SpringerWienNewYork

Interdisciplinary Studies in
Economics and Management

Vol. 5

Edited by the
Jubiläumsstiftung der
Wirtschaftsuniversität Wien

Alfred Taudes (ed.)

Adaptive Information Systems and Modelling in Economics and Management Science

SpringerWienNewYork

Univ.-Prof. Mag. Dr. Alfred Taudes
Abteilung für Produktionsmanagement, Wirtschaftsuniversität
Vienna, Austria

Printing supported by the Fonds zur Förderung der wissenschaftlichen Forschung
Vienna, Austria

SpringerWienNewYork is a part of Springer Science + Business Media
springeronline.com

Typesetting: Camera-ready by the authors
Printing and binding: Strauss GmbH, 69509 Mörlenbach, Germany

Printed on acid-free and chlorine-free bleached paper
SPIN: 10976005

With 68 Figures

CIP data applied for

ISSN 1615-7362
ISBN-10 3-211-20684-1 SpringerWienNewYork
ISBN-13 978-3-211-20684-3 SpringerWienNewYork

Acknowledgments, aims and scope of this research report

The Interdisciplinary Studies in Economics and Management series of books has been instrumental in reporting about the results generated in a joint effort of the group of researchers in mathematics and management science that joined in the Joint Research Program (Spezialforschungsbereich) on "Adaptive Systems and Modeling in Economics and Management Science" funded by the Austrian Science Foundation under grant SFB 010. The aim of the SFB reports in this series has been to present the joint findings of this group in a manner that both is interesting for readers with a background in economics and management and mathematics and statistics and allows non-expert readers to grasp the ideas of modern management science. Following the interdisciplinary dialogue that has been going on between the researchers both aspects are covered in an integrated way, hopefully providing a better access to modern topics in management.

So far, three volumes of SFB publications have appeared in this series:

- Josef A. Mazanec, Helmut A. Strasser: A Nonparametric Approach to Perception-Based Market Segmentation: Foundations. 2000.
- Christian Buchta, Sara Dolnicar, Thomas Reutterer: A Nonparametric Approach to Perceptions-Based Market Segmentation: Applications. 2000.
- Herbert Dawid, Karl Dörner, Georg Dorffner, Thomas Fent, Martin Feurstein, Richard Hartl, Andreas Mild, Martin Natter, Marc Reimann, Alfred Taudes: Quantitative Models of Learning Organizations. 2002.

This volume completes the previous volumes on market segmentation, product positioning, and target marketing and organizational learning with contributions of SFB 010 on modeling consumer behavior, modeling financial markets, agent-based simulation model, and statistical modeling and software development.

The Jubiläumsstiftung der Wirtschaftsuniversität Wien, founded in 1997 to mark the 100th anniversary of the WU Wien, is particularly engaged in fostering all types of research crossing disciplinary borders. By initiating a new series of publications devoted to these principles the WU-Jubiläumsstiftung wants to set an example that a major investment into an interdisciplinary style of research pays off. The authors left no stone unturned to fulfill this "benchmark" function.

A. Taudes, June 2005

Contents

List of Contributors

Christian Buchta, Josef Mazanec, Ulrike Schuster, Jügen Wöckl
Institute of Tourism and Leisure Studies,
Vienna University of Economics and Business Administration

David Meyer, Andreas Mild, Alfred Taudes
Department of Information Systems and Operations,
Vienna University of Economics and Business Administration

Alexander Pfister
Department of Economics, Vienna University of Economics and Business Administration

Kurt Hornik, Regina Tüchler, Achim Zeileis
Department of Statistics and Mathematics,
Vienna University of Economics and Business Administration

Friedrich Leisch, Alexandros Karatzoglou
Department of Statistics, Vienna University of Technology

Manfred Deistler, Eva Hamann, Thomas Ribarits, Wolfgang Scherrer,
Thomas Steinberger
Institute for Mathematical Methods in Economics, Vienna University of Technology

Tatiana Miazhynskaia, Leopold Sögner
Department of Managerial Economics, Vienna University of Technology

Georg Dorffner
Department of Medical Cybernetics and Artifical Intelligence, Medical University of Vienna

Engelbert Dockner, Rudolf Vetschera, Roland Bauer, Albert Schwingenschlögel
Department of Business Studies, University of Vienna

Sylvia Frühwirth-Schnatter
Department of Applied Statistics and Econometrics, Johannes Kepler University Linz

