 (2): 123–165.
Chintagunta, P., and R. Desiraju. 2005. Strategic pricing and detailing behavior in international markets. *Marketing Science* 24 (1): 67–80.
Chintagunta, P., R. Goettler, and M. Kim. 2012. New drug diffusion when forward-looking physicians learn from patient feedback and detailing. *Journal Marketing Research* 49 (6): 807–821.
Chintagunta, P.K., R. Jiang, and G.Z. Jin. 2009. Information, learning, and drug diffusion: The case of cox-2 inhibitors. *Quantitative Marketing and Economics* 7 (4): 399–443.
Cockburn, I., J. Lanjouw, and M. Schankerman. 2014. Patents and global diffusion of new drugs. NBER Working Paper 20492.
Coleman, J.S., E. Katz, and H. Menzel. 1966. *Medical Innovation: A diffusion study*. Indianapolis: Bobbs-Merrill.
Coscelli, A., and M. Shum. 2004. An empirical model of learning and patient spillover in new drug entry. *Journal of Econometrics* 122 (2): 213–246.
Danzon, P., R. Wang, and L. Wang. 2005. The impact of price regulation on the launch delay of new drugs: Evidence from twenty-five major markets in the 1990s. *Health Economics* 14 (3): 269–292.
Danzon, P., A. Towse, and J. Mestre-Ferrandiz. 2015. Value-based differential pricing: Efficient prices for drugs in a global context. *Health Economics* 24 (3): 294–301.
Datta, A., and D. Dave. 2013. Effects of physician-directed pharmaceutical promotion on prescription behaviors: Longitudinal evidence. NBER Working Paper 19592.
Dellande, S., M. Gilly, and J. Graham. 2004. Gaining compliance and losing weight: The role of the service provider in health care services. *Journal of Marketing* 68 (3): 78–91.
Desiraju, R., H. Nair, and P. Chintagunta. 2004. Diffusion of new pharmaceutical drugs in developing and developed nations. *International Journal of Research in Marketing* 21 (4): 341–357.
Dong, X., P. Chintagunta, and P. Manchanda. 2011. A new multivariate count data model to study multi-category physician prescription behavior. *Quantitative Marketing and Economics* 9 (3): 301–337.
Donohue, J., E. Berndt, M. Rosenthal, A. Epstein, and R. Frank. 2004. Effects of pharmaceutical promotion on adherence to the treatment guidelines for depression. *Medical Care* 42 (12): 1176–1185.
Duso, T., A. Herr, and M. Suppliet. 2014. The welfare impact of parallel imports: A structural approach applied to the German market for oral anti-diabetics. *Health Economics* 23 (9): 1036–1057.
EFPIA. 2015. *The pharmaceutical industry in figures*. Brussels, Belgium: European Federation of Pharmaceutical Industries and Associations.
Ekelund, M., and B. Persson. 2003. Pharmaceutical pricing in a regulated market. *The Review of Economics and Statistics* 85 (2): 298–306.
Fang, E., J. Lee, and Z. Yang. 2015. The timing of codevelopment alliances in new product development processes: Returns for upstream and downstream partners. *Journal of Marketing* 79 (1): 64–82.

Ferguson, J., R. Discenza, and J. Miller. 1987. Increasing the odds of patient compliance through prescription warning labels. *Journal of Health Care Marketing* 7 (1): 37–46.

Fischer, M., and S. Albers. 2010. Patient- or physician-oriented marketing: What drives primary demand for prescription drugs? *Journal of Marketing Research* 47 (1): 103–121.

Forissier, T., and K. Firlik. 2012. Estimated annual pharmaceutical revenue loss due to medication non-adherence. Capgemini Consulting & HealthPrize Technologies (https://www.capgemini.com/resource-file-access/resource/pdf/Estimated_Annual_Pharmaceutical_Revenue_Loss_Due_to_Medication_Non-Adherence.pdf/, Accessed Aug 2015).

Frank, R., and D. Salkever. 1997. Generic entry and the pricing of pharmaceuticals. *Journal of Economics and Management Strategy* 6 (1): 75–90.

Ganslandt, M., and K. Maskus. 2004. Parallel imports and the pricing of pharmaceutical products: Evidence from the European Union. *Journal of Health Economics* 23 (5): 1035–1057.

Ghinn, D. 2012. Activating digital opinion leaders (http://engagementstrategy.com/articles/the-changing-key-opinion-leader-and-advocate-environment/, Accessed Aug 2015).

Goldenberg, J., B. Libai, and E. Muller. 2002. Riding the saddle: How cross-market communications can create a major slump in sales. *Journal of Marketing* 66 (2): 1–16.

Gonzalez, J., C. Sismeiro, and S. Dutta. 2008. Can branded drugs benefit from generic entry? The role of detailing and price in switching to non-bioequivalent molecules. *International Journal of Research in Marketing* 25 (4): 247–260.

Gönül, F., F. Carter, E. Petrova, and K. Srinivasan. 2001. Promotion of prescription drugs and its impact on physicians' choice behavior. *Journal of Marketing* 65 (3): 79–90.

Grabowski, H., and J. Vernon. 1992. Brand loyalty, entry, and price-competition in pharmaceuticals after the 1984 drug-act. *Journal of Law and Economics* 35 (2): 331–350.

Grewal, R., A. Chakravarty, M. Ding, and J. Liechty. 2008. Counting chickens before the eggs hatch: Associating new product development portfolios with shareholder expectations in the pharmaceutical sector. *International Journal of Research in Marketing* 25 (4): 261–272.

Hahn, M., S. Park, L. Krishnamurthi, and A.A. Zoltners. 1994. Analysis of new product diffusion using a 4-segment trial-repeat model. *Marketing Science* 13 (3): 224–247.

Hellerstein, J. 1998. The importance of the physician in the generic versus trade-name prescription decision. *RAND Journal of Economics* 29 (1): 108–136.

Iizuka, T., and G. Jin. 2005. The effect of prescription drug advertising on doctor visits. *Journal of Economic Management Strategy* 14 (3): 701–727.

Iizuka, T., and G. Jin. 2007. Direct-to-consumer advertising and prescription choice. *The Journal of Industrial Economics*, Notes and Comments page Published at: http://www.essex.ac.uk/jindec/notes.htm.

Iyengar, R., C. Van den Bulte, and T. Valente. 2011. Opinion leadership and social contagion in new product diffusion. *Marketing Science* 30 (2): 195–112.

Iyengar, R., C. Van den Bulte, and J. Lee. 2015. Social contagion in new product trial and repeat. *Marketing Science* 34 (3): 408–429.

Jacobzone, S. (2000) Pharmaceutical policies in OECD countries: Reconciling social and industrial goals. OECD Labour Market and Social Policy—Occasional papers 40. Paris, France: OECD.

Janakiraman, R., S. Dutta, C. Sismeiro, and P. Stern. 2008. Physicians' persistence and its implications for their response to promotion of prescription drugs. *Management Science* 54 (6): 1080–1093.

Janakiraman, R., C. Sismeiro, and S. Dutta. 2009. Perception spillovers across competing brands: A disaggregate model of how and when. *Journal of Marketing Research* 46 (4): 467–481.

Kahn, B., and M. Luce. 2003. Understanding high-stakes consumer decisions: Mammography adherence following false-alarm test results. *Marketing Science* 22 (3): 393–410.

Kalra, A., S. Li, and W. Zhang. 2011. Understanding responses to contradictory information about products. *Marketing Science* 30 (6): 1098–1114.

Kappe, E., and S. Stremersch. 2016. Drug detailing and doctors' prescription decisions: The role of information content. *Marketing Science*, forthcoming.

Kolsarici, C., and D. Vakratsas. 2010. Category- versus brand-level advertising messages in a highly regulated environment. *Journal of Marketing Research* 47 (6): 1078–1089.
Kravitz, R., R. Bell, R. Azari, E. Krupat, S. Kelly-Reif, and D. Thom. 2002. Request fulfillment in office practice: Antecedents and relationship to outcomes. *Medical Care* 40 (1): 38–51.
Kremer, S., T. Bijmolt, P. Leeflang, and J. Wieringa. 2008. Generalizations on the Effectiveness of pharmaceutical promotional expenditures. *International Journal of Research in Marketing* 25 (4): 234–246.
Kyle, M. 2006. The role of firm characteristics in pharmaceutical product launches. *The Rand Journal of Economics* 73 (3): 602–618.
Kyle, M. 2007. Pharmaceutical price controls and entry strategies. *The Review of Economics and Statistics* 89 (1): 88–99.
Kyle, M., and Y. Qian. 2014. Intellectual property rights and access to innovation: Evidence from TRIPS, NBER Working Paper 20799.
Lamiraud, K., and P.-Y. Geoffard. 2007. Therapeutic non-adherence: A rational behavior revealing patient preferences? *Health Economics* 16 (11): 1185–1204.
Lanjouw, J. 2005. Patents, price controls and access to new drugs: How policy affects global market entry. NBER, Cambridge, MA: NBER Working paper 11321.
Law, M., S. Majumdar, and S. Soumerai. 2008. Effect of illicit direct to consumer advertising on use of etanercept, mometasone, and tegaserod in Canada: Controlled longitudinal study. *British Medical Journal* 337: a1055.
Lien, H.-M., M. Lu, C.-T. Ma, and T. McGuire. 2010. Progress and compliance in alcohol abuse treatment. *Journal of Health Economics* 29 (2): 213–225.
Ling, D., E. Berndt, and M. Kyle. 2002. Deregulating direct-to-consumer marketing of prescription drugs: Effects on prescription and over-the-counter product sales. *Journal of Law Economics* 45 (2): 691–723.
Liu, Q., and S. Gupta. 2011. The impact of direct-to-consumer advertising of prescription drugs on physician visits and drug requests: Empirical findings and public policy implications. *International Journal of Research in Marketing* 28 (3): 205–217.
Liu, Q., S. Gupta, S. Venkataraman, and H. Liu. 2015. An empirical model of drug detailing: Dynamic competition and policy implications. *Management Science*, forthcoming.
Lu, Z., and W. Comanor. 1998. Strategic pricing of new pharmaceuticals. *The Review of Economics and Statistics* 80 (1): 108–118.
Lundin, D. 2000. Moral hazard in physician prescription behaviour. *Journal of Health Economics* 19 (5): 639–662.
Manchanda, P., and P. Chintagunta. 2004. Responsiveness of prescription behavior to salesforce effort: An individual level analysis. *Marketing Letters* 25 (2–3): 129–145.
Manchanda, P., P. Rossi, and P. Chintagunta. 2004. Response modeling with nonrandom marketing-mix variables. *Journal of Marketing Research* 41 (4): 467–478.
Manchanda, P., Y. Xie, and N. Youn. 2008. The role of targeted communication and contagion in new product adoption. *Marketing Science* 27 (6): 961–976.
Mintzes, B., M. Barer, R. Kravitz, B. Bassett, J. Lexchin, A. Kazanjian, R. Evans, R. Pan, and S. Marion. 2003. How does direct-to-consumer advertising (DTCA) affect prescribing? A survey in primary care environments with and without legal DTCA. *Canadian Medical Association Journal* 169 (5): 405–412.
Mizik, N., and R. Jacobson. 2004. Are physicians easy marks? Quantifying the effects of detailing and sampling on new prescriptions. *Management Science* 50 (12): 1704–1715.
Montgomery, D., and A. Silk. 1972. Estimating dynamic effects of market communications expenditures. *Management Science* 18 (10): B485–B501.
Montoya, R., O. Netzer, and K. Jedidi. 2010. Dynamic allocation of pharmaceutical detailing and sampling. *Marketing Science* 29 (5): 909–924.
Nair, H., P. Manchanda, and T. Bhatia. 2010. Asymmetric social interactions in physician prescription behavior: The role of opinion leaders. *Journal of Marketing Research* 47 (5): 883–895.

Narayanan, S., R. Desiraju, and P. Chintagunta. 2004. Return on investment implications for pharmaceutical promotional expenditures: The role of marketing-mix interactions. *Journal of Marketing* 68 (4): 90–105.

Narayanan, S., and P. Manchanda. 2009. Heterogeneous learning and the targeting of marketing communication for new products. *Marketing Science* 28 (3): 424–441.

Narayanan, S., P. Manchanda, and P. Chintagunta. 2005. Temporal differences in the role of marketing communication in new product categories. *Journal of Marketing Research* 42 (3): 278–290.

Osinga, E., P. Leeflang, and J. Wieringa. 2010. Early marketing matters: A time-varying parameter approach to persistence modeling. *Journal of Marketing Research* 47 (1): 173–185.

Osinga, E., P. Leeflang, S. Srinivasan, and J. Wieringa. 2011. Why do firms invest in consumer advertising with limited sales response? shareholder perspective. *Journal of Marketing* 75 (1): 109–124.

Paruchuri, S., A. Nerkar, and D.C. Hambrick. 2006. Acquisition integration and productivity losses in the technical core: Disruption of inventors in acquired companies. *Organization Science* 17 (5): 545–562.

Parsons, L., and P. Vanden Abeele. 1981. Analysis of sales call effectiveness. *Journal of Marketing Research* 18 (1): 107–113.

Paterniti, D., T. Fancher, C. Cipri, S. Timmermans, J. Heritage, and R. Kravitz. 2010. Getting to "no": Strategies primary care physicians use to deny patient requests. *Archives of Internal Medicine* 170 (4): 381–388.

PhRMA. 2015a. *Pharmaceutical industry profile*. Washington, DC: Pharmaceutical Research and Manufacturers of America.

PhRMA. 2015b. 2015 Profile Biopharmaceutical Research Industry.

Prabhu, J.C., R.K. Chandy, and M.E. Ellis. 2005. The Impact of acquisitions on innovation: Poison pill, placebo, or tonic? *Journal of Marketing* 69 (1): 114–130.

Prigge, J.-K., B. Dietz, C. Homburg, W. Hoyer, and J. Burton. 2015. Patient empowerment: A cross-disease exploration of antecedents and consequences. *International Journal of Research in Marketing* 32 (4): 375–386.

Rao, A., and M. Yamada. 1988. Forecasting with a repeat purchase diffusion model. *Management Science* 34 (6): 734–752.

Rizzo, J. 1999. Advertising and competition in the ethical pharmaceutical industry: The case of antihypertensive drugs. *Journal of Law and Economics* 42 (1): 89–116.

Rosenthal, M.B., E.R. Berndt, J.M. Donohue, A.M. Epstein, and R.G. Frank. 2003. Demand effects of recent changes in prescription drug promotion. *Frontiers in Health Policy Research* 6 (1): 1–26.

Schwartz, J., M. Luce, and D. Ariely. 2011. Are consumers too trusting? The effects of relationships with expert advisers. *Journal of Marketing Research* 48 (Special Issue): S163–S174.

Scott, A., and A. Shiell. 1997. Do fee descriptors influence treatment choices in general practice? A multilevel discrete choice model. *Journal of Health Economics* 16 (3): 323–342.

Shah, M., J. Bentley, and D. McCaffrey III. 2006. Evaluations of care by adults following a denial of an advertisement-related prescription drug request: The role of expectations, symptom severity, and physician communication style. *Social Science and Medicine* 62 (4): 888–899.

Shankar, V., G. Carpenter, and L. Krishnamurthi. 1998. Late mover advantage: How innovative late entrants outsell pioneers. *Journal of Marketing Research* 35 (1): 54–70.

Shankar, V., G. Carpenter, and L. Krishnamurthi. 1999. The advantages of entry in the growth stage of the product life cycle: An empirical analysis. *Journal of Marketing Research* 36 (2): 269–276.

Sridhar, S., M. Mantrala, and S. Albers. 2014. Pharmaceutical detailing elasticities: A meta-analysis. In *Innovation and marketing in the pharmaceutical industry: Emerging practices, research, and policies*, ed. M. Ding, J. Eliashberg, and S. Stremersch. New York: Springer.

Strang, D., and N.B. Tuma. 1993. Spatial and temporal heterogeneity in diffusion. *American Journal of Sociology* 99 (3): 614–639.

Stremersch, S. 2008. Health and marketing: The emergence of a new field of research. *International Journal of Research in Marketing* 25 (4): 229–233.

Stremersch, S., V. Landsman, and S. Venkataraman. 2013. The relationship between DTCA, drug requests, and prescriptions: Uncovering variation in specialty and space. *Marketing Science* 32 (1): 89–110.

Stremersch, S., and A. Lemmens. 2009. Sales growth of new pharmaceuticals across the globe: The role of regulatory regimes. *Marketing Science* 28 (4): 690–708.

Stremersch, S., and W. Van Dyck. 2009. Marketing of the life sciences: A new framework and research agenda for a nascent field. *Journal of Marketing* 73 (4): 4–30.

The Economist. 2015. Drug pricing in America: Painful Pills. Sep 26.

Thirumalai, S., and K. Sinha. 2011. Product recalls in the medical device industry: An empirical exploration of the sources and financial consequences. *Management Science* 57 (2): 376–392.

Vakratsas, D., and C. Kolsarici. 2008. A dual-market diffusion model for a new prescription pharmaceutical. *International Journal of Research in Marketing* 25 (4): 282–293.

Vakratsas, D., and C. Kolsarici. 2014. How DTCA influences prescription pharmaceutical markets. In *Innovation and marketing in the pharmaceutical industry: Emerging practices, research, and policies*, ed. M. Ding, J. Eliashberg, and S. Stremersch. New York: Springer.

Valente, T.W. 1996. Social network thresholds in the diffusion of innovations. *Social Networks* 18 (1): 69–89.

Van den Bulte, C. 2000. New product diffusion acceleration: Measurement and analysis. *Marketing Science* 19 (4): 366–380.

Van den Bulte, C., and G. Lilien. 2001. Medical innovation revisited: Social contagion versus marketing effort. *American Journal of Sociology* 106 (5): 1409–1435.

Van den Bulte, C., and S. Stremersch. 2004. Social contagion and income heterogeneity in new product diffusion: A meta-analytic test. *Marketing Science* 23 (4): 530–544.

Venkataraman, S., and S. Stremersch. 2007. The debate on influencing doctors' decisions: Are drug characteristics the missing link? *Management Science* 53 (11): 1688–1701.

Verniers, I., S. Stremersch, and C. Croux. 2011. The global entry of new pharmaceuticals: A joint investigation of launch window and price. *International Journal of Research in Marketing* 28 (4): 295–308.

Williams, G., G. Rodin, R. Ryan, W. Grolnick, and E. Deci. 1998. Autonomous regulation and long-term medication adherence in adult outpatients. *Health Psychology* 17 (3): 269–276.

Wilkes, M., R. Bell, and R. Kravitz. 2000. Direct-to-consumer prescription drug advertising: trends, impact, and implications. *Health Affairs* 19 (2): 110–128.

Windmeijer, F., E. De Laat, R. Douven, and E. Mot. 2006. Pharmaceutical promotion and GP prescription behavior. *Health Economics* 15 (1): 5–18.

Wosinska, M. (2002) Just what the patient ordered? Direct-to-consumer advertising and the demand for pharmaceutical products. Working Paper Harvard Business School, 03-058.

Wosinska, M. 2005. Direct-to-consumer advertising and drug therapy compliance. *Journal of Marketing Research* 42 (3): 323–332.

Wuyts, S., S. Dutta, and S. Stremersch. 2004. Portfolios of interfirm agreements in technology-intensive markets: Consequences for innovation and profitability. *Journal of Marketing* 68 (2): 88–100.

Chapter 14
Marketing Models for Internet Advertising

Randolph E. Bucklin and Paul R. Hoban

14.1 Introduction

Advertisers are expected to spend $240 billion on internet advertising worldwide in 2019, which represents 36% of the global advertising industry (PricewaterhouseCoopers 2016a, b). In the United States, by far the largest advertising market, digital advertising has increased at a 17% annual rate over the past decade, reaching $59.6 billion in 2015. This exceeds the combined spending on print and radio advertising, $27.3 and $17.4 billion respectively, and lags only television advertising at $66.3 billion (PricewaterhouseCoopers 2016a, b). Due to its rapid growth, many projections have US internet advertising surpassing television in 2017 (Bruell 2014; PricewaterhouseCoopers 2016a, b).

Internet advertising continues to be dominated by two formats, paid search (i.e., sponsored search advertising or SSA) and display advertising (e.g., banner ads, digital video, rich media,[1] and sponsorships). In the United States, advertisers spent $29.4 billion on paid search in 2015, a 17.6% increase over 2014. Meanwhile, they spent $24.9 billion on display advertising, a 27% increase over the previous year. The remaining formats for internet advertising, such as classified ads and lead generation, had total spending of $5.3 billion in 2015. (PricewaterhouseCoopers 2016a, b). Given this spending pattern, it is not surprising that the bulk of academic attention to digital advertising has so far been focused on display and paid search. Accordingly, in this chapter we also focus our attention on recent developments in

[1]Rich media describes advertisements that contain video, audio, or other content that encourages user interaction with the ad.

R.E. Bucklin (✉)
UCLA Anderson School of Management, Los Angeles, USA
e-mail: randy.bucklin@anderson.ucla.edu

P.R. Hoban
University of Wisconsin-Madison, Madison, USA

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_14

quantitative marketing models in these two advertising domains. In what follows, our discussion will emphasize models which are most immediately relevant for decision making by advertisers and that can be implemented in practice.

The complexity of the digital advertising ecosystem creates challenges for both decision makers and modelers. In display advertising, for example, targeting algorithms have evolved to include individual characteristics such as age and gender and past behaviors such as browsing histories, email exchanges, shopping patterns, and responses to similar advertising. These are matched with characteristics of a particular ad and the website on which it may be embedded to determine when and to whom an impression is served. Advertising platforms, which are ultimately responsible for ad serving, consider these targeting algorithms proprietary, and offer managers little insight into the underlying processes. As we shall see, the prevalence of targeting creates challenges for modeling the effects of internet display advertising.

The challenges stemming from limited insight into targeting algorithms are compounded by the nature of the data generally available to managers and researchers. With a few exceptions, the models currently available for use in display advertising have been applied to site-centric clickstream data from one or more e-commerce websites. Such data provide detailed records of individual consumer exposure to advertising along with browsing behaviors at the focal e-commerce websites, including pages visited, time of visit, duration of visit, etc. However, there is generally limited, if any, information on consumer browsing behavior across other websites or competitor actions. For those looking to create and leverage modeling approaches appropriate for the display ad domain, these limitations create real hurdles.

In paid search advertising, the environment is similarly complex and data limitations also confront advertisers and modelers. Managing paid search advertising campaigns requires advertisers to make decisions across a wide range of variables, including which keywords to include in a campaign, how much to bid for ad placement on the search engine results page (SERP), selecting text ad content, and designating landing pages (i.e., the page on the advertiser's website that the user first sees after clicking on an ad). Data provided to advertisers by the major search engines (Google, Bing, and Yahoo), though voluminous and detailed, is limited to information about the ads placed by the firm and contains little, if any, detailed information on competitor bidding, placement, and performance. The complexities and incomplete information about the advertising environment also give rise to modeling challenges, many of which are the subject of recent work in the area.

The purpose of this chapter is to provide an overview of recent key developments and advances in the application of marketing models to aid decision making in internet advertising. Our coverage begins in the next section with models for display advertising. We then turn to models for paid search advertising, followed by a brief section on models that incorporate the effects of both display and search ads. We conclude with a summary, suggesting directions for future research in this fast-moving area.

14.2 Models for Internet Display Advertising

Internet display advertising encompasses a variety of image based formats, including banner ads, digital video, rich media, and sponsorships. Banner ads, rectangular images embedded within a webpage, are the most common form of display advertising. In the US, advertisers spent $7.7 billion on banner advertising during 2015. Digital video spending is considerably smaller, $4.1 billion in the US during 2015, but is also the fastest growing component of display advertising. Rich media and sponsorships round out the display advertising space, with advertisers spending just over $1.2 billion in the US during 2015 (PricewaterhouseCoopers 2016a, b).[2]

In the early days of display advertising, websites sold inventory directly to advertisers through an internal sales team. As the number of websites and devices increased dramatically, the volume and variety of available inventory made this system untenable, and ad networks came into favor. Ad networks aggregate advertising inventory from multiple websites and sell it to advertisers on either a cost per impression or performance based pricing model (e.g., cost per click, cost per lead, cost per conversion, etc.). From these ad networks, ad exchanges developed to sell inventory using an auction based system, with both ad networks and other media buyers actively bidding. In recent years, these exchanges have begun auctioning impression opportunities at the moment an individual visits a webpage, leading to so called programmatic or real time bidding (RTB). RTB allows advertisers, or their representatives, to bid for the right to show their ad to a specific individual at a given time in a precise position on a certain webpage. This has dramatically increased the specificity of targeting algorithms, and is quickly becoming the dominant display advertising pricing model. Programmatic was expected to comprise 67% of all US display ad spending in 2016 (eMarketer 2016).

These shifts in pricing have coincided with, and in some cases enabled, dramatic increases in the variety and complexity of targeting algorithms. When inventory was sold directly by publishers, targeting was largely based on aggregate audience demographics. As tracking technologies have improved, ad networks and media buyers have developed algorithms that take into account such varied information as the consumer's location, past browsing activity, social media posts, email messages, and responses to past, similar ads. With the development of RTB, media buyers are better able to control when a consumer sees an ad based on their recent activity. Thus, modern targeting algorithms take into account both time varying and time invariant consumer characteristics. It is only recently that modeling approaches have been developed to capture the effects of display advertising in the presence of targeting; we discuss these later in this section. In Table 14.1, we present a summary of the key aspects of the display advertising models we discuss below.

[2]Annual spending figures here exclude display ad spending targeted for mobile devices.

Table 14.1 Summary of models for display advertising

Article	Modeling objective	Types of model(s) and dependent variable(s)	Data	Key insights
Chatterjee et al. (2003)	Understand banner ad click-through	Binary logit; click-through	Site-centric clickstream	Extensive heterogeneity in click-through rates
Manchanda et al. (2006)	Estimate effects of banner ad exposure on online purchase timing	Hazard model; Purchase timing	Internal company data on banner ad exposure and online transactions	Concave response to banner ad exposure; elasticity of purchase with respect to banner ad exposure is 0.02
Rutz and Bucklin (2012)	Examine how banner advertising affects browsing	Multinomial logit; Webpage visited	Site-centric clickstream	Half of users do not respond to banner advertising; elasticity of within site page choice with respect to banner ad exposure is 0.2
Braun and Moe (2013)	Understand factors influencing display advertising wear-out	Zero inflated Poisson; Ad exposures Site visits Zero inflated Binomial; Conversions	Across site display advertising exposures and site-centric clickstream data	Significant wear-out from additional campaign exposures; Incremental impact from additional creative exposures
Hoban and Bucklin (2015)	Develop a method to accurately measure display advertising effectiveness in the purchase funnel	Binary logit; Website visit	Across site display advertising exposures and site-centric clickstream from a field experiment	Correlational estimates of display advertising effectiveness can be significantly biased; elasticity of between site page choice with respect to banner ad exposure is 0.1
Johnson et al. (2017)	Develop a method to accurately measure display advertising effectiveness when ad serving depends on prior ad responsiveness	Local Average Treatment Effect; Website visit	Field experiment data containing predicted and actual ad exposure between a treatment and control group	The experimental method can precisely measure ad effectiveness at a portion of the cost of public service announcements (PSA); Retargeting campaigns can be effective
Sahni (2015)	Examine the impact of temporal spacing on display advertising effectiveness	Binary logit; Page visits	Site-centric clickstream from a field experiment	Ad effects decay more quickly than a standard goodwill model implies

A. *Early Models for Click-through and Purchasing*

Display advertising can bring potential customers to an e-commerce site either directly via click-through or indirectly by impacting future searches and browsing. Click-through occurs when users, browsing elsewhere online, click on a display ad, activating an embedded link to the advertiser's website. In some of the earliest modeling of display advertising, Chatterjee et al. (2003) developed a model of banner ad click-through behavior based on a binary logit formulation. The model predicts the probability of an individual clicking on a banner ad exposure, given that the ad has not yet been clicked on by that user during the current browsing session.

Results from estimating the logit model showed that exposure to more banner ads within the same session was negatively related to click-through probability but at a decreasing rate. Thus, given that a user has not yet clicked on an ad in a session, the chances of doing so decline with further exposure. Across browsing sessions, the estimation results showed that time since last click, intersession time, and the cumulative number of banner exposures were all positively related to ad click-through, but the number of sessions at the site was negatively related to click probability, perhaps due to experience and learning effects among users.[3]

The Chatterjee et al. study used data from 1995. Today, click-through rates on banner ads are far lower, so subsequent modeling work has moved away from predicting click-through as the key outcome of interest. In an important experimental study, Drèze and Hussherr (2003) looked into the ramifications of low click-through on banner ads. Using eye-tracking, they found that about half of their subjects deliberately avoided looking at banner ads. Following up with a large-scale survey study, they found that banner ads were effective in boosting recall, recognition, and awareness. Drèze and Hussherr concluded that banner ads could be effective but that a broader set of measures—not just click-through rates—would be needed to gauge response.

One such broader measure may be to look directly at actual purchases made at an e-commerce site in response to banner ad exposure. Manchanda et al. (2006) modeled the impact of banner advertising on purchase behavior using a piecewise exponential hazard model where a purchase is treated as a "failure." The authors found that the log of the number of banner ad exposures was positively related to the probability of making a purchase on the web site. This work was an important development in the modeling of display advertising because it established an individual-level connection between exposure to banner ads and e-commerce transactions, regardless of the rate of click-through. [4]

[3]For a detailed discussion of this model, please see Bucklin (2008, pp. 332–334) in the first edition of this volume.

[4]This model is also discussed in detail in Bucklin (2008, pp. 334–335).

B. *Models for the Effects of Display Advertising*

In many situations, linking purchase directly to advertising is infeasible, if not impossible, or simply inappropriate (e.g., for an ad supported media company). For this reason, researchers have recently begun modeling the impact of display advertising on alternative, intermediate outcomes. They are also exploring better measures of display advertising exposure, and working to understand the factors driving display ad effectiveness.

One example of this approach is Rutz and Bucklin (2012), who developed a model of within site browsing behavior as a function of display advertising on the site itself. Using individual level clickstream data from a third-party automotive website, the authors tracked banner ad exposures and the pages visited by users. The data was collected over a four week period in February 2004, and contains a random sample of 250 individuals who initiated an online order during the last two weeks of the month. The ad exposure data contained the quantity and duration of ad exposures, as well as the vehicle make (e.g., Toyota, Ford, GMC, etc.) being advertised. Webpages were categorized based on the vehicle make about which they contained information, with each page containing information on at most one make. Using this data set, they modeled individual i's utility at time t from visiting a webpage for brand j as $u_{i,t,j}$ where:

$$u_{i,t,j} = c_{i,j} + \sum_{k=1}^{K^{Brand}} \alpha_{i,k} f\left(Brand_{i,t,j,k}\right) + \sum_{k=1}^{K^{Brand}} \beta_{i,k} f\left(Ad_{i,t,j,k}\right) + \varepsilon_{i,t,j} \tag{14.1}$$

where,

k	the session ID,
$c_{i,j}$	an individual specific brand intercept,
$\alpha_{i,k}$	an individual-specific response to brand exposures during session k,
$\beta_{i,k}$	an individual-specific response to ad exposures from session k,
$Brand_{i,t,j,k}$	the number of brand exposures during session k,
$Ad_{i,t,j,k}$	the exposure measure for ads from brand j during session k, and
$f(\cdot)$	the functional form by which $Brand_{i,t,j,k}$ and $Ad_{i,t,j,k}$ enter $u_{i,t,j}$.

The authors use a Bayesian Mixture multinomial logit model, which flexibly captures the pattern of differences in ad responsiveness across individuals when estimating parameters. They found evidence of two distinct customer segments of roughly equal size. The first segment browses considerably fewer pages, and is less influenced by exposure to display advertising. The second visits a larger number of pages, and reacts positively to banner advertising. The authors pointed out that more restrictive approaches to modeling user heterogeneity (e.g., latent class or unimodal hierarchical Bayesian approaches) would have produced results indicating that advertising had no impact on consumer browsing patterns.

The authors' specification also allowed them to examine several questions regarding how display advertising exposure should be quantified. They examined whether display advertising exposure $\left(AD_{i,t,j,k}\right)$ is better captured by exposure duration or impression quantity, finding that the former offers a better fit to their data. They also found decreasing marginal returns to the amount of ad exposure, with the logarithmic transformation of ad exposure producing a better fit than either the linear or squared specifications. Finally, they reported that the effects of display ad exposure were relatively short lived, with ads from prior browsing sessions having no discernible impact on current session page choices.

Whether or not internet display ads have carryover effects is an important issue in their efficacy. Though Rutz and Bucklin (2012) found no carryover, this was not the focus of their model. Braun and Moe (2013) further explored the nature of display advertising decay. They developed a model of the decreasing marginal returns and decay rate of display advertising, specifically capturing advertising wear-out and restoration effects. The volume and frequency of display ad impressions can vary dramatically across consumers. For those receiving a large number of impressions in a short period of time, creative and campaign level wear-out may be a significant concern. Wear-out can be mitigated through intelligent ad scheduling, either directly by reducing the ad frequency or indirectly by pausing a campaign and allowing for restoration effects to take hold. While many ad platforms allow advertisers to restrict individual impression rates or pause campaigns, optimally leveraging such settings requires knowledge of both the wear-out and restoration rates.

Using individual level clickstream data, Braun and Moe proposed a system of simultaneous regression equations to measure these effects. To capture the effect of creative and campaign level wear-out and restoration, they modeled advertising using an "ad stock" approach that allows for the accumulation and decay of advertising influence. Ad stock, A_{it}, for individual i at time t is defined recursively as:

$$\begin{aligned} A_{it} = \alpha A_{i,t-1} &+ AD\left[1 - \left(1 - \delta_1^{x_{it}}\right) + R_{1it}\left(1 - \delta_1^{x_{it}}\right)\right] \\ &+ \sum_j C_j\left[1 - \left(1 - \delta_2^{y_{ijt}}\right) + R_{2ijt}\left(1 - \delta_2^{y_{ijt}}\right)\right] \end{aligned} \tag{14.2}$$

where,

A_{it}	the accumulated ad stock for individual i at time t,
α	a geometric decay rate to be estimated,
AD	a baseline advertising campaign effect,
$\left(1-\delta_1^{x_{it}}\right)$	campaign level wearout,
$R_{1it}\left(1-\delta_1^{x_{it}}\right)$	restoration of any campaign level wearout,
C_j	a baseline creative effect,
$\left(1-\delta_2^{y_{ijt}}\right)$	creative specific wearout, and
$R_{2ijt}\left(1-\delta_2^{y_{ijt}}\right)$	restoration of any creative specific wearout.

The number of ad exposures (m_{it}) and site visits (v_{it}) for individual i at time t were modeled as zero-inflated Poisson processes, and conversions (s_{it}) were modeled using a zero inflated binomial distribution. To capture the effect of advertising on consumer behavior, visit rate (μ_{it}) and conversion probability (p_{it}) were modeled as a function of accumulated ad stock and weekly indicators (X_t), and impression rate (λ_{it}) is modeled as a function of weekly indicators (X_t), such that:

$$\log \lambda_{it} = \log \lambda_{0i} + \gamma_\lambda X_t, \tag{14.3}$$

$$\log \mu_{it} = \log \mu_{0i} + \beta_{\mu i} A_{it} + \gamma_\mu X_t, \text{ and} \tag{14.4}$$

$$\operatorname{logit} p_{it} = \operatorname{logit} p_{0i} + \beta_{pi} A_{it} + \gamma_p X_t, \tag{14.5}$$

where λ_{0i}, γ_λ, μ_{0i}, $\beta_{\mu i}$, γ_μ, p_{0i}, β_{pi}, and γ_p are parameters to be estimated.

Leveraging a Bayesian hierarchical structure, the authors allowed the baseline impression and visit rates, the baseline conversion probabilities, and the sensitivities to ad stock to be heterogeneous and correlated. Heterogeneity allowed them to identify individual differences in ad responsiveness versus baseline visit and conversion rates. By jointly modeling these latent parameters, the authors control for unobserved correlation in impression, visit, and conversion rates.

During a 10-week ad campaign for an automotive brand, Braun and Moe collected individual clickstream data on 5,803 randomly selected individuals who were either exposed to one or more ad impressions, visited the brand's website, or both. Within these data, they found substantial decay, such that 37.4% of ad stock effects are lost from week to week. They also found that each additional exposure in the campaign (regardless of creative) results in significant wear-out for all campaign creatives $(1 - \delta_1 = 0.778)$. This means that a second impression in the same week is only 22.2% as effective as the first, regardless of creative. Repeated exposure to the same creative created additional wear-out, but at a slower rate $(1 - \delta_2 = 0.403)$. This means that in addition to the campaign level effects, the second exposure to the same creative in a week is only 59.7% as effective as the first. These wear-out effects may be slowly offset by the restoration effects. During weeks in which an individual is not exposed to advertising, they found that 2.7% of campaign wear-out was restored, and 8.8% of creative wear-out was restored.

The authors concluded that varying creative content shown to individuals can increase both website visits and conversion rates, by 12.7% and 13.8%, respectively. The Braun and Moe modeling approach is useful for estimating the decay and restoration rates for a campaign so that the serving of individual impressions can be modified if deemed necessary to limit wearout. Further, they highlighted the importance of varying the creative content shown to individuals, as the results showed that such variation can increase both website visits and conversion rates.

Sahni (2015) also examined the impact of temporal spacing on display advertising effectiveness. Using data from online field experiments, he explored how the time between prior ad exposures influences the impact of current exposures. Because existing models of advertising carryover cannot capture such effects, he developed a memory based mathematical model of advertising effectiveness based on the Adaptive Control of Thought-Rational (ACT-R) model from cognitive psychology. This model is one description of how the brain is organized to think, and has been shown to result from an optimal information storage and retrieval process. In essence, memories that are more likely to be recalled in the future are stored in more accessible places. This recall rate is shown to be a function of frequency, recency, and temporal spacing. Sahni's research showed that there is significant carryover in display advertising and contemporaneous display impressions are more effective when prior impressions have been spaced farther apart. The author also demonstrated that the carryover effects would have been significantly overstated had temporal spacing not been incorporated in the model.

Sahni posits that the impact of a given impression, on memory and purchase behavior, is related to the memory strength associated with all past impressions from the same campaign. When memory strength for a given campaign is large, the impact of additional advertising decays more quickly. However, this memory strength is also positively related to purchase. The result is a tradeoff between the number of exposures in any given window and the persistence of future advertising effects. The memory strength A_{it}^{M} for individual i at time t is defined recursively as:

$$A_{it}^{M} = m_{it} = \ln\left(\sum_{k=1}^{t-1} (nExp_{ik} \times days_{ikt})^{\exp\left(b_i^M + c_i^M e^{m_{ik}}\right)}\right) \tag{14.6}$$

where,

$nExp_{ik}$	number of exposures in session k for individual i,
$days_{ikt}$	the number of days between session t and k for individual i,
m_{ik}	the memory strength for individual i at session k, and
b_i^M and c_i^M	individual specific parameters defining the relationship between memory strength and the decay rate.

Sahni estimated his model using data from an online field experiment at a restaurant search website. This site allowed users to search for restaurants meeting their criteria, and displayed information such as the menu, detailed ratings, and reviews. In the experiment, users were randomly assigned to either the treatment or control condition each time they started a new session on the website. In the treatment condition, every search page request had a random chance of displaying an ad for a focal restaurant among several other restaurant ads on the right hand side. In the control condition, the focal restaurant ad was replaced by one unrelated to any of the restaurants. The data cover a 3-month period in 2010 and contain 256,690 browsing sessions by 211,135 users. In terms of results, Sahni found that display advertising significantly impacted the probability of purchase. Moreover, accounting for memory strength was shown to be important as ad effects were found to decay more quickly than a standard goodwill or ad stock model would imply.

C. *Accounting for Targeting in Models of Display Ad Effectiveness*

To this point, none of the models we have discussed directly account for the potential selection bias that individually targeted ad impressions may induce. As mentioned above, ad platforms (i.e., ad networks, exchanges, and media buyers) continue to develop increasingly complex, individual level targeting algorithms. This creates two key challenges. First, when advertising is targeted to individuals based on their propensity to respond, correlational estimates of advertising effectiveness can be significantly biased. Second, media buyers generally consider their targeting algorithms proprietary, and provide advertisers with little specific information on their essential details. While an advertiser might be able to specify individual characteristics (e.g., consumer location, past webpages visited, etc.) to guide a campaign, they have little control over or insight into when and to whom specific impressions are served. This makes standard approaches to handling selection bias (e.g., instrumental variables, control functions, etc.) generally inapplicable.

The selection effects associated with unobserved targeting are compounded by the nature of clickstream data, frequently the only data source available to managers evaluating campaign effectiveness. While clickstream data contain detailed records of the interactions between a consumer and the firm's advertising and website, they generally lack information on competitor actions and consumer browsing behavior on third party sites. Lewis et al. (2011) point out that, in many cases, users who browse more pages are more likely to be exposed to a firm's advertising and to undertake many other online activities that may be of interest (e.g., site visit, purchase, etc.), even if the two actions are unrelated. Using data from Yahoo's ad network, they found that this unobserved browsing behavior can bias correlational estimates of display advertising effectiveness by two to three orders of magnitude.

The effectiveness of targeted display advertising has also been called into question. Goldfarb and Tucker (2011) explored the tradeoff between targeted and obtrusive display advertising campaigns. Targeted advertising was defined to be campaigns that matched the product advertised with webpage content. Obtrusive ads were defined as those that strove to be highly visible by leveraging video, creating a pop-up, or having the creative takeover the screen. They found that while using either of the two approaches increased stated purchase intent, when used in combination they were less effective. They attributed this effect to privacy concerns, finding that it was magnified for individuals who declined to provide income information and in categories where privacy is generally of concern (i.e., medical and financial).

Lambrecht and Tucker (2013) also examined targeting effectiveness, specifically focusing on whether retargeted ads are more effective when the content is dynamic or generic. Retargeting is the practice of disproportionately serving display ad impressions to consumers who have previously visited a specific webpage or website. Dynamic copy contains the precise products or services that a consumer viewed on that webpage while generic content only contains the website brand. They found that dynamically retargeted display ads generally underperformed their generic counterparts. However, once a consumer had visited a review site for the product of interest, which was interpreted as a refinement in product preferences, dynamic ads no longer underperformed.

As the foregoing should make clear, selection effects associated with targeting and unobserved browsing behavior can bias correlational estimates of display advertising. These biases can be either positive or negative, and are of unknown magnitude a priori. They also vary with the type of targeting being done, and the type of consumers being reached. Finally, the limited nature of clickstream data creates challenges in recovering unbiased estimates of advertising's impact using standard statistical approaches. This limitation has not gone unnoticed, and several researchers have recently proposed approaches to overcome these challenges.

Hoban and Bucklin (2015) develop one such approach, a model enhanced experimental method. In their experiment, users were randomly assigned to either the treatment or control group during their first digital interaction with the firm, which could occur either through ad exposure or site-visit. Targeting was consistent between the two groups, but the control group was shown ad copy for an unrelated charity in place of the firm's ads. This means that any apparent effect of display advertising in the control group would be a result of the unobserved targeting and browsing behavior. Within the treatment group, the apparent effect of display advertising includes these factors as well as the incremental effect of being exposed to the ad. Thus, display advertising's effect can be measured as the difference between the two groups in the probability of a given outcome (such as a purchase or a website visit).

Building on the findings of Lambrecht and Tucker (2013), Hoban and Bucklin examined display advertising effectiveness as users move through the purchase funnel. Using several funnel milestones, they segmented users into non-visitors, visitors, authenticated users, and converted customers. Non-visitors are users who have never been to the firm's website. Visitors have been to the firm's website, but declined to provide the personally identifying information necessary to create an account. Authenticated users have provided this information. Finally, converted customers have executed a purchase at the website.

Hoban and Bucklin selected site visit as their outcome of interest, though their approach generalizes to any observed, binary outcome. They model display advertising's impact using a binary logit response model containing covariates for the log transformed contemporaneous and lagged impression counts for both firm and charity ads, indicators for funnel stage, interactions between funnel stage and impression counts, and controls for ad timing. Letting $\Pr(visit_{t+k}|\cdot)$ represent the fitted probability of a visit at time $t+k$, display advertising's effect for n impressions, $\Delta_v(n)$, can be calculated as the following difference between the two treatment groups:

$$\Delta_v(n) = \bigcup_{k=0}^{K} Pr\left(visit_{t+k}|x_{t,f}=n, x_{t,c}=0, \beta, \Psi\right) - \bigcup_{k=0}^{K} Pr\left(visit_{t+k}|x_{t,f}=0, x_{t,c}=n, \beta, \Psi\right) \tag{14.7}$$

where,

$x_{t,f}$ the number of firm impressions at time t,
$x_{t,c}$ the number of charity impressions at time t,
Ψ the remaining model covariates,
β a vector of parameters to be estimated, and
K the number of lagged periods included.

Using clickstream data provided by an online financial services firm, the authors tracked 133,058 users over six weeks from February 19th, 2010 through April 2nd, 2010. Importantly, their modeling approach allows them to keep the control group small, accounting for only 1.6% of individuals and 1.5% of impressions. This is a critical consideration for managers, because of the opportunity cost of charity impressions served to the control group. Using these data, the authors found no discernible display ad impact on visitors (i.e., the group that did not progress to creating accounts) but significant effects for the other three purchase funnel stages—non-visitors, authenticated users, and converted customers.

Hoban and Bucklin extended their findings, using elasticities and the Dorfman-Steiner condition, to calculate optimal impression allocations across the funnel stages. This analysis predicted that the firm could increase site visits 10% by shifting ad impressions away from non-visitors and visitors and towards authenticated users. They also compare their results to correlational estimates, finding that the allocation recommended by such a model delivers only a 3% increase in site visits over the status quo. Their approach underscores the dangers of relying on correlational estimates, while offering a relatively inexpensive alternative based on data from a controlled experiment.

Johnson et al. (2017) proposed an alternative approach to measuring display ad effectiveness when ads are sold via exchanges and targeting and browsing behavior may bias correlational estimates. This approach relies on ad exchanges implementing new measurement technology which runs concurrent with their ad serving algorithms. Similar to Hoban and Bucklin, users are randomly assigned to treatment and control groups. The treatment ad is excluded from auctions for control group members, and they are instead served the ad that prevails among the remaining subset. Ideally, display advertising's effect could be measured as the difference in expected outcomes between treatment group users exposed to the ad and control group users who otherwise would have been.

Ad exchanges, however, do not control all aspects of ad exposure—i.e., winning an auction does not guarantee that an ad will be served to a user. For example, certain ads may be technologically incompatible with a given website, or websites may ban certain products or services from being advertised. Thus, to maintain symmetry between the treatment and control groups, the authors proposed a two-stage auction. In a first round, simulated auction, the researchers predict which users, in both the treatment and control groups, would be exposed to the treatment ad, and tag these users as being predicted ad recipients. In a second round auction, ad exposure is determined as described above.

Because the first round, simulated auction may not reflect true ad exposure, the difference in expected outcomes between treatment and control group members predicted to be exposed to the treatment ad is scaled by the probability of treatment ad exposure given that a treatment group member won the first round, simulated auction. Thus, the authors calculate the effect of display advertising as the local average treatment effect (LATE) among users who were predicted to view at least one ad:

$$PGA\,LATE = \frac{E[y|PGA = 1, T = 1] - E[y|PGA = 1, T = 0]}{\Pr[X = 1PGA = 1, T = 1]} \tag{14.8}$$

where,

PGA an indicator for predicted treatment ad exposure,
T an indicator for treatment group membership, and
X an indicator for exposure to the treatment ad.

Using data from an online retailer of apparel and sporting goods, Johnson et al. evaluated the effectiveness of a retargeting campaign that ran during a two week period during the winter of 2014. The campaign served 9 million ad impressions with a budget of $30,500. The experiment assigned 70% of the users to the treatment group, with the remaining 30% assigned to the control group. The authors found that the ad campaign increased site visits by 17.2% and sales by 10.8%, with both measures highly statistically significant.

14.3 Models for Paid Search Advertising

Search engine marketing (SEM) breaks down into two key parts, search engine optimization (SEO) and paid search advertising (paid search), also called sponsored search advertising (SSA). The goal of SEO is to optimize a firm's position in the so-called organic listings provided by search engines, i.e., the search results based on the algorithms search engines use. On the other hand, paid search allows firms to buy a placement in the so-called sponsored or paid listings of the search engine results page (SERP). This allows firms to appear on SERPs even if the search engine does not yet list an organic link, something which could be important for firms new to online marketing or those simply extending their business to new areas. Though SEO is quite important, marketing models developed to date have focused almost entirely on issues pertaining to paid search or SSA.

Our discussion of models for SSA begins with the general form of the profit optimization problem confronting the firm considering paid search advertising. From the profit function the issues that arise in response modeling and optimization can be viewed in practical context. We then turn our discussion to the major recent advances in marketing models for SSA. These can be classified into two groups. The first group of models focuses on the direct effects of paid search advertising in which the productivity of search ad spending is evaluated on its short-run ability to deliver conversions after a click. In the pay per click model of SSA, this is what incurs cost for the advertiser. A conversion can be either an online purchase or another desired outcome (such as a web site visit or sign-up). The second group of models seeks to examine longer-run or indirect effects of paid search. Typically, these are built from a broader perspective of the potential impact of a paid search ad campaign. Some of the indirect effects of paid search include spillover to future paid searches, bookmarking and direct type-in of URLs, and the generation of additional purchases from a newly acquired customer.

In SSA, firms bid to have their text ad appear in the sponsored ads section of the search results page which is displayed to the user in response to a search query. The ranking or position placement of the ad on the page is determined in a modified second price auction in which competing bids are each weighted by a "quality score." This score embodies the past performance of the ad in generating clicks as well as other ad related differences, such as campaign performance and fit of the ad to the search. In the auction, the first or highest position goes to the ad which will, in expectation, provide the search engine with the most revenue based on the expected number of clicks and the bid. Lower positions are then assigned in rank order of this expected value.

The auction mechanism at search engines is not entirely transparent to advertisers, which has become itself a source of major challenges to modelers in this area (e.g., Yao and Mela 2009). Furthermore, search engines provide little in the way of information regarding competing ads that appear on the SERP or what competitors are bidding. Thus, advertisers confront the problem of trying to optimize their paid search campaigns without key pieces of information.

When setting up a paid search campaign, the advertiser needs to make a series of decisions, each of which can affect the returns from SSA. First, the advertiser selects a set of relevant keywords, i.e., search terms, which it deems likely to be used by those potentially interested in the firm's product or service. Second, bids for each keyword need to be set, together with budgets for spending ceilings. Third, text ads—a headline, two lines of body, and a URL—are created. Fourth, the landing pages that are to be linked with the text ad via the URL are created or specified. Much of the modeling work on the direct effects of paid search is designed to aid advertisers in making these decisions. A more detailed discussion of the mechanics of paid search advertising may be found in Rutz and Bucklin (2013). In Table 14.2, we present a summary of models for sponsored search advertising that we discuss in the following sections.

Table 14.2 Summary of models for sponsored search advertising

Article	Modeling objective	Types of model (s) and dependent variable(s)	Data	Key insights
Ghose and Yang (2009)	Estimate impact of cost per click on position, click-through rate, and conversion rate	System model; clicks, conversions, cost, ad rank (position)	Six months of weekly data for a paid search campaign for a major retailer	Higher positions have higher click rates and conversion but cost more; middle positions (rank 4–6) may be best
Agarwal et al. (2011)	Estimate impact of position on click-through and conversion where bids vary exogenously	System model; clicks, conversions, position	Paid search data from a field experiment for an online retailer	Higher positions have higher click rates, not higher conversion; ads with top positions may not be best
Chan et al. (2011)	Incorporate customer lifetime value (CLV) into the analysis of paid search advertising	Pareto/NBD model for CLV; lifetime value, transaction rate, gross margins	Panel data of customers acquired by scientific supply company	Customers acquired through Google search advertising have substantially higher CLV than others
Rutz and Bucklin (2011)	Model and estimate the spillover effect of generic search on subsequent branded search	Dynamic linear model with latent awareness; clicks, conversions	Paid search data for the Google and Yahoo! search engines for a major hotel chain	Click-through on generic search ads spills over to brand searches, makes generic search ads more cost effective
Rutz and Trusov (2011)	Estimate effects of keyword and text ad features on click-through and conversion; use latent instrumental variables for position endogeneity	Binary logits with heterogeneity and correlated errors; clicks, conversions	Detailed ad and paid search campaign data for a ringtones provider	Textual content and design of ads affect click-through; targeting can improve profit
Skiera and Abou Nabout (2013)	Decision support system for keyword bidding	Profit optimization, regression; profit	Searches, clicks, conversions, costs for a set of keywords	Model improved ROI by 21%
Narayanan and Kalyanam (2015)	Study effects of changing ad position on clicks and conversion with full data on AdRank	Regression discontinuity; clicks, conversions	Paid search data across competitors bidding on the same keywords	Effect of higher position on click rate is positive and strongest from rank 2 to 1; little effect of position on conversion

A. *A Profit Function for Paid Search*

The direct or short-run returns of a paid search campaign can be evaluated based on a profit function which captures the margin obtained from user click-through and conversion, given ad impression, and the cost of the clicks. This direct profit may be written as follows:

$$\pi(Direct) = N \times CTR \times (CVR \times M - CPC) \tag{14.9}$$

where N is the number of impressions or ad exposures, CTR is the click-through rate on the impressions, CVR is the conversion rate given click-through, M is the contribution margin from the conversion, and CPC is the cost-per-click. The advertiser can increase or decrease the number of impressions (or times its ad appears on the SERP) by expanding or contracting the set of keywords included in the campaign. By raising bids on keywords, which increases CPC, the advertiser is likely to move its ads into higher ranked positions on the SERP. These higher positions, in turn, usually lead to an increase in CTR, thereby bringing more potential buyers to the advertiser's landing page. In order to realize the margin, M, the user must "convert" from site visitor to buyer, which occurs with the probability CVR. There is, in practice, a trade-off between CPC and CTR, as both are functions of ad position. Conversion rates may also be a function of ad position, though this relationship, if there is one, has been found to vary with the product or service and search terms involved.

The profit equation in (14.9) includes only the margin that can be linked directly to the search ad click. Even though a purchase conversion may not take place following the click-through, it is quite conceivable that it may lead to future purchases or other intangible benefits of value to the firm. The intangibles could include enhancements to brand image and word-of-mouth, for example. The need to expand the profit function to include these indirect returns from paid search has been widely recognized and may be concisely captured as follows (e.g., Skiera and Abou Nabout 2013):

$$\pi(Total) = \pi(Direct) + \pi(Indirect) \tag{14.10}$$

Here, π(*Indirect)* captures the profit contribution from paid search not attributable to direct click-through. As we will see later in this section, modelers have addressed indirect effects by focusing on one major issue at a time, often turning to different modeling approaches and utilizing additional data or insights into the firm's business model. While models for the direct effect are amenable to use across different firms and industries, models for indirect effects currently need more customization and cannot be as easily transferred across domains.

Returning to the direct profit equation in (14.9), optimizing keyword bidding for search ads appears, at first glance, to be a tractable problem. For example, a decision support system for this optimization problem has been developed by Skiera and Abou Nabout (2013). In their paper, they specified a similar profit function and used regression models, at the individual keyword level, to calibrate the relationship between bid and position as well as CTR and position.[5] This enables the optimization routine to adjust bids based on how the ad's position responds to the bid and how CTR responds to position. Skiera and Abou Nabout applied their approach to a set of 20 keywords in the paid search ad campaign of a small company and reported encouraging results in a field test. Returns were increased across the set of 20 keywords, primarily as a result of reduced bids and correspondingly lower CPCs.

B. *Models for the Direct Effect of Paid Search Advertising*

One of the challenges in specifying the response models needed for the optimization of the direct profit equation is that the rank or position of the text ad on the search results page is endogenous. In describing their approach, Skiera and Abou Nabout (2013) recognized this problem but point out the need for fast calibration of the model in order to facilitate real-time optimization of bidding. The endogeneity of position has several sources. First, position is determined as the result of the second-price auction and is not solely a function of the focal advertiser's bid and quality score. As noted earlier, search engines do not provide information on competitor ads or bidding, so it is not generally possible for models to include this information. This leads to the possibility of omitted variable bias. Second, as currently reported by search engines, position is an aggregation of the realizations for the ad, typically over the course of a day. This means that position is measured with error (see, for example, the extensive discussion of this problem by Abhishek et al. 2015). Lastly, advertisers operate within a feedback loop of bids influencing position and realized positions influencing bids, which may give rise to simultaneity problems.

Addressing the endogeneity of position has turned out to be major issue in properly modeling the CTR, CVR, and CPC components needed for the optimization of direct paid search profitability. Several approaches to this problem have been devised and appeared in the marketing literature. The first is to rely upon an exclusion restriction to identify a triangular system, which may then be estimated using standard econometric approaches (e.g., Greene 1999). The second is to develop instruments for position and use those in the estimation. The third is to harness a regression discontinuity approach. We now discuss each of these modeling approaches.

The triangular system of equations approach was first proposed by Ghose and Yang (2009) and also used subsequently by Agarwal et al. (2011). In this approach, identification is obtained when one of the equations does not depend upon an endogenous variable. Ghose and Yang specified a four equation system which they write in simplified form as follows:

$$p = f_1(Rank, X_1, \varepsilon_1) \tag{14.11}$$

[5]In practice, CPC is generally close to or equal to the bid.

$$q = f_2(Rank, X_2, \varepsilon_2) \tag{14.12}$$

$$CPC = f_3(X_3, \varepsilon_3) \tag{14.13}$$

$$Rank = f_4(CPC, X_4, \varepsilon_4) \tag{14.14}$$

Here, p is the click-through probability or CTR, modeled as a binary logit function of the display rank or position of the text ad, a set of exogenous covariates, $X1$, and an error term ε_1. A similar binary logit specification models the conversion probability or CVR (conditional upon click-through), q, also as a function of position and exogenous covariates, X_2. CPC is modeled as exogenously determined and is specified to be a linear function of ad position in the previous period, landing page quality, and other exogenous covariates capturing keyword-specific characteristics. Note that it is these assumptions that are the key to identifying the model in this approach. Lastly, position or *Rank* is a linear function of CPC and exogenous covariates. In each equation, the exogenous covariates include a time trend, the length of the keyword, and indicators for the presence of brand name or retailer name.

Ghose and Yang estimated their system of equations on 6 months of weekly data from a large nationwide retail chain that advertised on Google in 2007. As expected, they found that CTR is negatively related to position (or lower ad ranks which carry a higher position number). CVR was also found to be negatively related to position—i.e., text ads which appear closer to the top rank on the search results page convert to purchases at a higher rate. While Ghose and Yang did not use their model estimates for optimization, they did explore the implications for profitability of different ad ranks. Though conversion rates in their model are found to be highest in the top slot (rank 1), the CPC for the middle and bottom ranks dropped quickly. Thus, in their application, profits were found to be highest in the middle positions (ranks 4–6).

Agarwal et al. (2011) also employed the triangular system of equations approach to estimate a model of click-through, conversion, and position. While Ghose and Yang used the CPC equation as the basis for identification, Agarwal et al. took advantage of a field experiment setting in which bidding for keyword placement was randomized. This enabled them to treat CPC as exogenously determined in their data and to drop the CPC equation from the system. Following their notation, the three equation system is

$$U_{kt}^{CTR} = f\left(Position, X1, \varepsilon_{kt}^{\theta}\right), \tag{14.15}$$

$$U_{kt}^{CONV} = f\left(Position, X2, \varepsilon_{kt}^{\beta}\right), \tag{14.16}$$

$$Position = f\left(X3, \varepsilon_{kt}^{\alpha}\right). \tag{14.17}$$

Here, the first two equations are specified using the latent utility of click-through, U_{kt}^{CTR}, and the latent utility of conversion, U_{kt}^{CONV}, for keyword k at time t. The associated click and conversion probabilities are then modeled with standard binary logit functions as in Ghose and Yang (2009). The covariates in X1, X2, and X3 are exogenous explanatory variables, including quality score, day or week, and time. Position is modeled as a linear function of the (randomized) bid, the quality score, and exogenous covariates.

The model was estimated on paid search data from an online retailer of pet products using daily data for impressions, clicks, and purchases for 68 keywords over a 45-day period in 2009. Agarwal et al. found that CTR is negatively related to position. But, unlike Ghose and Yang, they found that CVR is positively related to position. The striking difference in the result for conversion and rank may be attributed to two factors. First, there are differences in the nature of the data sets. Ghose and Yang studied a large retailer selling across many categories which also had hundreds of bricks and mortar store locations and used weekly data; Agarwal et al. studied an online only retailer selling in a single category and used daily data. Second, Ghose and Yang studied a wide range of positions (from 1 to 131) whereas Agarwal et al. confined their analysis to the top seven positions. Relaxing this constraint in a second data set, Agarwal et al. reported that the strong increasing trend of conversion with ad position disappeared and that conversion became independent of position. It is also worth noting that Google chief economist Hal Varian has pointed out that conversion rates do not vary much with ad position (Friedman 2009). Like Ghose and Yang, Agarwal et al. discussed the profit implications of their model estimates, noting that the most profitable ad positions are not necessarily those at the top of the results page.

A classic approach to dealing with an endogenous predictor is to develop instruments for it—i.e., variables which are correlated with the predictor but uncorrelated with the error. Unfortunately, the nature of position, along with the limitations of the data provided to advertisers by search engines, makes it quite difficult to locate suitable observable instruments. To deal with this limitation, Rutz and Trusov (2011) proposed the use of latent instrumental variables (LIV), which does not require observed instruments to be available. In the LIV approach, latent variables are used to separate position into a portion uncorrelated with the errors and a portion possibly correlated with the errors. Rutz and Trusov define the LIV equation for position of keyword w at time t as follows:

$$pos_{wt} = \omega\gamma_{wt}^{c} + \zeta_{wt}^{LIV} \tag{14.18}$$

where γ_{wt}^{c} is a C × 1 binary vector of C – 1 zeros where the nonzero element designates that keyword w belongs to category c at time t, ω is a 1 × C vector of category weights to be estimated, and ζ_{wt}^{LIV} is the error term. The number of categories is determined in model estimation by examining the separation provided as well as the fit and performance of the model.

Rutz and Trusov used the LIV approach for handling position endogeneity to model click-through and conversion probabilities for a set of 80 keywords in the ringtones category advertised by a collaborating firm in 2007. In their application, they set C = 2, noting that, for C = 3, two of the latent categories lacked good separation. A comparison of results with and without the LIV treatment for position clearly indicated the need to address endogeneity in this key variable for modeling paid search. They also report an extensive set of findings regarding textual attributes of the search ads themselves. This opened up a new set of features over which advertisers can seek to optimize the productivity of paid search spending.

The LIV approach was also applied to handle position endogeneity by Rutz et al. (2012). The authors used a model of click-through and conversion, with position endogeneity handled by LIV in the same manner as Rutz and Trusov (2011). The goal of the Rutz et al. (2012) work was to assess the performance of individual keywords and to provide advertisers with the ability to improve keyword selection. The model was applied to a data set of 301 keywords used in the paid search campaign of a lodging chain in 2004. The authors report that position is negatively related to click-through as well as conversion (i.e., lower ranked ads were less likely to convert to a hotel reservation). Using the model-based estimates for conversion, the authors were also able to improve the assessment of keyword performance, especially when there was sparse conversion information at the keyword level. These estimates were then used as input into keyword selection. The LIV-based model recommended retaining 156 of the 301 keywords and dropping 145. The profit increase was approximately four percent over the status quo (retaining all 301 keywords). As the authors point out, this is only one component needed for the overall optimization of a paid search campaign (e.g., neither bid optimization nor the addition of new keywords were considered, and neither were changes to ad copy or landing pages).

The LIV approach to handling position endogeneity has been validated in more recent work by Abhishek et al. (2015). In this study the authors closely examined the problem of aggregation bias in the daily position information reported to advertisers by the search engines. Notwithstanding the other sources of endogeneity in position, the aggregation of realized position data from the individual impression level to the daily level sets up an errors in variables problem. Unfortunately, as the authors show, the measurement errors involved are not trivial and need to be taken into account in modeling paid search. As a solution to this problem, Abhishek et al. recommended the use of the latent instrumental variables approach to handle the errors-based source of endogeneity in ad position.

While both the triangular system and LIV approach have appeared multiple times in the literature, both have limitations. The triangular system relies upon assumptions that sufficient exogenous variation exists to model position (e.g., from randomized bidding). However, this approach does not address competing bids and their role in determining ad position nor does it account for measurement error in the position data reported by search engines. The LIV approach relies upon assumptions regarding the shape of the distribution of the endogenous variable and the outcome variables (clicks and conversions) which might not hold in practice.

Position effects are also modeled using a particular parametric specification which runs the risk that localized effects (such as moving from position 2 to 1 versus position 6 to 5) might be missed.

In an attempt to overcome the weaknesses of the triangular system and LIV approaches, Narayanan and Kalyanam (2015) modeled position effects using a regression discontinuity approach. Regression discontinuity (RD) estimates causal effects by taking advantage of the quasi experiment that results from forcing a continuous variable into discrete outcomes. Narayanan and Kalyanam recognized that the ordinal outcome of ad position is a good candidate for an RD approach because it is driven by the underlying continuous variable known as *AdRank*. Search engines compute *AdRank* for each advertiser's keywords using the quality score for the ad and the bid. The observed ad positions are determined by sorting the values of *AdRank* across competing advertisers that are bidding on a given user query. The authors proposed that the effect of a change in rank can be estimated by analyzing search ads in adjacent positions with AdRank values that are very close.

Narayanan and Kalyanam estimate the effect of a change in position on an outcome y_j, such as CTR, using the following regression equation:

$$y_j = \alpha + \beta \cdot 1\left(pos_j = i + 1\right) + \gamma_1 \cdot z_j + \gamma_2 \cdot z_j 1\left(pos_j = 1 + 1\right) + f(j \colon \theta) + \varepsilon_j. \quad (14.19)$$

In the above, $1\left(pos_j = i + 1\right)$ is an indicator for whether the ad is in the higher position of two matched ads. Thus, β captures the local position effect (e.g., of going from rank 2 to rank 1). The variable z_j is the difference between the *AdRank* values of the two ads in adjacent positions. The parameters γ_1 and γ_2 control for any variation in CTR that might occur due to differences in *AdRank*; note that the interaction effect allows this to differ on either side of the threshold. The $f(j \colon \theta)$ term includes fixed effects where θ is the parameter vector. A key element of any regression discontinuity approach is setting the amount of variation in the underlying forcing variable that will be allowed. Smaller values of z_j make it more likely that the ads are otherwise equivalent but, on the other hand, reduce the number of observations that can be included in the sample for analysis. In this case, the authors determine the allowable range using a "leave one out cross validation" approach which selects the range to minimize the squared error loss in prediction for the hold out observations.

To implement an RD design as specified by Narayanan and Kalyanam the modeler will need to have information on *AdRank* and position outcomes for competing advertisers. As noted above, this information is typically not available to advertisers, as search engines do not disclose competitor bids, quality scores, or report competitor positions. To estimate their RD model of position effects, the authors take advantage of a unique data set that includes information for four competing firms, comprising the bulk of the search advertising for a consumer durables category. (The data became available because one of the competitors acquired the other three firms and, in so doing, obtained the Google data on competing ads.) Due to the RD specification, the authors are able to estimate position effects on CTR, orders, and conversion across a range, looking at the effect of going from 2 to 1, 3 to 2, 4 to 3, and so forth. The estimation results for CTR are in line with previous findings and show a general decrease in CTR as position increases. This effect appears to be strongest, however, in going from position 1 to position 2. For orders and conversion, the authors report no significant effects of position change on these outcomes with the exception of going from position 6 to 5. The authors note that this latter effect may coincide with the page fold in the search results.

The regression discontinuity approach adds another tool in the modeler's ability to handle the challenges involved in understanding the effect of position on CTR and CVR. Because of the need for competing advertiser information, its usefulness to the individual advertiser for search campaign management is limited. On the other hand, the search engines need to understand the influence that position changes have on outcomes to refine their internal formulas for determining *AdRank*, the so-called "secret sauce" of paid search. Because the search engine has all of the required information, this may enable them to use the RD approach in lieu of costlier controlled experiments.

A. *Models for the Indirect Effects of Paid Search Advertising*

In the direct or short-term modeling approaches discussed above, a click that does not result in a conversion action incurs costs for the advertiser but provides no benefits. Taking a broader perspective on consumer search and purchase behavior opens possibilities for indirect benefits to be realized. For example, a consumer could decide to return to the site at a later point in time or purchase using a different channel, e.g., a brick-and-mortar store. The elements of the direct profit function specified above are standard and allow researchers to develop models and insights for optimization that transfer across product categories and firms. On the other hand, modeling the indirect effect of paid search has been, at least so far, the result of inquiry into specific situations. Unlike models for the direct effect, the generalizability of the approaches developed to model indirect effects has yet to be established.

Spillover from Generic Keywords to Branded Keywords

A branded keyword includes the trademarked brand name of the advertiser (e.g., Hilton Hotels), while a generic keyword does not (e.g., hotels Los Angeles). Searches for branded versus generic keywords are likely to reflect different immediate consumer goals and encounter different levels of competitive bidding. Rutz and Bucklin (2011) examined a dataset for a US hotel chain and found very different performance for branded and generic keywords. Branded keywords seemed to perform significantly better than generic keywords on all dimensions: click-through rates (13.7% vs. 0.26%), conversion rates (6.0% vs. 1.0%), CPC ($0.18 vs. $0.55) and cost-per-reservation ($2.94 vs. $55). This could be a misleading picture, however, if consumer search and decision making for lodging occurs over time. While a generic keyword search may not convert to a reservation, it does bring the consumer to the hotel's website where a subsequent reservation could be made later, perhaps following a branded keyword search to help the user navigate back to the web site.

Motivated by the hotel chain's problem, Rutz and Bucklin (2011) developed a model to capture the spillover from generic paid search to branded paid search using the daily data made available to advertisers by the search engines. Their approach harnesses the goodwill or ad stock specification (Nerlove and Arrow 1962) to capture the awareness that generic search activity creates, allowing it to, in turn, affect branded search activity in an econometric system model. They specify the awareness stemming from generic search activity using the following equation:

$$A_t = \beta^{gen} GEN_t + \alpha^A A_{t-1}, \tag{14.20}$$

where A_t is awareness at time t, GEN_t is a vector of generic search activity variables at time t and α^A is the carryover rate of awareness (higher values indicate slower decay). In their model, Rutz and Bucklin specify GEN_t to include daily impressions and clicks for generic keywords.

This awareness then enters equations for branded search impressions, click-through, and conversion in a system model that also includes equations for position and CPC. In this manner, the stock of awareness from generic search activity is allowed to influence branded search activity. The number of daily branded search queries involving the lodging chain's name, NS_t, were modeled as

$$NS_t = \beta^{NS} I_t - \alpha^{NS} NS_{t-1}^{imp} + \gamma^{NS} A_t + \varepsilon_t^{NS} \tag{14.21}$$

where I_t is a vector of indicator variables for day of week and month (to capture seasonality) and γ^{NS} captures the spillover effect from generic search via awareness. Analogous equations are specified for CTR and CVR.

Estimating the model as a Bayesian Dynamic Linear Model (DLM) on data from the Google search engine, the authors found that generic search activity significantly increased branded search activity via its effect on the latent awareness variable. Specifically, awareness was found to influence the number of branded searches, though it had no effect on click-through rate or conversion rate. In addition, the generic search activity that significantly influenced awareness came from clicks, not from impressions—i.e., users must click-through to the web site from a generic keyword search in order to trigger the spillover effect. The authors also reported similar results from estimating the model on data from Yahoo! Based on the model estimates of this process, adjustments to the value of generic search can be made which should be useful in bidding for generic keywords.

Spillover to Direct Type-In Future Site Visits

While Rutz and Bucklin (2011) focused on the spillover from generic paid search to branded paid search in generating future site visits, paid search activity can also spillover to future site visits if users return by directly typing the site URL into their browser or by clicking on a saved bookmark. These "direct type-in" visitors can be an important source of traffic for a web site, and, if paid search is responsible for initially bringing them to the site, such spillover should be accounted for in assessing the productivity of paid search ad spending. Rutz et al. (2011) study data from an automotive website which tracked, on a daily basis, the number of visits sourced directly versus those due to paid or organic search. The firm also tracked the specific keywords that led to visits sourced from paid search. This raised two modeling questions. First, could the proportion of direct type-in traffic due to previous paid search visits be estimated? Second, could the model link direct type-in visits to paid search visits at the keyword level so as to aid in keyword selection? The latter is especially challenging given that the firm maintained thousands of keywords but kept only a few months of historical data, setting up a "small n, large p" problem.

Rutz et al. (2011) show how to address these modeling issues using a time series approach accompanied by shrinkage estimation using a Bayesian Elastic Net. The key equation in their modeling approach specifies the number of direct type-in visits, $DirV_t$, as follows:

$$\begin{aligned} DirV_t = \alpha + \sum_{i-1}^{6} \beta_i^{day} I_i^{day} + \sum_{i-1}^{4} \beta_i^{month} I_i^{month} + \lambda DirV_t^{ESmo} + \delta OrgV_t^{ESmo} \\ + \sum_{k=1}^{K} \phi_k PsV_t^{ESmo,k} + \varepsilon_t. \end{aligned} \tag{14.22}$$

Here, $\lambda DirV_t^{ESmo}$ and $\delta OrgV_t^{ESmo}$ are the effects of exponentially smoothed ($ESmo$) values of previous direct type-in and organic search visits, respectively. The first two summation terms account for seasonality due to the day of week and month. The effect of previous paid search visits enters through the last term before the error, $\sum_{k=1}^{K} \phi_k PsV_t^{ESmo,k}$. Note that this is specified to allow for each individual keyword k to have a potentially different effect on $DirV_t$. Because there are far more keywords than data points in the daily time series, the authors turn to a shrinkage-based approach to develop estimates for the ϕ_k coefficients, and, in particular, employ a Bayesian Elastic Net (BEN).

The results from estimating the BEN time-series model show that paid search visitors do return to the site (often multiple times) and that keywords differ in their ability to generate such return traffic. For the keywords which are significant in generating return visits (599 out of 3186), the return visits averaged 3.3 per paid click. The best keywords for generating repeat visitors were the firm's brand name, car brand names, general terms relating to search (e.g., "search", "information", "comparison") and general terms related to web use (e.g., "online", "web"). Very specific keywords (e.g., "BMW 325i sports package") or those including general terms related to price and general terms related to used cars generated fewer return visits.

Paid Search and Customer Lifetime Value

For advertisers who seek to acquire new customers who are likely to buy repeatedly in the future and/or subscribe to a service, customer lifetime value (CLV) can be an important input into ad spending decisions. This raises the question whether customers acquired via paid search are more or less valuable to the firm than customers acquired through other advertising channels. This question was studied in depth by Chan et al. (2011) who examined data for a small U.S. B2B firm in the biomedical and chemical lab supplies business. The authors' dataset linked customer transactions with acquisition channels, enabling the CLV of customers acquired through paid search to be compared with those acquired through other channels. The authors point out that real acquisition costs for the paid search ad channel is not only what was spent on the click(s) that led to the customer's purchase conversion but also the costs for the clicks that did not convert.

Using a Pareto/NBD model to capture the "lifetime" value of each customer, Chan et al. estimate the CLV by customer and acquisition channel. The authors found that the "Google" customer dominates the "non-Google" customer from a CLV perspective. The authors also looked at CLV net of customer acquisition costs and account for offline purchases. It turned out that each customer acquired from Google nets the company an average of $1,280 on a lifetime basis. This confirmed that using paid search as a customer acquisition tool can be highly profitable and raises the question whether the costs of paid search ads might be bid up by competitors. Addressing this, the authors calculate a break-even CPC of $13.56 which was significantly greater than the CPC of $0.80 at the time of the study.

Field Experiments in Paid Search Advertising

Due in part to the challenges involved in estimating both the direct and indirect effects of paid search advertising, researchers have recently begun turning to large scale field experiments. This follows the already widespread use of such experiments to assess the effects of display advertising and takes advantage of random assignment to address the shortcomings—and endogeneity risk—in the data available to managers regarding the performance of their paid search ad campaigns. One straightforward way to set up a controlled experiment on sponsored search advertising is to pause an ad campaign in selected geographic regions using the location targeting options provided by the search engines. Another approach is to pause an ad campaign on one search engine while continuing it on others.[6] The results for the two conditions can then be compared to compute the lift attributable to paid search.

An example of the pause approach is the work by Blake et al. (2015), who reported findings from experiments conducted at eBay. They divided their results into those pertaining to searches for branded keywords (i.e., queries which included "eBay") and non-branded (or generic) keywords. Strikingly, their tests for branded keywords revealed essentially no difference in the number of customers reaching the eBay website; i.e., all of the traffic previously clicking-through on a paid link diverted to an organic link. This led eBay to discontinue all of its spending on

[6]Because search ad costs are triggered by customer queries and click-throughs, there is no easy way to use placebo ads as a control condition as is done with display ad tests.

branded keyword search. For generic keywords, the authors reported that the traffic lift due to paid search is far lower than the click-through counts implied by the search engine data and indeed not significantly different from zero on an overall basis. Investigating further, Blake et al. examined these effects for different segments of eBay users. Search advertising was most effective for customers new to eBay and for infrequent eBay purchases, supporting the notion that search advertising plays an informative role.

The Blake et al. work clearly demonstrates the value that search advertisers may gain from conducting experiments. But while it provided striking results, it is not clear to what extent the findings generalize beyond eBay. This question has been partially addressed by the recent large-scale experimental work on branded keyword search conducted by Simonov et al. (2015). In cooperation with Microsoft's Bing search engine, the experiment manipulated the number of sponsored ads to appear in response to a user query for a branded keyword and did so for several thousand advertisers. In one condition no sponsored ads were allowed to appear (the 0 condition) so users could only click-through to the site via an organic link. By comparing the 0 condition with the 1 condition (only the brand's sponsored link appeared), the experiment revealed the extent to which paid search for branded keywords provides incremental clicks versus organic-only links. The authors reported that the overall lift or increment is approximately 2–3%. While that is a positive result versus the null one from the eBay study, the authors also found that paid links "crowd out" about half of the click-through that would have otherwise occurred via organic links. Given the cost of those paid clicks, it may still be difficult to justify spending on branded keywords. An exception to this occurs for advertisers who face competing bidders for branded keywords (e.g., Alaska Airlines bidding on the Southwest Airlines keyword). In this case, it may be important to maintain spending on branded keywords to defend against competitive encroachment.

Still another type of experimental approach has been recently proposed by Kalyanam et al. (2015) in a study of paid search across 13 multi-channel US retailers. The experimental manipulation was to increase (or "heavy-up") the spending on paid search by expanding the number of keywords that each retailer bid on and maintaining those ads in the top ranked positions. The authors used a meta-analytic approach to analyze results across the various campaigns implemented through a Hierarchical Bayes regression. This enabled them to control for retailer-specific effects and allow for differences in precision in the results obtained for each retailer. Their findings show a strong and significant increase in overall retail sales for the participating retailers due to the increase in search advertising. While this experiment employed a "heavy-up" approach to search ad spending, presumably the method could also be used to analyze reductions in search ad spending, as opposed to resorting to complete pausing. In sum, the experimental approaches described in these papers provide valuable new tools to aid managers in managing paid search advertising and, like display advertising experiments, models will no doubt be play important roles in analyzing results from complex, large-scale experiments.

14.4 Models Integrating Search and Display Advertising

The modeling approaches we have discussed so far focus on either search or display advertising. Since most advertisers do not restrict their spending to either paid search or display alone, an important next step for modelers is to integrate the two forms of advertising into a single framework. Doing so should enable managers to better gauge the relative productivity of the two ad formats as well as potentially understand whether or not there are synergies or cross effects between search and display advertising.

Dinner et al. (2014) propose a modeling approach that brings search and display advertising together in an econometric framework. They study the effect of search and display ads on both online and offline sales while also incorporating potential cross effects (e.g., display advertising can affect paid search impressions and click-through) of the advertising. They specify a multi-equation system model of sales, both online and offline, as a function of online display, sponsored search, and traditional (offline) advertising and apply it to data from a high-end clothing and accessories retailer. Importantly, their econometric approach accounts for endogenous advertising decisions, dynamic advertising effects, autocorrelation in outcomes, and competitive effects.

The Dinner et al. model contains four principal equations, one each for online sales, offline sales, sponsored search impressions, and sponsored search click-through rates. Among other variables, online and offline sales are influenced directly by the accumulated ad stock, or goodwill, associated with traditional advertising, online display advertising, and search click-throughs. Analogous to work in display and search discussed above, the ad stock for medium j in market m at time t is modeled as:

$$AdStock_{jmt} = \lambda_j AdStock_{jmt-1} + \left(1 - \lambda_j\right) Advertising_{jmt}. \tag{14.23}$$

In their system, search impressions and click-throughs are a function of accumulated ad stock from traditional and online display ad spending, as well as contemporaneous spending on sponsored search advertising. Under this formulation, contemporaneous spending on sponsored search advertising indirectly affects sales via paid search impressions and clicks. Meanwhile, traditional and online display advertising are allowed to impact sales directly by bringing consumers to a store or website and indirectly via their effect on customer search and click-through rates.

Dinner et al. specify their system of four main equations as follows:

$$\begin{aligned} lnSearchImpressions_t &\\ = \beta_{1,0} &+ \beta_{1,1} TraditionalAdStock_t + \beta_{1,2} OnlineDisplayAdStock_t \\ &+ \beta_{1,3} OnlineSearchSpend_t + \beta_{1,4} ChristmasDummy_t + \beta_{1,5} Trend_t \\ &+ u_{1,t}, \end{aligned} \tag{14.24}$$

$$\begin{aligned} &lnSearchClickThroughRate_t \\ &\quad = \beta_{2,0} + \beta_{2,1} TraditionalAdStock_t + \beta_{2,2} OnlineDisplayAdStock_t \\ &\qquad + \beta_{2,3} OnlineSearchSpend_t + \beta_{2,4} ChristmasDummy_t + \beta_{2,5} Trend_t \\ &\qquad + u_{2,t}, \end{aligned} \tag{14.25}$$

$$\begin{aligned} &lnOfflineSales_t \\ &\quad = \beta_{3,0} + \sum_{m=1}^{M-1} \beta_{3,0,m} D_m + \beta_{3,1} TraditionalAdStock_{mt} \\ &\qquad + \beta_{3,2} OnlineDisplayAdStock_{mt} + \beta_{3,3} SearchClickthroughStock_{mt} \\ &\qquad + \beta_{3,4} CompetitorTraditionalAdvertising_{mt} + \beta_{3,5} CompetitorOnlineDisplayAdvertising_{mt} \\ &\qquad + \beta_{3,6} ChristmasDummy_t + \beta_{3,7} LargePromotion_{mt} \\ &\qquad + \beta_{3,8} SmallPromotion_{mt} + \beta_{3,9} Clearance_{mt} + \beta_{3,10} Trend_t \\ &\qquad + \beta_{3,11} UnemploymentRate_{mt} \\ &\qquad + \beta_{3,12} DummyTestBannerAdvertising_{mt} + u_{3,mt}, \text{ and} \end{aligned} \tag{14.26}$$

$$\begin{aligned} &lnOnlineSales_t \\ &\quad = \beta_{4,0} + \sum_{m=1}^{M-1} \beta_{4,0,m} D_m + \beta_{4,1} TraditionalAdStock_{mt} \\ &\qquad + \beta_{4,2} OnlineDisplayAdStock_{mt} + \beta_{4,3} SearchClickthroughStock_{mt} \\ &\qquad + \beta_{4,4} CompetitorTraditionalAdvertising_{mt} + \beta_{4,5} CompetitorOnlineDisplayAdvertising_{mt} \\ &\qquad + \beta_{4,6} ChristmasDummy_t + \beta_{4,7} LargePromotion_{mt} \\ &\qquad + \beta_{4,8} SmallPromotion_{mt} + \beta_{4,9} Clearance_{mt} + \beta_{4,10} Trend_t \\ &\qquad + \beta_{4,11} UnemploymentRate_{mt} \\ &\qquad + \beta_{4,12} DummyTestBannerAdvertising_{mt} + u_{4,mt}. \end{aligned} \tag{14.27}$$

In these equations, the ad stock variables are measured as noted above, and

$OnlineSearchSpend_t$	the amount spent on search advertising in period t,
$ChristmasDummy_t$	an indicator for the period from Thanksgiving to Christmas,
$Trend_t$	a linear trend term,
$CompetitorTraditionalAdvertising_{mt}$	traditional advertising spend by the firm's main competitor in market m at time t,
$CompetitorOnlineDisplayAdvertising_{mt}$	online display spend by the firm's main competitor in market m at time t,
$LargePromotion_{mt}$	an indicator for a large promotion was taking place in market m at time t,
$SmallPromotion_{mt}$	an indicator for a small promotion was taking place in market m at time t,

$Clearance_{mt}$	an indicator for a clearance sale in market m at time t,
$UnemploymentRate_{mt}$	the unemployment rate in market m at time t, and
$DummyTestBannerAdvertising_{mt}$	an indicator for a small number of market specific display advertising campaigns.

To recover consistent, unbiased estimates, the authors account for endogeneity in the advertising variables and potential correlation in the error terms, both over time and across outcomes. To handle endogeneity in advertising, the authors use ad expenditures from eight lower-end retailers as instruments for the focal retailer and its main competitor. They argue that changes in advertising costs will create exogenous variation in advertising spending, but that advertising for lower-end retailers is unlikely to influence sales at the higher-end focal firm. To control for autocorrelation in the outcomes, the authors allow the errors at time t to be correlated with the errors at time $t-1$ through the equation $u_{it} = \rho_i u_{it-1} + \varepsilon_{it}$ for Eqs. (14.24) and (14.25) and $u_{i,mt} = \rho_i u_{i,mt-1} + \varepsilon_{i,mt}$ for Eqs. (14.26) and (14.27). They control for potential contemporaneous correlations in the error terms of Eqs. (14.24) through (14.27) by allowing the error terms to covary such that $cov(\varepsilon_{i,t}, \varepsilon_{j,t}) = \sigma_{ij}, cov(\varepsilon_{i,t}, \varepsilon_{j,mt}) = \sigma_{ijm}$, and $cov\left(\varepsilon_{i,mt}, \varepsilon_{j,m't}\right) = \sigma_{ijmm'}$.

To estimate their model, Dinner et al. combined data from several sources covering the 2-year period from September 2008 to August 2010. They collected weekly, market-level sales data from the focal retailer in 25 markets. They combined this with weekly national advertising data that provides the number of paid search impressions and clicks, online display advertising spend, and offline advertising spend. They gathered weekly data on competitor advertising spend, both online and offline, from Kantar Media's Ad$pender. Finally, they collected market level economic data from the Bureau of Labor Statistics. The national advertising data was scaled proportionally to market level GDP to create 2,575 weekly, market-level observations (25 markets × 103 weeks).

The authors reported several important empirical findings. First, the long-term elasticities of online advertising (both display and paid search clicks) exceed those for traditional advertising. This holds for both online and offline sales. Second, cross-effect elasticities are large and significant, ranging from 37 to 87% of own-effect elasticities. Third, traditional advertising is negatively related to click-through rates, which the authors attribute to an information substitution effect. Finally, ignoring cross effects would have resulted in understating advertising ROI by 9% for traditional advertising, 88% for sponsored search advertising, and 91% for online display advertising. This highlights the importance of incorporating these factors into integrated digital advertising models. Failing to do so risks significantly understating the sales response to advertising, especially online advertising.

14.5 Conclusion

Though most of the marketing models available to aid managers in understanding and managing digital advertising have only recently appeared in the literature, substantial progress has already been made. In display advertising, researchers have developed approaches to capture the effects of banner ad exposure on both sales (Manchanda et al. 2006) as well as intermediate outcomes of interest such as page views or web site visitation (Rutz and Bucklin 2012). Because display advertising includes a creative component, it is not sufficient to merely examine exposures or spending levels but models now also can incorporate the effects of changing ad content and message strategy (Braun and Moe 2013). Modelers of display advertising effects have recently begun to address the widespread use of targeting and the likely correlations among display ad exposure and other online activities, both of which can create potentially severe selection bias in the estimation of display ad response. Future work is likely to make further progress in using models to better understand the outcomes of large-scale controlled field experiments. Harnessing models in this respect should also enable firms to control the cost of such experiments by minimizing the exposures needed in the control ad condition (Hoban and Bucklin 2015).

In sponsored search advertising a similar surge of recent progress has occurred. To model the direct effects of paid search advertising—and to move towards optimization on that basis—researchers have developed approaches to overcome some of the key limitations in the data that search engines provide to managers. In particular, the endogeneity in ad position has proved challenging to handle due to the absence of competitor information. Encouragingly, the use of triangular equation systems (e.g. Ghose and Yang 2009) and latent instrumental variables (e.g., Rutz and Trusov 2011) have both been shown to be successful in multiple studies involving each method. Regression discontinuity represents another rigorous approach to handling position endogeneity (Narayanan and Kalyanam 2015) but the data needed to implement it is, unfortunately, not routinely available. Nonetheless, it can serve as an important benchmark for and validation of the latent instrument and triangular system approaches.