Thomas Otter
Fisher College of Business, Ohio State University

Brian Sallans
Austrian Research Insitute for Artifical Intelligence

Lucas Zinner
Austrian Science Foundation

Introduction
General scientific concept: aims of SFB 010

Alfred Taudes

Special research area SFB 010 has been a major cooperative effort of researchers of Viennese universities and their partners to increase understanding of Adaptive Information Systems and Modeling in Economics and Management Science. Between 1997 and 2003, the research group aimed at improving

- quality of forecasting of economic time series
- capabilities for modelling consumer behaviour explanation of financial markets and investment decision processes
- explanation of financial markets and investment decision processes
- efficiency of organizational structures and planned organizational change
- adaptivity of management decision support systems
- theory of neural pattern recognition and the pattern recognition methodology for large-scale applications

through the development of an Artificial Economy (AE) to support strategic and tactical decisions of firms and to study industry evolution.

The concept of an AE has been used as a generic term, which can be interpreted at different levels as:

1. any mathematical model of economic phenomena
2. an economic model with learning agents
3. an economic model with learning agents in which the outcome of the interaction is determined by simulation.

Within the SFB all three levels of an AE have been considered and integrated. Research in the first two categories focused on data-analysis and analytical modelling, in the work done on the third level both the modelling of learning behavior and simulation were central. This definition of an AE relaxes standard assumptions like complete information and full rationality: rational decision makers are replaced by agents who have incomplete information, but are able to learn from empirical observations.

Introducing boundedly rational decision makers who update their believes on the basis of learning rules, opens the door for many ad hoc assumptions on both the level of ”irrationality” and the learning algorithm. In order to avoid any arbitrary approach, work within the SFB strictly followed the two principles:

1. any learning rules used are based on empirically observed behavior of real-world agents;

2. AE models are calibrated in this way that they generate artificial time series that mimic stylized facts observed in real markets.

These considerations motivated the SFB's broad definition of an AE given above: methods of advanced data analysis provide the basis for modelling markets and analytical/forecasting capabilites of agents, while agent-based simulation models allow the study of strategies and development of industries in an evolutionary setting.

In order to achieve the objectives set, expertise from both applied Initiatives and statistical and tools Initiatives is required and interaction among the initiatives is necessary. Thus, SFB 010 had an organizational structure covering three application-oriented initiatives focusing on marketing, production and strategy and finance, respectively, and two methodologically oriented initiatives (see Figure 1)

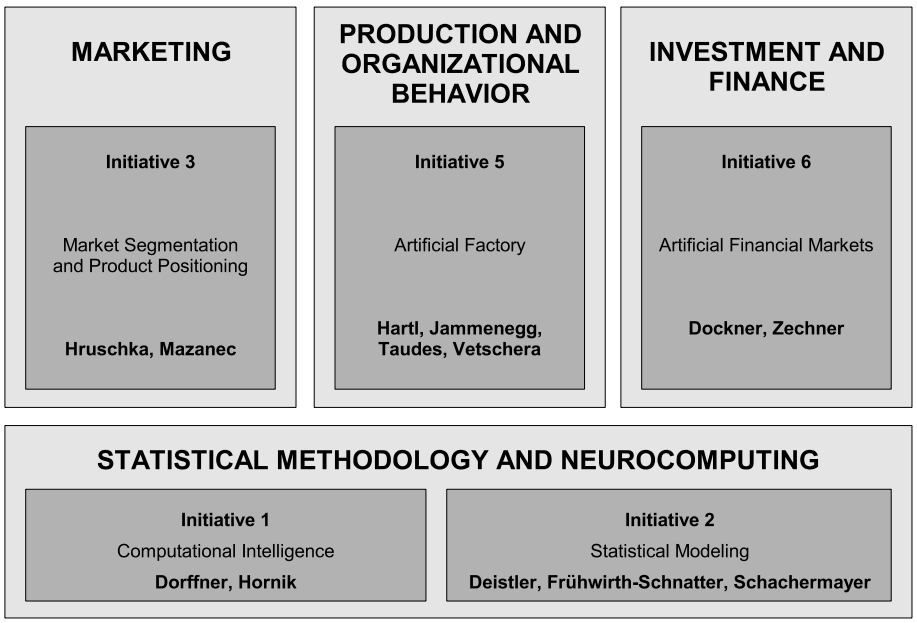

Figure 1: The SFB's Organizational Structure

The basis for the research performed in the individual projects has been the meta-level model of an AE shown in Figure 2. The meta-level of an AE developed by the SFB builds on the competencies of the researchers within the SFB and the state of knowledge on AEs in the literature. The AE consists of a product market, where several Artificial Firms (AFs) compete in an oligopolistic setting. Each AF consists of four adaptive agents: a corporate strategy agent, a production agent, a marketing agent and a corporate finance agent. Each agent builds a model of its respective environment

(technology space, customer perceptions and preferences, financial market) using a particular learning style, makes decisions regarding his functional strategy and interfaces with the other agents in order to integrate the domain-specific strategies into the AF's corporate strategy. Thus the AF determines its production program and market operation and implements the financial decisions.