Modeling indirect effects of paid search has also progressed rapidly and modelers have developed approaches to account for several important additional aspects. A click-through on a paid search ad is a choice action which brings consumers to the advertiser's web site—potentially exposing them to significant additional information. Though no concrete action may take place at that particular time, the consumer may seek to return in the future and can do so in various ways, including a search on the advertiser's name or directly typing (or having bookmarked) the URL into their browser (Rutz and Bucklin 2011; Rutz et al. 2011). The modeling work on these effects has shown that the indirect impact can be substantial. Also important is the ability of search advertising to bring new customers to the firm who, in turn, become repeat purchasers (Chan et al. 2011). Advertisers need to be able to incorporate these spillover effects to better gauge the productivity of search ad spending. Analogous to the recent experiment-based models in display advertising, researchers are also beginning to turn to controlled field experiments to assess the effects of paid search ads (e.g., Blake et al. 2015). With controlled experiments, investigators have an important alternative to models that are based solely on search engine data and another means to address the endogeneity problems. Though existing work is nascent, we expect to soon see more papers in paid search based on controlled experiments with models used to help interpret them.

Models of internet advertising also need to incorporate both search and display advertising—not to mention offline advertising and other digital media. The limited published modeling work that has addressed this need produced findings which highlight the crossover roles of traditional, display, and search advertising on both online and offline sales (Dinner et al. 2014). Though this study used weekly data collected at a high level of aggregation, it remains to be seen whether modelers can specify advertising measures using the daily and individual-level granularity that is available while also including them in the same model specification.

References

Abhishek, V., K. Hosanagar, and P.S. Fader. 2015. Aggregation bias in sponsored search data: The curse and the cure. *Marketing Science* 34 (1): 59–77.

Agarwal, A., K. Hosanagar, and M.D. Smith. 2011. Location, location, location: An analysis of profitability in position in online advertising markets. *Journal of Marketing Research* 48 (6): 1057–1073.

Blake, T., C. Nosko, and S. Tadelis. 2015. Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica* 83 (1): 155–174.

Braun, M., and W. Moe. 2013. Online display advertising: modeling the effects of multiple creatives on individual impression histories. *Marketing Science* 32 (5): 753–767.

Bruell, A. 2014. *Digital Ad Spending to Pass TV in U.S. by 2017, New Forecast Says*. 8 December 2014. http://adage.com/article/agency-news/digital-ad-spending-pass-tv-2017-magna-global/296090/. Accessed 31 Aug 2015.

Bucklin, R. 2008. Marketing models for electronic commerce. In *Handbook of marketing decision models*, ed. Berend Wierenga, 1st ed., 327–372. Springer.

Chan, T., C. Wu, and Y. Xie. 2011. Measuring the lifetime value of customers acquired from Google search advertising. *Marketing Science* 30 (5): 837–850.

Chatterjee, P., D.L. Hoffman, and T.P. Novak. 2003. Modeling the clickstream: Implications for web-based advertising efforts. *Marketing Science* 22 (4): 520–541.

Dinner, I.M., H.J. van Heerde, and S.A. Neslin. 2014. Driving online and offline sales: The cross-channel effects of traditional, online display, and paid search advertising. *Journal of Marketing Research* 51 (5): 527–545.

Drèze, X., and F.X. Hussherr. 2003. Internet advertising: Is anybody watching? *Journal of Interactive Marketing* 17 (4): 8–23.

eMarketer. 2016. *More Than Two-Thirds of US Digital Display Ad Spending Is Programmatic*. 5 April 2016. http://www.emarketer.com/Article/More-Than-Two-Thirds-of-US-Digital-Display-Ad-Spending-Programmatic/1013789. Accessed 20 June 2016.

Friedman, D. 2009. Conversion Rates Don't Vary Much by Ad Position. http://adwords.blogspot.com/2009/08/conversion-rates-dont-vary-much-with-ad.html. Accessed 16 Oct 2012.

Ghose, A., and S. Yang. 2009. An empirical analysis of sponsored search in online advertising. *Management Science* 55 (10): 1605–1622.

Goldfarb, A., and C. Tucker. 2011. Online display advertising: Targeting and obtrusiveness. *Marketing Science* 30 (3): 389–404.

Greene, W. 1999. *Econometric Analysis*, 4th ed. Upper Saddle River, NJ: Prentice Hall.

Hoban, P.R., and R.E. Bucklin. 2015. Effects of internet display advertising in the purchase funnel: Model based insights from a randomized field experiment. *Journal of Marketing Research* 52 (3): 375–393.

Johnson, G.A., R.A. Lewis and E.I. Nubbemeyer. 2017. Ghost Ads: Improving the Economics of Measuring Online Ad Effectiveness. *Journal of Marketing Research* forthcoming.

Kalyanam, K., J. McAteer, J. Marek, J. Hodges and L. Lin. 2015. Cross Channel Effects of Search Engine Advertising on Brick and Mortar Retail Sales: Meta-Analysis of Large Scale Field Experiments on Google.com, working paper, Santa Clara University, September.

Lambrecht, A., and C. Tucker. 2013. When does retargeting work? Information specificity in online advertising. *Journal of Marketing Research* 50 (5): 561–576.

Lewis, R.A., J.M. Rao, and D.H. Reiley. 2011. Here, There, and Everywhere: Correlated Online Behaviors. In *Proceedings of world wide web conference 2011 research papers*, 157–166. Hyderabad, India: ACM.

Manchanda, P., J.P. Dube, K.Y. Goy, and P. Chintagunta. 2006. The effect of banner advertising on internet purchasing. *Journal of Marketing Research* 43 (1): 98–108.

Narayanan, S., and K. Kalyanam. 2015. Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Science* 34 (3): 388–407.

Nerlove, N., and K. Arrow. 1962. Optimal advertising policy under dynamic conditions. *Economica* 29 (114): 129–142.

PricewaterhouseCoopers. 2016a. *IAB Internet Advertising Revenue Report 2015 Full Year Results,* New York: Interactive Advertising Bureau, 21 April. http://www.iab.net/media/file/IAB_Internet_Advertising_Revenue_FY_2014.pdf. Accessed 20 June 2016.

PricewaterhouseCoopers. 2016b. *IAB Internet Advertising Revenue Full Year 2015,* Interactive Advertising Bureau, 21 April. http://www.iab.com/wp-content/uploads/2016/04/FY2015-PwC-Matt-Hobbs-Presentation-Global-Trends.pdf. Accessed 21 June 2016.

Rutz, O.J., and R.E. Bucklin. 2011. From generic to branded: A model of spillover in paid search advertising. *Journal of Marketing Research* 48 (1): 87–102.

Rutz, O.J., and R.E. Bucklin. 2012. Does banner advertising affect browsing for brands? Clickstream model says yes, for some. *Quantitative Marketing and Economics* 10 (2): 231–257.

Rutz, O., and R.E. Bucklin. 2013. Paid search advertising. In *Advanced database marketing: Innovative methodologies & applications for managing customer relationships*, ed. K. Coussement, K.W. De Bock and Scott A. Neslin, 229–243. Gower.

Rutz, O.J., and M. Trusov. 2011. Zooming in on paid search ads—an individual-level model calibrated on aggregated data. *Marketing Science* 30 (5): 789–800.

Rutz, O.J., M. Trusov, and R.E. Bucklin. 2011. Modeling indirect effects of paid search advertising: which keywords lead to more future visits? *Marketing Science* 30 (4): 646–665.

Rutz, O.J., R.E. Bucklin, and G.P. Sonnier. 2012. A latent instrumental variables approach to modeling keyword conversion in paid search advertising. *Journal of Marketing Research* 49 (3): 306–319.

Sahni, N. 2015. Effect of temporal spacing between advertising exposures: evidence from an online field experiment. *Quantitative Marketing and Economics* 13 (3): 203–247.

Simonov, A., C. Nosko, and J. Rao. 2015. Competition and Crowd-out for Brand Keywords in Sponsored Search, working paper, University of Chicago (Booth), September 30.

Skiera, B., and N. Abou Nabout. 2013. Practice prize paper—PROSAD: A bidding decision support system for profit optimizing search engine advertising. *Marketing Science* 32 (2): 213–220.

Yao, S., and C. Mela. 2009. Sponsored search auctions: Research opportunities in marketing. *Foundations and Trends in Marketing* 3 (2): 75–126.

Chapter 15
Advertising Effectiveness and Media Exposure

Peter J. Danaher

15.1 Introduction

In the first edition of this handbook, Danaher (2008) contributed a chapter on advertising models. In that chapter a number of concerns about advertising were listed and predicted to worsen over the next decade. Well, that decade has arlmost finished and so we begin this chapter by repeating the concerns about advertising that were written in 2007 and published in 2008. These concerns were:

1. accurately gauging the effectiveness of advertising at generating sales or market share response
2. decreasing television viewing to major networks which reduces the ability to reach a mass audience
3. increasing advertising avoidance by consumers
4. an increasing number of available marketing communication channels.

Without wishing to claim clairvoyance, almost a decade later these concerns have proven to eventuate. Television audiences have declined, as have newspapers. Ad avoidance has increased (Schweidel and Kent 2010) and there are many additional marketing communication channels (Danaher et al. 2015). On a positive note, there have been several illuminating studies that have examined advertising effectiveness. Consequently, most of this chapter is devoted to advertising effectiveness studies from the past 10 years. These studies are those that cover multiple media, as few campaigns today are in a single medium. The first subsection considers models that link advertising to sales using single-source individual-level data, as proposed by Danaher and Dagger (2013). With the growth in multichannel buying, in particular offline and online sales, there is also increasing interest in how

P.J. Danaher (✉)
Marketing Department, Monash University, Melbourne, Australia
e-mail: peter.danaher@monash.edu

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_15

different advertising media channels perform in terms of driving sales in different channels (Dinner et al. 2014). Therefore, we conclude this chapter by looking at advertising models for multiple media, multiple sales channels and multiple brands.

15.2 Multimedia Advertising

15.2.1 The New Media Landscape

In the past 10 years the advertising landscape has changed enormously. The rapid and sustained rise of the internet has opened up many new advertising media. The old model of marketing communication is largely a firm-to-consumer concept, whereby companies send messages to current and prospective customers. The predominant traditional media, such a television, newspapers, radio and magazines were rightly called "mass media" because they had large audiences and so could deliver high reach and frequency. The frustration for advertisers is that with mass media they have no way of telling whether a person is actually exposed to their advertising. An advertisement placed within a mass media vehicle becomes an "opportunity-to-see" rather than a definite exposure. For this reason the credibility of traditional advertising has eroded due to the difficulty of measuring its effectiveness and ROI (Srivastava and Reibstein 2005).

By contrast, the internet, beginning with online display (banner) ads provides advertisers with the much-needed customer exposure feedback they desire. If a person clicks on a banner ad that is reasonably strong evidence they have seen the ad. Moreover, if the click results in a sale then that sale can be directly linked to the banner ad. In the early part of this millennium banner ads had click-through rates of up to 5%, but within five years this dwindled to less than 0.5% (Manchanda et al. 2006), and have more recently been reported at 0.01% for a travel website (Lambrecht and Tucker 2013). Discouraged by declining click-through rates, advertisers using the internet turned to search engine marketing (SEM). The attraction of this advertising channel is the low relative cost, but its appeal primarily arises from the realization that customers were now requesting to see ads. This contrasts with the old advertising doctrine of pushing ads onto customers. Furthermore, the costing model for SEM is pay-per-click, so payment is only required when a customer actually clicks on the sponsored search ad (this raises concerns about click fraud (Wilbur and Zhu 2009), but that's another issue). Pay-per-click costing models are also offered by some banner ad distributors, although cost-per-thousand (CPM) pricing is more common, whereby a fixed fee is paid for delivering a certain number of ad impressions, where there is no guarantee of exposure (Danaher et al. 2010), which is more congruent with traditional advertising. This is largely because of the very low click-through rates and ad avoidance issues with banner ads (Dreze and Hussherr 2003). Whether ad impression is a valid criteria for the effectiveness of banner advertising has attracted considerable research attention, as will be shown below.

The other new development in media advertising in the past decade has been social media, particularly from Facebook. Although Facebook had a rocky start subsequent to its IPO, it now attracts substantial ad revenues. It's two key appeals are (i) low cost, and (ii) the ability to narrowly target prospective customers based on their demographics, interests and any information they post (e.g., recent vacations, moving house or changing jobs). Firms have quickly developed their own Facebook pages and encourage customers to "like" their pages so as to receive further messages. This has regenerated interest in customer engagement (Verhoef et al. 2010). Now advertising is not always about firm-to-customer messages, it is sometimes about an on-going relationship enabled though social media, blogs and reviews; although there is now some evidence of the impact of social media on sales (Kumar et al. 2016).

The most recent development in the media landscape has been the rapid uptake of smart phones. This turned an ordinary mobile phone into a mobile computer, enabling internet browsing and email to be easily available to users at any time and any place. Moreover, the location information about mobile users has suddenly become another key dimension for targeting advertisements and promotions (Danaher et al. 2015; Fong et al. 2015).

The recent growth in media channels, all being digital, has come at a time in which the world economy has had some severe ructions. The 2008 Global Financial Crisis had an immediate negative effect on ad spend, although there has since been growth (Danaher et al. 2015). Many firms have either slightly reduced their overall ad budgets or stayed about the same. Thus, there has been growth in the number of media channels, but no additional money to spend in these new channels. This has forced advertisers to make tough decisions about how to allocate their ad budget across the many possible media channels. Moorman's (2012) survey of marketing managers shows that many will reduce their spend in traditional media, but increase their spend in digital media. In general, digital media are cheaper than traditional media, which makes them more appealing. However, a key issue today is the relative effectiveness of all these media (Danaher and Dagger 2013; Dinner et al. 2014). The rush to invest ad spend in social media is largely without justification, as few studies have conclusively shown it to be effective, and now even practitioners are questioning its value (Brooke 2016). This contrasts with many studies demonstrating the effectiveness of television and print advertising (Sethuraman et al. 2011).

15.2.2 Prior Multimedia Advertising Literature

There is now ample evidence that people frequently consume several media simultaneously (Lin et al. 2013) and advertisers exploit this by using multiple channels to reach their target group (Godfrey, Seiders and Voss 2011). An additional motivation for using several media is the possibility of realizing "media synergy," which Naik and Raman (2003, p. 375) define as occurring when "the combined effect of multiple [media] activities exceeds the sum of the individual effects."

Table 15.1 lists most of the prior multimedia advertising studies. Much of the early work on media synergy examines just two traditional media (e.g., Edell and Keller 1989 look at TV and radio, while Naik and Raman 2003 focus on TV and print). More recent attention has turned to the interplay between traditional and new (i.e., online) media, starting with Chang and Thorsen (2004), who find synergy between TV and website banner advertising. Naik and Peters (2009) review all prior studies of multimedia synergy, especially for traditional and new media, and find inconsistent results. For example, Havlena et al. (2007) and Dijkstra et al. (2005) both report positive synergy between TV and print advertising, but little or no synergy effects when online banner advertising is added to the media mix. Godfrey et al. (2011) look at customer sales resulting from phone, email and direct mail contacts for customers of a vehicle service dealership and find that too much contact can actually harm sales. This leads them to comment that "When firms use different communication channels in combination, it is a matter of speculation whether the effects are additive or multiplicative and, if multiplicative, whether the interaction enhances or diminishes customer response" (Godfrey et al. 2011, p.95).

Dinner, van Heerde and Neslin (2014) study the effects of online and offline advertising on online and offline sales, but do so with aggregate data. They do not address synergies among media, but they do find cross channel elasticities whereby online advertising influences offline sales. Danaher and Rossiter (2011) examine 11 media, but they also do not address possible synergy effects and their effectiveness measure is just purchase intention subsequent to a survey respondent being exposed to a hypothetical advertising scenario. Furthermore, banner advertising, paid search and social media are not among their 11 media. Naik and Peters (2009) study 6 media (TV, radio, magazines, newspaper, online and direct mail) to see how effective they are at attracting visits to a car dealership and also visits to the dealership's website to create car configurations. They find positive synergy between traditional and new media, although more in a collective offline/online framework as opposed to all media pairings synergistically working together. Last, Danaher and Dagger (2013) study 10 media (television, radio, newspaper, magazines, online display ads, sponsored search, social media, catalogs, direct mail and email) for a large department store retailer in Australia. They find that traditional media and sponsored search perform very well, while online display and social media advertising are not effective at generating sales. However, their particular retailer had only a very small online sales presence and did not permit promoted items to be purchased online. They also considered a very specific time period when the retailer had a month-long promotional sale. They tested for synergy among their 10 media, but that model performed worse than a model with no media synergy. A recent study by Srinivasan et al. (2015) has linked Facebook likes to sales for a packaged good product, while Kumar et al. (2016) demonstrate the influence of firm-generated-content, distributed by social media, on sales for a wine retailer.

Table 15.1 Previous multimedia advertising effectiveness studies

Reference	Media	No. media	Data	Outcome measure
Edell and Keller (1989)	TV, radio	2	Experiment with student samples	Ad recall
Naik and Raman (2003)	TV, print	2	Monthly time series	Sales
Chang and Thorsen (2004)	TV, banner	2	Experiment with student samples	Attention, message credibility
Lin, Venkataraman and Jap (2013)	TV, newspaper, radio, banners	4	Time-use surveys	Choice of media channels
Dijkstra, Buijtels and van Raaij (2005)	TV, print, banner	3	Experiment with student samples	Brand recall and brand claim recall
Danaher and Rossiter (2011)	TV, radio, magazines, newspapers, catalogs, direct mail (personally and generically addressed), e-mail, SMS, door-to-door visits and telemarketing	11	Consumer surveys and hypothetical advertising scenarios	Purchase intentions
Naik and Peters (2009)	TV, radio, magazines, banners, newspaper, direct mail	6	Time series	Website visits and store visits
Godfrey, Seiders and Voss (2011)	Phone, email, direct mail	3	Individual-level panel data for 13 consecutive quarters	Repurchase visits and repurchase spending.
Prins and Verhoef (2007)	Phone, TV, radio, print, outdoor	5	Individual-level panel data for 24 consecutive months	Adoption of a new service
Danaher and Dagger (2013)	TV, radio, newspaper, magazines, online display ads, paid search, social media, catalogs, direct mail and email	10	Individual-level for a 26 day ad campaign	Offline sales and profit

(continued)

Table 15.1 (continued)

Reference	Media	No. media	Data	Outcome measure
Dinner, van Heerde and Neslin (2014)	Traditional (aggregate spend on newspapers, magazines, radio, TV, billboard), banners, paid search	3	Time-series over 103 weeks	Online and offline sales for 25 markets
Wiesel, Pauwels and Arts (2011)	Fax, flyer, paid search, catalog, email	5	Time-series over 876 days	Online sales and profit
Bollinger, Cohen and Lai (2013)	TV, banners, social	3	Individual-level	Sales
Joo, Wilbur and Zhu (2013)	TV, paid search	2	Individual-level	Sales
Zantedeschi et al. (2016)	Email, catalog	2	Individual-level	Online and offline sales
Frison, Dekimpe, Croux and De Maeyer (2014)	TV, radio, newspaper, magazine, billboard, cinema	6	Time series over 80 months	Offline sales
Srinivasan, Rutz and Pauwels (2015)	Paid search, social, TV	3	Time series over 40 weeks	Offline sales
Kumar et al. (2016)	TV, email, social	3	Individual-level panel data for 85 weeks	Offline sales
Danaher and Danaher (2016)	Email, catalog, organic search, paid search, banners, social	6	Individual-level panel data over 2 years	Online and offline sales for 3 brands

In sum, much prior research has looked at one purchase channel only (mostly offline), and has used awareness, recall or purchase intentions as an effectiveness measure. Sethuraman, Tellis and Briesch (2011) emphasize the importance of gauging how advertising affects purchase behavior, as judged by advertising elasticity. A number of the more recent studies do look at advertising elasticities (e.g., Danaher and Dagger 2013; Dinner et al. 2014; Frison et al. 2014; Srinivasan et al. 2015; Kumar 2016), which is encouraging. Also, the interplay between offline and online sales and the various media channels has not been explored much, with a notable exception being Dinner et al. (2014) and Danaher and Danaher (2016). In the final section of this chapter we present a model that can handle multiple media, sales channels and brands using individual-level panel data.

15.3 Multimedia Advertising Elasticities

Using single-source data is regarded as the most rigorous method to gauge advertising effectiveness (Assael and Poltrack 1993; Danaher and Dagger 2013; Lodish et al. 1995; Sethuraman et al. 2011). Table 15.1 lists many prior multimedia studies, but many of them use time series data. Advertising elasticities can be estimated from time series data, but single-source data is better. A recent study by Danaher and Dagger (2013) develops a method to link media exposure to sales at the individual level. A key feature of their study is that it examines 10 media, rather than just 2–3, as has been common for most prior multimedia studies.

15.3.1 Model for Multimedia Advertising Using Individual-Level Data

As Danaher and Dagger (2013) point out, one of the difficulties of obtaining single source data with multiple media is that it is challenging to measure exposure to many media for the same people. This is rarely done by commercial media research firms and some industry initiatives (like Project Apollo) have failed (Magnostic 2008). Consequently, they develop a media measurement survey that relies on self-reports of media exposure. The task is made easier for respondents by only asking about media in which the advertiser is known to have placed ads, as opposed to all media vehicles.

When purchasing items a person must firstly visit a retailer's offline or online store or call the telephone 1-800 number (purchase incidence), then decide what and how much to buy (purchase outcome). Purchase outcome can be any one of quantity or dollar-sales. A natural model for this two-step process is the Tobit type II model (Fox et al. 2004; Van Heerde et al. 2008). This model assumes a probit model for purchase incidence and a Tobit model for purchase outcome (which is observed only for those making at least one purchase during the sale).

Danaher and Dagger (2013) define their probit model for purchase incidence as

$$z_i = \begin{cases} 1 & \text{if} \quad z_i^* > 0 \\ 0 & \text{otherwise} \end{cases},$$

where z_i^* is a latent variable constructed from a linear model as

$$z_i^* = w_i^{'}\alpha + \xi_i. \tag{15.1}$$

Here, $w_i^{'}$ is a vector of independent variables which potentially influence a person's purchase incidence during a promotional sale period, such as exposure to various media, and ξ_i is a random error.

Conditional on purchase incidence, Danaher and Dagger (2013) use a Tobit model for the log of the purchase outcome. This Tobit model allows for left censoring at 0, which is important in their application as 45% of respondents buy nothing in the sale period. The latent variable for the log of the purchase outcome is also modeled by a linear model as

$$y_i^* = x_i^{'}\beta + \varepsilon_i, \tag{15.2}$$

and $y_i = y_i^*$ is observed only when $z_i = 1$, otherwise $y_i = 0$. The setup for the Tobit type II model is completed by assuming a covariance matrix, denoted Σ, to link the error terms in Eqs. (15.1) and (15.2), defined as

$$\Sigma = \begin{bmatrix} 1 & \rho\sigma_\varepsilon \\ \rho\sigma_\varepsilon & \sigma_\varepsilon^2 \end{bmatrix}. \tag{15.3}$$

The parameter ρ is a measure of the correlation between purchase incidence and outcome.

Danaher and Dagger (2013) show that the advertising elasticity for the ith person and the kth advertising media, denoted η_{ik}, is

$$\eta_{ik} = \beta_k \Phi(w_i^L\alpha) + \alpha_k \phi(w_i^L\alpha)[(x_i^L\beta) - \rho\sigma_{i\varepsilon}(w_i^L\alpha)]. \tag{15.4}$$

Equation (15.4) captures the interplay between advertising's influence on the decision to visit a store to make a purchase and how much to buy once arriving at the store.

Table 15.2 gives the estimated ad elasticities for the 10 media in the Danaher and Dagger (2013) study. For sales and profit, the most effective medium is catalogs, followed closely by direct mail, then television. Email, paid search, radio and newspapers are also effective, but not as strong as for the top-three media. What is interesting about Table 15.2 is that both social media and banner (online display) ads have nonsignificant ad elasticities. The Danaher and Dagger (2013) findings

Table 15.2 Advertising elasticities for a multimedia campaign, as reported in Danaher and Dagger (2013)

Media	Outcome measure	
	Dollar sales	Profit
TV	0.078*** (0.0230)[a]	0.042** (0.0166)
Newspaper	0.011* (0.0065)	0.010** (0.0045)
Radio	0.022** (0.0102)	0.019** (0.0074)
Magazine	−0.004 (0.0103)	−0.001 (0.0072)
Online display	−0.011 (0.0132)	−0.009 (0.0093)
Search	0.036* (0.0210)	0.014 (0.0146)
Social media	0.020 (0.0357)	0.021 (0.0250)
Catalog	0.104*** (0.0350)	0.052** (0.0253)
Mail	0.095* (0.0527)	0.067** (0.0336)
Email	0.054* (0.0328)	0.038* (0.0233)

[a]Standard error in parentheses
*Significant at the 10% level; **Significant at the 5% level; ***Significant at the 1% level

show there isn't much support for the rapid growth in these two media as advertising media. In fact, it seems that traditional media work better in their context.

15.3.2 Model for Multimedia Advertising Using Time Series Data

One of the weaknesses of the Danaher and Dagger (2013) study is that all sales were in-store, whereas many retailers today sell merchandise in-store and online. Dinner, van Heerde and Neslin (2014) develop a comprehensive model for in-store and online sales with three media. They had three media, comprised of traditional (total spend on newspapers, magazines, radio, TV, billboard), banner advertising and paid search. In their case about 85% of sales are in-store, with the remainder being online. They have 103 weeks of data, and 25 separate markets for a clothing retailer in the US.

Their conceptual model has banner advertising and paid search driving search impressions and search click-through rate. In turn, these drive offline and online sales. They postulate that traditional advertising influences sales directly. Dinner, van Heerde and Neslin's (2014) model is a system of four equations

$$\begin{aligned}\ln SearchImpressions_t = \beta_{1,0} &+ \beta_{1,1} TraditionalAdStock_{1,t} \\ &+ \beta_{1,2} OnlineDisplayAdStock_{1,t} \\ &+ \beta_{1,3} OnlineSearchAdStock_{1,t} \\ &+ Season\ and\ trend\ controls_{1,t} + \nu_{1,t}\end{aligned} \tag{15.5}$$

$$\begin{aligned}\ln SearchClickThroughRate_t = \beta_{2,0} &+ \beta_{2,1} TraditionalAdStock_{2,t} \\ &+ \beta_{2,2} OnlineDisplayAdStock_{2,t} \\ &+ \beta_{1,3} OnlineSearchAdStock_{2,t} \\ &+ Season\ and\ trend\ controls_{2,t} + \nu_{2,t}\end{aligned} \tag{15.6}$$

$$\begin{aligned}\ln OfflineSales_{mt} = \beta_{3,0} &+ \beta_{3,1} TraditionalAdStock_{3,mt} \\ &+ \beta_{3,2} OnlineDisplayAdStock_{3,mt} \\ &+ \beta_{3,3} SearchClickthroughStock_{3,mt} \\ &+ CompetitorAdvertising_{mt} + PromotionControls_{mt} \\ &+ Market\_dummies_m + \nu_{3,mt}\end{aligned} \tag{15.7}$$

$$\begin{aligned}\ln OfflineSales_{mt} = \beta_{4,0} &+ \beta_{4,1} TraditionalAdStock_{4,mt} \\ &+ \beta_{4,2} OnlineDisplayAdStock_{3,mt} \\ &+ \beta_{4,3} SearchClickthroughStock_{4,mt} \\ &+ CompetitorAdvertising_{mt} + PromotionControls_{mt} \\ &+ Market\_dummies_m + \nu_{4,mt}\end{aligned} \tag{15.8}$$

Because Dinner, van Heerde and Neslin (2014) formulate their model in log-log form, it is easy to obtain elasticities from Eqs. (15.7) and (15.8). Ad stock is the exponential-smoothed weighted average of current—and past-period advertising, defined in the same way as Danaher et al. (2008). Dinner, van Heerde and Neslin's (2014) full model accounts for endogeneity, advertising carryover, contemporaneous correlation and autocorrelation, and they estimate their model with 3SLS.

Dinner, van Heerde and Neslin (2014) find that online display and paid search advertising have higher elasticities than traditional advertising. They also report a number of significant cross elasticities whereby online advertising influences offline sales. The total sales elasticity in their case is 0.071 for traditional advertising, which is close to the 0.078 figure for TV advertising obtained by Danaher and Dagger (2013), as reported here in Table 15.2. However, Dinner, van Heerde and Neslin (2014) report much higher elasticities for online display (banner) advertising (0.118) and paid search (0.14) than did Danaher and Dagger (2013). Part of the reason for this is likely to be contextual, as Danaher and Dagger's (2013) retailer has no online sales, and it would make sense for digital media to work better for online sales. We explore this more in the next subsection.

15.3.3 Model for Multiple Advertising and Sales Channels and Several Brands Using Individual-Level Panel Data

Our final model in this chapter is one that allows for online and offline sales, multimedia advertising and multiple brands, as described in more detail by Danaher, and Zhang (2016). Data are at the individual-level and span a period of 2 years. The motivation for this model comes from data provided by a North American specialty retailer that sells mostly apparel, as supplied by the Wharton Customer Analytics Initiative. About 80% of sales are in-store and the rest are online. The retailer has three relatively distinct brands that operate independently. However, they collect customer data in a combined database and so have information on customer sales to each of their brands.

The data cover the period from July 2010–June 2012, which we aggregate into 24 monthly periods for each individual. There are 42,000 panelists in total, but 8% buy nothing and some people join the panel part way through the data period. After eliminating these people, our final dataset has about 26,000 customers. For each of these people we have the number of times they were exposed to 6 media (catalog, email, banner ads, organic search, paid search, social media – Facebook and Twitter). In addition, we monitor the number of times each customer visits each brands' website each month, what might be termed "customer initiated website visits." The reason for this is customers may visit a brand's website even without being prompted by marketing communications from the firm. The dominant forms of marketing communication by the firm are printed catalogs sent via the post and email.

Danaher and Zhang's (2016) model is a further multivariate extension of the Tobit Type II model, as proposed by van Heerde et al. (2008). We first model purchase incidence with a probit model, defined as

$$z_{ibct} = I_{\left\{z^*_{ibct} > 0\right\}}, \tag{15.9}$$

where z^*_{ibct} is an associated latent variable described by the following equation

$$z^*_{ibct} = \alpha_{0i} + w'_{ibct}\alpha + u_{ibct}. \tag{15.10}$$

Here $z_{ibct} = 1$ when person i $(i = 1, \ldots, n)$ purchases brand b $(b = 1, \ldots, B)$ from channel c $(c = 1, \ldots, C)$ at time t $(t = 1, \ldots, T)$. The vector w_{ibct} contains observed covariates that are expected to influence purchase, with associated parameter vector α. Note that the intercept is a random effect with distribution $N(\bar{\alpha}_0, V_\alpha)$ and $u_{ibct} \sim N(0, 1)$, as is required for a probit model.

Conditional on a purchase $(z_{ibct} = 1)$, we model y_{ibct}, the amount spent by individual i on brand b in channel c at time t as follows:

$$y_{ibct} = \beta_{0i} + x'_{ibct}\beta + \varepsilon_{ibct}, \tag{15.11}$$

Let $u_{it} = (u_{i1t}, \cdots, u_{ikt}, \cdots, u_{iKt})'$ and $\varepsilon_{it} = (\varepsilon_{i1t}, \cdots, \varepsilon_{ikt}, \cdots, \varepsilon_{iKt})'$, where the subscript k denotes the combination of the subscript pair b, c. Now let the error terms in Eqs. (15.10) and (15.11) come from a covariance matrix Σ that can be partitioned as follows:

$$\sum = \begin{bmatrix} \sum_{11} & \sum_{12} \\ \sum'_{12} & \sum_{22} \end{bmatrix}, \tag{15.12}$$

where $\sum_{11} = E(\varepsilon_{it}, \varepsilon'_{it})$, $\sum_{12} = E(\varepsilon_{it}, u'_{it})$, and $\sum_{22} = E(u_{it}, u'_{it})$. Note that, due to the identification constraint of the probit model, $\sum_{22}$ is a correlation matrix with ones on the diagonal and all off-diagonal elements are within the range $(-1, 1)$. The overall model is therefore

$$\begin{pmatrix} y_{ibct} \\ z^*_{ibct} \end{pmatrix} \sim N_{2K}\left(\begin{matrix} \beta_{0i} + \beta_{B1,R}x_{B1,R,ibct} \\ \beta_{0i} + \beta_{B1,D}x_{B1,D,ibct} \\ \vdots \\ \beta_{0i} + \beta_{B3,D}x_{B3,D,ibct} \\ \alpha_{0i} + \alpha_{B1,R}w_{B1,R,ibct} \\ \alpha_{0i} + \alpha_{B1,D}w_{B1,D,ibct} \\ \vdots \\ \alpha_{0i} + \alpha_{B3,D}w_{B3,D,ibct} \end{matrix}, \sum \right), \tag{15.13}$$

where B1–B3 denotes the three brands, and R and D denote in-store (retail) and direct (online) channels. Danaher and Zhang (2016) discuss endogeneity, heterogeneity, the specification and other elements of this model in more detail. Danaher and Zhang (2016) calculate ad elasticities in a similar way to those for the univariate Tobit Type II model and we reproduce them in Fig. 15.1. There is quite a lot of variation in advertising elasticity across brands and sales channels. However some trends are apparent. The first is that catalogs for B1 and B3 do better for in-store sales. Conversely, banner ads for B2 and B3 do better for online sales. Broadly speaking, traditional direct catalogs sent through the post are more effective at driving in-store rather than online sales. The opposite is true for banner (online display) ads. Email and organic and paid search are less differentiated by purchase channel, but are generally quite effective. Social media is effective only for B1 and B3 in-store sales and is probably the weakest of the advertising media. Nevertheless, this is one of the first studies to show that social media ad elasticities are statistically significant (see also, Kumar et al. 2016).

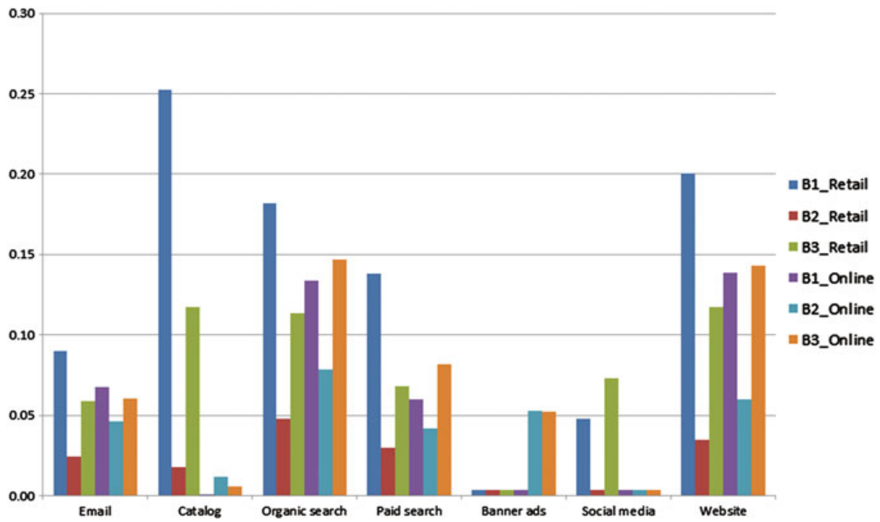


Fig. 15.1 Advertising elasticities grouped by purchase channel

15.3.4 Models for Attributing Media to Purchases

The growing use of digital media combined with the ability to track both customer ad exposure and purchase behavior has resulted in a new approach to assessing advertising effectiveness. Essentially what this does is to work backwards from a purchase incident and find all the "touchpoints" a customer traverses on their path to purchase. Touchpoints might include visits to a retailer's website, email/catalogs sent to the customer by the retailer, broadcast advertising, organic or paid search and social media likes or posts. The weight applied to each touchpoint is known as media attribution. Initial efforts at attributing a purchase to a medium were rather crude, with a popular method being to ascribe all the weight to the most recent touchpoint, known as "last touch attribution" (Li and Kannan 2014). This clearly favors high volume and inexpensive media like email and works against low frequency media like catalogs. It also ignores carryover effects, which are known to be important in advertising (Dinner et al. 2014). Recent work by Abhishek et al. (2014) and Anderl et al. (2014) has looked to improve on the existing naïve methods for media attribution. Danaher and van Heerde (2016) prove that there is a mathematical link between attribution and advertising elasticity, which is a major advance because until now these concepts have been treated in isolation. They also show that media allocation based on attribution is inferior to allocation based on ad elasticities.

15.4 Advertising Media Selection Models

Danaher (2008) reviewed media planning models for estimation of the exposure distribution. This is an important first step in media planning. A subsequent step is the allocation of advertising exposures across different media vehicles. This is especially relevant today with the internet, which has millions of different websites and therefore poses significant computation problems for media allocation optimization.

15.4.1 Traditional Set up for Media Scheduling

A typical mathematical set up of an advertising scheduling problem is to maximize target audience reach, subject to a budget constraint. Formally this can be written as follows.

$$\text{Maximize } 1 - f(X=0), \text{ subject to } 0 \le X_i \le k_i \quad \text{and} \quad \sum_{i=1}^{m} c_i X_i \le B,$$

where $X = \sum_{i=1}^{m} X_i$, k_i is a possible upper limit on the number of ad insertions in media vehicle i, c_i is the cost per advertising insertion and B is the total budget.

15.4.2 Models for Media Channel Selection

Over the past decade there has been a rapid increase in the number of media channels. Even within the television environment, the dominance of the networks has eroded considerably in the face of numerous cable stations (Krugman and Rust 1993). Audiences have fragmented as a result, and this has reduced the ability of TV networks to deliver a mass audience quickly, as they were previously able to. Even more rapid than television channel proliferation, has been the rise of the internet. While advertisers were initially unsure about the effectiveness of the internet (Smith 2003), today ad spend on the web is substantial and sustained (Bucklin and Hoban 2016). Other digital communication channels, such as SMS text messaging, have also emerged as possible advertising media (Danaher et al. 2015).

Up until 10 years ago, it was relatively easy to reach a mass audience with just one medium, but this is becoming increasingly difficult. As a consequence, one of the most pressing advertising issues for today is which portfolio of media is best to achieve maximum effectiveness. To cope with this growing issue, models are required that can assess the relative effectiveness of alternative media, in much the same way that a marketing mix model can assess the relative importance of alternative marketing instruments. Danaher (2008) reviews models that have attempted to combine different media, and we now discuss such models particularly related to the internet.

15.4.3 A Model for Multiple Websites

As Danaher (2007) demonstrates, a promising model for handling multimedia applications, for media planning at least, is based on the Sarmanov (1966) distribution. The general form of the Sarmanov bivariate distribution for (X_1, X_2) is

$$f(X_1, X_2) = f_1(X_1) f_2(X_2)[1 + \omega \phi_1(x_1)\phi_2(x_2)], \tag{15.14}$$

where $f_i(X_i)$ is the marginal distribution for random variable X_i and $\phi_i(x_i)$ are called "mixing functions", with the requirement that $\int \phi_i(t) f_i(t) dt = 0$. Park and Fader (2004) use the Sarmanov distribution in (15.17) to model the bivariate visit-time distribution between two websites.

For multiple media, the Sarmanov distribution generalizes to

$$f(X_1, X_2, \ldots, X_m) = \left\{\prod_{i=1}^{m} f_i(X_i)\right\}\left[1 + \sum_{j_1 < j_2} \omega_{j_1, j_2} \phi_{j_1}(x_{j_1})\phi_{j_2}(x_{j_2}) + \ldots + \omega_{1,2,\ldots,m} \prod \phi_i(x_i)\right]. \tag{15.15}$$

Danaher et al. (2010) show that if the univariate marginal distributions are assumed to be negative binomial, which is well-suited to the internet websites, then

$$\phi_i(x_i | t_i) = e^{-x_i} - \left(\frac{\alpha_i}{t_i(1 - e^{-1}) + \alpha_i}\right)^{r_i} \tag{15.16}$$

is the appropriate mixing function.

15.4.4 Optimizing the Audience for Multiple Websites

Danaher et al. (2010) discuss issues that differentiate audience optimization for the internet compared with traditional media. These include a potentially limitless number of exposures/impressions per person and the fact that advertisers might limit those exposures by introducing frequency capping. Web publishers, too, might want to limit the number of banner ads from just one advertiser. This introduces the concept of a share of impressions, whereby an advertiser has only a probability of having their impression displayed when a web page is served. This differs from traditional mass media advertising where an ad is delivered to the entire audience view/reading/listening at the time. Consequently, optimizing reach (or another audience measure) is slightly different for the internet. Danaher et al. (2010) show that when applying the share of impressions adjustment and the budget constraint as well as the reach model from Eqs. (15.17) and (15.18), the online advertising optimization problem for maximizing reach by varying $(s_1, s_2, \ldots, s_m)$ can be stated formally as follows.

Maximize

$$\begin{aligned}
\text{Reach} &= 1 - f(X_1 = 0, X_2 = 0, \ldots, X_m = 0 | t_1, t_2, \ldots, t_m, s_1, s_2, \ldots, s_m) \\
&= 1 - \left\{ \prod_{i=1}^{m} f_i(x_i = 0 | r_i, \alpha_i, t_i, s_i) \right\} \left[1 + \sum_{j_1 < j_2} \omega_{j_1,j_2} \left(1 - \left(\frac{\alpha_{j_1}}{s_{j_1} t_{j_1} (1 - e^{-1}) + \alpha_{j_1}} \right)^{r_{j_1}} \right) \right. \\
&\quad \times \left(1 - \left(\frac{\alpha_{j_2}}{s_{j_2} t_{j_2} (1 - e^{-1}) + \alpha_{j_2}} \right)^{r_{j_2}} \right) + \sum_{j_1 < j_2 < j_3} \omega_{j_1,j_2,j_3} \left(1 - \left(\frac{\alpha_{j_1}}{s_{j_1} t_{j_1} (1 - e^{-1}) + \alpha_{j_1}} \right)^{r_{j_1}} \right) \\
&\quad \left. \times \left(1 - \left(\frac{\alpha_{j_2}}{s_{j_2} t_{j_2} (1 - e^{-1}) + \alpha_{j_2}} \right)^{r_{j_2}} \right) \left(1 - \left(\frac{\alpha_{j_3}}{s_{j_3} t_{j_3} (1 - e^{-1}) + \alpha_{j_3}} \right)^{r_{j_3}} \right) \right],
\end{aligned} \tag{15.17}$$

subject to $0 \leq s_i \leq 1$, and

$$\frac{N}{1000} \sum_{i=1}^{m} \frac{s_i c_i r_i t_i}{\alpha_i} \leq B. \tag{15.18}$$

If frequency capping is to be used the only change to the optimization problem set up is to replace s_i with $s_i / \Pr(X_i \leq cap)$, with the denominator probability obtained from the univariate negative binomial distribution.

Danaher et al. (2010) provide an example of the use of this optimization scheme. They show that it is better for online media planners to attempt to buy a set number of impressions in a fixed time period, where the number of impressions is determined by the media planner, not the web publisher. They call this fully “flexible scheduling.” Danaher et al. (2010) also find that frequency capping always increases the reach, but another, less obvious way of increasing reach is to lengthen the campaign duration.

Finally, with the increasing number of websites and demand for internet advertising, the potential number of websites is huge, which places enormous demands on computation time. Future work in this area needs to address the computation time issue. For example, Danaher and Smith (2011) show that a multivariate exposure model using copulas is much quicker than the Sarmanov model. Some exciting recent research that optimizes website audiences across thousands of websites has been reported by Paulson, Luo and James (2014).

15.5 Conclusion

These are exciting, but also challenging times, for advertisers. The rapidly-increasing ways in which firms can communicate with customers, and vice versa, opens up new opportunities for marketing managers. However, the availability of new media channels does not necessarily make them effective. The studies presented here and elsewhere find a high level of consumer conservatism towards digital media, favoring traditional advertising media, especially in the offline purchase channel. However, where digital media are finding a place is for generating sales in online purchase channels. Digital advertising also has an indirect effect in generating sales in-store, particularly when combined with traditional media.

The purpose of this chapter has been to present several alternative models and methods that can be used to calculate ad elasticities for multimedia campaigns, and also across multichannel purchase platforms. These models should have wide applicability. No doubt, over the next decade, there will be continued development of models and methods in this important domain for marketing.

References

Abhishek, Vibhanshu, Peter S. Fader, and Kartik Hosanagar. 2014. Media exposure through the funnel: A model of multi-stage attribution. Working paper, Heinz College, Carnegie Mellon University.

Anderl, Eva, Ingo Becker, Florian V. Wangenheim, and Jan H. Schumann. 2014. Mapping the customer journey: A graph-based framework for online attribution modeling. Working paper, Passau University, Germany.

Assael, Henry, and David Poltrack. 1993. Using single source data to select tv programs. *Journal of Advertising Research* 33 (1): 48–56.

Bollinger, Bryan K., M.A. Cohen, and J. Lai. 2013. Measuring asymmetric persistence and interaction effects of media across platforms. Working paper, NYU. http://ssrn.com/abstract=2342349.

Brooke, Zach. 2016. Marketers report big increases in social media spending with little to show in return. American Marketing Association, Marketing News Weekly. https://www.ama.org/publications/eNewsletters/Marketing-News-Weekly/Pages/Social-Media-Spending-Worth-It.aspx#sthash.xTZgYFE2.dpuf. Accessed 29 Feb 2016.

Bucklin, Randolph E. and Paul R. Hoban. 2016. Marketing models for internet advertising. In *Handbook of marketing decision models*, 2nd ed, eds. Berend Wierenga and Ralf van der Lans. New York, NY: Springer.

Chang, Yuhmiin, and Esther Thorson. 2004. Television and web advertising synergies. *Journal of Advertising* 33 (2): 75–84.

Danaher, Peter J. 2007. Modeling page views across multiple websites with an application to internet reach and frequency prediction. *Marketing Science* 26 (3) (May/June): 422–437.

Danaher, Peter J. 2008. Advertising models. In *Handbook of marketing decision models*, ed. Berend Wierenga, and New York, 81–106. NY: Springer.

Danaher, Peter J., and Tracey S. Dagger. 2013. Comparing the relative effectiveness of advertising channels: A case study of a multimedia blitz campaign. *Journal of Marketing Research* 50 (4) (August): 517–534.

Danaher, Peter J., and Tracey S. Danaher. 2016. Multimedia advertising effectiveness in a multichannel and multibrand environment. Working paper, Monash University, Australia.

Danaher, Peter J., Janghyuk Lee, and Laoucine Kerbache. 2010. Optimal internet media selection. *Marketing Science* 29 (2) (March): 336–347.

Danaher, Peter J., G. Mullarkey, and S. Essegaier. 2006. Factors affecting website visit duration: A cross-domain analysis. *Journal of Marketing Research* 43 (May): 182–194.

Danaher, Peter J., Andre Bonfrer, and Sanjay Dhar. 2008. The effect of competitive advertising interference on sales for packaged goods. *Journal of Marketing Research* 45 (April): 211–225.

Danaher, Peter J., and John R. Rossiter. 2011. Comparing perceptions of marketing communication channels. *European Journal of Marketing* 45 (1/2): 6–42.

Danaher, Peter J., and Mike Smith. 2011. Modeling multivariate distributions using copulas: Applications in marketing (with discussion and rejoinder). *Marketing Science* 30 (1): 4–21.

Danaher, Peter J., Mike Smith, Kulan Ranasinghe, and Tracey S. Danaher. 2015. Where, when and how long: Factors that influence the effectiveness of mobile phone coupons. *Journal of Marketing Research* 52 (5) (October): 710–725.

Danaher, Peter. J., and Harald van Heerde. 2016. The delusion of attribution: Caveats in attributing multimedia exposures to purchases. Working paper, Monash University, Australia.

Danaher, Peter J., and Rong Zhang. 2016. Modeling the inter-brand, purchase channel and cross promotional effects for luxury clothing goods. Working paper, Monash University, Australia.

Dinner, Isaac, Harald J. van Heerde, and Scott A. Neslin. 2014. Driving online and offline sales: The cross-channel effects of traditional, online display, and paid search advertising. *Journal of Marketing Research* 51 (5): 527–545.

Dijkstra, Majorie, Heidi E. Buijtels, and W. Fred van Raaij. 2005. Separate and joint effects of medium type on consumer responses: A comparison of television, print and the internet. *Journal of Business Research* 58: 37–386.

Dreze, Xavier, and Francois X. Hussherr. 2003. Internet advertising: Is anybody watching? *Journal of Interactive Marketing* 17 (4): 8–23.

Edell, Julie, and Kevin Lane Keller. 1989. The information processing of coordinated media campaigns. *Journal of Marketing Research* 26 (May): 149–163.

Fong, Nathan, Zheng Fang, and Xueming Luo. 2015. Geo-conquesting: Competitive locational targeting of mobile promotions. *Journal of Marketing Research*.

Fox, Edward J., Alan L. Montgomery, and Leonard M. Lodish. 2004. Consumer shopping and spending across retail formats. *Journal of Business* 77 (2): S25–S60.

Frison, Steffi, Marnik G. Dekimpe, Christophe Croux, and Peter De Maeyer. 2014. Billboard and cinema advertising: Missed opportunity or spoiled arms? *International Journal of Research in Marketing* 31: 425–433.

Godfrey, Andre, Kathleen Seiders, and Glenn B. Voss. 2011. Enough is enough! The fine line in executing multichannel relational communication. *Journal of Marketing* 75 (July): 94–109.

Havlena, William, Robert Cardarelli, and Michelle de Montigny. 2007. Quantifying the isolated and synergistic effects of exposure frequency for tv, print and internet advertising. *Journal of Advertising Research*: 215–221.

Joo, Mingyu, Ken C. Wilbur, Bo Cowgill, and Yi Zhu. 2013. Television advertising and online search. *Management Science* 60 (1): 56–73.

Krugman, Dean M., and Roland T. Rust. 1993. The impact of cable and VCR penetration on network viewing: assessing the decade. *Journal of Advertising Research* 33 (January): 67–73.

Kumar, Ashish, Ram Bezawada, Rishika Rishika, Ramkumar Janakiraman, and P.K. Kannan. 2016. From social to sale: The effects of firm-generated content in social media on customer behavior. *Journal of Marketing* 80 (1): 7–25.

Lambrecht, Anja, and Catherine Tucker. 2013. When does retargeting work? Information specificity in online advertising. *Journal of Marketing Research* 50 (October): 561–576.

Li, Hongshuang (Alice) and P.K. Kannan. 2014. Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research* (51) (February): 40–56.

Lin, Chen, Sriram Venkataraman, and Sandy D. Jap. 2013. Media multiplexing behavior: Implications for targeting and media planning. *Marketing Science* 32 (2): 310–324.

Lodish, Leonard M., Magid Abraham, Stuart Kalmenson, Jeanne Livelsberger, Beth Lubetkin, Bruce Richardson, and Mary Ellen Stevens. 1995. How TV. Advertising works: A Meta-analysis of 389 real world split cable tv advertising experiments. *Journal of Marketing Research*, (37) (May): 125–139.

Magnostic. 2008. Marketing measurement misplay: Project apollo is dead. http://magnostic.wordpress.com/2008/02/25/marketing-measurement-misplay-project-apollo-is-dead/. Accessed 15 June 2012.

Manchanda, Puneet, Jean-Pierre Dube, Khim Yong Goh, and Pradeep K. Chintagunta. 2006. The effect of banner advertising on internet purchasing. *Journal of Marketing Research* 43 (February): 98–108.

Moorman, Christine. 2012. The CMO survey: Highlights and insights, February 2012. http://cmosurvey.org/files/2012/03/The_CMO_Survey_Highlights_and_Insights_Feb-2012_Final.pdf. Accessed 27 June 2012.

Naik, Prasad A., and Kay Peters. 2009. A hierarchical marketing communications model of online and offline media synergies. *Journal of Interactive Marketing* 23: 288–299.

Naik, Prasad A., and Kalyan Raman. 2003. Understanding the impact of synergy in multimedia communications. *Journal of Marketing Research* 40 (November): 375–388.

Park, Young-Hoon, and Peter S. Fader. 2004. Modeling browsing behavior at multiple websites. *Marketing Science* 23(3) (Summer): 280–303.

Paulson, Courtney, Lan Luo, and Gareth James. 2014. *Optimal large-scale internet media selection, paper presented at the marketing science conference*. Atlanta, Georgia: Emory University.

Prins, Remco, and Peter C. Verhoef. 2007. Marketing communication drivers of adoption timing of a new e-service among existing customers. *Journal of Marketing* 71 (April): 169–183.

Sarmanov, O.V. 1966. Generalized normal correlations and two-dimensional frechet classes Doklady. *Soviet Mathematics* 168: 596–599.

Schweidel, David A., and Robert J. Kent. 2010. Predictors of the gap between program and commercial audiences: An investigation using live tuning data. *Journal of Marketing* 74 (May): 18–33.

Sethuraman, Raj, Gerard J. Tellis, and Richard A. Briesch. 2011. How well does advertising work? Generalizations from meta-analysis of brand advertising elasticities. *Journal of Marketing Research* 48 (June): 457–471.

Smith, David L. 2003. Online Reach and Frequency: An Update, April. http://www.mediasmithinc.com/white/msn/msn042003.html.

Srinivasan, Shuba, Oliver Rutz, and Koen Pauwels. 2015. Paths to and off purchase: Quantifying the impact of traditional marketing and online consumer activity. *Journal of the Academy of Marketing Science*.

Srivastava, Rajendra, and David Reibstein. 2005. Metrics for linking marketing to financial performance. Marketing science institute report no. 05–200.

Heerde, Van, Els Gijbrechts Harald, and Koen Pauwels. 2008. Winners and losers in a major price war. *Journal of Marketing Research* 45 (October): 499–518.

Verhoef, Peter C., Werner J. Reinartz, and Manfred Krafft. 2010. customer engagement as a new perspective on customer management. *Journal of Service Research*, 13 (3) 247–252.

Wiesel, Thorsten, Koen Pauwels, and Joep Arts. 2011. Marketing's profit impact: Quantifying online and offline funnel progression. *Marketing Science*, 30 (4), 604–611.

Wilbur, Kenneth C., and Yi Zhu. 2009. Click Fraud. *Marketing Science* 28 (2): 293–308.

Zantedeschi, Daniel, Eleanor McDonnell Feit, and Eric T. Bradlow. 2016. Measuring multi-channel advertising response. SSRN working paper.

Chapter 16
Social Media Analytics

Wendy W. Moe, Oded Netzer and David A. Schweidel

16.1 Introduction

One of the most significant developments in recent years in the domain of marketing involves the proliferation of user-generated content, particularly online social media. Social media has created a power shift in the relationship between consumers and brands, providing consumers more power by allowing them to easily broadcast their views and opinions about brands to a large audience. At the same time social media has opened a window for firms into the voice of the consumer. Previously, marketers had to employ costly and time-intensive marketing research methods such as interviews, focus groups and surveys to better understand how consumers perceive their brands. Now, consumers voluntarily turn to social media and share their opinions publically for other customers as well as for the brand managers to see.

This new medium provides wealth of data from which marketing researchers can extract customer insights. However, in order to analyze and leverage social media data, we must first understand the behavior that generates the data. Thus, the first part of this chapter will discuss **online opinion behavior**, the process by which user generated content (UGC) is created, and its implications for deriving insights from such data. We then discuss **social media as a source for marketing research** and describe some of the models that have been developed for social media data mining. There are several challenges involved in converting the vast volumes of social media data to useful managerial insights. Key amongst these challenges is the fact

W.W. Moe (✉)
University of Maryland, College Park, MD, USA
e-mail: wendy_moe@rhsmith.umd.edu

O. Netzer
Columbia University, New York, NY, USA

D.A. Schweidel
Georgetown University, Washington, DC, USA

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_16

that most social media data are unstructured and textual in nature. With a firm understanding of the consumer and the appropriate methodologies, marketers can then begin to use social media to understand and influence their customers. This leads us to a second, but equally important, function of social media, which we discuss in the third part of the chapter—**social media as a communications channel**.

16.2 Understanding the Behavior of Social Media Content Generators

Researchers have studied offline consumer word-of-mouth behavior for decades.[1] Westbrook (1987) identified three motivations that drive consumers to spread word-of-mouth: product involvement, self-involvement and altruism. Anderson (1998) found a relationship between satisfaction with the product and the likelihood of engaging in word-of-mouth behavior, where highly dissatisfied customers were more likely to share their opinions.

In the context of online social media environments, Hennig-Thurau et al. (2004) proposed a taxonomy of online word-of-mouth motivations. They propose that consumers are motivated to participate in online word-of-mouth for a variety of reasons including users seeking assistance, expressing negative feelings, helping others, self-enhancement, social benefits, economic incentives, aiding the firm and seeking advice. Berger and Milkman (2012) find that social transmission is affected in part by the arousal of content. Toubia and Stephen (2013) examined the motivations underlying consumers' posting on Twitter and differentiate between posters who derive intrinsic utility from posting on social media and those who derive image-related utility. Lovett et al. (2013) combine multiple sources of data to compare the underlying motivation and characteristics of online and offline word-of-mouth. They find that social and functional drivers are more prominent in online word-of-mouth, whereas emotional drivers were most important in offline word-of-mouth. Furthermore, brand differentiation plays an important role in online word-of-mouth but less so in offline word-of-mouth.

Whereas Anderson (1998) and others have found that offline word-of-mouth was predominantly negative as dissatisfied customers were more likely to engage in word-of-mouth activities, online word-of-mouth tends to be predominantly positive. This positivity bias has been documented across multiple studies (Chevalier and Mayzlin 2006; Godes and Mayzlin 2004; Resnik and Zeckhauser 2002). One of the most robust findings in product reviews, one form of online word-of-mouth, is that of a J-shaped relationship between frequency of posts and satisfaction with the product (Moe and Schweidel 2012). Figure 16.1 shows the J-shape curve proposed by Moe and Schweidel (2012). The J-shape relationship suggests that while those with

[1]For a review of research on word-of-mouth, we refer interested readers to Berger (2014).

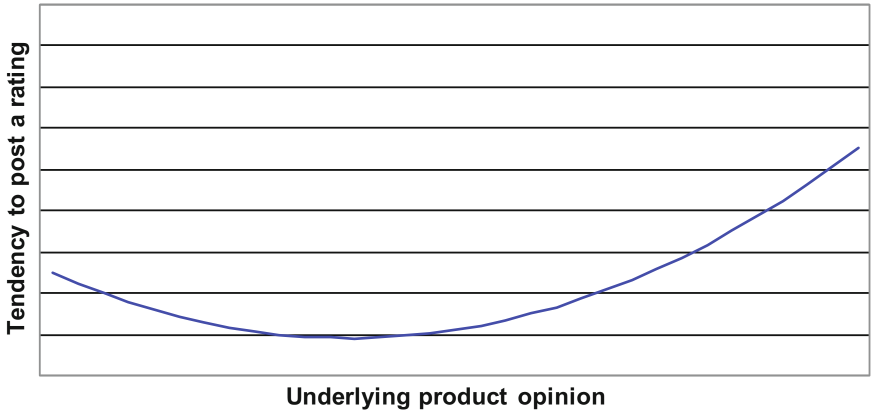


Fig. 16.1 J-shaped relationship between product opinion and ratings incidence

negative opinions are more likely to share an opinion than those with a moderate opinion (as is the case in offline word-of-mouth), those with positive opinions are even more likely to share online. While the positivity bias of online product ratings has been well documented, no explanation has yet been provided for this noticeable difference between online and offline word-of-mouth behavior.

Beyond the positivity bias in online product ratings, researchers have shown that online product ratings and reviews have a distinct downward trend over time. A number of theories have been provided for this trend including product life cycle effects, where later adopters are less satisfied with their purchase decision based on the evaluations of innovators who may hold very different preferences (Li and Hitt 2008), and preference matching effects, where consumers have more difficulty sifting through the posted ratings as the number of ratings available increases (Godes and Silva 2012).

Moe and Schweidel (2012) propose an online social dynamics account to explain the downward trend in online opinion. In the context of online product ratings, they differentiate between (1) the consumer's decision whether to post a rating and (2) the consumer's decision of what rating to post and examine how the opinions of others already posted to the same forum affects each of these two decisions. They contrast two types of consumers: activists and low-involvement consumers. The activists are frequent posters who are more likely to provide critical evaluations and make efforts to differentiate their posted opinions from those of others. The low-involvement consumers, in contrast, post less frequently and hold more positive opinions. However, the low-involvement consumers are unique in that they are less likely to post word-of-mouth when the opinions posted previously are highly varied. This is in stark contrast to the activists who thrive in these environments and are actually more likely to post when there is disagreement in previous reviews. These two contrasting behaviors of the activists and the low-involvement posters lead to a very interesting dynamic. As the number of opinions posted in the forum increases and disagreement among the posters emerge, the low-involvement consumers begin to withdraw from the conversation and refrain from sharing their opinions. The activists, on the other hand, relish these dissentious environments and freely contribute their critical opinions. Over time, the minority activists dominate opinion environments and their opinions are disproportionately represented. Because activists posted opinions tend to be more negative, this leads to a downward trend in posted opinions over time.

What does this mean for social media data and analytics?

These two empirical findings of J-shape relationship between the frequency of product reviews and satisfaction with the product and the downward trend in ratings and reviews over time have important implication for social media analytics. The explanations provided to both of these findings suggest that there may be self-selection in the consumers' decision to participate in the online conversation. That is, the downward trend in reviews over time is not necessarily a reflection of any real decline in consumers' opinions of product quality. This is an important consideration for brand managers and social media analysts who may be considering product modifications or other changes to their overall marketing mix, as online comments may not be representative of the perceptions held by the broader customer base. Similarly, the J-shaped distribution of opinions posted may not be representative of the underlying customer population. Differences may be due to the differences between the overall customer base and those who post on social media or it may be a result of a systematic shift in the opinion posted due to the social dynamics described above. Either way, in analyzing user generated content, the researcher needs to consider such biases. That being said, because user generated content affects purchase decision of a much larger population than those who generated it (Ludwig et al. 2013), these, possibly self-selected, opinions may affect actual purchases of a much larger and more representative population.

Offline researchers have also identified an *audience effect* that can affect the opinions one share with others. For example, Fleming et al. (1990) showed that an individual's evaluation changed depending on who he/she believed would be the audience of the evaluation. In the online environment, Schlosser (2005) showed that product evaluations changed depending on whether or not the individual providing the opinion believed it would be made public. The existing research is clear in that audience effects can alter individual posting behavior. In aggregate, this leads to a bias in the social media data that is dependent on the nature of the audience that a social media user expects to face. For example, a user who is posting to a community of close friends and family (e.g., on Facebook) is likely to express him or herself differently than a user who is posting to a professional network of colleagues (e.g., on LinkedIn). Thus, because various social media venues tend to attract different audiences, we should expect systematic differences across various social media venues.

Schweidel and Moe (2014) directly examine the differences in social media posting behavior across various social media venues such as, forums, blog and micro blogs. They found that opinions shared on different venues systematically vary due to the unique audience that each venue attracts and the social dynamics that each venue encourages. For example, discussion forums allow for the most social interaction. As a result, opinions posted to these forums tend to be more negative and experienced the most severe downward trend over time (consistent with the findings of Moe and Schweidel 2012). Blogs facilitated the least social interactions while allowing the poster to provide more depth in their comments. As a result, opinions expressed in blogs were more moderate and did not experience any downward trends over time.

The discussion above suggests that aggregation of social media data over time and/or venue may mask or even bias the results of the analysis performed on that data. Furthermore, as consumers self-select themselves to participate in the discussion, the data available for analytics and modeling may not be necessarily representative of the underlying opinions of the entire customer base. In analyzing social media data we encourage researchers to develop models that explicitly acknowledge the individual level behaviors and the factors that influence them. In the next section, we review a number of models with applications for marketing research, with particular focus on models that de-bias social media data for social dynamics and venue effects for the purposes of brand tracking (e.g., Schweidel and Moe 2014) and methods to analyze the text that dominates so much of social media data (e.g., Netzer et al. 2012).

16.3 Using UGC for Marketing Research

16.3.1 From Numerical Ratings to Textual Information

One of the main difficulties in utilizing consumer-generated content for quantitative analysis is that the data are primarily qualitative in nature. In discussing future directions for social interaction research, Godes et al. (2005) note that one of the difficulties in tapping into UGC is the ability to analyze the content. Similarly, Liu (2006) analyzed messages posted on Yahoo Movies message board and reported "an extremely tedious task" in mechanically analyzing over 12,000 movie review messages using human reviewers.

Because of the difficulties associated with analyzing text, many researchers have resorted to characteristics of the consumer-generated data such as product ratings to represent the content of the consumers' opinions. Three common measures of product ratings have been primarily used in the literature: volume, valence and variance. These measures have been used by numerous researchers when investigating the relationship between ratings and sales (e.g., Chevalier and Mayzlin 2006;

Dellarocas et al. 2007; Moe and Trusov 2011) as well as the dynamics in the social media discussion (e.g., Godes and Mayzlin 2004; Moe and Schweidel 2012).

In a study investigating how online conversations affect television ratings, Godes and Mayzlin (2004) operationalized variance (or dispersion) using a measure of entropy to reflect the extent to which conversations occur in different online newsgroups or discussion forums:

$$Entropy_{it} = \begin{cases} -\sum_{n=1}^{N} \frac{POST_{it}^{n}}{POST_{it}} \log\left(\frac{POST_{it}^{n}}{POST_{it}}\right) & \text{if } POST_{it} \geq 0 \\ 0 & \text{if } POST_{it} = 0 \end{cases}$$

where $POST_{it}^{n}$ is the total number of posts on newsgroup n about program i between the airing times of episodes t and $t + 1$. $POST_{it}$ is the total number of posts about program i between the airing of episodes t and $t + 1$ airing, aggregated across all newsgroups. Entropy is maximized when the number of posts is equally distributed across the N newsgroups. As Godes and Mayzlin (2004) note, in contrast to variance, entropy does not depend on the total volume of posts as it instead relies on the share of posts appearing in each newsgroup. The authors find that increased entropy is associated higher television ratings, which may reflect reaching a broader audience.

Dellarocas and Narayan (2006) provide a volumetric measure of product reviews that combines social media data with sales data. They propose a measure of *density,* operationalized as the ratio of the number of people posting online comments during a fixed period of time to the number of people purchasing the product.

In many contexts, users directly provide a measure of valence in the form of an ordinal rating, such as 1–5 star ratings in Amazon.com product reviews. However, many social media environments, such as blogs, forums and social networks, do not include quantitative summary of the consumers' evaluations. Instead, posts to these venues are comprised solely of text. In these contexts, valence measures are often constructed using sentiment analysis (e.g., Pang and Lee 2008) where words associated with positive, neutral or negative emotions are extracted to capture the overall sentiment expressed in the message.

Richer content analyses beyond just valence measures have also been employed. However, as the field of text mining is still relatively new, the degree of expertise needed to develop and employ a solid sentiment analysis tool is quite high, and the accuracy of such automated processes is still evolving. Accordingly, many researchers still employ manual coding. For example, Schweidel and Moe (2014) have used a commercial firm to manually code the sentiment and product attribute mentions of over 7,500 messages across three social media venues. Nevertheless, as social media data is often voluminous, it is clear that a more automatic and multifaceted view of the online text can help generate meaningful insights.

16.3.2 Text Mining

Text mining (sometimes called knowledge discovery in text) refers to the process of extracting useful information from unstructured text (Fellbaum 1998; Feldman et al. 1998; Feldman and Sanger 2006). For example, Swanson and colleagues found relationships between magnesium and migraine (Swanson 1988) and between biological viruses and weapons (Swanson and Smalheiser 2001) by text mining disjoint literatures and uncovering words common to both literature bases.

Text mining has become particularly popular and successful in fields in which meaningful information must be extracted from "mountains" of data in a relatively short period of time. Such fields include security and intelligence organizations looking for signs of irregular activity in the stream of public and less public written media (Fan et al. 2006) or doctors searching for biomedical information in the superabundance of medical information (Rzhetsky et al. 2004). Academics have used text mining in order to automatically meta-analyze the knowledge base on a particular topic (Börner et al. 2003). With the increasing availability of voluminous digitized data sources, the business world started to take notice of the opportunities offered by text-mining tools to automatically analyze the infinite stream of financial report data, to open a window to consumer online discussions and to collect competitive intelligence information. Many companies are offering text-mining services to businesses to help them with these tasks.

Computer scientists and information system researchers have made the greatest leap in developing advanced text-mining apparatuses. Collaborations between computer scientists or information systems researchers and business researchers helped facilitate the dissemination (albeit limited) of these tools to business research (e.g., Das and Chen 2007; Feldman et al. 2008; Ghose et al. 2012; Lee and Bradlow 2011; Netzer et al. 2012). In marketing, the first attempts to text-mine UGC used manual text-mining involving humans reading the messages and judging their content (e.g., Godes and Mayzlin 2004; Liu 2006). This inefficient and inaccurate methodology was described by the authors as a "tedious task" and a "costly and noisy process." Computer scientists such as Dave et al. (2003), Hu and Liu (2004), Liu et al. (2005) and Feldman et al. (2007) offered a solution to the tedium by building apparatuses that could automatically summarize and quantify consumer reviews. These advances have facilitated a fast and continuing diffusion of text mining applications to research in marketing (e.g., Decker and Trusov 2010; Ghose et al. 2012; Lee and Bradlow 2011; Netzer el al. 2012) we further describe these and other applications of text mining in marketing later on in this chapter.

16.3.3 Text Mining Approaches

It is beyond the scope of this chapter to provide a detailed and exhaustive description of different text-mining tools and the methodology involved. We refer the interested reader to books that specialize in text mining (e.g., Feldman and Sanger 2006). Instead, our objective is to describe, at a high level, the most commonly used types of text-mining analyses in marketing and some of the considerations one should be aware of in applying such tools in marketing contexts.

At the most basic level text mining has been used in marketing to extract individual entities such as brands, product attributes, emotions and adjectives used to describe products. Numerous commercial companies are offering buzz monitoring services, tracking how frequently a brand is being mentioned across alternative social media. Similarly, academic researchers have looked at how often brands are mentioned in social media venues, which emotions are being mentioned (e.g., Berger and Milkman 2012; Ludwig et al. 2013), or which attributes are being mentioned in a review related to a particular product (e.g., Lee and Bradlow 2011). Netzer et al. (2012) note that the task of accurately identifying brands (e.g., Audi or Volvo) is easier than identifying product models (e.g., Audi A4 or Volvo S6). In the context of cars they report F1 accuracy levels[2] of 98.1% for car brands and 91.6% for car models. Extracting more difficult entities such as adverse drug reactions (ADRs) led to lower F1 accuracy levels of 81.6%, all within acceptable ranges in the text mining literature. Dictionaries such as WordNet (Fellbaum 1998) and the Linguistic Inquiry and Word Count (LIWC; Pennebaker et al. 2001) have been used as easy tools to conduct such basic text mining analysis. However, as we discuss later for most text mining application more advanced tools such as natural language processing (NLP) are needed.

A slightly more advanced set of tools is needed if one is interested in capturing the sentiment of a particular textual unit such as a product review (See Pang and Lee 2008 for a review of methods). Most of the sentiment analysis tools (sometimes called opinion mining) rely on NLP, statistics tools, or machine learning.[3] Often a combination of approaches is used together with a sentiment dictionary (e.g., Sentiwordnet—sentiwordnet.isti.cnr.it or Sentistrength—sentistrength.wlv.ac.uk) to obtain more accurate sentiment analysis. For example, Ghose et al. (2012) used part-of-speech tagging combined with crowd sourced Amazon mechanical Turks scoring of adjectives to derive their sentiment tool. Netzer et al. (2012) used a machine-learning approach, combined with a sentiment dictionary and human coded rules to identify common problems reported in various car models. Das and Chen (2007) used a statistical approach involving classifiers to capture sentiment for stocks from message boards. Tirunillai and Tellis (2012) used a combination of statistical classifier and a support vector machine approach to capture the valence of UGC about products. The accuracy level of sentiment analysis methods is still limited and the sentiment tool often needs to be tailored to the specific domain of analysis. For example, the word "high" would be considered as positive sentiment

[2]F1 is measured as the harmonic mean of the levels of recall and precision, where recall is the proportion of instances that were identified, and precision is the proportion of correctly identified instances of the set of identified instances.

[3]For more information about NLP, we refer interested readers to Manning and Schütze (1999).

in the context of stock prices but negative sentiment in the context of blood pressure. Thus, marketing researchers are advised to use domain-specific tools or tools that can be adapted to a specific domain. Furthermore, it is recommended to manually examine the level of accuracy of the tool using human coders on a sample of textual units.

At the next level of complexity of analysis lies the process of relation extraction. Relation extraction refers to the process of identifying textual relationships among extracted entities. For example, in the context of pharmaceutical drugs, Netzer et al. (2012) and Feldman et al. (2015) identified the textual relationships between drugs and ADRs that imply that drug X causes ADR Y. Such textual relationship extraction often requires linguistic analysis, usually involving NLP, to allow the text-mining algorithm to understand the textual context of the sentence. Netzer et al. (2012) reported F1 accuracy of 73.6% in identifying the relationship between drugs and ADRs. Applications of such relation extractions are still few in marketing, primarily due to the text mining complexity involved in accurately making such relational inferences from unstructured data. However, we believe this area is one of the most promising directions for future work. Marketing researchers are often more interested in extracting the relationships between products, attributes and the sentiment or context of the relationship between them, than simply measuring the volume of mentions of a brand or even the overall sentiment about a particular brand.

Once the relationships between entities have been extracted one needs to summarize the co-occurrences of entities in the textual corpora. One may be tempted to simply report co-occurrences of how often any pair of entities appears together in the text. For example, how many times did the brand BMW appear with the term "sporty"? The problem with reporting simple co-occurrence is that if an entity appears very frequently in the text it will also co-occur with more entities than an entity that occurs less frequently. Thus, one should normalize the measure of co-occurrence for how often each entity appears in the text independently, to capture how often two entities occur in the text over and beyond chance. Various such "normalization" approaches have been proposed.

One of the most commonly used measures in the text-mining literature is the term frequency–inverse document frequency (tf-idf) weighting. tf-idf is used to weigh the occurrence of each term by its role in the document. The term frequency for term j in document m is defined by $tf_{jm} = X_{jm}/N_m$, where X_{jm} is the number of times term j appeared in document m, and N_m is the number of terms in document m. $idf_j = \log\left(D/M_j\right)$, where D is the total number of documents and M_j is the total number of documents where term j appeared. tf-idf is given by $tf - idf_{jm} = tf_{jm} \times idf_j$. The term frequency component captures how prominent a particular term is in the document. For example, if the word "sporty" appeared three times in a car review that include 10 words the prominence of sporty is higher than a similar but much longer review that includes three mentions of the word sporty in a review that includes 100 words. However, if all reviews include the words sporty three times, the appearance of three mentions of sporty in a particular review is less informative. Accordingly, the inverse document frequency "normalizes" the term frequency measure to how likely we are to see the term in a typical document. Thus, multiplying term frequency by inverse document frequency gives us an overall measure of prominence of a term in a particular document, after controlling for the likelihood of the term to appear in the entire textual corpora.

Another classic measure of normalized co-occurrence commonly appearing in the co-word analysis literature as well as the market basket literature is the measure of lift. Lift is the ratio of the actual co-occurrence of two terms to the frequency with which we would expect to see them together. The lift between terms A and B can be calculated as

$$\mathrm{Lift}(A,B) = \frac{P(A,B)}{P(A)\times P(B)},$$

where $P(A)$ is the probability of occurrence of term A in a given textual unit, and $P(A, B)$ is the probability that both A and B appear in a given textual unit. A lift ratio of less than (more than) 1 suggests that the two terms appear together less than (more than) one would expect by the mere occurrence of each of the two terms in the text separately. Other frequently used measures of co-occurrence are the Salton's cosine similarity and the Jaccard index (e.g., Toubia and Netzer 2017).

The process of text mining often involves five steps. In the context of identifying products and product attributes and the textual relationships among them the process can be defined as:

1. *Downloading*: The textual information is downloaded (often in an html format). This process can be either done manually using a pre-specified set of URLs or using a Web scraper that searches the Web for instances of a particular topic or product.
2. *Cleaning*: html tags and non-textual information such as images and commercials are cleaned from the downloaded files.
3. *Information Extraction*: Entities such as products and product attributes are extracted from the messages. The researcher may use an *n*-gram approach to extract entities that include more than one word. In the *n*-gram approach the text-mining algorithm extracts all possible sequences of up to *n* words found in the text. Additionally, the researcher may wish to use stemming algorithms to reduce and combine words into their word stem or root (e.g., using stemming, the words, "run," "ran," "running," would all be captured by the entity "run"). In some cases stop words such as "a" or "the" are also removed in this step.
4. *Chunking*: The textual parts are divided into informative units such as threads, messages, and sentences.
5. *Semantic relationships*: The linguistic algorithm identifies the co-occurrence of entities in the same textual unit (e.g., two cars that are mentioned together in the same forum message). At a deeper level of textual relationship, the researcher

may want to not only identify that the two entities (e.g. a drug and an ADR) were mentioned together, but also what is the nature of the textual relationship between the two entities (e.g., the drug causes the ADR).[4]

When text-mining UGC the researcher is often faced with a problem of high dimensionality. That is, the text-mining process often results in thousands of possible unique words that appeared in the mined corpora. The problem becomes even more severe when one uses an n-gram approach. This dimensionality problems leads to both statistical and interpretation difficulties. From a statistical point of view if one wishes to use the terms extracted as independent variables or predictors to predict some market outcome, estimating a model with hundreds or thousands of predictors is difficult. From an interpretation point of view, deriving meaningful insights from such large space of variables is challenging. Stemming the words to their stem or root helps to reduce the dimensionality of the entity space as multiple words are combined to their common stem. However, the number of derived stems is still often highly unwieldy. The simplest approach to reduce the dimensionality of the problem is to trim the list of entities to only entities that appeared at least a certain number of times. However, determining the threshold from which to remove entities is often ad hoc and this process may still leave the researcher with a large number of entities.

A more statistically driven approach to reduce the dimensionality and derive insights from textual information is to use latent Dirichlet allocation (LDA), often called topic modeling (Blei et al. 2003). The idea behind LDA is that each document (e.g., product review) contains a mixture of topics. The distribution of topics is assumed to have a Dirichlet prior. Each topic is then (probabilistically) associated with a set of words. The advantage of the LDA approach is that one can automatically infer which topics were most likely to be mentioned in each document. The set of topics could be either learned in a fully unsupervised manner (i.e., purely informed from the data) or in a supervised or semi-supervised manner (i.e., informed fully or partially by the researcher). Tirunillai and Tellis (2014) have demonstrated the potential of using LDA in marketing by using unsupervised LDA to capture the latent dimensions underlying product quality in product reviews. For example, for mobile phones the authors report that the top six dimensions of quality (topics) in order of importance are “ease of use,” “secondary features,” “performance,” “visually appealing,” “reliability” and “customer service.” For footwear, on the other hand, the most important dimension was “physical support” and the second most important “visually appealing.” Some of the limitations of the standard LDA approach is that it does not consider the order in which words appear in a document or sentence structure. More generalized approaches have been developed that take into account sentence structure (e.g., Büschken and Allenby 2016) and the identity of the author (e.g., Rosen-Zvi et al. 2004).

[4]If the researcher is interested in sentiment analysis or other types of output rather than relationship extraction, Step 5 could be replaced with the eventual goal of the text-mining task. For example, if the researcher is interested in understanding which topics were mention in a review, step 5 may be replaced with a topic modeling approach.

Software packages such as the *tm* package in *R* (Feinerer and Hornik 2015) and *NLTK* in *Python* (Bird et al. 2009) have made text mining more accessible. That being said, in employing text mining techniques, researchers should exercise care to ensure that the tools is appropriate for the problem at hand and the underlying data generating mechanism, and that the tools is properly executed for the idiosyncrasies of the specific problem it is used for.

16.3.4 Applications of Text Mining in Marketing

One can, broadly speaking, divide the applications of text mining UGC data in marketing into two main groups based on the goals of the analysis: (1) using UGC to *describe* and monitor markets, and (2) using UGC to *predict* relevant market outcomes.

1. *Text mining UGC to describe markets*—by text mining UGC companies can listen to and monitor consumer discussions and opinions about their own products as well as the competition. Furthermore, because the UGC data stream keeps updating in real time, one can monitor the changes in consumer perceptions over time. In that sense one can think of text mining UGC data as leveraging social media as a marketing research playground or as an almost infinite size, and re-occurring, focus group. Accordingly, researchers have explored the type of insights that can be generated from mining UGC data as a descriptive listening tool. For example, Schweidel and Moe (2014) analyzed approximately 7,500 user-posted text across multiple social media venues. The authors measured sentiments both at the overall brand level and at product-specific and attribute-specific levels. They define a single measure of brand health as well as sentiment measures of specific aspects of their product.

   Because UGC provides firms with a window into the discussion about their own and their competitive products, one can use UGC to assess the competitive market structure. For example, Netzer et al. (2012) looked at the co-occurrence between pairs of cars in the same message in nearly 900,000 messages from the sedan cars forum Edmunds.com to create a competitive market structure map of the car industry, simultaneously analyzing nearly 170 different car models. Leveraging the longitudinal nature of the data the authors show that, following a marketing campaign, Cadillac re-positioned itself away from the group of American brands and towards the luxury import brands. Similarly, Lee and Bradlow (2011) used product reviews to extract the attributes and attribute levels that were mentioned with each brand of digital cameras to create market structure maps based on the similarity in attribute mentions across digital camera brands.

   Tirunillai and Tellis (2014) used LDA to describe the dimensions of product quality mentioned in product reviews across five product categories. By tracking the quality dimensions over time the authors explore the competitive brand positions on the quality dimensions. For example, in the context of computers, the authors found that Dell's perception of "ease of use" were highly volatile over time, whereas those of Hewlett Packard were relatively steady. Similarly, Zhang et al. (2015), used dynamic topic modeling to monitor the evolution of the competitive landscape.

These studies highlight the value text-mining tools provide in converting a largely qualitative source of data such as UGC, to a quantitative data and the opportunity to generate useful descriptive and perspective information and insights for business decision making.

2. *Text mining UGC to predict market outcomes*—in addition to leveraging text mining of UGC data to listen to consumers' discussions, researchers have also explored relating the text-mined information to market outcomes such as consumer choices, aggregate sales or stock prices.

   Several studies have shown that textual analysis of UGC can predict stock prices. For example, Tirunillai and Tellis (2012) have linked the volume and valence (as measured by sentiment analysis) of UGC about a firm to the firm's stock performance. The authors find positive relationship between the volume of the chatter about the brand and the brand's stock return. The authors report that negative UGC can lead to negative stock returns but positive UGC had little effect on the stock prices. Yu et al. (2013) used sentiment analysis across multiple social media sources such as blogs, forums and Twitter to demonstrate relationship between social media sentiment and stock prices. Similarly, Bollen et al. (2011), showed that the inferred mood from Twitter posts can predict overall stock market performance.

   Another important outcome measure is firm sales. Archak et al. (2011) demonstrate that the textual information in product reviews can help extract consumer preferences for product attributes, which in turn adds predictive power in predicting sales over and beyond the reviews' numerical ratings. Similarly, Decker and Trusov (2010), mine product reviews to estimate consumer preferences and predict the overall evaluations of products. Ludwig et al. (2013) find that positive affective content in book reviews affects conversion rates for books in Amazon.com. At the individual level, Ghose et al. (2012) used a text mining analysis of both product reviews and hotel descriptions, together with image recognition, and crowd sourcing to predict consumers' choices among hotels and design a ranking algorithm for hotels.

   Mining UGC has been also shown to predict other business relevant outcomes such as diffusion of information and success of ideas and movies. Berger and Milkman (2012) used a combination of automated sentiment analysis, the LIWC dictionary, and manual coding to assess the drivers and predict the sharing of New York Times articles. Toubia and Netzer (2017) show that automatically mining the text of individual ideas generated by consumers, can help flagging promising ideas and recommend, in real time, words for consumer to improve their ideas. Text mining non-UGC type of data, Eliashberg et al. (2007), demonstrate that text mining movie scripts using NLP and statistical learning tools can help predict the success of the movies.

Looking at social welfare outcomes, Feldman et al. (2015), demonstrated that by mining UGC medical forums one could predict drug label changes as early as 10 years prior to the label change. Similarly, Culotta (2010) showed that one can detect influenza epidemics by mining Twitter messages. Twitter messages have been also shown to predict overall political opinion. O'Connor et al. (2010), found correlations of as high as 80% between the political sentiment of Twitter messages and the political public opinions measured in surveys. Taken together, we see that the rich textual information available in UGC data can help predict consumer preferences and evaluations, aggregate firm sales, stock prices, major events such as drug label change and success of ideas and information goods.

Moving forward, we expect that advances in text mining tools will allow researchers and marketers to go beyond capturing volume, valence, or particular words that can describe markets and predict outcomes, and towards understanding the textual relationships and deeper information content expressed in UGC. Combined with content analysis applied to other aspects of UGC, such as hyperlinks (e.g., Liu 2007) and tags (e.g., Nam and Kannan 2014), such analyses can aid in, for example, detecting problems with products based on consumer discussions in consumer forums or understanding the comparative language consumers use to describe competitive products.

16.4 Social Media as a New Marketing Channel

In addition to being a new source of marketing data that can be used to generate consumer insights, social media provide an important channel through which consumers can communicate with one another and with organizations, and organizations can communicate back with consumers. Indeed, several researchers have demonstrated that consumer discussions in social media as a standalone channel, have an effect on sales (e.g., Chevalier and Mayzlin 2006; Godes and Mayzlin 2009; Moe and Trusov 2011). Furthermore, researchers have also shown how social media can be leveraged by firms as part of a broader marketing strategy (e.g., Stephen and Galak 2012; Srinivasan et al. 2015).

16.4.1 Social Media's Effect on Sales

Social media has a created a major shift in power between consumers and firms. It provided to consumers a vehicle with a wide reach to easily and publicly express their liking and disliking for products. Additionally, consumers frequently consult social media and UGC venues such as forums, blogs and product reviews before making a purchase. Accordingly it is likely that the content of social media would causally affect sales. While we have mentioned previously that several studies have demonstrated the ability of UGC data (e.g., product review ratings) to predict sales (e.g., Godes and Mayzlin 2004; Moe and Trusov 2011) arguing for a causal relationship is much more difficult. Vector autoregressive (VAR) models (e.g., Stephen and Galak 2012; Srinivasan et al. 2015) econometric approaches (e.g., Anderson and Magruder 2012; Mayzlin et al. 2014), and field experiments (Godes and Mayzlin 2009) were proposed to isolate and identify the impact of social media activity on performance.[5]

The seminal study in establishing the relationship between product review ratings and product sales was conducted by Chevalier and Mayzlin (2006). In their research, they examined how star ratings, provided by customers, impacted book sales at both Amazon.com and Barnesandnoble.com. Chevalier and Mayzlin used the fact that the inherent quality of the books on both websites is likely to be the same, thus differences in sales variation over time between Amazon.com and Barnesandnoble.com for any single book can likely be attributed to changes in the online ratings that are unique to each site. This study was the first to establish the role that user-provided product ratings have on product sales. Interestingly, their study also found that the impact of 1-star reviews exceeds that of 5-star reviews, suggesting that negative user-generated content may be more impactful than positive user-generated content.

Anderson and Magruder (2012), used a clever approach of regression discontinuity approach to establish causal relationship between Yelp ratings and restaurant reservations. The authors found that increase in star rating causes increase in restaurant reservations, which results in a higher likelihood of the restaurant being fully booked.

Moe and Trusov (2011) further investigated the effects of product ratings on product sales by considering both direct and indirect effects. Specifically, they measure the effects that previous ratings have on the arrival of subsequent ratings as well as on product sales. In other words, posted ratings can directly affect product sales; they can also affect subsequent ratings thereby indirectly affecting future product sales. Methodologically, the authors develop an exponential hazard model for the arrival of each ratings level for a given product and model the effects of lagged measures of valence (the average ratings for the product), variance (the variance in ratings for the product), and volume (the total number of ratings for the product).

The authors deconstruct the impact of product ratings on sales into components attributable to baseline effects, social dynamics and idiosyncratic error. This

[5]For a review of the impact of online WOM on sales, we refer readers to the meta analyses conducted by Babic et al. (2016) and You et al. (2015).

decomposition is found to provide superior model fit compared to benchmarks that rely directly on summary measures of social media (i.e., the volume, valence and variance).

The impact of UGC on sales is likely to differ based on the volume, valence and variance of the UGC, as well as across product categories and the review platform. For, example, Sun (2012) combines a theoretical model and empirical evidence to show that high variance of product reviews is better for products that were rated low, as the high variance signals a niche product.

The findings presented thus far suggest that user-generated content can positively affect key metrics such as sales. Thus, from a managerial perspective, it would be very tempting to try and manipulate the ratings environment to improve sales. Dellarocas (2006) investigated the outcome of such strategic manipulation by the firm and proposed a game theoretic model in which profit-maximizing firms seek to manipulate ratings by contributing fake anonymous product ratings. Dellarocas (2006) describes an outcome in which firms expend resources to artificially inflate their perceived quality through online ratings, but such behavior is expected by consumers. In this game theoretic outcome, consumers will assume that manipulation of online forums occurs and thus discount the quality signal they obtain from online forums. Given this eventual outcome, firms will thus choose to conduct a minimum level of manipulation since there is no long-term benefit of such behavior. Mayzlin et al. (2014) argue and demonstrate that independent single property hotels are more likely to participate in generating deceptive hotel reviews than branded chain hotels.

This is not to say that all firm-generated word-of-mouth is intended to deceive customers. Godes and Mayzlin (2009) conduct a field test to understand the impact of firm-generated word-of-mouth on sales. The authors worked with a marketing agency focused on creating word-of-mouth communications for its clients by providing consumers with small incentives to spread word-of-mouth about its clients. For each incident of word-of-mouth that is spread for the client, the consumers spreading word-of-mouth is asked report their relationship to the recipient. Godes and Mayzlin (2009) showed that exogenously generated word-of-mouth positively affects week-to-week sales. The authors' results also suggest that firms may be better suited to recruit less loyal customers to spread word-of-mouth. Doing so may enable the firm to reach consumers who are unaware of the firm's offerings and whose opinions may be more malleable compared to those consumers reached by more loyal customers.

16.4.2 Social Media as Part of the Broader Marketing Strategy

While the above discussion establishes the relationship between social media word-of-mouth and firm performance, social media is not a channel that operates in isolation of other elements of the marketing mix. In addition to user-generated social media, marketers also manage earned media and paid advertising activities.

Stephen and Galak (2012) investigate the effects of traditional earned media (e.g., media mentions in traditional outlets of newspapers, magazines, television and radio) relative to social earned media (e.g., blogs, forum posts, and new members registering for the forum). Using a zero-inflated autoregressive double Poisson model for the marginal distributions, the authors employ a multivariate normal copula to correlate the marginal distributions for sales and media variables. They find evidence to suggest that both traditional and social earned media positively impact sales. Their results also indicate that, in addition to the direct impact of earned social media on sales, Social earned media may indirectly affect sales by increasing the amount of traditional earned media.

Gopinath et al. (2014) go beyond looking at the volume of online word of mouth to investigate how the content of word of mouth, along with advertising, impacts sales. In addition to considering the volume of online activity, the authors account for the valence specific to comments that focus on recommendations, attribute features and emotional attachment. They find that it is the topic-specific valence measures and advertising that directly impact sales, while advertising also exhibits an indirect effect by impacting the topic-specific valence measures and the volume of online word of mouth.

Srinivasan et al. (2015) extend the work of Stephen and Galak (2012) by considering the impact of marketing mix activity on consumers' online activity and, ultimately, sales. Using a vector autoregressive (VAR) model, the authors examine the impact of price and distribution channels, television advertising, paid search clicks, website visits, and Facebook activity on sales. While the elasticities of price and distribution on sales have the highest magnitudes, consumers' online activities including paid search, website traffic and Facebook likes have larger elasticities compared to television advertising. Importantly, the authors illustrate that these activities are inter-related. For example, paid search is affected by television advertising and Facebook likes, while it impacts distribution, site visits, Facebook likes and sales.

Taken together, this stream of research highlights the inter-connected nature of UGC in organizations' marketing efforts. In evaluating the impact of marketing efforts on performance measures, we must be cognizant to account for both the direct impact of marketing efforts on performance, as well as the indirect effect of marketing efforts through the production of UGC. Viewed in this light, UGC may provide an early indication of marketing effectiveness.

16.5 Conclusion

User-generated content (UGC) has been investigated from different perspectives by marketing scholars. Some have investigated the process by which it is produced and diffused to better understand the phenomenon. Others, looking at it as the digital manifestation of the voice of the consumer, have employed it as a rich and economical means of conducting marketing research, from tracking brand health to inferring brand associations and the competitive landscape. We have also seen it interpreted and employed as a new tool for marketing to consumers. As these streams of research have largely evolved independently of each other, one of the goals of this chapter is to provide an integrated perspective on the evolution and future of UGC in marketing.

New methodology, like deep learning, is being developed in related disciplines and gradually being adopted by marketing researchers in academia and practice. As these methodologies evolve, coupled with techniques developed by marketing academics, our ability to characterize social media and incorporate it into subsequent analysis improves. For example, while familiar metrics like valence and volume still have a place in marketing models, our understanding of consumers is much richer thanks to the content of the UGC and topics identified through text analytic methods.

As we discussed earlier, several papers have investigated the integrity of UGC data and the risk of deceptive UGC (Anderson and Simester 2014; Dellarocas 2006; Mayzlin et al. 2014). However, possibly due to the difficulty in credibly identifying deceptive reviews, we do not have a firm assessment of the degree of this phenomenon and its effect on reliably using UGC as a marketing research or a predictive tool. We encourage future researchers to tackle this important problem.

We encourage future research to investigate targeting individual consumers based on UGC and its propagation. For example, can social media posts reveal consumers' interests? What messages resonate most with them? Who is influential within their social networks? Such analyses may prove beneficial for marketers, as marketing messages can be delivered to those who will be most responsive to them. In doing so, researchers may examine the extent to which insights gleaned from social media analytics complement what can be learned about consumers using other data sources. For example, to what extent does the incorporation of social media activity affect perceptions based on analyzing transactional activity? Beyond informing the likelihood of future transactions (e.g., Schweidel et al. 2014), merging social media data with CRM data could provide information regarding the likely categories in which customers will be most likely to purchase in the future.

From the perspective of brand management, UGC can provide insights into how the brand is perceived in the market place. But, these perceptions are likely to vary considerably across consumers and markets. Future work may build on the text analytic work (e.g., Netzer et al. 2012; Tirunillai and Tellis 2014) to examine other forms of unstructured data in which brands appear and understand brand associations using such data. In particular, research may consider the development of predictive models for the trajectory of conversations.

As research into social media continues to develop, academic researchers should be mindful of the potential application of their research. Popular social media monitoring platforms often lack advanced analytic tools, in part because of the magnitude of the data and the corresponding computation resources needed to apply the most advanced analytic tools. Given the keen interest in mining social media to understand the voice of the consumer, we encourage researchers to engage in research that has the potential to be deployed at scale, without sacrificing the rigor of their work.

References

Anderson, E.W. 1998. Customer satisfaction and word of mouth. *Journal of Service Research* 1 (1): 5–17.

Anderson, E.T., and D.I. Simester. 2014. Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research* 51 (3): 249–269.

Anderson, M., and J. Magruder. 2012. Learning from the crowd: Regression discontinuity estimates of the effects of an online review database. *The Economic Journal* 122 (563): 957–989.

Archak, N., A. Ghose, and P.G. Ipeirotis. 2011. Deriving the pricing power of product features by mining consumer reviews. *Management Science* 57 (8): 1485–1509.

Babic, A., F. Sotgiu, K. de Valck, and T.H. Bijmolt. 2016. The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research* 53 (3): 297–318.

Berger, J. 2014. Word of mouth and interpersonal communication: A review and directions for future research. *Journal of Consumer Psychology* 24 (4): 586–607.

Berger, J., and K.L. Milkman. 2012. What makes online content viral? *Journal of Marketing Research* 49 (2): 192–205.

Bird, S., E. Loper, and E. Klein. 2009. *Natural language processing with Python*. O'Reilly Media Inc.

Blei, D.M., A.Y. Ng, and M.I. Jordan. 2003. Latent dirichlet allocation. *The Journal of Machine Learning Research* 3: 993–1022.

Bollen, J., H. Mao, and X. Zeng. 2011. Twitter mood predicts the stock market. *Journal of Computational Science* 2 (1): 1–8.

Börner, K., C. Chen, and K.W. Boyack. 2003. Visualizing knowledge domains. *Annual Review of Information Science and Technology* 37 (1): 179–255.

Büschken, J., and G.M. Allenby. 2016. Sentence-based text analysis for customer reviews. *Marketing Science* 35 (6): 953–975.

Chevalier, J.A., and D. Mayzlin. 2006. The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research* 43 (3): 345–354.

Culotta, A. 2010. Towards detecting influenza epidemics by analyzing Twitter messages. In *Proceedings of the first workshop on social media analytics*, 115–122. ACM.

Das, S.R., and M.Y. Chen. 2007. Yahoo! for Amazon: Sentiment extraction from small talk on the web. *Management Science* 53 (9): 1375–1388.

Dave, K., S. Lawrence, and D.M. Pennock. 2003. Mining the peanut gallery: Opinion extraction and semantic classification of product reviews. In *Proceedings of the 12th international conference on World Wide Web*, 519–528. ACM.

Decker, R., and M. Trusov. 2010. Estimating aggregate consumer preferences from online product reviews. *International Journal of Research in Marketing* 27 (4): 293–307.

Dellarocas, C. 2006. Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science* 52 (10): 1577–1593.

Dellarocas, C., and R. Narayan. 2006. A statistical measure of a population's propensity to engage in post-purchase online word-of-mouth. *Statistical Science* 21 (2): 277–285.

Dellarocas, C., X.M. Zhang, and N.F. Awad. 2007. Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing* 21 (4): 23–45.

Eliashberg, J., S.K. Hui, and Z.J. Zhang. 2007. From story line to box office: A new approach for green-lighting movie scripts. *Management Science* 53 (6): 881–893.

Fan, W., L. Wallace, S. Rich, and Z. Zhang. 2006. Tapping the power of text mining. *Communications of the ACM* 49 (9): 76–82.

Feinerer, I., and K. Hornik. 2015. tm: Text Mining Package. R package version 0.6-2. http://CRAN.R-project.org/package=tm.

Fellbaum, C. 1998. *WordNet*. Blackwell Publishing Ltd.

Feldman, R., and J. Sanger. 2006. Information extraction. In *The text mining handbook: Advanced approaches in analyzing unstructured data*, 94–130.

Feldman R., M. Fresko, J. Goldenberg, O. Netzer, and L. Ungar. 2007. Extracting product comparisons from discussion boards. In *Proceedings of the 7th IEEE international conference data mining 2007,* 469–474. Piscataway, NJ: IEEE.

Feldman R., M. Fresko, Y. Kinar, Y. Lindell, O. Liphstat, M. Rajman, Y. Schler, and O. Zamir. 1998. Text mining at the term level. In *Principles of data mining knowledge discovery*, 65–73. Berlin: Springer.

Feldman R., M. Fresko, J. Goldenberg, O. Netzer, and L. Ungar. 2008. Using text mining to analyze user forums. In *Proceedings of the 2008 international conference on service systems service management*, 1–5. Melbourne, VIC: IEEE Systems, Man, and Cybernetics Society.

Feldman, R., O. Netzer, A. Peretz, and B. Rosenfeld. 2015. Utilizing text mining on online medical forums to predict label change due to adverse drug reactions. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 1779–1788.

Fleming, J.H., J.M. Darley, J.L. Hilton, and B.A. Kojetin. 1990. Multiple audience problem: A strategic communication perspective on social perception. *Journal of Personality and Social Psychology* 58 (4): 593.

Ghose, A., P.G. Ipeirotis, and B. Li. 2012. Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science* 31 (3): 493–520.

Godes, D., and D. Mayzlin. 2009. Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Science* 28 (4): 721–739.

Godes, D., and D. Mayzlin. 2004. Using online conversations to study word-of-mouth communication. *Marketing Science* 23 (4): 545–560.

Godes, D., and J.C. Silva. 2012. Sequential and temporal dynamics of online opinion. *Marketing Science* 31 (3): 448–473.

Godes, D., D. Mayzlin, Y. Chen, S. Das, C. Dellarocas, B. Pfeiffer, B. Libai, S. Sen, M. Shi, and P. Verlegh. 2005. The firm's management of social interactions. *Marketing Letters* 16 (3–4): 415–428.

Gopinath, S., J.S. Thomas, and L. Krishnamurthi. 2014. Investigating the relationship between the content of online word of mouth, advertising, and brand performance. *Marketing Science* 33 (2): 241–258.

Hennig-Thurau, T., K.P. Gwinner, G. Walsh, and D.D. Gremler. 2004. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing* 18 (1): 38–52.

Hu, M., and B. Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the 10th ACM SIGKDD international conference on knowledge discovery and data mining*, 168–177. ACM.

Lee, T., and E.T. Bradlow. 2011. Automated marketing research using online customer reviews. *Journal of Marketing Research* 48 (5): 881–894.

Li, X., and L.M. Hitt. 2008. Self-selection and information role of online product reviews. *Information Systems Research* 19 (4): 456–474.

Liu, Y. 2006. Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing* 70 (3): 74–89.

Liu, B. 2007. *Web data mining: Exploring hyperlinks, contents, and usage data*. Springer Science & Business Media.

Liu, B., M. Hu, and J.Cheng. 2005. Opinion observer: Analyzing and comparing opinions on the Web. In *Proceedings of the 14th international conference World Wide Web*, 342–351. Chiba, Japan: Association for Computer Machinery.

Lovett, M.J., R. Peres, and R. Shachar. 2013. On brands and word of mouth. *Journal of Marketing Research* 50 (4): 427–444.

Ludwig, S., K. De Ruyter, M. Friedman, E.C. Brüggen, M. Wetzels, and G. Pfann. 2013. More than words: The influence of affective content and linguistic style matches in online reviews on conversion rates. *Journal of Marketing* 77 (1): 87–103.

Manning, C.D., and H. Schütze. 1999. *Foundations of statistical natural language processing*, vol. 999. Cambridge: MIT press.

Mayzlin, D., Y. Dover, and J. Chevalier. 2014. Promotional reviews: An empirical investigation of online review manipulation. *American Economic Review* 104 (8): 2421–2455.

Moe, W.W., and D.A. Schweidel. 2012. Online product opinions: Incidence, evaluation, and evolution. *Marketing Science* 31 (3): 372–386.

Moe, W.W., and M. Trusov. 2011. The value of social dynamics in online product ratings forums. *Journal of Marketing Research* 48 (3): 444–456.

Nam, H., and P.K. Kannan. 2014. The informational value of social tagging networks. *Journal of Marketing* 78 (4): 21–40.

Netzer, O., R. Feldman, J. Goldenberg, and M. Fresko. 2012. Mine your own business: Market-structure surveillance through text mining. *Marketing Science* 31 (3): 521–543.

O'Connor, B., R. Balasubramanyan, B.R. Routledge, and N.A. Smith. 2010. From tweets to polls: Linking text sentiment to public opinion time series. *ICWSM* 11 (122–129): 1–2.

Pang, B., and L. Lee. 2008. Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval* 2 (1–2): 1–135.

Pennebaker, J.W., M.E. Francis, and R.J. Booth. 2001. *Linguistic inquiry and word count: LIWC 2001*, 71. Mahway: Lawrence Erlbaum Associates.

Resnik, P., and R. Zeckhauser. 2002. Trust among strangers in internet transactions: Empirical analysis of eBay's reputation system. In *The economics of the internet and e-commerce,* vol. 11, 127 Advances in applied microeconomics.

Rosen-Zvi, M., T. Griffiths, M. Steyvers, and P. Smyth. 2004. The author-topic model for authors and documents. In *Proceedings of the 20th conference on uncertainty in artificial intelligence*, July 7, 487–494. AUAI Press.

Rzhetsky, A., I. Iossifov, T. Koike, M. Krauthammer, P. Kra, M. Morris, H. Yu, P.A. Duboué, W. Weng, W.J. Wilbur, and V. Hatzivassiloglou. 2004. GeneWays: A system for extracting, analyzing, visualizing, and integrating molecular pathway data. *Journal of Biomedical Informatics* 37 (1): 43–53.

Schlosser, A.E. 2005. Posting versus lurking: Communicating in a multiple audience context. *Journal of Consumer Research* 32 (2): 260–265.

Schweidel, D.A., and W.W. Moe. 2014. Listening in on social media: A joint model of sentiment and venue format choice. *Journal of Marketing Research* 51 (4): 387–402.

Schweidel, D.A., Y.-H. Park, and Z. Jamal. 2014. A multi-activity latent attrition model for customer base analysis. *Marketing Science* 33 (2): 273–286.

Stephen, A.T., and J. Galak. 2012. The effects of traditional and social earned media on sales: A study of a microlending marketplace. *Journal of Marketing Research* 49 (5): 624–639.

Srinivasan, S., O.J. Rutz, and K. Pauwels. 2015. Paths to and off purchase: Quantifying the impact of traditional marketing and online consumer activity. *Journal of the Academy of Marketing Science* 1–14.

Sun, M. 2012. How does the variance of product ratings matter? *Management Science* 58 (4): 696–707.

Swanson, D.R. 1988. Migraine and magnesium: Eleven neglected connections. *Perspectives in Biology and Medicine* 31 (4): 526–557.

Swanson, D.R., and N.R. Smalheiser. 2001. Information discovery from complementary literatures: Categorizing viruses as potential weapons. *Journal of American Society for Information Science and Technology* 52 (10): 797–812.

Tirunillai, S., and G.J. Tellis. 2012. Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science* 31 (2): 198–215.

Tirunillai, S., and G.J. Tellis. 2014. Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research* 51 (4): 463–479.

Toubia, O., and O. Netzer. 2017. Idea generation, creativity and prototypicality. *Marketing science* 36 (1):1–20.

Toubia, O., and A.T. Stephen. 2013. Intrinsic vs. image-related utility in social media: Why do people contribute content to twitter? *Marketing Science* 32 (3): 368–392.

Westbrook, R. A. 1987. Product/consumption-based affective responses and postpurchase processes. *Journal of marketing research* 258–270.

Yu, Y., W. Duan, and Q. Cao. 2013. The impact of social and conventional media on firm equity value: A sentiment analysis approach. *Decision Support Systems* 55 (4): 919–926.

You, Y., G.V. Gautham, and A.M. Joshi. 2015. A meta-analysis of electronic word-of-mouth elasticity. *Journal of Marketing* 79 (2): 19–39.

Zhang, H., G. Kim, and E.P. Xing. 2015. Dynamic topic modeling for monitoring market competition from online text and image data. In *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*, 1425–1434. ACM.

Chapter 17
Integrating Social Networks into Marketing Decision Models

Xi Chen, Ralf van der Lans and Michael Trusov

17.1 Introduction

The rise of online social networks has been the most significant development on the web in the last decade. It has not only transformed how consumers interact with each other, but also affected the way companies communicate with their customers. Social networks such as Facebook (launched in 2004), Twitter (launched in 2006), Sina Weibo (launched in 2009), WhatsApp (launched in 2009), and WeChat (launched in 2011) are now standard platforms for consumers to share experiences, communicate with friends and follow celebrities and brands. In combination with the rising popularity of smart phones, consumers are continuously online, with Facebook having almost 1 billion users that visit the social network on a daily basis (Facebook Reports 2015). According to an *eMarketer* report, over one in four people worldwide use social networking sites in 2015,[1] and the penetration rates in North America and Western Europe are nearly 50%. These developments have provided opportunities for new marketing strategies, such as viral marketing

[1]See http://www.emarketer.com/Articles/Results?t=1045&p=5 for more details.

X. Chen (✉)
Rotterdam School of Management, Erasmus University Rotterdam,
3062 Rotterdam, Netherlands
e-mail: chen@rsm.nl

R. van der Lans
School of Business and Management, Hong Kong University
of Science and Technology, Hong Kong, China
e-mail: rlans@ust.hk

M. Trusov
Robert H. Smith School of Business, University of Maryland,
College Park, MD 20742, USA
e-mail: mtrusov@umd.edu

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_17

(Van der Lans et al. 2010), crowdsourcing in new product (Stephen et al. 2016), and crowdfunding of new projects and ventures (Lin et al. 2013; Lin and Viswanathan 2016; Ordanini et al. 2015). In addition, similar to the introduction of scanner data in the 80's, the emergence of online social networks has provided a wealth of data for marketers to better understand social interactions between consumers and information to optimize marketing decisions. For instance, Hinz et al. (2011) found that using social network data to optimize a referral program of a mobile service provider was able to almost double the number of newly registered customers compared to a strategy that ignored social network data. Despite these successful applications of integrating social network data into marketing decision making, the analysis of social network data has brought many new challenges. The goal of this chapter is to provide a modeling framework for incorporating social network data and to discuss the major challenges and potential solutions.

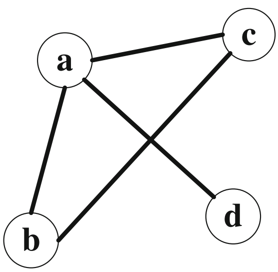
Although the increasing availability of social network data is new, the analysis of social networks is not (Wasserman and Faust 1994). Marketing and sociology have a long tradition to understand how consumers and companies interact with each other and what the consequences of these interactions are (Iacobucci and Hopkins 1992). The structure of the interactions and relationships among consumers and companies (i.e., the structure of the social network), may have important consequences, such as the success of finding jobs (Granovetter 1973), and survival of firms (Uzzi 1996). Not surprisingly, marketing researchers have always been interested in understanding social interactions between consumers. For instance, Reingen (1984) studied the social network of sorority members of a university and how the social network structure affected choice decisions. The social network was obtained through surveys. Similarly, Coleman et al. (1966) collected the social network of physicians through surveys to study social influence on adoption decisions of new drugs. These studies have provided many useful theoretical insights into the structure of social networks and how they shape consumer decision making. However, these studies generally involve small networks collected by surveys, consisting of at most a few hundred customers, which is in stark contrast with today's massive social networks consisting of thousands and even millions of consumers. Applying these methodologies to large social networks and integrating them into marketing decision models is challenging. Other areas in marketing and management have focused on larger scale networks, such as citation networks (Goldenberg et al. 2010), networks of alliances between firms (Baum et al. 2010), and networks of products (Oestreicher-Singer et al. 2013). These network analyses are interested in the general structure of the network and how individuals (i.e., journals or firms) are positioned in these networks compared to others (Iacobucci and Hopkins 1992). These network analyses do not link the structure of these networks to behavioral outcomes, which is important for optimizing marketing strategies.

In this chapter, we do not attempt to review comprehensively the literature on social networks. Instead, we aim to provide a framework for researchers and practitioners to understand social networks and how to assimilate them into the development of new models for marketing decision making. This chapter is structured as follows. In Sect. 17.2, we provide a brief overview of how social networks can be represented mathematically as adjacency matrices and how to derive some key properties, such as centrality measures from these matrices. Section 17.3 is the core of our chapter and provides four modeling approaches to incorporate social networks into marketing decision models. Section 17.4 discusses important challenges of social network analysis, and finally Sect. 17.5 closes with a discussion.

17.2 A Brief Introduction to Social Network Analysis

Figure 17.1 illustrates a social network of four individuals, a to d. As illustrated in the graph, these four individuals are tied through four connections or links. In this example, individual a is connected to all other individuals in the network, while individuals b and c are connected to each other as well as individual a. Finally, individual c is only connected to individual a. In addition to representing this social network as a graph with nodes for individuals and links for relationships (Fig. 17.1—left panel), the social network can also be represented mathematically as an adjacency matrix A (Fig. 17.1—right panel). The adjacency matrix in this example is a square matrix consisting of four columns and rows, corresponding to each individual a to d, respectively. Each element a_{ij} in the adjacency matrix A indicates whether two individuals i and j are connected. If $a_{ij} = 1$, this means that individual i and j are connected, and it equals zero otherwise. Moreover, because in this example, connections are mutual $a_{ij} = a_{ji}$. Thus, the adjacency matrix in this example contains eight elements that equal one, corresponding to the four connections in the network (for instance, $a_{12} = 1$, corresponding to the connection between individuals a and b). By convention, the diagonal elements of the adjacency matrix are equal to zero, such that an individual cannot be connected to itself.

Based on this example, more generally, we can define a social network by a group of N individuals (or other social units) connected through links. An individual may for example represent a user on Facebook, but also a brand or company,



Graph

⟺

	a	b	c	d
a	0	1	1	1
b	1	0	1	0
c	1	1	0	0
d	1	0	0	0

Adjacency Matrix

Fig. 17.1 Graph and matrix representation of a social network while a link may represent a certain type of social relation between two individuals (e.g., friendship on Facebook, but also colleagueship or advisor-advisee relationships). A social network can be represented by a graph or by $N \times N$ adjacency matrix A, as explained above. Although the social network in the example above is undirected, resulting in a symmetric adjacency matrix in which $a_{ij} = a_{ji}$, and that all connections are equally important (i.e. $a_{ij} \in \{0, 1\}$), this is not necessarily the case. Some networks may be directed (i.e., $a_{ij} \neq a_{ji}$), while connections could have different weights with $a_{ij} \in (-\infty, +\infty)$. For instance, the social network of Twitter is directed, as following other members in this social network is not neccessarily mutual. Similarly, not all connections between individuals in a network may be equally important, which can be represented by a weighted network with $a_{ij} \in (-\infty, +\infty)$ representing the tie strength between two individuals. Finally, a social network may consist of multiple connections, which can be represented by K different adjacency matrices A_k, with $k \in \{1, 2, \ldots, K\}$. For example, individuals can be connected through links of both friendship and colleagueship, which is called a "multiplex" or "multi-relational" network.

The adjacency matrix A is useful to extract important network properties (see Table 17.1 for frequently used examples, and Wasserman and Faust 1994; Watts and Strogatz 1998 for a more detailed overview). These network properties describe how individuals fit together to form a social network. These networks measures can be divided into two levels: macro-, and micro-level measures. Macro-level measures describe aggregated characteristics of the social network or important subcomponents of it. Micro-level measures describe differences among individual's connections compared with other individuals in the social network. Macro-level measures look at the structural properties of the overall network, such as size, number of links, and distribution of centralities of individual network members. Table 17.1 lists three important macro measures. Density could serve as a proxy of the overall social influence traversing the network. Reciprocity and transitivity indices reflect how closely individuals on the social network are connected. Reciprocity is a useful measure in directed networks (such as Twitter) as it calculates the percentage of relationships between people that are mutual. Transitivity (or global clustering coefficient) indicates to what degree individuals in a social network tend to cluster.

Micro-level measures focus on individuals in the network and their positions. Researchers use centrality measures to capture the importance of individuals in the network, and these can be used for targeting strategies. Examples of centrality measures are provided in Table 17.1. Degree centrality counts the number of connections an individual has. Betweenness and closeness centrality reflect individuals that are positioned on the shortest paths connecting other individuals on the social network. These individuals are important bridges between different parts of the network. Individuals have high eigenvector centrality if they have many connections and when these connections are relatively important as well (i.e., these individuals have many friends, whose friends also have many connections, etc.). The clustering coefficient reflects how strongly connected the connections of an individual are, and is related to the macro-level transitivity measure.

Table 17.1 Descriptions of selected network measures

Network measures	Formulae	Explanations
Macro-level measures		
Density	$\frac{\sum_i \sum_j a_{ij}}{N \times (N-1)}$	It captures the sparsity or density of the network, with value $(0, 1)$
Reciprocity	$\frac{\sum_i \sum_j a_{ij} \times a_{ji}}{M}$	It calculates the proportion of links for which a link in the opposite direction exists, i.e. $a_{ij} = a_{ji} = 1$. For undirected network, this value equals to one
Transitivity	$\frac{\sum_{i \neq j,k} \sum_{j \neq k} \sum_k a_{ij} a_{ik} a_{jk}}{C_n^3}$	It captures the basic idea of transitivity or the extent of "friends of my friends are also my friends"
Micro-level measures		
Degree centrality	$d_i = \sum_j a_{ij}$	It counts the number of direct connections (i.e., friends) of an individual
Betweenness centrality	$\sum_{j \neq i,k} \sum_{k \neq i} \frac{ns_{jk}^i}{ns_{jk}}$	ns_{jk} is the number of shortest paths between j and k; and ns_{jk}^i the number of shortest paths between j and k via i. It reflects the centrality node i when items are transferred along (shortest) paths
Closeness centrality	$\frac{1}{\sum_{j \neq i} ls_{ij}}$	ls_{ij} is the distance between i and j, measured by length of shortest path between them. It captures global centrality of a node i, or how close node i is to any other node in the network
Eigenvector centrality	$x^{(i)}$, where $Ax = \lambda_{\max} x$	$\lambda_{\max}$ is the maximum eigenvalue of A, and x is the associated eigenvector. It captures the idea that node i is central if it's connections are also central
Local clustering coefficient	$\begin{cases} \frac{\sum_j \sum_k a_{jk}}{d_i \times (d_i - 1)}, \text{if } A \text{ is undirected} \\ \frac{\sum_j \sum_k a_{jk}}{2 \times d_i \times (d_i - 1)}, \text{if } A \text{ is directed} \end{cases}$	Where j and k are i's neighbors. It measures how closely connected an individual's neighbors are

17.3 Integrating Social Networks into Marketing Models

To integrate social networks into marketing decision models of social influence, we first propose a generalized framework. Using the generalize framework, we then describe four modeling perspectives of integrating social networks into marketing models: (1) network measures, (2) statistical models, (3) economics models and (4) agent-based models. The generalized framework contains two levels, a "data model" and a "network process model", detailed as follows:

$$\begin{cases} \text{Data Model: } Y = f\left(X_1, Z \middle| \beta\right) & (17.1) \\ \text{Network Process Model: } Z = g\left(A, X_2 \middle| \theta\right) & (17.2) \end{cases}$$

In the above two equations, Y is the marketing variable of interest, which may capture individual consumer decisions (e.g., adoption, choice, forwarding information) or aggregated patterns of consumer decisions (e.g. penetration rates, diffusion patterns, sales, the reach of viral marketing campaigns). The dependent variables Y are assumed to depend on a set of control variables X_1 and social influence Z. The type and structure of social influence, represented by Z, on the marketing variable of interest Y is determined by the network process model (Eq. 17.2). Function $f(\cdot)$ links control variables X_1 and social influence Z to marketing variable Y, and function $g(\cdot)$ links adjacency matrix A, potential other control variables X_2 to social influence variable Z, with β and θ as associated parameter vectors.

In our formulation, marketing decision variables such as advertising, sales promotions, or seeding strategies can be part of both sets of control variables X_1 and X_2. In case marketing decision variables are captured by control variables X_1, the goal of including social influence in the Data Model (17.1) is to obtain more accurate estimates of the effects of marketing decisions on marketing outcomes Y. For instance, Yang and Allenby (2003) showed that for explaining Japanese car choices (i.e., Y variable), the parameter estimates for marketing instruments, such as price, were more accurate after controlling for network effects (Z). Alternatively, marketing decision variables may be part of X_2 in the network process model. In this case, it is possible to determine how marketing decision variables indirectly affect marketing outcomes through social influence. For instance, Hinz et al. (2011) studied the effects of seeding on attracting new customers for a mobile service provider. In this case, X_2 represents the set of customers that have been targeted to initiate the viral marketing process. Targeting different groups of customers leads to different viral processes and thus different customer acquisition results.

Next, using Eqs. (17.1) and (17.2) we explain how the four modeling perspectives: network measures, statistical models, economic models and agent-based models, integrate social networks into marketing decision models. The key difference between these four perspectives is the network process model, i.e. how social influence is derived from the social network structure.

17.3.1 The Network Measures Approach

Most research in marketing uses the network measures approach. This approach directly derives social influence from the adjacency matrix using the tools described in Sect. 17.2. Therefore, the process model in Eq. 17.2 represents a series of formulae for network measures. For instance, if degree centrality or closeness centrality is considered to influence consumer decisions, the process model represents the formulae for these centrality level measures. In addition to micro-level measures that influence individual consumer decisions, medium- and/or macro-level measures could be used to describe aggregate level marketing variables. For instance, researchers could use the size of the network or the number of hubs in the network to predict the reach of a viral marketing campaign. The most commonly used network measures in previous marketing research are centrality measures, such as degree centrality and eigenvector centrality (see Table 17.1 for definition).

If the network measures do not depend on unknown parameters, the implementation of the network measures approach is relatively straightforward. Given the network measures of interest, researchers may directly implement these measures to represent social influence variables Z in the data model (Eq. 17.1). The choice of network measures is of course an important question and can be approached theoretically as well as empirically. Empirically, researchers could select network measures based on the fit of the data model, such as R^2, BIC (Bayesian Information Criterion), or LMD (Log Marginal Density). Alternatively, there is a rich discussion in the sociology literature about network measures and how they relate to different types of social influence. For example, an important question in sociology is whether social contagion in adoptions stems from information transfer or normative social influence (Burt 1987; Van den Bulte and Lilien 2001). The answer to this question can be derived from the importance of different network measures in adoption decisions. If social contagion stems from information transfer, degree centrality should significantly influence adoption. Otherwise, if social pressure is the driving force of social contagion, structural equivalence[2] should affect adoption.

The network measures approach has been particular popular in the new product diffusion literature. For instance, Iyengar et al. (2011) studied how opinion leadership and social contagion within social networks affect the adoption of drugs by physicians. They related the adoption decision of physicians to several network measures, such as the number of incoming and outgoing connections in different networks, as well as whether these connections had earlier adopted the drug. In contrast to Van den Bulte and Lilien (2001), they found evidence of contagion operating over network ties, even after controlling for marketing effort. Moreover, using their model estimates, they found that network interventions to promote a new drug (such as stimulating opinion leaders to influence their peers), are more effective than traditional detailing activities. In a follow-up study, Iyengar et al. (2015) found that network effects are different for adoptions and repeated subscriptions. Katona et al. (2011) studied the diffusion of an online social network. They found that the adoption probability to register to the social website was higher for consumers who were connected to many adopters (i.e., degree centrality) and when these adopters formed a dense sub-network (i.e., clustering effect).

[2]Structural equivalence refers to the extent to which two consumers are connected to the same or similar others. One frequently used formula is: $SE_{ij} = \frac{d_{ij}}{d_i + d_j - d_{ij}}$, where $d_{ij} = \sum_{k \neq i,j} a_{ik} a_{jk}$.

Goldenberg et al. (2009) studied how the structure of hubs, sub-networks where individuals with much higher degree centrality, drives adoption and the diffusion of online goods on a social platform. They found that hubs tend to adopt earlier in the diffusion process, and that innovative hubs have a greater impact on the speed of the adoption process, and follower hubs have a greater impact on market size. Hinz et al. (2011) compared different seeding strategies (i.e. selecting a group of consumers to initiate the viral marketing campaigns) to attract new customers to a mobile service provider. They found that selecting well-connected consumers (i.e. high in degree centrality) in a social network resulted in a much higher conversion of new customers. Compared to a random seeding strategy, in which network measures were ignored, the conversion rate was doubled by targeting well-connected consumer. Tucker (2008) investigated which adopters of a video-messaging technology caused the strongest effects of network externalities on other potential adopters. She compares four centrality measures and found that adopters in boundary spanning positions caused the strongest effects on network externalities.

Aside from diffusion-related settings, the network measures approach has also been used to study the drivers of user-generated content (UGC) (Mallapragada et al. 2012; Ransbotham et al. 2012; Yoganarasimhan 2012), idea generation (Stephen et al. 2016), e-commerce (Stephen and Toubia 2010) and customer retention (Nitzan and Libai 2011). For instance, Yoganarasimhan (2012) studied how the size (degree centrality and second-order degree, i.e., friends of friends) and structure of the local network (betweenness and clustering) around the poster of a YouTube video affects its propagation. She demonstrated that both the size and structure of a poster's local network is a significant driver of the number of video views. Mallapragada et al. (2012) examined the effects of the positions of founders in a social network on the release timing of open source projects. They found that a more central position of founders in the developer community could reduce the time to product release by up to 31%. Ransbotham et al. (2012) found that the value of collaborative UGC is a function of both direct efforts of its contributors and their embeddedness (as measured by degree and closeness centrality) in the content-contributor network. Stephen et al. (2016) studied the creativity of customers in a social network as a function of their position (degree centrality and clustering). They found that customers become less creative if their local network is highly interconnected, given they rely on their peers as a source of information. Stephen and Toubia (2010) examined the economic value implications of a social network between sellers in a large online social commerce marketplace. They found that increasing the connectivity of a seller's network makes them more accessible to buyers and consequently generates considerable economic value. Nitzan and Libai (2011) explored the role of customers' social network in their defection from a service provider. They found that exposure to a defecting neighbor is associated with an increase of 80% in the defection hazard, after controlling for other factors such as satisfaction and economic incentives.

The studies described above demonstrate how the introduction of social network measures in the data model (Eq. 17.1) can help further our understanding about the role of network structure in various marketing phenomena. The previous studies all assumed that the network measures do not depend on unknown parameters θ in the process model (Eq. 17.2). However, it is also possible that social influence Z derived from the adjacency matrix A depends on unknown parameters θ. Some network measures, such as Bonacich centrality, depend on unknown parameters (Bonacich 1987). While previous research often fixes the parameters of these centrality measures to pre-specified values, it is not always clear whether such centrality measures accurately describe social influence. For instance, by setting the β parameter of Bonacich centrality to the inverse of the largest eigenvalues of the adjacency matrix, one obtains eigenvector centrality. Eigenvector centrality implies that individuals are more central if they have many connections, and their connections also have many connections. Gelper et al. (2015) demonstrated that this measure did not accurately predict the spread of information in viral marketing campaigns. In a large-scale viral marketing campaign on an online social network similar to Facebook, they showed that the estimation of the Bonacich centrality parameters in the network process model led to substantially different insights and provided more accurate forecasts. They found that the most influential individuals have many connections, but those connections only have few friends, which is represented by a negative value of the β parameter in the Bonacich measure. In a holdout sample, they demonstrated that seeding network members using the Bonacich centrality with estimated β resulted in a 70% higher, compared to a seeding strategy that selected individuals with the highest degree (and 230% higher compared to a random seeding strategy that ignores the social network structure). In another study, Chen et al. (2017) provided a Bayesian estimation framework to estimate the parameters of the network process model. They considered that individuals are connected through relationships with different characteristics and that the importance of different relationship characteristics can be integrated into one weighted network, with the weights estimated in the process model. In two empirical applications, they found that the importance of relationship characteristics vary substantially in the diffusion of information, which has important implications for the selection of seeds in viral marketing campaigns. In the first empirical application, involving the diffusion of a microfinance program in small Indian villages, they found that adoptions can be increased by 10% if the importance of relationships is taken into account. Even more strikingly, in their second empirical application, involving a large online social network in which individuals shared messages about brands, the reach of the campaign could be improved by up to 92%, compared with a seeding strategy that ignored the importance of relationship characteristics.

Although the network measures approach is simple and powerful, it essentially summarizes the whole network structure with only a few measures. In some situations, this may lead to the loss of important information. The statistical approach is a possible solution, which aims to integrate the *whole* network into the data model. Next we discuss this approach.

17.3.2 The Statistical Approach

Another way to integrate social networks into marketing decision models is to use the statistical approach, which models the social network as a stochastic spatial process. For instance, Yang and Allenby (2003) modeled interdependence of consumers' decisions using a spatial auto-regression specification, where the spatial weighting matrix is replaced with multiple (weighted) social networks. Moon and Russell (2008) developed a product recommendation model based on the principle that customer preference similarity stemming from prior purchase behavior is a key element in predicting current product purchase. Using an auto-logistic model, they demonstrated that the proposed approach provided better recommendations compared to benchmark models that ignore social factors for a customer database provided by an insurance firm. Choi et al. (2010) studied the imitation effect among new buyers on an e-commerce website with a spatio-temporal model, where interdependence of consumers is considered in a similar way as Yang and Allenby (2003). More recently, Wang et al. (2013) employed a Markov Random Field (MRF) to study how consumers' product choices were influenced by the product choices of their peers and how the influence mechanism differed for fashion- versus technology-related products. As expected, they found that for fashion-related products, popular individuals (with more desirable characteristics) were more influential, while for technology-related products they found that experts (with more credibility) were more influential. To illustrate the statistical approach, we describe the model introduced by Yang and Allenby (2003) in more detail.

Using Eqs. (17.1) and (17.2), the spatial autoregressive model proposed by Yang and Allenby (2003) can be presented as follows:

$$\begin{cases} \text{Data Model: } Y = X\beta + Z + \varepsilon & (17.3) \\ \text{Network Process Model: } Z = \rho A Z + \mu & (17.4) \end{cases}$$

In this model, Y represent latent preferences in a *Probit* model, such that consumers make a choice if $Y > 0$, and $X\beta$ captures the effects of covariates on choice. The vector of error terms ε is assumed to be *i.i.d.*, following a standard normal distribution. As noted by Yang and Allenby (2003), consumers latent preferences Y can be correlated as their preferences are interdependent through their relationships in the social network. This correlation is captured by the autoregressive parameter Z, which is determined by the Network Process Model, a spatial autoregressive process depending on adjacency matrix A. The error term μ is assumed to be normally distributed. An important parameter in this model is ρ, which captures the strength of the spatial auto-correlation and the correlation of latent preferences in the network. As we discussed above, the adjacency matrix is allowed to be a weighted matrix. Yang and Allenby (2003) specified it as a weighted sum of different adjacency matrices, with the weights estimated by the model. This model captures the complete interdependence of consumer preferences in their social network and can be estimated using Bayesian techniques. Yang and Allenby (2003) applied this model to purchases for Japanese-made cars to understand to what extent preferences are interdependent among consumers. They found that consumer preferences are indeed interdependent and that geographic reference groups were stronger determinants of these interdependencies, compared to demographic reference groups. Moreover, allowing for these interdependences led to more accurate parameter estimates as well as holdout sample forecasts that are managerially relevant.

A limitation of the spatial autoregressive model is that the covariance matrix of the error terms is restrictive and only depends on parameter ρ and the adjacency matrix. It is possible to allow a more flexible structure using Markov Random Field (MRF). The assumption in Markov Random Field is that the preferences of two individuals i and j are conditional independent if i and j are not connected. In practice, a commonly used MRF is the Gaussian MRF, which can be presented as follows:

$$\begin{cases} \text{Data Model: } Y = X\beta + Z & (17.5) \\ \text{Network Process Model: } Z \sim N(0, \Sigma) & (17.6) \end{cases}$$

where, $\left(\sum^{-1}\right)_{ij} = 0$ if and only if $a_{ij} = 0$. This model integrates the whole network by linking the precision matrix $\sum^{-1}$ to the adjacency matrix. Interested readers are referred to Cressie (1991) for more detailed discussions.

The key difference between the "statistical approach" and the "network measures approach" is that the former integrates the whole network into the data model, while the latter may lose a certain level of information. Therefore, the statistical approach could lead to better model fit and forecasting performance. On the other hand, the network measures approach may be attractive in practice as various network measures are generally linked to different well-established sociology theories and are easier to implement and estimate. A commonality of both approaches is that they focus on statistical relationships between variables, and these relationships may not hold if the structure of the model changes due to policy or regime shifts (Lucas 1976). To accommodate such structural changes, it is necessary to model the underlying decision process of consumers in the social network. Both the agent-based and economics approach try to integrate assumptions of underlying consumer decisions in the model, which we describe next.

17.3.3 The Economics Approach

The "economics approach", sometimes called "structural approach", accounts for how consumers adjust their decisions in response to policy/strategy changes. Specifically, this approach falls into the category of network games in economics (Galeotti et al. 2010). Network games study social/economic interactions where consumers' well-being depends on their own actions as well as on the actions taken by their peers. The modeling framework can be calibrated on actual data for policy or strategy evaluation (e.g. pricing, promotion) (Jackson 2014). In marketing, Hartmann (2010) was the first to use the economics approach to study how consumers coordinate their choices in small groups. Coordination of choices is common when eating together at restaurants, watching movies together, or shopping together. Using a dataset recording group membership of golfers during their purchase decisions, he was able to identify social influence in coordinated decisions. These findings have important implications for targeting decisions, as valuable customers do not necessarily purchase frequently, but may instead exert strong influence on others. Specifically, Hartmann found that on average 35% of the value of customers could be attributed to how they affected other customers in the group. While Hartmann focused on coordinated decisions in small groups of not more than four golfers, Chen et al. (2016) integrated a large scale social network into the choice model of consumers for which the choice utility depends on decisions of their peers, but their choices are not coordinated. For instance, when donating to charity, choosing the time to go to work or a supermarket to avoid traffic, buying apparel or luxury products, or deciding when to play an online game to encounter desirable opponents, consumers often do not coordinate their decision, but their utility depends on decisions of others. They highlighted that in customer base analysis it is important to integrate social influence to determine the value of a customer to the firm. Application of their model to login decisions in a social network similar, they found that on average 20% of the value of consumers can be attributed to their social influence on other members in the network. Moreover, they found significant heterogeneity across consumers, such that the social value of some consumers exceeded their own direct value and other consumers even had negative social value. Next, we will briefly review models from these two papers.

Hartmann (2010) investigated consumers' decision-making with social interaction, employing a discrete and static game of complete information (i.e. coordination game) that can be summarized as follows:

$$\begin{cases} \text{Data Model:} & y_i = \begin{cases} 1 \text{ if } V\left(x_i;\beta\right) + S\left(y_{-i};\gamma_{ig},\alpha_{ig}\right) + \varepsilon_i > 0 \\ 0 \text{ otherwise} \end{cases} & (17.7) \\ \text{Network Process Model:} & S\left(y_{-i};\gamma_i,\alpha_{ig}\right) = \gamma_{ig} \sum_{k=1}^{z_i} \alpha_{ig}^{z_i-1}, \text{ with } z_i = \sum\nolimits_{j \neq i} A_{ij}^{g} y_j & (17.8) \end{cases}$$

Here, y_i is a binary decision variable and y_{-i} summarizes the decisions of other customers. $V(x_i;\beta)$ and $S(y_{-i};\gamma_{ig},\alpha_{ig})$ are the private and social component of utility as in Brock and Durlauf (2001), and ε_i is an error term that is unobserved to the econometrician but observable to all consumers in the group (i.e. common knowledge), resulting in a complete information game. The network process model focuses on group g that consumer i belongs to when making the decision. This group g is reflected by sub-network A^g, which is a complete network such that all customers in the group are connected. In this model, consumer i is assumed to coordinate choice and derive utility from the aggregate decisions of the group z_i. Parameters γ_{ig} and α_{ig} capture the strength of the social interactions, where α_{ig} captures possible nonlinear effects in the number of group members that make similar decisions. A challenge in estimating network games, including the proposed model by Hartmann is that there exist multiple decisions y that satisfy the system of equations, leading to multiple equilibria. Hartmann solved this problem by selecting the Pareto dominant equilibrium. Moreover, the model is tailored to decision making in small groups (consisting of not more than four customers) in which everyone is connected to each other, and incorporating large social networks is nontrivial.

Chen et al. (2016) developed a choice model with social interactions that is able to incorporate large scale social networks. Different from Hartmann (2010), they assumed consumers only know the distribution but not the realized values of their peers' random utility shocks. This results in an incomplete information game in which consumers make decisions based on their expectations about peer decisions, instead of actual decisions as in Hartmann (2010). Their model can be summarized as follows:

$$\begin{cases} \text{Data Model:} \quad y_i = \begin{cases} 1 \text{ if } V(x_i;\beta) + S(y_i, E_i(y_{-i});\gamma_i, A) + \varepsilon_i > 0 \\ 0 \text{ otherwise} \end{cases} & (17.9) \\ \text{Network Process Model:} \quad S(y_{-i}, E_i(y_{-i});\gamma_i, A) = \gamma_i E_i\left[1 - \frac{1}{d_i}\sum_{j=1}^{N}(y_i - y_j)^2\right] & (17.10) \end{cases}$$

A key difference in this formulation compared to Hartmann is that social influence $S(y_i, E_i(y_{-i});\gamma_i, A)$ depends on consumer i's expectations about peers' decisions $E_i(y_{-i})$ and that the full social network A is taken into account. The network process model (17.10) assumes that social influence is determined by the percentage of connections that are expected to make a similar decision as customer i. This leads to a system of equations that under certain conditions has a unique Bayesian-Nash equilibrium. Chen et al. (2016) developed a stochastic Bayesian estimation that efficiently reduces the computation of the model's equilibrium, allowing application of this model to large scale networks.

Similar to the spatial statistics approach, the economics approach incorporates the entire social network. In Hartmann's case, the whole network is divided into many sub-networks where everyone is linked to everyone. Chen et al. (2016)

extended this to the complete network, which may be directed and/or weighted. A key feature of the economics approach is the incorporation of an equilibrium, which results from social interaction. This allows researchers to do policy/strategy evaluations through counterfactual studies. Both Hartmann (2010) and Chen et al. (2016) quantified the social value of customers with counterfactual studies assuming, respectively, that a consumer left the group or became inactive on the social network.

17.3.4 The Agent-Based Approach

The three approaches reviewed above are able to derive many insights, but their analysis could be difficult as simple interactions often lead to complex aggregate outcomes. The agent-based approach is an alternative approach that simulates interactions on social networks from basic principles of consumers' behaviors and aggregates the outcomes for further analysis or theory testing (Rand and Rust 2011). This approach is particularly useful when closed-form expressions for Network Process and Data Models are difficult (if not impossible) to obtain. For example, by using the agent-based approach, the relationship between a behavioral outcome Y (e.g., aggregated product adoptions) and social influence Z can be explored by assuming some agent-level behavioral model that accounts for peer influence effects and then simulating agent decisions on the network defined by A while manipulating parameters of interest (e.g., network structure or inputs to the behavioral model).

In the last decade, the agent-based approach has gained popularity in the marketing literature, especially to better understand diffusion processes (Garber et al. 2004; Goldenberg et al. 2002; Watts and Dodds 2007). For instance, Garber et al. (2004) used cellular automata (a special case of agent-based models) to explain the effects of spatial clustering of early sales on the diffusion of products. Based on these simulations, the authors propose a spatial distance measure to predict the early success of product launches. Application of this measure to new health and personal hygiene products sales in supermarkets as well as home furniture sales, demonstrated that the newly proposed measure was able to predict products success accurately.

While Garber et al. (2004) used the aggregate outcomes of the simulation to validate real-world phenomena and to come up with new measures to predict product success, researchers have also used the agent-based approach directly in empirical models (Dover et al. 2012; Trusov et al. 2013). Empirical data collected on a micro (e.g., individual consumer) and/or a macro-level (e.g., aggregated sales) can be used in a model validation process that includes grounding, calibration, verification and harmonization (Garcia et al. 2007). In terms of the basic model structure (Eqs. 17.1–17.2), the network process model corresponds to the agent-based model that simulates individual decisions on a social network. Subsequently, this information is used in the data model to fit aggregate diffusion data.

Using this procedure, it is possible to determine the parameters of the agent-based model that best fits the data model (e.g., using standard fit statistics such as R^2). In order to integrate these two steps, in some cases researchers need to first develop the agent-based model analytically. For instance, Dover et al. (2012) developed a network-based growth model and solved for growth and decline slopes, which served as basis for the empirical testing. Specifically, they used this approach to obtain early forecasts of diffusion processes over a social network, while not observing the actual network. Across 17 data sets involving adoption processes across many different industries, they demonstrated that a network's degree distribution has a significant impact on the contagion properties of the subsequent adoption process. Their agent-based approach was able to recover the actual degree distribution without observing the actual network, which in turn allowed them to improve diffusion forecasts.

While Dover et al. (2012)'s method used early observations to learn network characteristics, which are then used to improve forecast accuracy for later stages of the adoption process, Trusov et al. (2013)'s approach focused on the product *pre-introduction* forecast. They argued that the product diffusion process was affected by both systematic and non-systematic factors. Hence, a model calibrated on a single diffusion instance may face transferability issues when applied to a product that is introduced to a market under different non-systematic conditions. Their approach focused on identifying market conditions that are relatively stable (or common to all products) by exploring systematic patterns *across* diffusions rather than *within* a diffusion. Trusov et al. (2013) extracted the systematic effect from the historical data by matching the distribution of historical diffusions with a set of synthetic footprints obtained through simulations on a set of networks of known types and used this as an input for Bayesian inference models. They applied their method to the adoption of applications on Facebook and found that their method was able to uncover the true underlying structure of the social network, resulting in more accurate prelaunch prediction of the diffusion process. This methodology bridges the gap between the disciplines of Bayesian statistics and agent-based modeling, and their approach shows superior performance compared to the models that ignore a network component.

17.4 Modeling Challenges and Potential Solutions

As in any empirical application, the analysis of social networks in marketing faces challenges regarding data collection, data availability, and unobserved variables that may lead to endogeneity problems. First, marketing researchers generally analyze sampled networks that are relatively small in size and with clear spatial or social boundaries, for instance the network of physicians in specific cities. However, social networks generally consist of millions of consumers and ignoring parts of the network may affect the behavior of consumers in the sample. Selecting a proper sample from large networks such that the most important properties of the original global network are preserved remains a challenging issue. Second, the relationships between social networks evolve over time and are usually weighted and/or directed. In most research, social network data represents the situation at a specific moment in time and only links between consumers are accessible to researchers and not their weight or importance. Thus, social network data is generally censored (relationship strength is unobserved) and truncated (the network is static), which may lead to biased parameters estimates. A third challenge in accurately recovering the effects of social networks is due to endogeneity. People tend to befriend similar others, and connected people may therefore be similar based on observed and unobserved characteristics. Failing to control for unobserved similarities between connected people may result in endogeneity problems, as effects of unobserved similarities on behaviors may be mistakenly attributed to the effects of social influence. These challenges are not unique to social networks, for instance sampling, measurement errors and endogeneity are common issues in empirical models. However, approaching these issues in social network settings requires new understandings and methods. Overall, researchers are just beginning to understand and address these issues (Jackson 2014), and some potential solutions thus being more a conceptual framework than a practical solver.

17.4.1 Sampling from Social Networks

As mentioned above, the size of a social network is typically very large and may even be unknown to researchers. As a consequence, sampling is necessary for most studies concerning social networks. However, obtaining a valid sample from social networks might be challenging. As a first obstacle, the definition or criteria for a valid sample is obscure. Unlike in settings that do not involve social networks, it is fairly difficult to construct a series of criteria to evaluate a sampling scheme or the obtained samples. In addition, the validity of a sample often depends on the purpose of the research, whether to recover the topology of the global network (Ebbes et al. 2013) or the accurate evaluation of social inter-correlations (Chen et al. 2013). Here, we review some of the interesting findings from two pioneering papers on this front in marketing.

Ebbes et al. (2013) evaluated the efficacy of nine different sampling methods in recovering the underlying structural characteristics of global networks. Some important sampling procedures are random node sampling, forest fire sampling, random-walk sampling and snowball sampling. The random node sampling procedure is similar to the simple random procedure in which individuals are randomly selected from the network. In forest fire, random-walk and snowball sampling, researchers first randomly select a seed individual from the network and then randomly select their neighbors and their neighbors' neighbors etc., using different procedures. In forest fire sampling, each neighbor is sampled with a certain burn probability, in random-walk sampling only one neighbor is selected and in snowball sampling all neighbors are selected. In their comparison, Ebbes et al. (2013)

focused on the recovery of four characteristics: the degree distribution, clustering coefficient, betweenness centrality, and closeness centrality, each of which is relevant for certain marketing processes. Via extensive simulations, they found that sampling methods differ substantially in the ability to recover global network characteristics. Traditional sampling procedures, such as random node sampling, produced poor results. When the focus of a marketing research project is on understanding peer influence (e.g., degree distributions and clustering coefficients), *forest fire sampling* with a medium burn rate performs the best. When the focus is on broader network effects (e.g., speed of diffusion, or the "multiplier" effects of network seeding), then *random-walk sampling* (i.e., forest-fire sampling with a low burn rate) performs the best, as it is most effective in recovering the distributions of betweenness and closeness centrality. Also, of great relevance for marketers, sample size has only a minimal impact on sampling performance, unless the sample is very small relative to population size. These results have been validated using different real-world networks, including a Facebook network and a co-authorship network.

Chen et al. (2013) compared sampling methods from a different angle. While Ebbes et al. (2013) aimed to recover structural properties of the global social network, Chen et al. (2013) investigated the recovered social auto-correlation with sampled sub-networks using a spatial auto-correlation model similar to Yang and Allenby (2003). Using simulations of various networks and sampling methods, they demonstrated that sampling methods that preserve the network structure (such as snowball sampling and forest-fire sampling with high forward-burning probability) perform better in recovering the social auto-correlations. They also pointed out that for scale-free networks (i.e. social networks with power-law degree distribution), it is more difficult to obtain a valid sample to recover the social auto-correlation. This is because in power-law networks there are only a few individuals with a large number of connections and these individuals tend to have a too small probability of being included in the sample. This led to an under-estimation of the magnitude of social auto-correlations in the sampled networks.

These two studies highlighted that the "best" sampling method depends on the goals of the research as well as the structure of the social network. However, they provide some important guidelines. From these studies, it seems that traditional random node or random link sampling methods generate biased samples. The graph exploration sampling procedures that include snowball, random-walk and forest fire sampling seem to be able to generate satisfactory samples. In practice, however, researchers are recommended to test the robustness of their findings on different networks, obtained through different sampling techniques. In addition, from our experience, it is better if researchers can obtain a giant component of the global network that includes all connections of each individual in the network. As researchers can gain access to higher computing power, it is possible to analyze giant components consisting of tens of thousands of consumers in empirical social networks models.

17.4.2 Censored and Truncated Social Network Data

As argued by Granovetter (1973) in his seminal work, the strength of relationships in social networks varies depending on the duration of the relationship, emotional intensity, and the type of information that is exchanged between people. Empirical research in marketing also finds consistently that the strength of relationships varies in social networks depending on the type of relationship (i.e., friend, roommate, colleague, acquaintance) (Brown and Reingen 1987; De Bruyn and Lilien 2008; Reingen 1984), and that this is moderated by the type of information exchanged (Schulze et al. 2014). However, in practice, the strength of a relationship in a social network is censored information and researchers only observe a binary network, indicating whether relationships are present or absent. The dichotomous specification of social networks may lead to biased inference of social effects and suboptimal suggestions for marketing strategies (Chen et al. 2017). We term this challenge as a censoring problem, because the dichotomous networks are a censored version of actual weighted counterparts. Another related problem is the so-called truncation of social network data. The truncation comes from the fact that social networks evolve over time as new relationships are created and old ones are severed. However, in current state-of-the-art network models in marketing, most studies assume a static (and dichotomous) network. If social networks are dynamic in nature and when there are lagged effects of social structure on the behavioral outcome, ignoring dynamics may result in biased conclusions.

To deal with censored and truncated network data, researchers could either try to collect additional sources of information or extend their model to recover the weights of relationships and time-varying connections. One possible way to incorporate the strength of relationships in the analysis of social networks is to use additional information as proxy for tie strength. For instance, Goldenberg et al. (2001) worked on a mobile network, where call frequency between people were observed and used as proxy for tie strength. Ansari et al. (2011) used music downloads between independent artists as the information source to infer tie strength. Another way to incorporate the strength of relationships in social networks is to infer or identify it with a formal model. Trusov et al. (2010) devised a Bayesian shrinkage approach to distinguish strong vs. weak ties. Their approach uses egocentric networks of individuals and infers the strength of each relationship by focusing on repeated decisions of individuals (in their application they used login decisions on a social network) and how these decisions are related to those of their connections. To reduce the computational burden of the large number of connections, they apply a shrinkage approach that allows pooling of information across connections. The model by Chen et al. (2017) as described in Sect. 17.3.1, infers the importance of relationship characteristics by estimating the weights of different social networks. They showed that modeling relationship characteristics significantly improves forecasting accuracy and allows for better seeding decisions in viral marketing campaigns. An alternative approach to recover tie strength is to model the underlying latent space of the social network (Hoff et al. 2002). This approach uses the observation that individuals with similar characteristics are more likely to connect in a social network. Using this idea, it is possible to project the social network based on the underlying latent space derived from individual characteristics. Ansari et al. (2011) extended this approach to multiplex and weighted networks, where they were able to successfully predict the incidence of relationships and also intensity of individual connections. As argued by Braun and Bonfrer (2011), this approach could also potentially be used to describe evolving social networks. The latent space approach predicts the probability for each pair of individuals to be connected. Hence, individuals that are not connected but have a high probability to be connected are potential candidates for future connections, which could be used to model an evolving network. However, researchers are just beginning to understand these issues and developing appropriate models for censoring and truncation is an important direction for future research.

17.4.3 Endogeneity in Models of Social Networks

One prominent feature of social networks is homophily, which means similar individuals tend to connect to each other and form relatively dense subgroups (also referred to as assortativity). In practice, however, not all characteristics are observed by researchers. Moreover, some unobserved characteristics could also influence consumers' behaviors on top of shaping network structure. Thus, the correlations between consumers' behaviors may be at least in part attributable to common, but unobserved characteristics (Manski 1993). More specifically, consumers who are peers may be similar in terms of a set of unobserved characteristics that also affect behavioral outcomes, generating correlation or spurious causality of peer influence. Van den Bulte and Lilien (2001) nicely illustrated the consequences of ignoring unobserved variables in the analysis of drug adoption by physicians. Their research shows evidence for social influence in the adoption decision of physicians if marketing activities of pharmaceutical companies are ignored. However, there is no evidence for social influence if marketing activities, which are common to all physicians in the social network, are controlled for. Ruling out the effects of unobserved characteristics and, thus, controlling for potential endogeneity problems are important, but challenging. Ignoring these effects may lead to incorrect conclusions and inappropriate policy and marketing strategy decisions.

The most accurate approach to address potential endogeneity problems is to collect experimental data in which confounding factors are randomized across experimental conditions. The emergence of online social networks increasingly allows researchers to do large scale field experiments. For instance, Aral and Walker (2014) developed a large scale controlled field experiment to identify social influence of Facebook users by randomly manipulating the forwarding behavior of adopters in a viral marketing campaign. Bapna and Umyarov (2015) designed a large scale field experiment on the online Last.fm social network to study the effect of peer influence on the decision to subscribe as premium user of the music website.

By randomly upgrading 1,000 free users in the network to premium users, they observed that peer influence caused an increase of over 60% in the odds to upgrade to premium user. On Sina Weibo, a Chinese social network similar to Twitter, Gong et al. (2015) studied the effects of a media company's tweets and re-tweets of influential network members on the demand for television programs. By experimentally manipulating tweets, they found that re-tweets are especially important to bring in new customers to the company. As an alternative to large scale field experiments, researchers may also try to identify external shocks to the environment that causes a natural experimental setting (Chen et al. 2011). For instance, Tucker (2008) used the variation in the value of watching TV across time and regions as natural experimental conditions to identify network externalities in the adoption decision of a video messaging technology adopted by employees in a social network.

If obtaining experimental data is not possible, there are several modeling approaches that aim to correct for potential endogeneity biases. The first approach involves the construction of valid instrumental variables for social interaction. One possibility is to explore the structure of the social network, using the characteristics of indirect friends or friends of friends as instruments. The underlying assumption in this approach is that the formation of a link between two consumers depends only on the characteristics of these two consumers (Bramoullé et al. 2009). This assumption could be reasonable in some situations. However, the typical scale-free or "the-rich-get-richer" nature of many real-world social networks seems to contradict this assumption (Barabási and Albert 1999). That is, when people make friends, they pay attention to the status or connectivity of their potential friends. Hence, to justify applicability of this approach, a researcher needs to establish that the formation of the social network is as assumed. Otherwise it is important to find other variables that are exogenous to the network structure, but related to the behavioral outcome (see Tucker 2008 for an example).

A second modeling approach to account for unobserved characteristics is to build a model with assumptions about the formation of the network. This approach is also valuable to address the censoring and truncation problem, as described above. This approach is related to the instrumental variable approach, as it creates social influence measures Z that are related to the underlying network, but not to the unobserved characteristics. It, thus, creates a cleaned version of social influence, not attributable to unobserved characteristics ξ that affect both the social network as well as the dependent variable Y. As seen from Eq. (17.12) below, Layer 2 of the network process model serves to exhume unobserved characteristics ξ from the network structure and feed it to the data model via Eq. (17.11) for bias correction. In this way, the unobserved characteristics are formally controlled. The insight for this approach comes from Goldsmith-Pinkham and Imbens (2013). Specifically, they observed that homophily causes pairs of friends to be more similar than pairs of non-friends in terms of their observed and unobserved traits. Thus, if one observes two friends that are dissimilar along the observed traits, then they are more likely to be similar along unobserved traits. On the other hand, if two non-friends are similar along observed traits, then they are more likely to be dissimilar along unobserved traits. Using this observation, it is possible to uncover unobserved characteristics from the network structure and observed characteristics.

$$\text{Data Model: } Y = f(X_1, Z, \xi | \beta) \tag{17.11}$$

$$\text{Network Process Model: } \begin{cases} \text{Layer 1: } Z = g(A|\theta) \\ \text{Layer 2: } A = h(X_2, \xi | \vartheta) \end{cases} \tag{17.12}$$

A network formation model can be incorporated into Layer 2 of the network process model, where the formation process may depend on observed individual characteristics X_2, possible unobserved individual characteristics ξ as in Goldsmith-Pinkham and Imbens (2013), and a parameter vector ϑ. The network formation model can follow two basic approaches. The first naïve approach assumes that consumers are not strategic in network formation (i.e. they do not foresee the influence of their befriending behavior on possible future behavioral outcomes). The other, more sophisticated approach, allows consumers to be strategic. The naïve approach uses random graph models to model the formation of the social network. One of most discussed models in recent years is the exponential random graph model (ERGM). These network formation models assume that the observed network is drawn from a distribution of random graphs. By allowing the network to be random, researchers introduce a procedure similar to robustness checks. For more detailed discussions of ERGMs, interested readers are referred to Chatterjee and Diaconis (2013), Goodreau (2007) and Robins et al. (2007). In contrast to the naïve approach, the strategic approach tries to incorporate microeconomic principles in the formation of the social network. Several recent papers are exploring this approach to understand the economic formation of networks (Jackson and Watts 2002; Mele 2013). However, the integration of such models to understand behavioral outcomes into the data model (Eq. 17.1) has received limited attention (Badev 2013). The computational demands of integrating network formation processes into the model are a major challenge.

17.5 Conclusion

The emergence of massive social network datasets has provided valuable new insights into marketing theories of social interactions as well as better normative models for marketing strategy development. From a methodological perspective, the incorporation of social network data into marketing decision models has created new challenges for academics as well as practitioners. In this chapter, we provided a framework for the incorporation of social network data into decision models and discuss some of the major challenges in this area of research. We summarized four approaches to integrate social networks into marketing decision models: (1) the network measures approach, (2) the statistical approach, (3) the economics approach, and (4) the agent-based approach. The choice of the approach depends on the goals of the researcher and availability of data. In general, the network measures approach is the easiest to implement and allows the comparison of the effect of several network measures, such as degree and eigenvector centrality. This approach is useful to predict individual-level decisions, such as adoption decisions or sharing in viral marketing. However, this approach requires a comprehensive understanding of the underlying social influence pattern to select appropriate network measures and detailed panel data of the effects of social influence on consumer decisions (i.e., adoptions over time, who shares information with whom). If researchers require more flexibility in the underlying structure of social influence, statistical models are recommended to integrate social networks into marketing decision models. The economics approach is recommended if researchers do not directly observe the underlying processes of how social influence drives consumer behavior (for instance, information sharing needs to be inferred from actual consumer decisions, such as adoptions), but understanding these underlying processes is imperative for policy or strategy evaluation. A disadvantage of the economics approach is that it is difficult to apply these models to large datasets, as they often involve multiple equilibria. In these circumstances, agent-based models may be valid alternatives, as these are especially powerful in complex settings where closed-form solutions are impossible to derive.

In addition to selecting one of the four approaches to integrate social networks into marketing decision models, researchers need to determine the sensitivity of their results to: (1) sampling of social network data, (2) censored and truncated social network data, and (3) endogeneity. While these are important challenges of social network analysis, we are sure that many of these issues will be resolved in the years ahead. We hope that this chapter provides inspiration for both academics and practitioners to incorporate social networks into their analysis and to further our knowledge into this exciting research area.

References

Ansari, Asim, Oded Koenigsberg, and Florian Stahl. 2011. Modeling multiple relationships in social networks. *Journal of Marketing Research* 48 (4): 713–728.

Aral, Sinan, and Dylan Walker. 2014. Tie strength, embeddedness, and social influence: A large-scale networked experiment. *Management Science* 60 (6): 1352–1370.

Badev, Anton. (2013). Discrete games in endogenous networks: Theory and policy. PSC Working paper series, PSC 13–05.

Bapna, Ravi, and Akhmed Umyarov. 2015. A randomized field experiment on peer influence in online social networks. *Management Science* 61 (8): 1902–1920.

Barabási, Albert-László, and Réka, Albert. (1999), Emergence of scaling in random networks. *Science*, 286(5439): 509–12.

Baum, Joel A.C., Robin Cowan, and Nicolas Jonard. 2010. Network-independent partner selection and the evolution of innovation networks. *Management Science* 56 (11): 2094–2110.

Bonacich, Phillip. 1987. Power and Centrality: A Family of Measures. *American Journal of Sociology* 92 (5): 1170–1182.

Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin. 2009. Identification of peer effects through social networks. *Journal of econometrics* 150 (1): 41–55.

Braun, Michael, and André Bonfrer. 2011. Scalable inference of customer similarities from interactions data using Dirichlet processes. *Marketing Science* 30 (3): 513–531.

Brock, William A., and Steven N. Durlauf. 2001. Discrete choice with social interactions. *The Review of Economic Studies* 68 (2): 235–260.

Brown, Jacqueline Johnson, and Peter H. Reingen. 1987. Social ties and word-of-mouth referral behavior. *Journal of Consumer Research* 14 (3): 350–362.

Burt, Ronald S. 1987. Social contagion and innovation: cohesion versus structural equivalence. *American Journal of Sociology* 92 (6): 1287–1335.

Chatterjee, Sourav, and Persi Diaconis. 2013. Estimating and understanding exponential random graph models. *The Annals of Statistics* 41 (5): 2428–2461.

Chen, Xi, Ralf, van der Lans, and Michael Trusov. 2016. Scaling social influence games to massive networks: implications for customer base analysis. Working paper.

Chen, Xi, Ralf van der Lans, and Tuan Q. Phan. 2017. Uncovering the importance of relationship characteristics in social networks: Implications for seeding strategies. *Journal of Marketing Research*, 54 (2): 187–201.

Chen, Xinlei, Yuxin Chen, and Ping Xiao. 2013. The impact of sampling and network topology on the estimation of social intercorrelations. *Journal of Marketing Research* 50 (1): 95–110.

Chen, Yubo, Qi Wang, and Jinhong Xie. 2011. Online social interactions: A natural experiment on word of mouth versus observational learning. *Journal of Marketing Research* 48 (2): 238–254.

Choi, Jeonghye, Sam K. Hui, and David R. Bell. 2010. Spatiotemporal analysis of imitation behavior across new buyers at an online grocery retailer. *Journal of Marketing Research* 47 (1): 75–89.

Coleman, James S., Elihu Katz, and Herbert Menzel. 1966. *Medical innovation: A diffusion study*. Indianapolis, Indiana US: The Bobbs-Merrill company.

Cressie, N.A. 1991. *Statistics for spatial data*. New York, NY (EUA): Wiley.

De Bruyn, Arnaud, and Gary L. Lilien. (2008). A multi-stage model of word-of-mouth influence through viral marketing. *International Journal of Research in Marketing* 25(3): 151–63.

Dover, Yaniv, Jacob Goldenberg, and Daniel Shapira. 2012. Network traces on penetration: Uncovering degree distribution from adoption data. *Marketing Science* 31 (4): 689–712.

Ebbes, Peter, Zan Huang, and Arvind Rangaswamy. (2013). Subgraph sampling methods for social networks: The good, the bad, and the ugly. HEC Paris research paper no. MKG-2014-1027.

Facebook Reports. July 2015. Facebook Reports Second Quarter 2015 Results. http://investor.fb.com/releasedetail.cfm?ReleaseID=924562.

Galeotti, Andrea, Sanjeev Goyal, Matthew O. Jackson, Fernando Vega-Redondo, and Leeat Yariv. 2010. Network games. *The Review of Economic Studies* 77 (1): 218–244.

Garber, Tal, Jacob Goldenberg, Barak Libai, and Eitan Muller. 2004. From density to destiny: Using spatial dimension of sales data for early prediction of new product success. *Marketing Science* 23 (3): 419–428.

Garcia, Rosanna, Paul Rummel, and John Hauser. 2007. Validating agent-based marketing models through conjoint analysis. *Journal of Business Research* 60 (8): 848–857.

Gelper, Sarah, Ralf, van der Lans, and Gerrit van Bruggen. (2015). Competition for attention in online social networks: Implications for viral marketing. Working paper.

Goldenberg, Jacob, Sangman Han, Donald Lehmann, and Jae Hong. 2009. The role of hubs in the adoption processes. *Journal of Marketing* 73 (2): 1–13.

Goldenberg, Jacob, Barak Libai, and Eitan Muller. 2002. Riding the saddle: How cross-market communications can create a major slump in sales. *Journal of Marketing* 66 (2): 1–16.

Goldenberg, Jacob, Barak Libai, and Eitan Muller. 2001. Talk of the network: A complex systems look at the underlying process of word-of-mouth. *Marketing Letters* 12 (3): 211–223.

Goldenberg, Jacob, Barak Libai, Eitan Muller, and Stefan Stremersch. 2010. Database submission-the evolving social network of marketing scholars. *Marketing Science* 29 (3): 561–567.

Goldsmith-Pinkham, Paul, and Guido W. Imbens. 2013. Social networks and the identification of peer effects. *Journal of Business & Economic Statistics* 31 (3): 253–264.

Gong, Shiyang, Juanjuan, Zhang, and Ping Zhao. (2015). Tweeting increases product demand. Working paper.

Goodreau, Steven M. 2007. Advances in exponential random graph (p*) models applied to a large social network. *Social Networks* 29 (2): 231–248.

Granovetter, M. S. 1973. The strength of weak ties1. *American Journal of Sociology* 78 (6): 1360–1380.

Hartmann, Wesley R. 2010. Demand estimation with social interactions and the implications for targeted marketing. *Marketing Science* 29 (4): 585–601.

Hinz, Oliver, Bernd Skiera, Christian Barrot, and Jan U. Becker. 2011. Seeding strategies for viral marketing: An empirical comparison. *Journal of Marketing* 75 (6): 55–71.

Hoff, Peter D., Adrian E. Raftery, and Mark S. Handcock. 2002. Latent space approaches to social network analysis. *Journal of the American Statistical Association* 97 (460): 1090–1098.

Iacobucci, D., and Hopkins, N. 1992. Modeling dyadic interactions and networks in marketing. *Journal of Marketing Research* 5–17.

Iyengar, Raghuram, Christophe Van den Bulte, and Jae Young Lee. 2015. Social contagion in new product trial and repeat. *Marketing Science* 34 (3): 408–429.

Iyengar, Raghuram, Christophe Van den Bulte, and Thomas W. Valente. 2011. Opinion leadership and social contagion in new product diffusion. *Marketing Science* 30 (2): 195–212.

Jackson, Matthew O. 2014. Networks in the understanding of economic behaviors. *The Journal of Economic Perspectives* 28 (4): 3–22.

Jackson, Matthew O., and Alison Watts. 2002. The evolution of social and economic networks. *Journal of Economic Theory* 106 (2): 265–295.

Katona, Zsolt, Peter Pal Zubcsek, and Miklos Sarvary. 2011. Network effects and personal influences: The diffusion of an online social network. *Journal of Marketing Research* 48 (3): 425–443.

Lin, Mingfeng, Nagpurnanand R. Prabhala, and Siva Viswanathan. 2013. Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. *Management Science* 59 (1): 17–35.

Lin, Mingfeng, and Siva Viswanathan. 2016. Home bias in online investments: An empirical study of an online crowdfunding market. *Management Science* 62 (5): 1393–1414.

Lucas, Robert E. (1976), Econometric policy evaluation: A critique. In *Carnegie-Rochester conference series on public policy*, vol. 1. Elsevier.

Mallapragada, Girish, Rajdeep Grewal, and Gary Lilien. 2012. User-generated open source products: Founder's social capital and time to product release. *Marketing Science* 31 (3): 474–492.

Manski, Charles F. 1993. Identification of endogenous social effects: The reflection problem. *The review of economic studies* 60 (3): 531–542.

Mele, Angelo (2013). A structural model of segregation in social networks. Available at SSRN 2294957.

Moon, Sangkil, and Gary J. Russell. 2008. Predicting product purchase from inferred customer similarity: An autologistic model approach. *Management Science* 54 (1): 71–82.

Nitzan, Irit, and Barak Libai. 2011. Social effects on customer retention. *Journal of Marketing* 75 (6): 24–38.

Oestreicher-Singer, Gal, Barak Libai, Liron Sivan, Eyal Carmi, and Ohad Yassin. 2013. The network value of products. *Journal of Marketing* 77 (3): 1–14.

Ordanini, Andrea Michael, Trusov P.K. Kannan, Yogesh Joshi, and Lei Wang. (2015). Co-investment network dynamics in crowdfunded ventures. Working paper. http://ssrn.com/abstract=2657674.

Rand, William, and Roland T. Rust. 2011. Agent-based modeling in marketing: Guidelines for rigor. *International Journal of Research in Marketing* 28 (3): 181–193.

Ransbotham, Sam, Gerald C. Kane, and Nicholas H. Lurie. 2012. Network characteristics and the value of collaborative user-generated content. *Marketing Science* 31 (3): 387–405.

Reingen, Peter H. 1984. Brand congruence in interpersonal relations: a social network analysis. *Journal of Consumer Research* 11 (3): 771–783.

Robins, Garry, Pip Pattison, Yuval Kalish, and Dean Lusher. 2007. An introduction to exponential random graph (p*) models for social networks. *Social Networks* 29 (2): 173–191.

Schulze, Christian, Lisa Schöler, and Bernd Skiera. 2014. Not all fun and games: Viral marketing for utilitarian products. *Journal of Marketing* 78 (1): 1–19.

Stephen, Andrew T., and Olivier Toubia. 2010. Deriving value from social commerce networks. *Journal of Marketing Research* 47 (2): 215–228.

Stephen, Andrew T., Peter Pal Zubcsek, and Jacob Goldenberg. 2016. Lower connectivity is better: The effects of network structure on redundancy of ideas and customer innovativeness in interdependent ideation tasks. *Journal of Marketing Research* 53 (2): 263–279.

Trusov, Michael, Anand V. Bodapati, and Randolph E. Bucklin. 2010. Determining influential users in internet social networks. *Journal of Marketing Research* 47 (4): 643–658.

Trusov, Michael, William Rand, and Yogesh V. Joshi. 2013. Improving prelaunch diffusion forecasts: Using synthetic networks as simulated priors. *Journal of Marketing Research* 50 (6): 675–690.

Tucker, Catherine. 2008. Identifying formal and informal influence in technology adoption with network externalities. *Management Science* 54 (12): 2024–2038.

Uzzi, Brian. 1996. The sources and consequences of embeddedness for the economic performance of organizations: The network effect. *American Sociological Review* 61 (4): 674–698.

Van den Bulte, Christophe, and Gary L. Lilien. (2001). Medical innovation revisited: Social contagion versus marketing effort1. *American Journal of Sociology*, 106 (5): 1409–35.

Van der Lans, Ralf, Gerrit Van Bruggen, Jehoshua Eliashberg, and Berend Wierenga (2010), A viral branching model for predicting the spread of electronic word of mouth. *Marketing Science* 29(2): 348–65.

Wang, Jing, Anocha Aribarg, and Yves F. Atchadé. 2013. Modeling choice interdependence in a social network. *Marketing Science* 32 (6): 977–997.

Wasserman, Stanley, and Katherine Faust. 1994. *Social network analysis: methods and applications (structural analysis in the social sciences) New York*. NY, US: Cambridge University Press.

Watts, D.J., and S.H. Strogatz. 1998. Collective dynamics of'small-world'networks. *Nature* 393 (6684): 440–442.

Watts, Duncan J., and Peter Sheridan Dodds. 2007. Influentials, networks, and public opinion formation. *Journal of consumer research* 34 (4): 441–458.

Yang, Sha, and Greg M. Allenby. 2003. Modeling interdependent consumer preferences. *Journal of Marketing Research* 40 (3): 282–294.

Yoganarasimhan, Hema. 2012. Impact of social network structure on content propagation: A study using YouTube data. *Quantitative Marketing and Economics* 10 (1): 111–150.

Chapter 18
Morphing Theory and Applications

Gui B. Liberali, John R. Hauser and Glen L. Urban

18.1 The Morphing Concept

As electronic commerce often matches or exceeds traditional bricks-and-mortar commerce, firms seek to optimize their online marketing efforts. When feasible, these firms customize marketing efforts to the needs and desires of individual consumers, thereby increasing click-through-rates (CTR) and conversion (sales). When done well, such customization enhances consumer relationships and builds trust.

A/B testing is a popular means to optimize marketing efforts. The firm compares two or more communications vehicles, say two banner advertisement or two website implementations. For example, potential consumers (website visitors) are randomly assigned to two banners—one might emphasize general brand image and one might emphasize the comparative advantage of a product's features. The firm measures response in the form of CTRs or conversion to identify the better banner. The better banner is then used in day-to-day website operations. A/B testing can be used with multiple marketing instruments or with aspects of marketing instruments

G.B. Liberali (✉)
Rotterdam School of Management, Erasmus University, 50 Burgemeester Oudlaan,
3062 PA Rotterdam, The Netherlands
e-mail: liberali@rsm.nl

J.R. Hauser
MIT Sloan School of Management, Massachusetts Institute of Technology,
E62-538, 77 Massachusetts Avenue, Cambridge, MA 02142, USA
e-mail: hauser@mit.edu

G.L. Urban
MIT Sloan School of Management, Massachusetts Institute of Technology,
E62-536, 77 Massachusetts Avenue, Cambridge, MA 02142, USA
e-mail: glurban@mit.edu

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3_18

that are mixed and matched in an experimental design. A/B testing has proven effective and has increased the profitability of many marketing instruments.

Morphing improves A/B testing in many ways. First, morphing uses optimal adaptive experimentation. For example, as the morphing system begins to observe consumer response it allocates sample to A versus B to learn efficiently. Morphing trades off learning about consumer response (learn) with using that knowledge to display the best banner for the consumers (earn). The learn-while-earning process allocates same to different banners to maximize long-term profits. For example, if a morphing system learns that a particular banner is unlikely to be the best banner, it ceases to assign consumers to see that banner. If a morphing system learns that a particular banner is especially promising it automatically and optimally allocates more consumers to that banner.

Second, morphing automatically identifies the latent segment to which each consumer belongs. Morphing detects a consumer's segment from the clicks that the consumer makes on the firm's website (or from tracking the consumer prior to visiting the firm's website). For example, a consumer with a more-verbal cognitive style might click more often on text-based descriptions than on pictures, whereas a consumer with a more-visual cognitive style might click more often on pictures. Alternatively, a consumer who is beginning his or her search for automobiles might click on comparison charts while a consumer who is ready to buy might click on dealer-location or special-deal pull-down menus.

Third, morphing matches marketing instruments to each consumer' segment, and does so optimally. Because morphing identifies latent segments automatically, morphing can use optimal experimentation for each segment to learn the best marketing instrument for that segment. For example, if the consumer has a verbal cognitive style, then the look and feel of the website can "morph" to feature more verbal content. If the consumer is in the buying stage for an automobile, then the website can help the consumer find dealers or cars with specific features. It might even offer an incentive for a test drive.

Fourth, because morphing identifies the best marketing instrument for each segment from those that are tried, it provides rich information for further development and design of those instruments. Indeed, in our experience, this organizational learning has proven to be critical to enhanced outcomes for the firm.

18.1.1 Morphing Overview

In this chapter we review almost 10 years of morphing experience. To date, most of the contributions have been proof-of-concept research projects, but, increasingly, firms are beginning to adopt and test morphing capabilities. We begin with a brief overview on the steps in a prototypical morphing application.

Morphing, as first proposed by Hauser, Urban, Liberali, and Braun (HULB 2009), consists of the following steps:

1. Clicks on a website are monitored and, from those clicks, algorithms automatically infer the likelihood of the consumer segment to which the consumer belongs.
   a. Websites, banners, or other marketing materials may be designed so that they appeal (potentially) to different segments of consumers.
   b. Consumers in a calibration study visit example websites and provide data by which to identify their segments.
   c. The calibration data provides a model of how browsing behavior differs by segment.
2. Marketing materials, such as banners, are provided to consumers to maximize goals such as profit, sales, or click-through rates.
   a. The system learns as is goes.
   b. Learning is automatic and near optimal.
   c. The goal of learning is to match the marketing materials to the consumer's segment to maximize the firm's goals.

Ideally, morphing targets and learns from each and every consumer and does so in real time. However, some systems now "batch" learning in the sense that the rules in the second step are updated periodically rather than for every consumer (e.g., Bertsimas and Mersereau 2007; Schwartz et al. 2016). Recently, morphing has been extended to automatically determine when is the best time to morph the marketing materials (e.g., Hauser et al. 2014).

18.1.2 Morphing Example

Figure 18.1 illustrates the general concept of morphing with a stylized example from banner advertising. This concept is not limited to banner advertising; morphing applies to a wide range of marketing materials. For example, HULB morph the look and feel of the website.

The set of numbers in the upper left of Fig. 18.1 are the firm's best guess at the segment to which the consumer belongs. These estimates are based on the consumer's clickstream up to this point. For example, the firm might believe that this consumer's cognitive style is most likely to be verbal-impulsive (65%), but there is a lesser chance that the cognitive style might be visual-impulsive (15%), verbal-deliberative (10%), or visual-deliberative (10%). The morphing algorithm uses Bayes Theorem to estimate these probabilities from the clicks that the consumer has already made on the website.

We provide details on the method later; we provide here the intuition. Suppose that, in an earlier calibration study, we measured consumer's cognitive styles using traditional methods. For example, we might ask the consumer to answer a bank of questions, the answers to which indicate the consumer's cognitive style. Suppose

Fig. 18.1 The morphing concept—an example with four styles and four morphs (stylized illustration)

further that we observed that consumers with verbal-impulsive cognitive styles clicked on action-oriented textboxes, but consumers with other cognitive styles clicked on other portions of the website. Then if we observe that consumer clicks on many textboxes and prefers short action-oriented descriptions rather than longer fact-based descriptions, that consumer may more likely to be verbal-impulsive than one of the other cognitive styles. The actual probability then is proportional to the percent of consumers with verbal-impulsive cognitive styles (known from the calibration study) times the likelihood that a person who clicks on action-oriented verbal textboxes is verbal-impulsive (also known from the priming study). Because a website is likely to offer a large number of click choices, we describe each click by its characteristics to reduce dimensionality. The end result, which is updated when the consumer provides more clicks, are the percentages in the upper left corner of Fig. 18.1.

Before we describe the learn-while-earning aspect of morphing, it is easier if we consider a more traditional situation in which we have observed a large number of consumers of each cognitive style for each of the potential banners. For this situation, the table in the lower right side of Fig. 18.1 contains outcome probabilities for each banner-segment combination. For example, if a consumer has a cognitive style that is "verbal-impulsive" and that consumer is given the "Buy! versus Learn More" banner, the likelihood that consumer would click on the banner is 0.10. If that consumer were given instead the "Emotional" banner, then the likelihood of a click would increase to 0.20. These probabilities are based on prior consumers, with those cognitive styles, who have been shown each of the banners. If this were an A/B test, the banners would have been randomly assigned until we had sufficient precision on the outcome-probability estimates.

If the firm had perfect information about the segment to which the consumer belonged and if it knew the outcome probability perfectly, the firm would select the banner with the highest outcome probability by looking up the highest outcome probability in the row corresponding to the consumer's segment. For example, if the firm knew the consumer was verbal-impulsive and it knew the outcome probabilities, it would provide the consumer with the "Emotional" banner because it has the highest outcome probability in the verbal-impulsive row.

However, firms do not know the consumer's segment with certainty. Instead, based on the consumer's clicks, firms have estimates of the likelihood that the consumer belongs to each of the four cognitive-style segments. If the system had completed its learning, the best banner would be the banner that maximizes the firm's immediate goals such as CTR. In Fig. 18.1, the best banner for the consumer is the Emotional (in column 3) because it has the highest expected reward given our estimate of the consumer's cognitive style. We obtain that estimate by multiplying the probabilities the consumer belongs to each cognitive-style segment times the probabilities that a consumer in that segment clicks through when shown a given banner. For example, the likelihood that the consumer clicks through when given the Emotional banner is 0.17 obtained as $0.17 = 0.65 \times 0.20 + 0.15 \times 0.13 + 0.10 \times 0.10 + 0.10 \times 0.13$.

Emotional is the best banner to provide to the consumer *after* the system has completed its learning, but Emotional may or may not be the best banner to provide to the consumer while the system is still learning the outcome probabilities. For example, it might be best if the system were to occasionally try other banners so that it can learn probabilities for other banners. Furthermore, the system might be able to make the best decision even if it does not know the outcome probabilities with certainty. For example, it might be able to eliminate banners with extremely low outcome probabilities and not waste sample consumers on those banners. This is a dilemma. If the firm tries other banners it sacrifices its goals for the immediate consumer, but the system learns what is best for the next consumer and all subsequent consumers with the same cognitive style. The morphing system automatically assigns banners to consumers with (near) optimal experimentation by balancing the cost of experimentation with the value of learning about the outcome probabilities.

The mathematics used to balance earning and learning are sophisticated, but the implementation is relatively straightforward using a mathematical concept called "Gittins' indices (GIs)." At any given time there is a GI for each cognitive-style-banner combination. Each GI represents both the value the firm can gain from that consumer (the current best estimate of the outcome probability) and the option value to the firm for learning more about outcome probabilities. We provide more details later. Gittins (1979) proved that there is a rule based on GIs that provide optimal experimentation when consumers can be assigned to consumer segments without error. The rule is simple, provide the morph with the highest GI. Because the rule is optimal, a firm using GIs can expect higher profits than it would earn with naïve A/B testing.

However, as Fig. 18.1 indicates, we do not know the consumer's segment with certainty. But we have probabilities, based on the consumer's clicks, that the consumer belongs to a segment—e.g., a 65% chance that the consumer's segment is verbal-impulsive. Krishnamurthy and Mickova (1999) demonstrated that if one were to use GIs rather than the best estimates of outcome probabilities, then the multiple-latent-segment experimentation would be near optimal. This is exactly what is done in morphing. We replace the outcome probabilities with GIs and compute the expected value—the expected Gittins' index (EGI). We assign to the consumer the banner with the highest EGI.

Morphing has an additional advantage over standard A/B testing. Because the system is always in learn-while-earning mode, the GIs automatically and near optimally pick up any changes in the underlying outcome probabilities. For example, if consumers' tastes change and a banner is no longer as effective, the GIs will begin to shift automatically to take into account the option value of learning more about that banner. Similarly, if a new banner is added to the mix, the GIs for that banner start at prior beliefs, but the morphing system quickly and optimally learns the true outcome probability.

Morphing uses information from prior consumers to learn the updated value of the outcome probabilities for each banner-segment combination. This value is updated after each user is exposed to a banner—after we observe whether or not the consumer clicks-through or makes a purchase (conversion). The value of each banner-segment combination GI is based on our current estimate of the choice probability for that banner-segment combination *plus* an option value that reflects the value of learning more about that banner-segment combination. As the system learns from many observations, the option value is decreases. For the banner that is best for a consumer segment, the value of the banner-segment combination converges to the predicted purchase value. For other banner-segment combinations, the system might cease to allocate sample because it is not profitable to do so, even considering the option value of learning.

The rate at which convergence is achieved is an empirical question, and is based on the expected traffic of the website. For example, it is optimal for websites with millions of visitors per day to learn at a different rate than websites with only a few thousand visitors per day. In practice, the degree to which each individual observation changes beliefs about the best morph depends on how many observations we expect to observe during the relevant time period.

Morphing is not limited to banners—it can be used to match consumers to website designs, call-center scripts, or any marketing instrument. For the remainder of this chapter, when it is clear in context, we use the general term "morph" instead of banner to reflect the generality of potential applications.

Morphing is based on continuously improving initial knowledge about each consumer's segment, using that information to assign the best morph for a consumer, and learning from the outcome of the morph assignment. As shown in the next section, morphing has been applied in a variety of situations, including applications to improve performance for firms which had previously run randomized controlled experiments (A/B tests).

18.2 Optimal Online Experimentation

18.2.1 From Learn-then-Earn to Learn-While-Earning: The Multi-armed Bandit Problem

Randomized controlled experiments are the cornerstone of causal inference. Firms run hundreds, or even thousands, of online randomized experiments every day, in what is often referred to as A/B testing. Typically, an A/B test is based on random assignment of treatments to website visitors and is continued until sufficient statistical power is achieved such that reliable conclusions can be made regarding the effect of a specific website configuration on sales or other variables of interest. For example, one firm may run an online A/B test to learn whether it is more effective to present information about its product in a 2- or 3-column format. The dependent variable is typically CTR or online purchase (conversion).

Because of randomization and statistical power, traditional A/B tests tend to follow the learn-then-earn paradigm. During the testing phase the focus is on *learn*, i.e., estimating the effect of each treatment on consumer behavior. Once the estimates are obtained, the focus changes to outcome maximization (‘*earn*’), when the firm deploys the winning treatment on large scale.

The traditional *learn-then-earn* paradigm of A/B testing has two major weaknesses. First, it is based on the responses of average consumers; it ignores heterogeneity in consumer preferences. It does not take into account that different consumers may respond differently to the marketing instrument(s) when the firm deploys the winning treatment (marketing instrument) to all consumers. When consumers are not all the same, ignoring individual differences can be costly. For example, in Fig. 18.1 it is better that the verbal-impulsive consumer get an Emotional banner while the visual-deliberative consumer get an Informative banner.

Morphing addresses this issue using consumer segment probabilities (for latent consumer segments) to handle heterogeneity when computing the optimal treatment for each consumer. Morphing enables each consumer (or each segment) to get the best morph based on the latest information about the behavior of the consumer’s segment. “Best” takes into account both learning and earning.

Second, the learning phase in traditional A/B tests is inefficient, leading to wasted resources because it invests the same amount of resources on good and bad treatments. Typically, A/B tests assign the same sample size to each cell during the learning phase, which means that the precision of the estimates of good treatments is the same as the precision of the estimates of the bad treatments. However, as the firm learns quickly that a treatment is suboptimal, it wastes resources when it assigns more consumers to a suboptimal morph in order to make its estimate of the outcome probability more precise for that treatment.

Morphing invests sample size in those cells that most clarify which marketing instruments to give to which consumers. Because traditional A/B testing continues to invest sample in learning about suboptimal treatments, the firm loses revenue every time a treatment is assigned to a cell that the firm already knows has a low probability of leading to a good outcome (click or conversion). In morphing, once the firm is confident that a marketing instrument is best for a consumer segment, it optimally assigns that marketing instrument to almost all subsequent consumers in the segment.

Solving this learn-while-earning problem is not easy. Obtaining better estimates about the effect of marketing instruments (*learn*) is costly in the short-term, but leads to higher revenue on the long term. On the other hand, using current estimates to assign marketing instruments to consumers (*earn*) avoids the short-term cost of learning, but suffers from higher opportunity costs. The firm misses future sales because it does not learn which marketing instrument is really best for each consumer segment. For example, if there is no exploration, then, if current estimates suggest that a 3-column design does not have the highest conversion rate for a specific segment, the 3-column design will never be shown again to any consumers in that segment. This loss of future potential can loom large, particularly if some "shock" changes outcome probabilities.

The learn-while-earning problem is at the heart of morphing. This problem is in the class of "multi-armed bandit" problems. When segments are known, website morphing methods provide an optimal solution to this problem in real-time (HULB). When segments must be inferred, the solution is not provably optimal, but is extremely close to optimal. Morphing dynamically—and near optimally—allocates larger sample to the best treatment-segment combination based on the solution of the learn-versus-earning formulation originally developed by Gittins (1979).

18.2.2 From Learning About Designs to Learning About Consumers

Adapting a website to each consumer involves a fundamental change in the philosophy of A/B testing. Typically, A/B testing assigns banners or website variations to consumers on a random basis. As a result, an A/B test identifies the marketing instrument that is best on average, not the marketing instrument that is tailored to each consumer. In some cases, the marketing instrument might be best for no one. For example, suppose that consumers are either Type X or Type Y and suppose there are three morphs, A, B, and C. Suppose that the outcome probabilities for Type X consumers are 0.9, 0.5, and 0.0 for A, B, and C, respectively. Suppose they are 0.0, 0.5, and 0.9 for Type Y consumers. On average, the best morph is B, with average outcome probability 0.5. However, if we could assign A to Type X consumers and B to Type Y consumers, we would achieve an improved outcome probability of 0.9. Customization matters.

Morphing changes the A/B testing logic fundamentally. Instead of testing marketing instruments that apply to all consumers, morphing learns and selects the best marketing instrument for each consumer. Instead of randomly assigning marketing instruments to a test or control treatment, morphing optimally assigns consumers to marketing instruments. As more and more consumers are run through a morphing system, the algorithm identifies the best allocation of morphs to consumers to maximize the outcome variable (such as conversion).

Changing A/B test focus from A versus B marketing instruments to a focus on consumers is a major shift for most firms doing A/B testing. The change in focus has two practical implications. First, most firms and A/B software assign incoming consumers randomly to test or control cells. The software is not designed to learn about consumers and then assign consumers to different marketing instruments based on information about that consumer and the accumulated experience from other consumers.

Second, morphing requires tracking consumer-level information. Most large firms today use software packages that act as a layer isolating managers from the raw data. Reports are produced automatically with summary statistics showing which marketing instrument is the best on average, and at which *p*-value. Obtaining reports based on individual consumers (instead of marketing instruments) requires access to and analysis of raw data, something that is often a formidable task for most firms.

Morphing requires firms to change radically the way they design and run their A/B tests, and the way they use information about website visitors (consumers). In our illustration in Fig. 18.1, we defined consumer segments by cognitive styles. This is illustrative only. We can define consumer segments by the stage in the buying process, interest in the category, cultural styles, cognitive styles, source (whether the consumer is coming from an online search or a referral), personas, devices (tablet, desktop/laptop or mobile), purchase tendencies, or any other variable that can be observed in a calibration study.

18.2.3 Handling Consumer Differences: The Case for Cognitive Styles

Although consumer segments can be defined in a variety of ways, one of the most frequent ways to segment website consumers is based on the way they interact with websites and other morphs. The way consumers respond to websites is heavily related to how they gather, process, and evaluate information—their *cognitive styles* (Hayes and Allinson 1998). A cognitive style reflects "individual differences in how we perceive, think, solve problems, learn and relate to others (Witkin et al. 1977, p. 15)." Examples of dimensions of cognitive styles include impulsive-deliberative, visual-verbal, and analytical-holistic (for more examples, please refer to the online appendix in HULB).

If measured well, a consumer's cognitive style is stable over time, so there are no history-dependent interactions (Markovian structure) which would make it difficult for the morphing algorithm to converge to true outcome probabilities. Decades of research in psychology suggest that people develop cognitive styles over the years, and that their preferences for cognitive styles change slowly.

Cognitive styles are easily interpretable and actionable from a managerial point of view. Designers can easily relate to cognitive styles, such as verbal-visual, to develop website designs, banners, or other marketing instruments that are likely to be suited well for one style rather than others. The more a morph is tailored to a specific style, the more likely a consumer using that cognitive style will relate to and feel comfortable with the way the website, banner, or other marketing instrument communicates with the consumer. Strong prior beliefs help morphing converge faster, but, even if the designers' prior beliefs are wrong, morphing learns optimally the best morph-to-segment assignments based on consumer reaction. The GIs converge automatically to the best assignments *based on consumer response* even if the initial designers guess incorrectly.

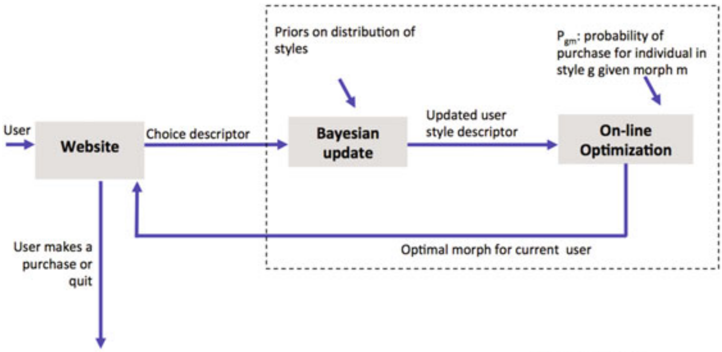
18.3 Learning Loops: Consumer Segments and Morph × Segment Assignments

For ease of exposition we illustrate morphing with an application based on cognitive styles.

The morphing process has two learning loops. In the first learning loop, the morphing system observes clicks from each consumer and, after sufficiently many clicks, updates its estimates of the probability that consumer belongs to each segment. In earlier versions (Morphing 1.0), the number of clicks was set exogenously. Later versions (Morphing 2.0) choose the number of clicks endogenously and near optimally for every consumer.

In the second learning loop, the system learns the outcome probabilities *across consumers* in a segment as those consumers are exposed to morphs (treatments) and respond with successes (such as a click through or a purchase conversion) or failures (a non-click-through or a non-purchase). Figure 18.2 illustrates the process. For ease of exposition we reduced the number of cognitive styles from four in Fig. 18.1 to two in Fig. 18.2. The basic concepts apply to as many cognitive styles as can be defined, but when more cognitive styles are used, more data (consumers) are needed for the morphing system to work well.

The process starts when a consumer comes to the website. At this point (typically, on a landing page), the consumer may be exposed to a morph-independent stimulus. We observe the first click (or set of clicks). A click (or set of clicks) can be thought of as a choice among various links, all of which have cognitive cues. When the consumer chooses one of the links, we gain data about the consumer's cognitive style. For example, one consumer may decide to go to the virtual advisor area of a website by clicking on a verbal description instead of clicking on an image. Using this information about the consumer's click choices, the learning-about-the-consumer loop updates prior beliefs about the consumers'



Illustrative example

	Banner Ads ->	1	2	3	4
Cognitive Style	Impulsive	0.14	0.1	0.11	0.13
	Deliberative	0.1	0.11	0.2	0.13

Fig. 18.2 Morphing and learning (The Gittins Index numbers (GI) in the table are for illustration only)

cognitive style using the Bayes Theorem. The resulting posterior beliefs become the updated probabilities that the consumer has either an impulsive or a deliberative cognitive style.

Next, the morphing system uses the information about the consumer's cognitive style to look-up the GIs in an optimality table, as in the lower right of Fig. 18.2. Recall that the GIs indicate the value of each morph (highest earn-vs.-learn value) for each segment. The GIs are larger than the expected outcome probabilities because they include the option value of learning. As we did for Fig. 18.1, we use the probabilities that the consumer belongs to each segment to compute an expected Gittins index. We choose the morph with the highest expected Gittins index, breaking ties randomly. The consumer is exposed to the morph with the highest Gittins index. For example, suppose that the learning-about-the-consumer loop predicts that there is an 80% probability that the consumer is deliberative. Then the expected Gittins' index, EGI, for the third banner advertisement would be 0.00352 = (0.20)(0.11) + (0.80)(0.20). This is higher than the EGI for Banner 1 (0.00224), Banner 2 (0.00176), or Banner 4 (0.002704).

The consumer continues to browse until he or she either clicks-through (if CTR is the outcome measure) or buys (if conversion is the outcome measure) or leaves the website without clicking-through or buying. Based on the observation of the consumer's response to the third banner advertisement, the optimality table is updated accordingly as described in the next section. (Because there was still some uncertainty in identifying the consumer's segment, the response probability for both the impulsive and deliberative cognitive styles would be updated, albeit the deliberative style more so than the impulsive style. There would be no updates for the first, second, and fourth banners.)

Recall that the GIs are not outcome probabilities. If we were to assign morphs to consumers based on outcome probabilities, we would assign the morph with the largest expected outcome. Such a rule would never assign any other morphs and we would never improve our knowledge about outcome probabilities associated with those morphs. This non-assignment problem is known as the *curse of serendipity*, or lack of exploration.

The informal intuition behind Gittins' solution to this dilemma is that the GIs summarize the value of earning-versus-learning as an optimality index. In particular, a GI equals our estimate of the outcome probability augmented by the value of exploration. A GI is computed for every cognitive-style-morph combination and allows the system to explore and serve morphs to learn how to assign the best morph for each consumer. As morphs are served and outcomes observed, we 'spend' the exploration value. When all exploration value is spent, only the true outcome probability remains in the updated table. Section 18.4 presents a more formal description of this approach. An appendix provides the analytical expressions used to implement the expected Gittins index solution.

18.3.1 Steps in a Morphing Project

Prior to implementing the morphing algorithm in day-to-day operations, parameters must be estimated in a calibration study. See Fig. 18.3. It is also feasible to use data from ongoing day-to-day operations to updated the parameters further, however, to date, parameterization has always been based on a calibration study.

In a calibration study, a firm recruits a sample of consumers to answer questions before and after visiting the website(s). The firm observes respondents' clickstreams on the website(s). Additional questions enable the firm to assign each consumer to the consumer's cognitive style (or other segmentation scheme). Because we want variation in clickstreams, morphs are assigned randomly to respondents in the

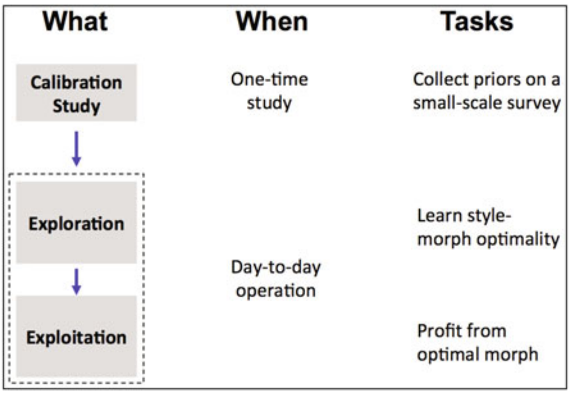


Fig. 18.3 Morphing phases

calibration study. The sample size for the calibration study is usually much smaller than would be typical for a randomized A/B experiment. One advantage of automatic learning is that, after the initial calibration study, new morphs can be inserted into the morphing system. The morphing system will automatically, and near optimally, allocate consumers to those morphs to learn consumer response to the new morphs.

Calibration-study questions enable the morphing-system developer to estimate a model, conditioned on segment membership, which predicts consumers' click-alternative choices as a function of the characteristics of the click alternatives. Once the model is calibrated, morphing uses standard Bayesian methods to estimate day-to-day consumers' latent segment probabilities. Day-to-day operations use the Bayesian model and the GIs to control the amount and speed of learning.

18.4 The Analytics of Morphing

The Morphing 1.0 algorithm was first published in HULB. The Bayesian updating effects a rapid assessment of consumers to segments. A dynamic program produces the GIs to select the best morph for a segment, An expectation over the GIs identifies the best morph. The Morphing 2.0 algorithm was published in Hauser et al. (2014). In this section we provide the basic morphing concepts. An appendix provides the formal notation and equations.

18.4.1 Learning the Consumer's Segment from the Consumer's Clicks

The Bayesian model in morphing was motivated by a Bayesian advisor that identified which vehicle (car or truck) to recommend to a consumer (Urban and Hauser 2004). The recommendation was based on that consumer's answers to a series of questions about the potential consumer's use of the vehicle. In our case, we use the consumer's clicks on the website to identify the likelihood that the consumer belongs to a particular consumer segment rather than to identify the best recommendation.

Let n index consumers, r index segments, m index morphs, and t index clicks for each consumer. At any point in the consumer's visit to the website, the consumer has a choice of which click to make next. We characterize the probability of any click for a consumer with a particular cognitive style. To do this, we describe each click by a set of characteristics. For example, we might observe basic dimensions such as graphical vs. verbal, functional characteristics such as "use an analytic tool" or "read a post," or website areas such as "virtual advisor" or "learning center." HULB use eleven website characteristics for a broadband-sales website.

In the calibration study, morphing analysis estimates a choice model that includes weights for each characteristic. Because we know the characteristics of all possible clicks every time a consumer makes a click, we compute the "utility" of a given click as a function of the to-be-estimated weights for the characteristics. A logit model assumes that the consumer maximizes his or her "utility" as given by the click characteristics and an error term. We use the calibration data and standard maximum-likelihood or Bayesian methods to estimate the weights.

After we estimate the weights, we use these weights in day-to-day website operation to compute the probability, for each cognitive style, that the consumer will choose a given click. In stylized symbols, we compute: $Prob\{click \mid cognitivestyle\}$. The likelihood of a particular clickstream is just the product of these probabilities multiplied for all the clicks made by the consumer. Equation 18.1 in the appendix provides details of the logit model likelihood. Note that, although we observe the clicks, we still need to compute the probability, as conditioned on cognitive style, for the next step in morphing.

From the calibration study, or from the history of consumers visiting the website, we form prior beliefs about the segment to which the consumer belongs. For example, we might believe that 25% of all consumers are verbal-impulsive. Call this probability, $Prob\{cognitivestyle\}$. We want to compute the probability of a cognitive style based on the observed click stream. We do this with Bayes Theorem recognizing that

$$Prob\{cognitivestyle|clickstream\} \propto \\ Prob\{clickstream|cognitivestyle\} * Prob\{cognitivestyle\}$$

where $\propto$ means proportional. To compute the actual probability we normalize the expression so that $Prob\{cognitivestyle|clickstream\}$ adds up to 1.0 when summed over cognitive styles. Equation 18.2 in the appendix provides details. Fortunately, Eq. 18.2 involves relatively fast calculations so that cognitive styles can be determined almost instantaneously between clicks on a website

18.4.2 Learning How to Assign Morphs to Segments Optimally

We represent our knowledge about outcome probabilities by a function that we can update quickly. In particular, we choose a "beta distribution." The beta distribution has two parameters that depend upon the consumer's segment, r, and the morph, m. These parameters are α_{rm} and β_{rm}. If p_{rm} is the probability that a consumer in segment r, who was shown morph m, clicks through (or converts), then, for the beta distribution, the mean outcome probability is $E[p_{rm}] = \alpha_{rm}/(\alpha_{rm} + \beta_{rm})$. Larger values of the parameters mean less uncertainty in our beliefs about the outcome probabilities.

The beta distribution allows fast updating. For example, if the cognitive style where known, then α_{rm} increases by 1.0 for every success (click-through or conversion) and β_{rm} increases by 1.0 every time the consumer leaves without a "success," e.g., without a click through or a conversion.

Assigning the optimal morph to the *n*th consumer is more complicated than simply maximizing the immediate reward. The GI includes the option value. It does more than simply maximize $E[p_{rm}]$. Because each outcome improves our knowledge about p_{rm}, the updated distribution enables us to make better decisions in the future. The dynamic decision problem balances immediate rewards with the knowledge gained that enables better decisions in the future.

This dynamic problem for this type of multi-armed bandit was first formulated in the 1940s and, for many years, considered to have no simple solution. However, in the late 1970s, John Gittins had a seminal insight that he could compare the decision problem for each "arm of a bandit problem" to an equivalent fixed outcome. He could then compare the equivalent fixed outcomes from many arms and choose the outcome that was best. The value of the fixed outcome became known as a Gittins' index. The concept was generalized to many problems. Today, if a problem can be solved with indices, it is said to be indexable. When cognitive styles are known, the morphing problem is indexable.

The basic dynamic program is formulated as a recursion known as a Bellman equation. The "state" of the problem is the current values of α_{rm} and β_{rm}, as well as a discount factor, a. The discount factor indicates how much the morphing algorithm should discount the future. For example, if a website has 100,000 visitors spread equally throughout the year, then HULB suggest that $a = 0.999999$.

The recursion recognizes that, for any given rm combination, the best strategy is to choose the larger of the fixed outcome or to keep experimenting. Gittins' proved that once the fixed outcome is chosen, the best strategy is to continue choosing the fixed outcome. (This makes sense, α_{rm} and β_{rm} are not changing if there is no experimentation. When there is no experimentation on an arm, choice among alternatives does not change and neither does the solution.) If the algorithm chooses to try that morph for that cognitive style, then we get to observe an outcome—either a success or a failure. But we know the likelihood of a success, $\bar{p}_{rm} = E[p_{rm}]$.

With probability, $\bar{p}_{rm}$, we observe a success and reap our reward, which we set to 1.0. With probability $1 - \bar{p}_{rm}$, we observe a failure and reap no reward. In each case we get to continue playing the game with the updated α_{rm} and β_{rm}. Equation 18.3 in the appendix provides the details. We provide here the recursion in words. Let V indicate the value of continuing to play with a given set of parameters. Then, the recursion is:

$$
\begin{aligned}
&V(current\alpha_{rmn}, \beta_{rmn}, a) \\
&\quad = \max\left\{ \begin{array}{c} \frac{GI_{rmn}}{1-a}, \bar{p}_{rmn}(1 + V(updatedbasedonsuccess) + \\ (1 - \bar{p}_{rmn})V(updatedbasedonfailure) \end{array} \right\}
\end{aligned}
$$

The first term reflects the value of continuing to $t = \infty$ with a discount factor of a. In this equation we added the subscript n to indicate that these values change after each consumer. To use the recursion, we solve the equation with an iterative search for all of the α's and β's we expect in practice. We table the GI's so that the GI's can be assessed quickly.

When a consumer's segment is not known with certainty, the dynamic program becomes a partially-observable Markov decision process (POMDP). In general POMDPs are difficult to solve, but this particularly POMDP has a near-optimal solution that runs in real time between clicks (Krishnamurthy and Mickova 1999). Specifically, the revised algorithm replaces the Gittins' index with the expected Gittins' index, EGI, as illustrated in Sect. 18.3. Equation 18.4 in the appendix provides the details.

Finally, we update beliefs about the parameters of the beta distribution. The challenge is that we do not know the consumer's latent segment with certainty. We only know the probabilities that the consumer belongs to each of the consumer segments. The true Bayesian updating formulae are no longer easy, but we can use a trick. When the segments are latent, we can update if we consider "fractional observations" using an analogy to the standard likelihood function. (Fractional updates represent a pseudo-Bayesian updating that provides estimates that work extremely well for morphing. See formal analyses and simulations in Hauser et al. 2014.)

If we observe a success, conditioned on the consumer having seen morph m, we consider this as a fractional success for each latent consumer segment, r. The fractional of the success is probability that the consumer was in consumer segment r. The binomial distribution is well defined for fractional observations and naturally conjugate to the beta distribution, so we use the same formulae, except with fractional observations. Equation 18.5 in the appendix provides the details. Updating occurs when the consumer leaves the website. The fractional-observation formulae enable the morphing algorithm to run in real time between a consumer's clicks on the website.

18.4.3 When Do We Know Enough to Find the Optimal Morph?

In Morphing 1.0, the algorithm morphed after a fixed number of clicks by the consumer on the website. For example, in HULB's application to a BT Group website that sold broadband service, a morph was considered after the 10th click. In Urban et al. (2014)'s application to banner advertising on CNET, the banners were morphed after the 5th click. In both cases, the time to morph was set by the researchers' judgment based on simulated performance. We can do better by choosing the click on which to morph.

In choosing the time to morph, we address the tradeoff between exposure and precision. We gain greater exposure of the best morph to the consumer by presenting the optimal morph as early as possible in the consumer's website visit. Doing so exposes the consumer to the best morph for the longest amount of time possible. We gain greater precision by identifying the best morph as late as possible in the consumer's website visit, because doing so uses better consumer-segment estimates to find the best morph. To address this trade-off, Morphing 2.0 uses a second recursion.

The generalized morphing algorithm, published in Hauser et al. (2014), solves an embedded dynamic program that enables the optimal trade-off between exposure and precision for each consumer. To formulate the dynamic program, the authors had to first address three issues in consumer response to morphing.

First, to evaluate the impact of every morph that the consumer sees, the algorithm must explicitly track how long a consumer is exposed to each morph, and decide how to attribute credit to each morph seen. Second, the algorithm must allow the system to change morphs as often as necessary because, as more information on the cognitive style becomes available from clicks, beliefs about the true cognitive style become closer to the true cognitive style. This introduces memory into the algorithm because the algorithm must keep track of how many clicks were made for each morph exposure for the consumers. This challenge recognizes that the optimal morph after many clicks may be different from what was thought to be the optimal morph when the algorithm had less information about the consumer cognitive style (fewer clicks). Third, if the algorithm allows multiple morph changes and allows the system to decide when to morph, consumers might experience cognitive load from seeing multiple morph. This cognitive load induces potential cognitive switching costs that must be modeled. By modeling switching costs, the algorithm only changes morphs if the gains from changing morphs are greater than the cognitive costs of switching morphs.

18.4.3.1 Every Morph Seen Matters: The Attribution Problem

If we allow a consumer to be exposed to more than one morph during a website visit(s), we need to attribute credit regarding the observation (a success or a failure) to each morph seen. For example, assume a consumer saw Morph A during the first five clicks, then saw Morph B during the last ten clicks, and then made a purchase. Which morph should get the lion's share of the credit for this success? Should it be the first morph because "first impression lasts?" Or, should it be the second morph because it was seen for longer, or perhaps because of recency effects?

We address this attribution problem by specifying attribution weights, w_t's, for each time period, t, when computing value of a morph for consumer n. The weights, w_t's, are measured, judged, or estimated empirically in each application, and used as parameters of the model. Because w_t is applied to each morph seen at every time t, it spreads the credit through all morphs seen. To keep the number of w_t's small, we allow t to index observation periods that may be one click or more than one click. We normalize the impact weights so that they sum to 1.0 over clicks (or observation periods).

18.4.3.2 Changing Morphs: Switching Costs

It is reasonable to expect that consumers may experience cognitive load if the website design changes too often or too dramatically. The costs of switching tasks have been extensively studied in psychology starting with Jersild (1927) and Spector and Biederman (1976), and more recently in Meiran (2000). In marketing, switching costs are well-established (e.g., Weiss and Anderson 1992; Jones et al. 2000, 2002). Researchers have also studied how consumers react to switching costs when browsing websites (Balabanis et al. 2006, and Johnson et al. 2003).

Additive switching cost are common in the multi-armed bandit literature and algorithms exist (e.g., Banks and Sundaram 1994; Dushochet and Hongler 2006; Jun 2004), but additive switching costs require that we keep track of the timing of all switches for a consumer. This path dependence makes it more difficult to solve the optimization problem. Because of this difficulty, additive switching costs make algorithms infeasible for real-time morphing.

On the other hand, a multiplicative switching cost can be factored out in a recursive equation that optimizes the time to morph. Multiplicative switching costs are more intuitive because their effect is proportional to the likelihood of purchase. Not only do multiplicative switch costs assure that all probabilities remain defined between zero and one, but we expect that the amount by which a low probability is lowered by a switching cost would be less than the amount by which a high probability is lowered by a switching cost. For example, suppose a switch lowers p_{rmn} from 0.800 to 0.700. Comparable proportional cost would lower p_{rmn} from 0.090 to 0.070 while a comparable additive cost would lower p_{rmn} from 0.090 to less than 0.000. To date, we have not tested the multiplicative assumption, but it seems to be a more-reasonable representation of switching costs than an additive assumption.

For both practical and theoretical reasons, we solve a problem with multiplicative switching costs and do so in real time. Specifically, we assume that a switch in a morphs lowers the consumer's purchase probability. The switch lowers the purchase probability by a (switching) factor of γ where $\gamma \leq 1$. In theory, γ can be determined by experimental means in a priming study. However, to date, γ has been set by managerial judgment. Hauser et al. (2014) explore the sensitivity of γ between 0.80 and 1.00.

18.4.3.3 Putting It Together

The switching factor (γ) and a period-weight (w_t) are tuning parameters that must be selected *before* the algorithm is used to morph a website (in day-to-day

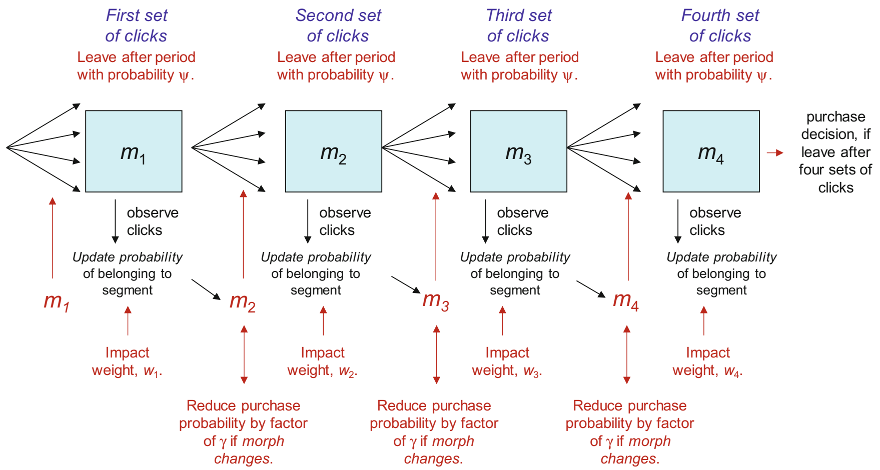


Fig. 18.4 The when-to-morph problem (modified from Hauser et al. 2014)

operations). The tuning parameters require either managerial judgment or experiments during the calibration study. In a calibration study, segment membership is measured directly, therefore the true consumer segment is known among calibration respondents. To estimate tuning parameters, the calibration study would also assign switches randomly at different time periods. With sufficiently many observations in the calibration study, γ and the w_t's can be identified.

Figure 18.4 illustrates the conceptual decision problem for the case where the consumer makes a purchase (or leaves the website) after four observation periods. Specifically, during the first observation period, the website displays a morph. The respondent makes clicks while exploring the website and we update our beliefs about the consumer's segment, Using the new information, and anticipating more information from subsequent decision periods, we decide which morph to display in the second decision period if the consumer stays on the website. (The consumer may decide to leave after a decision period. For example, the consumer might leave after the tth observation period with probability, ψ_t.) If the morph in the second period is different from the morph in the first period, the consumer incurs a multiplicative switch cost, γ.

This process continues until the consumer reaches the fourth period at which time the consumer either purchases or leaves without purchasing. Figure 18.4 illustrates the process as if the consumer makes a decision after the fourth period. However, in practice, the consumer can make a decision at any period and/or continue beyond the fourth period. (The recursive equation in the appendix allows random exits.)

In general, the when-to-morph decision problem is coupled with the learn-while-earning decision problem where the learn-while-earning algorithm experiments with different morphs for each segment. For example, if we show morph m to a consumer in the consumer segment r for more data periods, we learn more about the response probability for that segment-morph combination (p_{rm}). Fortunately, the dynamics of the two decision problems happen on two very different scales. The "which-morph-for-which-segment" learn-while-earning decision problem is solved from observations based on success over failure over thousands of consumers. On the other hand, the when-to-morph decision problem is solved between clicks for one consumer at a time.

Because of these differing dynamics we decouple the two problems. In particular, we use the Gittins' indices (GIs) to represent the value of showing morph m to a consumer in segment r and we use the concept of the expected Gittins' index (EGI) to decide the best morph. The GIs are updated *between* consumers and are held constant when the when-to-morph decision problem that is solved between clicks by the current consumer.

Putting this altogether we obtain a recursive relationship that must be solved between clicks for the current consumers. Unfortunately, this recursive relationship does not appear to be indexable. The conceptual recursive equation is the following, where t indicates the observation period. To keep this recursion simple, we have not specified random exit. Equation 18.6 in the appendix provides greater details.

$$V_t\left(m_t^*, m_{t-1}, clicks\right) = \max_{\mathrm{m_t}}\left\{w_t EGI + \sum_s Prob\{segment = s\} V_{t+1}\left(m_{t+1}^*, m_t, clicks|segment = s\right)\right\}$$

The equations that we used in Morphing 1.0 assumed that only the last morph seen by the consumer affects the probability of a successful outcome for that consumer. When we solve the when-to-morph dynamic program, we must generalize the fractional observation updating procedure. In particular, we keep track of which morph was shown to the consumer in each observation period. The fractional observation is now the probability (based on all observed clicks for that consumer) modified by the w_t's. For example, if the consumer saw morph m_1 for the first period and morph m_2 for the second, third, and fourth periods, then, if the w_t's are normalized to 1.0, the fraction assigned to morph m_1 for segment r is the (terminal) probability that the consumer is in segment r times w_1. The fraction assigned to morph m_2 for segment r is the (terminal) probability that the consumer is in segment r times $w_2 + w_3 + w_4$. Equation 18.7 in the appendix provides details.

18.4.4 On-Going Extensions and Other Methods

Shortly after the HULB was published, morphing was extended to handle longitudinal interventions. The extended algorithm was tested in a field experiment matching AT&T banner advertising with cognitive styles identified from clicks consumers made on CNET.com. See Urban et al. (2014). Other methods have been published addressing the application of multi-armed bandit ideas to morphing. Table 18.1 summarizes a few of these applications.

Table 18.1 Examples of multi-armed bandit algorithms for online experimentation

	Focus (what is it learning about)	General or industry-specific	Considerations on optimality
Hauser et al. (2009, 2014), Urban et al. (2014)	Consumer	General	Optimal for indexable problems (know consumer segments). Near-optimal for partially observable consumer segments
Scott (2010), Schwartz et al. (2016)	Creative	General	Designed to run in batches. Asymptotically optimal. Arms pulled proportionally to posterior probability of being optimal
Bertsimas and Mersereau (2007)	Creative	General	Designed to run in batches. Lagrangian decomposition and asymptotic approximations
Chung et al. (2009, 2015)	Consumer	Industry-specific	Promotes explorative search with a rejuvenation heuristic step

In general, we see that some methods, such as Thompson Sampling, focus on aggregate or batched data and on non-consumer-specific marketing instruments. Interestingly, Schwartz et al. (2016) test alternative multi-armed bandit solutions in counterfactual synthetic-data experiments and suggest that Gittins-based strategies often outperform Thompson-sampling-based strategies even in batched applications. Other heuristics also do well. The website morphing papers and Chung et al. (2009) learn at the level of the individual consumer to enable the system to match marketing instruments to consumers efficiently.

18.5 Applications of Morphing: Evidence from the Field

The first application of morphing, as reported in HULB, was a research collaboration with the BT Group, formerly British Telecom (BT). In this project, the data indicated that, had morphing been implemented system-wide, the lift in BT's online sales of broadband plans would have increased by 20%—about $80 million in additional revenue. These data were analyzed further in Hauser et al. (2014). Their counterfactual synthetic-data experiments suggested that the improved Morphing 2.0 methods would have outperformed the original Morphing 1.0 methods by 69%. The gains reflect the ability of Morphing 2.0 to handle switching costs, attribution, and the optimal time to morph. The Morphing 2.0 algorithm was applied to a website selling card loans in Japan. An initial study of 1,395 consumers provided data for counterfactual experiments that predicted a 63% improvement—substantially more than the Morphing 1.0 algorithm (Hauser et al. 2014).

Urban et al. (2014) applied morphing to banner advertising. In a field experiment on CNET.com, banner morphing achieved a 97% lift in click-through rates for context-matched banners relative to a no-morph control group. CNET.com is a high-traffic website that hosts display banner advertising; it was not feasible in the field experiment to track online sales (conversion). To examine the impact on conversion as well as click-through rates, Urban et al. ran a longitudinal field study with General Motors' Chevrolet Division. The field experiment documented that matching banners to the stage of the consumer's buying process, body-type preference, and cognitive style significantly increased click-through rates, brand consideration, and purchase likelihood relative to a control.

We are aware of several morphing applications that are now being developed. For example, one application has begun a proof-of-concept test using traffic from a major telecom provider in The Netherlands. The calibration study has been completed and the cognitive-style Bayesian loop has also been coded. This application includes four cognitive styles, three morphs, and several funnel-stage outcomes. It includes controls so that morphing can be evaluated.

A second morphing application is being developed in collaboration with an online marketplace in The Netherlands. This application morphs the automotive section of the online market place, and uses two consumer-knowledge segments instead of cognitive styles. The online marketplace is expanding its assignment mechanism to allocate consumers to test and control using the morphing algorithm. A third application is starting at a disruptive financial-products-comparison portal with operations in various countries in Asia. While none of these applications have yet gone live, they indicate the feasibility of developing morphing websites and morphing banners across a wide variety of applications.

18.6 Design and Implementation Decisions in a Morphing Project

Morphing methods substantially increase click-through and conversion rates because they fundamentally change the way website design, banner advertisements, and other marketing instruments are tested. Conversion managers and the IT teams involved in a morphing project benefit from the managerial and technical implications of such changes. This section provides an overview of key changes based on our experience with morphing projects in a variety of firms.

18.6.1 From Managing Aggregate Data to Handling Consumer-Level Data

Perhaps the most unexpected practical challenge companies face when considering morphing is the unprecedented need to handle data that is tagged to individual consumers. This operates on multiple levels.

- Morphing requires that the firm track and update its estimate of the probability that each consumer belongs to a segment. These updates might be based on data from the firm's website, but advanced applications base these updates on consumers' activities on many channels such as clicks on the website, posts on social media, or call-center input.
- Morphing requires that firms track, at least temporarily, which consumers are exposed to which morphs. Fortunately, the system needs only to maintain parameter updates and indices, not the entire morph-to-consumer history, but many websites must be modified to maintain even this level of information.
- Morphing requires that consumers be assigned to A/B cells dynamically. It is no longer sufficient to assign consumers randomly to A/B cells. Rather morphing bases these (near-optimal) assignments based on balancing immediate profit and long-term learning.
- Some firms may wish to test morphing itself versus a control such as random assignment or a fixed-morph control. In this case, random assignment of consumers to treatments occurs at a higher conceptual level. Consumers are assigned to strategies (morphing vs. a control) rather than marketing instruments. To the best of our knowledge, no off-the-shelf software has the capability of assigning consumers to strategies.

These four challenges require a major technological shift, because standard randomized-assignment code needs to be updated, reporting systems need to adapt, and firms need to rethink their policies on A/B testing. Morphing experiments identify which designs are best for which consumers, but, often, morphing also provides the organization with a new way to think about website, banner, and marketing-instrument design. Morphing provides a new way to manage click-through rates, conversion, and other funnel measures.

18.6.2 Consumer Segments, Marketing Instruments, and Outcome Variables

Website morphing integrates three foundational elements of e-commerce: demand (e.g., consumer segments), supply (e.g., marketing instruments used by a firm), and online transactions (e.g., conversion, a request for information, or a click-through). This section provides an overview of what needs to be done for each element.

Consumer segments. While cognitive styles remain one of the best segmentation variables, morphing can be applied to a variety of segmentation variables including country of residence, personas, source (referral or not), device being used (mobile, computer, etc.), stage in the buying process, etc. The only real requirements are that consumers do not switch segments during a session and that segments can be identified in the calibration study.

Marketing instruments. Morphing can apply to any marketing instrument that can be tested with traditional A/B testing. Marketing instruments include website designs, banner advertising, call-center scripts, product recommendations, price levels, promotional coupons, etc. Marketing instruments (morphs) can also be defined at a higher level of abstractions, such as an advertising campaign that is implemented in several different online channels. For example, one morph for a telecom firm could be a campaign focused on emotional content—the campaign might present its services as a way to keep close to family and friends. A second morph could be a campaign focused on informational content—the campaign might show how the quality of service is better than the competition. Both campaigns could run in parallel and be implemented in various media channels. Morphing would identify which segments of consumers relate best to which of the two campaigns. (Of course, in this case, consumers would need to be tracked across channels.) Based on consumers' clickstreams, elements of the best campaign can be targeted to the right consumers. The elements might even be channel dependent.

Click-through rates, conversion rates, and other funnel measures. Click-through rates versus conversion rates are not conflicting goals, but they are often distant in time in the purchase funnel (e.g., Hongshuang and Kannan 2014). Analyses must be done carefully to infer causality when the temporal distance between exposure, consideration, and purchase is too long. Purchases in many categories do not happen immediately; purchases may happen weeks, or even months, after exposure to a marketing instrument. In the interim, there are often other changes in the website and in the environment that also affect sales. Either these changes must be modeled or managers must recognize the inherent uncertainty in end-of-the-funnel measures. Project leaders need to carefully clarify what are realistic optimization goals and then choose the outcome variable (funnel measure) accordingly. The choice of the outcome variable is crucial for the mechanics of morphing (what to optimize), for what firms learn, and for how the success of morphing is evaluated.

18.6.3 A Roadmap to Implement Morphing

Each morphing project has several tasks and milestones that need to be achieved.

1. Select the segmentation criteria, e.g., cognitive styles or other variables.
2. Select the morphs, e.g., marketing instruments such as website design or banners.

3. Select the outcome variable, e.g., click-through rates, awareness, trial, or conversion.
4. Determine the webpages and links to monitor. Perhaps design the webpages so that segments are easy to identify (next-generation websites).
5. Assess a categorization of each monitored link using a panel of judges for use in the Bayesian model (the website characteristics).
6. Run a calibration study to observe consumer's clicks and assign representative consumers to segments. In the calibration study consumers are assigned through direct questions.
7. Using the data from the calibration study, estimate a model that predicts click preferences for each consumer segment.
8. Pre-compute the click likelihoods, the probability of a click given the characteristics of the click (and competing clicks) and the consumer's segment. Do this for each link and each consumer segment so that segment likelihoods can be obtained quickly with the Bayesian model.

Coding

9. Implement the consumer inference system. This is the real-time Bayesian-loop inference code running on the webserver.
10. Define the control cell, e.g., random assignment, status-quo method, fixed-morph, or best-guess?
11. Decide whether to have a single test cell or multiple test cells. In a single-test-cell design, morphing chooses the optimal morph to show. One could potentially decide to run two or more test cells in parallel, each running a different set of morphs, segments, and/or outcome variables.
12. Adapt the existing A/B system to randomly assign (and keep track of) consumers to test and control cells.
13. Implement the system that receives the morph assignments and selects the morph to serve to the current consumer.
14. Adapt the reporting system to report consumer outcomes (click-throughs, conversions, or other funnel measures) based on the selected outcome variable. The system should report at the consumer-level for each morph the consumer received. A Morphing 2.0 system may also need to record the number of observation periods (sets of clicks) for each morph seen by the consumer.
15. Implement code that delivers the best morph to a consumer based on the morphing optimization.

18.6.4 Priors and Convergence

There are two sets of priors used to initialize the morphing system. The first prior represents initial beliefs about the consumer segment, before any click is observed. This is typically selected to be either flat (equal probabilities to each consumer segment) or equal to the observed percentages of consumer segments in the calibration study. The decision depends on sample size, precision, and reliability of the estimates in the calibration study. It is relatively easy to update this prior after sufficiently many consumers have been observed in day-to-day operations. This prior is important because it affects which consumers get which morphs. Often, inferences about a consumer's segment must be made after a relatively few clicks by the consumer.

The second prior is the prior beliefs about the outcome probabilities. The Gittins' indices are calculated for the first customer based on priors that reflect the strength of our beliefs about the initial outcome probabilities for every morph-segment cell. Typically, the prior beliefs are based on observed outcomes of morph × segment probabilities in the calibration study. In some cases, the morphs may still be under development during the calibration study. In these cases, it is reasonable to start with flat priors (expected baseline click-through rate or another appropriate measure). Fortunately, in day-to-day operations, the performance of the morphing algorithm is relatively robust with respect to this prior on outcome probabilities. As consumers visit the website, the data on their clicks soon overwhelms the prior beliefs about outcome probabilities.

Typically we expect to see a pattern of transition from learning to earning that is somewhat similar to Fig. 18.5. Figure 18.5 applies to a single consumer segment in a website that receives approximately 100,000 annual visitors. It documents how the GI for each morph changes as more consumers are exposed to morphs and as more outcomes are observed. Notice that, after a few thousand visitors, all indices have converged to the true morph × segment outcome probabilities. At about the 2,500th visitor the system briefly experiments with morph 3 due to random variation but soon learns that morph 2 is the one with highest outcome probability. When the segment is not known but inferred using probabilities of the consumer belonging to a segment, convergence is not as rapid, but still occurs.

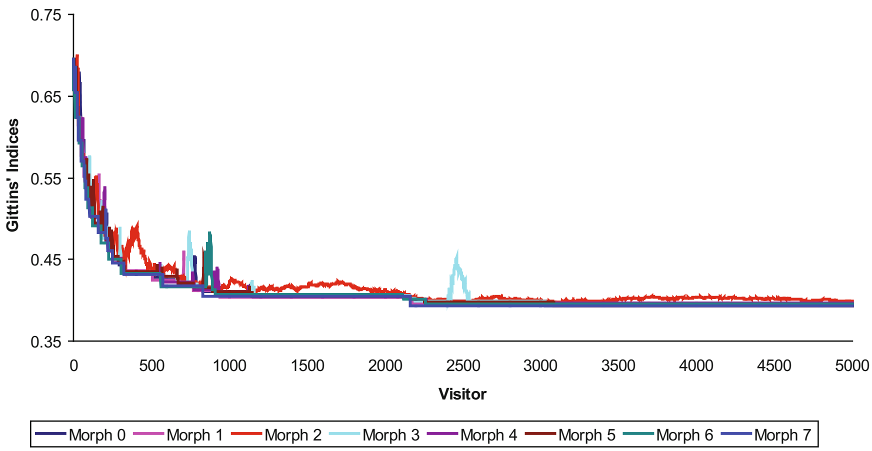


Fig. 18.5 Convergence of morphing algorithm for a cognitive style (illustrative data from HULB)

The rate at which the GIs converge is controlled by the discount factor, the volume of data, and accuracy of the cognitive-style posteriors. Applications with less concentrated cognitive-style posterior probabilities (probabilities that are closer to equally-likely) tend to converge more slowly than applications with concentrated probabilities (probabilities that are close to 0.0 or 1.0). Convergence is slower when consumer segments must be inferred because fractional updating spreads outcome observations over multiple consumer segments. The right discount factor enables the GIs to be matched appropriately to the firm's cost of capital.

18.7 Do's and Don'ts of Morphing and Organizational Impact

Although morphing makes morph assignments in real time while consumers click on websites, firms may decide to update the outcome probabilities for morph × segment combinations in batches. While this is possible, project leaders should be aware that it is only by multiple iterations of updates on that table (see Fig. 18.1) that the system will learn optimally. In extreme cases, the learning loop can be run offline, but an offline learning loop is not an efficient use of resources. An offline system learns from past data, but does not realize gains from optimal experimentation. Similarly, if the Bayesian loop is run offline, the morphing system does not have the ability to match morphs to consumer segments.

Morphing tends to cross organizational silos in large, traditional corporations. A project typically requires efforts from the engineering team (to code the systems listed in the roadmap), sales teams (morphs are marketing instruments), web designers, reporting (to develop the APIs that need to be integrated with cloud servers), and consumer experience teams. As in any project involving change, support form the highest level of the organization is crucial for success. If the firm does not have an established culture of A/B experimentation, morphing is likely to require additional steps. In such cases, our experience suggests that a pilot project can be implemented to optimize marketing-instrument A/B testing without the Bayesian loop, or to allocate marketing instruments to consumers without the second learning loop (the GIs). After A/B or consumer-level testing has been completed successfully, the full system will be easier to sell within the organization.

From a computational point of view, there are two major considerations. First, there is a need for real-time inference of consumer segments. Our formulae allow for rapid computation, but algorithms must be coded and implemented and may require re-training so that the web developers gain experience with Bayes-based algorithms. Second, the performance of the data transfer between the morphing servers (based on the cloud) and the firm's traditional webservers must be tested extensively to make sure performance is appropriate. Computations are designed to be rapid and the traffic that flows between servers is designed to be light (just a few bytes per click), but the connection between servers must have high levels of data reliability and speed.

From a purely methodological point of view, the use of Gittins' indices is based on a technical assumption—that the multi-armed bandit problem is indexable. Indexability in the canonical problem is usually expressed as a requirement that the arms of the multi-armed bandit are independent. For further details refer to Gittins et al. (2011). When the arms are not independent, the problem may still be indexable, but Gittins' indices may no longer be optimal. (When the "arms" change by external means, the multi-arm bandits are called restless bandits. Indices such as Whittle's index may need to be used—Whittle 1988.)

Morphing is flexible, but designers should be aware of a key tradeoff. The dimensionality of the optimality table grows proportionally to the number of morphs times the number of consumer segments. Successful applications balance relevance and speed. Simulations are valuable because they provide benchmarks before the firm chooses the number of consumer segments and the number of morphs.

From the viewpoint of website development, webpages can be designed so that consumer segments are easy to identify. Links and other content can be planned in a way to maximize the information obtained from each click, reducing the number of clicks needed to learn the consumer's segment. We call such websites generation 2 (Gen-2) or next generation (next-Gen) websites.

18.8 Open Questions and Relevant Challenges

Morphing theory and methods provide opportunities for further research in substantive, conceptual, and methodological areas. From a substantive point of view, there are opportunities for new applications of morphing using other marketing instruments, such as price levels, promotion types, retention policies, call centers, and product bundles. Morphing is also feasible when using different devices or combinations of devices, namely desktop computers, tablets, and smartphones.

We are not aware of projects based on morphs built across media channels, but we believe that a consumer's clicks across online channels may substantially improve consumer-segment inference. There is effort to measure media across multiple channels, but that does not include morphing. There are opportunities for morphing to coordinate actions that blend direct human action, such as calls, with automated actions, such as product recommendations. Because morphing affects organizational culture, there are opportunities to explore the feedback from morphing results to the creative process at agencies. Creative teams responsible for the development of website designs and banners obtain new creative insights by understanding which consumer segments respond best to which marketing instruments.

Morphing 2.0 provides a structure to model switching costs, attribution, and random exit. This structure opens opportunities in the measurement of switching costs, in the study of attribution, and in the modeling of exit probabilities. There are challenging issues in how to aggregate lessons learned over multiple A/B tests. Schwartz et al. (2016) provide a means to address these issues within batch-processed A/B testing, but challenges remain for allocating marketing instruments to individual consumers in non-batch modes. Advances in multi-armed bandit research provides many opportunities such as correlated-arms bandits (as in Keller and Oldale 2003) and restless bandits (using the Whittle index as in Lin et al. 2015).

See Urban et al. (2009) for more managerial issues. The online appendix of HULB provides additional insights on the development of cognitive styles and the appendices in Hauser et al. (2014) provide details on a number of technical issues, including fractional updating.

Acknowledgements The on-going experiments are supported by the Erasmus Research Institute in Management (ERIM).

Appendix: The Equations of Morphing

Let n index consumers, r index segments, m index morphs, and t index clicks for each consumer. Capital letters, R, M, and T_n denote totals. Let c_{tn} denote the tth click by nth consumer and $\vec{c}_{tn} = \{c_{1n}, c_{2n}, \ldots, c_{tn}\}$ denote the vector of clicks up to an including the tth click. At each click choice, the consumer faces J_{tn} click alternatives as denoted by c_{tnj} where j indexes click alternatives. Let $c_{tnj} = 1$ if consumer n clicks the jth click alternative on the tth click; $c_{tnj} = 0$ otherwise. Let $\vec{x}_{tnj}$ denote the characteristics for click-alternative j faced by consumer n on the tth click. Let $\vec{X}_{tn}$ be the set of $\vec{x}_{tnj}$'s up to an including the tth click for all $j = 1$ *to* J_{tn}. Let $\tilde{u}_{tnj}$ be the utility that consumer n obtains from clicking on the jth click alternative on the tth click. Let $\vec{\omega}_r$ be a vector of click-alternative-characteristic preferences for the rth consumer segment and $\tilde{\varepsilon}_{tnj}$ be an extreme value error such that $\tilde{u}_{tnj} = \vec{x}'_{tnh}\vec{\omega}_r + \tilde{\varepsilon}_{tnj}$. Let Ω be the matrix of the $\vec{\omega}_r$'s. Let $\delta_{mn} = 1$ if the nth consumer makes a purchase after seeing morph m; $\delta_{mn} = 0$ otherwise.

The likelihood that the nth respondent chooses clicks $\vec{c}_{T_n n}$ given the consumer belongs to segment r is given by:

$$\Pr\left(\vec{c}_{T_n n} | r_n = r, \Omega, \vec{X}_{tn}\right) = \Pr(\vec{c}_{T_n n} | r_n = r) = \prod_{t=1}^{T_n} \prod_{j=1}^{J_{tn}} \left(\frac{\exp\left[\vec{x}'_{tnj}\vec{\omega}_r\right]}{\sum_{\ell=1}^{J_t} \exp\left[\vec{x}'_{tn\ell}\vec{\omega}_r\right]} \right)^{c_{ntj}} \tag{18.1}$$

We estimate Ω from the calibration study by forming the likelihood over all respondents and by using standard maximum-likelihood methods or Bayesian methods. Denote these estimates by $\widehat{\Omega}$.

In Morphing 1.0, we observe the consumer's clickstream up to the τ_oth click. The unconditional prior probabilities, $\Pr_0(r_n = r)$ are observed in the calibration study or from website experience. Bayes Theorem provides:

$$q_{rn\tau_o}(\vec{c}_{\tau_o n}, \widehat{\Omega}, \vec{X}_{\tau_o n}) \equiv \Pr\left(r_n = r | \vec{c}_{\tau_o n}, \widehat{\Omega}, \vec{X}_{\tau_o n}\right) = \frac{\Pr\left\{\vec{c}_{\tau_o n} | r_n = r, \widehat{\Omega}, \vec{X}_{\tau_o n}\right) \Pr_0(r_n = r)}{\sum_{s=1}^{R} \Pr\left\{\vec{c}_{\tau_o n} | r_n = s, \widehat{\Omega}, \vec{X}_{\tau_o n}\right) \Pr_0(r_n = s)} \tag{18.2}$$

For ease of exposition, we temporarily add the r subscript to δ_{rmn} to indicate a situation in which the segment, r, is known. Let p_{rmn} be the probability that consumer n in segment r, who experienced morph m, will make a purchase (or other success criterion). This probability is distributed: $f_n(p_{rmn} | \alpha_{rmn}, \beta_{rmn}) \sim p_{rmn}^{\alpha_{rmn}-1} (1 - p_{rmn})^{\beta_{rmn}-1}$ where α_{rmn} and β_{rmn} are parameters of the beta distribution. Updating implies $\alpha_{rm,n+1} = \alpha_{rmn} + \delta_{rmn}$ and $\beta_{rm,n+1} = \beta_{rmn} + (1 - \delta_{rmn})$. Normalizing the value of a purchase to 1.0, the expected immediate reward is $E[p_{rmn} | \alpha_{rmn}, \beta_{rmn}] = \alpha_{rmn} / (\alpha_{rmn} + \beta_{rmn})$.

Let G_{rmn} be the Gittins' index for the mth morph for consumers in segment r, let $a \leq 1$ be the discount rate from one consumer to the next, and let $V_{Gittins}(\alpha_{rmn}, \beta_{rmn}, a)$ be the value of continuing with parameters a, α_{rmn}, and β_{rmn}. We table G_{rmn} by iteratively solving the Bellman equation.

$$V_{Gittins}(\alpha_{rmn}, \beta_{rmn}, a) = \max\left\{ \begin{array}{c} \frac{G_{rmn}}{1-a}, \frac{\alpha_{rmn}}{\alpha_{rmn}+\beta_{rmn}} [1 + aV_{Gittins}(\alpha_{rmn}+1, \beta_{rmn}, a] \\ + \frac{\beta_{rmn}}{\alpha_{rmn}+\beta_{rmn}} aV_{Gittins}(\alpha_{rmn}, \beta_{rmn}+1, a) \end{array} \right\} \tag{18.3}$$

When consumer segments are latent, we replace the Gittins' index with the expected Gittins' index, EGI_{mn}.

$$EGI_{mn} = \sum_{r=1}^{R} q_{rn\tau_o}(\vec{c}_{\tau_o n}, \widehat{\Omega}, \vec{X}_{\tau_o n}) G_{rmn}(\alpha_{rmn}, \beta_{rmn}, a) \tag{18.4}$$

For latent segments, the updating equations are based on "fractional observations." Details are available in Hauser et al. (2014).

$$\begin{aligned} \alpha_{rm,n+1} &= \alpha_{rmn} + q_{rnT_n}(\vec{c}_{T_n n}, \widehat{\Omega}, \vec{X}_{T_n n}) \delta_{mn} \\ \beta_{rm,n+1} &= \beta_{rmn} + q_{rnT_n}(\vec{c}_{T_n n}, \widehat{\Omega}, \vec{X}_{T_n n})(1 - \delta_{mn}) \end{aligned} \tag{18.5}$$

For the Morphing 2.0 extension, let w_t be the weight for observation period t and let γ be the multiplicative switching cost. We add a t subscript to morphs such that m_{tn} indicates the morph seen by consumer n in the tth observation period. To keep track of morph changes, we define an indicator variable such that $\Delta_{m'_{tn} tn} = 1$ if we *change* to morph m'_{tn} for consumer n in period t; $\Delta_{m'_{tn}tn} = 0$ otherwise. Because the consumer may see many morphs, we drop the m subscript from δ_{mn} such that $\delta_n = 1$ if the consumer makes a purchase; $\delta_n = 0$ otherwise.

To determine when to morph, we solve a Bellman equation by backward recursion for each consumer. The immediate reward is the γ-discounted, weighted expected Gittins' index. The expectation uses $q_{rn}(\vec{c}_{t-1,n}, \widehat{\Omega}, \vec{X}_{t-1,n})$ because this inferred probability represents our expectations over all future clicks. The segment-conditional continuation reward is $V_\tau\left(m^*_{\tau n}, m_{\tau-1,n}, \vec{c}_{\tau-1,n}, \widehat{\Omega}, \vec{X}_{\tau-1,n}\right| r_{true,n} = s\big)$. It is computed by keeping track of morph changes for $\tau \geq t$. We take the expectation with respect to the probability of observing each consumer segment to obtain the unconditional reward. Let ψ_t be the probability of exit after the tth observation period and let $\Psi(S|t-1) = E_n[\prod_{s=t}^{S}(1-\psi_s)]$, Then the Bellman equation is:

$$V_t(m^*_{tn}, m_{t-1,n}, \vec{c}_{t-1,n}, \widehat{\Omega}, \vec{X}_{t-1,n}) = \max_{m_{tn}} \left\{ \begin{array}{c} \gamma^{\Delta_{m_{tn}}} w_t \sum_r q_{rn}\left(\vec{c}_{t-1,n}, \widehat{\Omega}, \vec{X}_{t-1,n}\right) G_{rm_{tn}n} \Psi(t|t-1) + \\ \sum_s \left[q_{sn}\left(\vec{c}_{t-1,n}, \widehat{\Omega}, \vec{X}_{t-1,n}\right) V_{t+1}\left(m^*_{t+1,n}, m_{tn}, \vec{c}_{t-1,n}, \widehat{\Omega}, \vec{X}_{t-1,n}, r.e.|s\right)\right] \Psi(t+1|t) \end{array} \right\} \tag{18.6}$$

We let $\eta_{mnt} = 1$ if consumer n saw morph m during the tth observation period; $\eta_{mnt} = 0$ otherwise. Generalized fractional-observation updating becomes:

$$\alpha_{rm,n+1} = \alpha_{rmn} + q_r\left(\vec{c}_{T_n n}, \widehat{\Omega}, \vec{X}_{T_n n}\right)\gamma^{N_{T_n}}\left(\sum_{t=1}^{T_n} \eta_{mnt} w_t\right)\delta_n \tag{18.7}$$

$$\beta_{rm,n+1} = \beta_{rmn} + q_r\left(\vec{c}_{T_n n}, \widehat{\Omega}, \vec{X}_{T_n n}\right)\gamma^{N_{T_n}}\left(\sum_{t=1}^{T_n} \eta_{mnt} w_t\right)(1-\delta_n)$$

References

Bertsimas, D., and A.J. Mersereau. 2007. A learning approach for interactive marketing to a customer segment. *Operations Research* 55 (6): 1120–1135.

Chung, T.S., R.T. Rust, and M. Wedel. 2009. My mobile music: An adaptive personalization system for digital audio players. *Marketing Science* 28 (1): 52–68.

Chung, T.S., M. Wedel, and R.T. Rust. 2015. An adaptive personalization using social networks. *Journal of the Academy of Marketing Science* 42 (1): 1–22.

Dusonchet, F., and M.-O. Hongler. 2006. Priority index heuristic for multi-armed bandit problems with set-up costs and/or set-up time delays.*International Journal of Computer Integrated Manufacturing* 19 (3): 210–219.

George, Balabanis, Nina Reynolds, and Antonis Simintiras. 2006. Bases of e-store loyalty: Perceived switching barriers and satisfaction. *Journal of Business Research* 59 (2): 214–224.

Gittins, J.C. 1979. Bandit processes and dynamic allocation indices. *Journal of the Royal Statistical Society, Series B* 41 (2): 148–177, plus commentary.

Gittins, J.C., K. Glazebrook, and R. Weber. 2011. *Multi-armed bandit allocation indices*. London: Wiley.

Hayes, J., and C.W. Allinson. 1998. Cognitive style and the theory and practice of individual and collective learning in organizations. *Human Relations* 31 (7): 847–871.

Hauser, J.R., G.L. Urban, G. Liberali, and M. Braun. 2009. Website morphing. *Marketing Science* 28 (2): 202–224.

Hauser, J.R., G. Liberali, and G.L. Urban. 2014. Website morphing 2.0: Switching costs, partial exposure, random exit, and when to morph. *Management Science* 60 (6): 1594–1616.

Hongshuang, L., and P.K. Kannan. 2014. Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research* 51: 40–56.

Jersild, A.T. 1927. Mental set and shift. *Archives of Psychology* 21 (89): 1–92.

Jeffrey, S.B., and R.K. Sundaram. 1994. Switching costs and the gittins index. *Econometrica* 62 (3): 687.

Keller, G., and A. Oldale. 2003. Branching bandits: A sequential search process with correlated pay-offs. *Journal of Economic Theory* 113: 302–315.

Krishnamurthy, V., and J. Mickova. 1999. Finite dimensional algorithms for the hidden Markov model multi-armed bandit problem. *IEEE International Conference on Acoustics, Speech, and Signal Processing* 5: 2865–2868.

Lin, Song, Juanjuan Zhang, and John R. Hauser. 2015. Learning from experience, simply. *Marketing Science* 34 (1): 1–19.

Meiran, N. 2000. Modeling cognitive control in task switching. *Psychological Research* 63: 234–249.

Michael, A Jones, David L Mothersbaugh, and Sharon E Beatty. 2000. Switching barriers and repurchase intentions in services. *Journal of Retailing* 76 (2): 259–274.

Michael, A Jones, David L Mothersbaugh, and Sharon E Beatty. 2002. Why customers stay: Measuring the underlying dimensions of services switching costs and managing their differential strategic outcomes. *Journal of Business Research* 55 (6): 441–450.

Schwartz, E.M., Bradlow, E.T., and Fader, P.S. 2016. Consumer acquisition via display advertising using multi-armed bandit experiments. *Marketing Science*.

Scott, S.L. 2010. A modern Bayesian look at the multi-armed bandit. *Applied Stochastic Models Business and Industry* 26 (6): 639–658.

Spector, A., and I. Biederman. 1976. Mental set and shift revised. *American Journal of Psychology* 89: 669–679.

Tackseung, Jun. 2004. A survey on the bandit problem with switching costs. *De Economist* 152 (4): 513–541.

Urban, G.L., and J.R. Hauser. 2004. "Listening-in" to find and explore new combinations of customer needs. *Journal of Marketing* 68: 72–87.

Urban, G.L., J.R. Hauser, G. Liberali, M. Braun, and F. Sultan. 2009. Morphing the web—Building empathy, trust, and sales. *MIT Sloan Management Review* 50 (4): 53–61.

Urban, G.L., G. Liberali, E. MacDonald, R. Bordley, and J.R. Hauser. 2014. Morphing banner advertising. *Marketing Science* 33 (1): 27–46.

Weiss, A.M., and E. Anderson. 1992. Converting from independent to employee salesforces: The role of perceived switching costs. *Journal of Marketing Research.* 29 (1): 101–115.

Witkin, H.A., C. Moore, D. Goodenough, and P. Cox. 1977. Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research* 47: 1–64.

Whittle, P. 1988. Restless bandits: Activity allocation in a changing world. *Journal of Applied Probability* 25 (2): 287–298.

About the Authors

Chapter 1

Berend Wierenga is Professor of Marketing at the Rotterdam School of Management, Erasmus University (RSM). His main research areas are marketing models, managerial decision making in marketing, and marketing management support systems. He has widely published on these topics in journals such as *International Journal of Research in Marketing*, *Journal of Marketing*, *Journal of Marketing Research*, *Management Science, Marketing Science*, and many others. His books include *Marketing Management Support Systems: Principles, Tools and Implementation* (2000) and *Marketing Decision Making and Decision Support* (2010) (both with Gerrit van Bruggen). Berend Wierenga is the Founding Editor of the *International Journal of Research in Marketing* (journal of the European Marketing Academy-EMAC) and is the Editor of the first edition of the *Handbook of Marketing Decision Models* (2008). Before joining Erasmus University, he was a faculty member at Wageningen University where he also obtained his Ph.D. He has held visiting positions at Stanford University, INSEAD, and the Wharton School of the University of Pennsylvania. At Rotterdam School of Management, Berend Wierenga has been Chairman of the Marketing Department, Scientific Director of ERIM (Erasmus Research Institute of Management), and Dean of the School. He is the recipient of the inaugural EMAC Distinguished Marketing Scholar Award, the Ad Fontes Penning of the Erasmus University, and he was appointed Officer in the Order of Oranje-Nassau.

Ralf van der Lans is Associate Professor of Marketing at Hong Kong University of Science and Technology, Hong Kong. He has a M.Sc. in econometrics from Erasmus University Rotterdam and a Ph.D. (both cum laude) from Tilburg University, the Netherlands. His main research interests are in eye-tracking, social networks and choice for which he develops statistical and econometric models to improve marketing decisions in these areas. His research appeared in journals such as *Marketing Science*, *Journal of Marketing Research*, *International Journal of Research in Marketing,* and *Journal of the American Statistical Association*. Van der Lans is a member of the editorial board of the *International Journal of Research in Marketing* and has won several research and teaching awards.

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3

Chapter 2

Harald van Heerde is Research Professor of Marketing and MSA Charitable Trust Chair in Marketing at Massey University, Auckland, New Zealand. He has an MSC in Econometrics (cum laude) and a Ph.D. (cum laude) from the University of Groningen, the Netherlands. He has published in the major marketing journals about measuring marketing effectiveness, with a special emphasis on dynamic effects and time-varying parameters. Van Heerde serves as an Associate Editor at *Marketing Science* and at the *Journal of Marketing Research* and as an editorial board member at the *Journal of Marketing* and the *International Journal of Research in Marketing*.

Scott A. Neslin is the Albert Wesley Frey Professor of Marketing at the Tuck School of Business, Dartmouth College. He has been a visiting scholar at the Yale School of Management, the Fuqua School of Business, and Columbia Business School. Professor Neslin's expertise is in sales promotion, advertising, and database marketing. He is co-author with Robert C. Blattberg of the book, *Sales Promotion: Concepts, Methods, and Strategies*, and author of the Marketing Science Institute monograph, *Sales Promotion*. Professor Neslin has served as President of the INFORMS Society for Marketing Science (ISMS) and is an ISMS Fellow. Website: http://mba.tuck.dartmouth.edu/pages/faculty/scott.neslin/.

Chapter 3

Tingting Fan is an assistant professor at Department of Marketing, Business School of The Chinese University of Hong Kong. She got her Ph.D. from Stern School of Business, New York University. Her research interest includes quantitative marketing, consumer experience with high-technology products and services, digital marketing (mobile phone, video games, and online social network), branding, and empirical IO (industrial organization). Her research has won research grants from Wharton Customer Analytics Initiative (WCAI), Marketing Science Institute (MSI), and Product Development Management Association (PDMA). She is now teaching Marketing Management and New Product Management.

Peter N. Golder is Professor of Marketing at the Tuck School of Business at Dartmouth. His research on market entry timing, new products, long-term market leadership, and quality has won more than 10 best paper or best book awards including several of the most prestigious awards in the field: O'Dell (Journal of Marketing Research), Maynard (Journal of Marketing), INFORMS Long Term Impact Award (Marketing Science), Bass (Marketing Science), and Berry Book Prize (American Marketing Association).

Donald R. Lehmann is George E. Warren Professor of Business at Columbia University Graduate School of Business. He has a B.S. degree in mathematics from Union College, Schenectady, New York, and an M.S.I.A. and Ph.D. from the Krannert School of Purdue University. His research interests include modeling choice and decision making, meta-analysis, the introduction and adoption of innovations, and the measurement and management of marketing assets (customers and brands). He has taught courses in marketing, management, and statistics at

Columbia, and has also taught at Cornell, Dartmouth, and New York University. He has published in and served on the editorial boards of Journal of Consumer Research, Journal of Marketing, Journal of Marketing Research, Management Science, and Marketing Science, and was the founding editor of Marketing Letters and editor of the International Journal of Research in Marketing. In addition to numerous journal articles, he has published several books: including Market Research and Analysis, Analysis for Marketing Planning, Product Management, Meta Analysis in Marketing, and Managing Customers as Investments. He has won best paper awards from several journals, multiple lifetime achievement awards, and is a Fellow of both the Association for Consumer Research and the Informs Society for Marketing Science. Professor Lehmann has served as Executive Director of the Marketing Science Institute and as President of the Association for Consumer Research.

Chapter 4

Dominique Hanssens is Distinguished Research Professor of Marketing at the UCLA Anderson School of Management. From 2005 to 2007 he served as Executive Director of the *Marketing Science Institute.* A Purdue University Ph.D. graduate, Professor Hanssens's research focuses on strategic marketing problems, in particular marketing productivity. He is a Fellow and winner of the Buck Weaver award of the INFORMS Society for Marketing Science, and a recipient of the Churchill and Mahajan awards from the American Marketing Association. He is a founding partner of MarketShare, a global marketing analytics firm headquartered in Los Angeles.

Marnik G. Dekimpe (Ph.D., University of California, Los Angeles) is Research Professor of Marketing and Head of the Marketing Department at Tilburg University (The Netherlands) and Professor of Marketing at KU Leuven (Belgium). He has won best-paper awards in *Marketing Science*, the *Journal of Marketing Research*, the *Journal of Marketing*, the *International Journal of Research in Marketing*, and *Technological Forecasting and Social Change*. In 2016, he won the EMAC Distinguished Marketing Scholar Award, a life-time achievement award granted by the European Marketing Academy. In 2010–2012, he served as editor-in-chief of the *International Journal of Research in Marketing*, the leading European marketing journal. He is currently an academic trustee with AiMark.

Chapter 5

Tammo H.A. Bijmolt is Professor in Marketing Research at the University of Groningen, the Netherlands. His research interest include loyalty programs, retailing, e-commerce, advertising, and meta-analysis. His publications have appeared in leading international journals, such as Journal of Marketing, Journal of Marketing Research, Marketing Science, Journal of Consumer Research, International Journal of Research in Marketing, and Psychometrika. He won the best paper award in 2007 of the International Journal of Research in Marketing, and in 2011 of the Journal of Interactive Marketing. He is vice-president of the European Institute of Advanced Studies in Management (EIASM, Brussels).

Peter C. Verhoef (1972) is Professor of Marketing at the Department of Marketing, Faculty of Economics and Business, University of Groningen, The Netherlands. His publications appeared in journals, such as *Journal of Marketing, Journal of Marketing Research, Marketing Science and International Journal of Research in Marketing*. His research has received multiple awards. He functions as a senior editor for the *International Journal of Research in Marketing* and as area editor for the *Journal of Marketing Research* and *Marketing Science*. He is the founder of the *Customer Insights Center*. He has published a book on big data analytics in marketing: http://www.bigdatasmartmarketing.com.

Chapter 6

Pradeep K. Chintagunta is the Joseph T. and Bernice S. Lewis Distinguished Service Professor of Marketing at the Booth School of Business, University of Chicago. He is interested in empirically studying consumer, agent and firm behavior, and more recently, "development marketing"—studying the role of marketing in economic development. He graduated from Northwestern University and has also served on the faculty of the Johnson School, Cornell University. He has been a visiting professor at Harvard and Stanford universities. He is currently the Coordinating editor of the journal "Quantitative Marketing & Economics."

Chapter 7

Greg Allenby is the Kurtz Chair in Marketing at Ohio State University. He is a Fellow of the INFORMS Society for Marketing Science and the American Statistical Association. He is also the 2012 recipient of the AMA Parlin Award for his contributions to the field of marketing research and also a recipient of the ISMS Long-Term Impact award. He is past editor of *Quantitative Marketing and Economics*, and past area/associate editor for *Marketing Science*, *Journal of Marketing Research*, the *Journal of Business and Economic Statistics*.

Jaehwan Kim is a professor of marketing and associate dean at Korea University Business School, Seoul, Korea. He obtained a Ph.D. from the Ohio State University. His research is focused on microeconomic model for consumer demand including multiple-discreteness and constraint-based decision. His research has been published in Marketing Science and Journal of Econometrics.

Peter E. Rossi is James Collins Professor of Marketing, Economics and Statistics at the Anderson School of Management, UCLA. He received his Ph.D. from University of Chicago and BA from Oberlin College. He has published widely in marketing, economics, statistics and econometrics with more than 14,000 Google Scholar cites. He is a co-author of *Bayesian Statistics and Marketing*, John Wiley Series in Probability and Statistics (2005) and author *of Bayesian Semi and Non-parametric Methods in Marketing and Micro-Econometrics*, Princeton University Press (2014). He is author of the R package, bayesm. A fellow of the American Statistical Association and the *Journal of Econometrics*, he is founding editor, *Quantitative Marketing and Economics*.

Chapter 8

Andrew Ching is a Professor of Marketing and Economics at University of Toronto (Rotman School of Management and Department of Economics). He is an Associate Editor for *Management Science*, and a member of editorial board for *Marketing Science* and *Journal of Marketing Research*. He has received several major research grants from *Social Sciences and Humanities Research Council*. He has published research articles in *Econometrica, Marketing Science, Management Science, Quantitative Marketing and Economics, Journal of Banking and Finance, International Economic Review, Journal of Applied Econometrics, International Journal of Industrial Organization, Journal of Choice Modelling, Marketing Letters*.

Tülin Erdem is the Leonard N. Stern School Professor of Business and Professor of Marketing at the Stern School of Business, NYU. Before joining Stern in 2006, she has also been the E.T. Grether Professor of Business Administration and Marketing at the Haas School of Business, University of California at Berkeley. She has received best paper awards, as well major research grants, including two major National Science Foundation grants. She served as an Area Editor at several top journals. She was the editor-in-chief of the *Journal of Marketing Research* (2009–2012). She also served as the President of INFORMS Marketing Society.

Michael Keane is the Nuffield Professor of Economics at Oxford, and a Chief Investigator of CEPAR at UNSW. He previously held positions as Professor of Economics at Minnesota, NYU, Yale and UTS. Professional honours include: Fellow of the Econometric Society (2005), Australian Federation Fellow (2005), Council of the Econometric Society (2009), Australian Laureate Fellow (2012) and Fellow of the Journal of Econometrics (2016). Keane is best known for work on simulation methods (e.g., the "GHK algorithm") that have made estimation of complex discrete choice models more feasible, and for contributions to the theory and application of dynamic discrete choice models.

Chapter 9

Hans Baumgartner (Ph.D., Stanford University) is Chair of the Marketing Department and Smeal Professor of Marketing in the Smeal College of Business at the Pennsylvania State University. His research interests are in the areas of consumer behavior and research methodology. He has published articles on these topics in the *Journal of Consumer Research*, *Journal of Marketing Research*, *Marketing Science*, *Journal of Marketing*, *Journal of Consumer Psychology*, *International Journal of Research in Marketing*, *Organizational Behavior and Human Decision Processes*, *Psychological Methods*, and *Journal of Economic Literature*, among others.

Bert Weijters (Ph.D. in Marketing; Ghent University 2006) does research on survey methods, with a focus on response bias (response styles, context effects and comprehension difficulties) and methods for dealing with such bias through questionnaire design and data analysis. Bert also studies consumer behavior, specifically technology-enabled consumption (self-service technologies, digital music consumption), brand perceptions (brand personality, prototypical brands), and green consumption. His work has been published in the *Journal of Consumer Research, Journal of Marketing Research, Psychological Methods, International Journal of Research in Marketing, Journal of Business Ethics, Journal of the Academy of Marketing Science, Applied Psychological Measurement,* and other outlets.

Chapter 10

Eva Ascarza is an Associate Professor of Marketing at Columbia Business School. She is a marketing modeler who uses tools from statistics, economics and machine learning to answer relevant marketing questions. Her main research areas are customer analytics and customer relationship management, with special attention to the problem of customer retention. She uses field experimentation (e.g., A/B testing) as well as econometric modeling and machine learning tools not only to understand and predict patterns of behavior, but also to optimize the impact of firms' interventions.

Peter Fader is the Frances and Pei-Yuan Chia Professor of Marketing at the Wharton School of the University of Pennsylvania. His expertise centers around the analysis of behavioral data to understand and forecast customer shopping/purchasing activities. Some of these insights are reflected in his book, "Customer Centricity: Focus on the Right Customers for Strategic Advantage," while others are brought to life via Zodiac (http://www.zodiacmetrics.com/), a tech firm (which he co-founded) that aims to make top-notch customer valuation models (and their applications) accessible to a broad array of data-driven organizations.

Bruce Hardie is a Professor of Marketing at London Business School. His research and teaching interests lie in the areas of customer and marketing analytics. Much of his research focuses on the development of stochastic models of buyer behaviour for customer-base analysis, with an emphasis on ones that can be implemented in Excel, thus lowering the barriers to adoption by firms.

Chapter 11

Ralf van der Lans, see Chap. 1.

Michel Wedel is the Pepsico Chaired Professor of Consumer Science at the Robert H. Smith School of Business, and a Distinguished University Professor, at the University of Maryland. He was ranked the most productive marketing researcher in the world. He has been a consultant for many companies like Microsoft, Procter & Gamble, Mars, Unilever, Daimler, McKinsey, Nielsen, Gfk and TNS amongst others. Professor Wedel has fundamentally improved the understanding of consumer behavior and marketing decision making by developing and applying advanced statistical and econometric methods. He is a pioneer in the development of eye-tracking methodology and is considered a worldwide expert on visual marketing. Professor Wedel has published more than 175 peer reviewed articles and over 20 book chapters which have been cited over 16,000 times. He is area editor of *Marketing Science,* the *Journal of Marketing Research* and the *Journal of Marketing*. Wedel's work has received several awards. He received the Muller award for outstanding contributions to the social sciences from the Royal Dutch Academy of the Sciences, the Churchill award for lifetime contributions to the study of marketing research from the American Marketing Academy, and is a fellow of the American Statistical Association and the Institute for Operations research and Management Science.

Chapter 12

Marnik Dekimpe, see Chap. 4.

Barbara Deleersnyder is an Associate Professor of Marketing at the Tilburg University, the Netherlands. She teaches courses in strategic marketing management and advanced research techniques. Her research interests are in the area of branding and new product development in retailing, the impact of general economic conditions on consumer and firm behavior, market(ing) performance measurement and management of change. Her studies have appeared in international marketing journals under which the *Journal of Marketing*, the *Journal of Marketing Research*, the *International Journal of Research in Marketing*, and *Quantitative Marketing and Economics*.

Chapter 13

Vardan Avagyan is assistant professor of marketing at Erasmus School of Economics, Erasmus University Rotterdam. He received his Ph.D. in business administration and quantitative methods with a major in marketing from University of Carlos 3, Madrid. His research primarily focuses on innovation, including topics such as innovation decisions at the fuzzy front-end and innovation commercialization. His work appeared in journals such as *International Journal of Research in Marketing* and *European Journal of Operational Research*.

Vardit Landsman is an Associate Professor at the Erasmus School of Economics, Erasmus University Rotterdam and a senior lecturer at the Coller School of Management, Tel-Aviv University. Her research focuses on the implementation of new modeling approaches and advanced estimation methods to the study of marketing phenomena. Her recent work on the pharmaceutical industry deals with physician-patient relationship structures, physicians' characteristics as a source of heterogeneity in market behavior and the effect of regulations on market outcomes for pharmaceutical drugs. Her research was published in top marketing journals such as *Marketing Science*, the *Journal of Marketing* and *Management Science*.

Stefan Stremersch holds the Desiderius Erasmus Distinguished Chair of Economics and a Chair of Marketing, both at Erasmus School of Economics, Erasmus University Rotterdam, the Netherlands and is professor of Marketing at IESE Business School, Barcelona, Spain. His main research interests are innovation, marketing of technology and science, and pharmaceutical marketing. Stremersch has won several awards, such as the Harold H. Maynard Best Paper Award of the

Journal of Marketing (2002), the IJRM Best Paper Award (2012 & 2014), the JC Ruigrok Prize for the most productive young research in the Netherlands (2005; awarded only once every 4 years to an economists), the Rajan Varadarajan Early Career Award of the American Marketing Association (2008), the American Marketing Association's Award for Global Marketing (2006) and he was a finalist for the Informs Marketing Science Long Term Impact Award in 2009, 2010, 2013 and 2014. In 2015, Ghent University and the Francqui Foundation awarded him the honorary International Francqui Chair, selected across all sciences.

Chapter 14

Randolph E. Bucklin is Professor of Marketing at the UCLA Anderson School where he holds the Peter W. Mullin Chair. Professor Bucklin's research focuses on the quantitative analysis of customer behavior. He specializes in models using historical records of customer transactions from bar-code scanner and digital marketing data. His recent work has been recognized with the Paul Green Award (2011 and 2012) and the William F. O'Dell Award (2015 and 2016). Professor Bucklin served as co-editor of the *Journal of Marketing Research* (2014–16) and co-editor of *Marketing Letters* (2006–10). He was faculty chairman and deputy dean at the Anderson School from 2012 to 2015.

Paul R. Hoban is an Assistant Professor of Marketing at the Wisconsin School of Business at the University of Wisconsin—Madison. His research leverages quantitative methods to study consumer response and firm decision making. His work has been published in the Journal of Marketing Research, and he has served as a reviewer for Marketing Science and the Journal of Marketing Research. Professor Hoban earned his undergraduate degree in marketing from Michigan State University before earning his Ph.D. from the UCLA Anderson School of Management. Before his academic career, he worked as an analyst and consultant in the e-commerce space.

Chapter 15

Peter Danaher is Professor of Marketing and Econometrics at Monash University in Australia. Formerly, he was Coles Myer Chair of Marketing and Retailing at the Melbourne Business School, and was previously at the University of Auckland. He has held visiting positions at London Business School, The Wharton School, NYU and MIT. He is an Area Editor for the *Journal of Marketing Research*, and also serves on the Editorial Boards of *Marketing Science, IJRM* and the *Journal of Service Research*. His primary research interests are media exposure distributions, advertising effectiveness, television audience measurement and behavior, internet usage behavior, customer satisfaction measurement, forecasting and sample surveys, resulting in many publications in journals such as the *Journal of Marketing Research, Marketing Science, Journal of the American Statistical Association* and the *Journal of Marketing*.

Chapter 16

Wendy Moe is Professor of Marketing and Director of the Masters of Science in Marketing Analytics at the University of Maryland's Smith School of Business. She is an expert in online and social media marketing, focusing on analytics. Professor Moe has been recognized by the *American Marketing Association* and the *Marketing Science Institute* as a leading scholar in her field with the Howard Award, the Young Scholar Award, the Erin Anderson Award and the Buzzell Award. She holds a Ph.D., M.A. and B.S. from the Wharton School at the University of Pennsylvania as well as an MBA from Georgetown University.

Oded Netzer is the Associate Professor of Business at Columbia Business School and an affiliate with the Columbia Data Science Institute. He received an M.Sc. in Statistics and a Ph.D. in Business, from Stanford University. Professor Netzer's research centers on one of the major business challenges of the 21st century: developing quantitative methods that leverage data to gain a deeper understanding of customer behavior and guide firms' decisions. His research has been published in top journals. Professor Netzer won multiple awards for his research and teaching. He is the winner of the ISMS long-term award, best paper awards, outstanding dissertation award, and Columbia Business School Award for teaching excellence.

David A. Schweidel is Goizueta Term Chair Professor, Caldwell Research Fellow and Associate Professor of Marketing at the Goizueta Business School of Emory University. Schweidel received his B.A. in mathematics, M.A. in statistics, and Ph. D. in marketing from the University of Pennsylvania. Prior to joining Emory in 2012, he was on the faculty of the Wisconsin School of Business at the University of Wisconsin-Madison. Schweidel is an expert in the areas of customer relationship management and social media analytics. His research focuses on the development and application of statistical models to understand customer behavior and inform managerial decisions.

Chapter 17

Ralf van der Lans, see Chap. 1.

Xi Chen is Assistant Professor of the Marketing Management Department at Erasmus University, where he has been since 2014. He received his Ph.D. in Marketing from Hong Kong University of Science and Technology in 2013. His research interests span both marketing and network science. Much of his work has been on generating new marketing insights and improving marketing performance with network theory. His work has been published in *Journal of Marketing Research*.

Michael Trusov is Associate Professor of Marketing at the Robert H. Smith School of Business at the University of Maryland. His research interests include social media, search engines, social networks, e-commerce, recommendation systems, user-generated content, text analysis, eye-tracking and machine learning. His research has won several awards, including the O'Dell Award, the Paul Green

Award, the Donald Lehmann Award, the SMA Emerging Scholar Award, the Emerald Management Reviews Citation of Excellence Award, and the MSI's Alden Clayton Award. Professor Trusov has extensive industry experience in the area of software development and IT consulting specializing in marketing automation, database management, Internet applications, and e-commerce.

Chapter 18

Gui Liberali is associate professor at the Erasmus University (RSM). His research has been published in Marketing Science, Management Science, IJRM, EJOR, and Sloan Management Review. Gui is co-editor of an IJRM special issue, co-chair of the 1st AMA and EMAC joint symposium, finalist of the John Little award, and founder of the Erasmus Centre for Optimization of Digital Experiments. His research focused on optimization, multi-armed bandits, morphing, and online experimentation. Gui was a visiting scholar at the MIT Sloan School of Management for several years. He holds a Doctorate in Marketing, and a B.Sc. in Computer Science.

John R. Hauser is the Kirin Professor of Marketing at the MIT Sloan School of Management. He has co-authored textbooks on product development, is a former editor of *Marketing Science*, a founder and principal at Applied Marketing Science, Inc., a former trustee of the Marketing Science Institute, a fellow of INFORMS and ISMS, and past-President of ISMS. He has won the Converse, Parlin, and Buck Weaver Awards. He teaches marketing management and research methods and enjoys research on a variety of practical and sometimes less practical topics.

Glen Urban is the *David Austin Professor in Management (Emeritus) at the MIT Sloan School of Management.* He has been a member of the MIT Sloan School of Management faculty since 1966, was Deputy Dean at the school from 1987 to 1991, and Dean from 1993 to 1998. Urban is co-author of seven books and numerous papers. His papers have won several prestigious awards, including two O'Dells. He has received the Paul D. Converse, Charles Parlin, and Buck Weaver Awards for lifetime achievement in marketing research. His current research is on Deep Learning applied to marketing decisions.

Author Index

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3

L

Subject Index

B. Wierenga and R. van der Lans (eds.), *Handbook of Marketing Decision Models*, International Series in Operations Research & Management Science 254, DOI 10.1007/978-3-319-56941-3