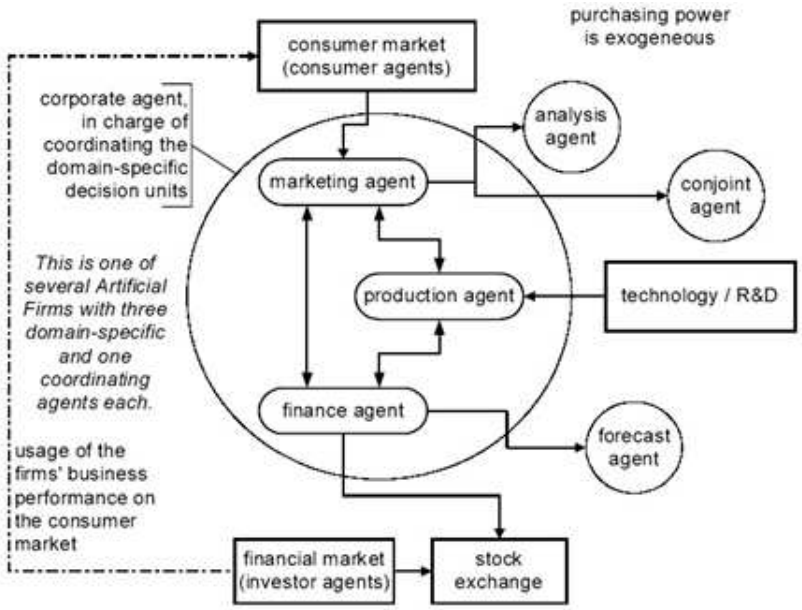

Figure 2: The SFB's Artificial Economy Meta Model

Through market research data the marketing agent infers profitable positions in the space of customer perceptions and preferences, derives sales forecasts and suggests technical specifications of products to be offered. The task of the production agent is to estimate costs of product development and production programs. It builds its expectations based on knowledge about a space of cost functions characterizing the available technologies, where it has to choose between learning by doing, experimentation and imitation (benchmarking). Both the marketing and the production agents provide time series of sales and cost estimates for the corporate finance agent, who calculates the Net Present Value of the strategies thus defined. It also makes financial decisions based on its observation of the financial market (such as risk-adjusted interest rates). Based on the firm's overall strategy in terms of a harmonized production, marketing and financial program, the actual sales, costs and profits are calculated on the basis of the competitors' actions and provided as feedback to the AF. Then, it is natural to analyze what different strategies might emerge under different environmental settings and which strategies are successful under what conditions.

The motivations for this type of model of an AE are as follows:

Focus on managerial issues involving different function areas: This design decision has been based on the nature of the common projects, the integration of previous work and the competencies of the researchers within the SFB. A complete model of an AE, encompassing detailed models of, e.g., a labor market or government intervention has not been judged feasible, but rather a model of Artificial Markets (AM). We feel that this focus has been a unique selling proposition to the SFB, as most other works on AEs have a macroeconomic policy focus and deal with firms' strategies in a rather simple way, e.g. by assuming no product innovation at all.

Use of adaptive models: It is another unique feature of the SFB's AE model as compared to previous approaches that the AF's agents are capable of inventing new strategies through learning. The appropriate modelling has been done in cooperation between the methodologically oriented and the application-oriented Initiatives. Of particular interest are experiments with different levels of "bounded rationality", ranging from simple imitative or adaptive response behavior to rationality based on heuristics developed in management science.

Development of a common computing environment: A common framework to facilitate distributed software development and to enable simulations of the AM model is indispensable for such a project. The simulation environment has been implemented using an object-oriented style of programming, allowing for distributed objects (over a cluster of workstations or even the Internet).

Definition of a realistic, data-driven environment: The empirical validation of results obtained via the simulation of Artificial-Life models is a difficult task. This topic has been partially addressed through a careful definition of the AF's environment – via the integration of knowledge about the topology of the respective search spaces gained in empirical studies using advanced data-analytic methods. The goal has been to develop models that are simple enough to allow a controlled simulation and validation but nevertheless realistic enough to derive findings transferable to real-world settings. Another criticism of Artificial-Life simulation models is the lack of reliable data on micro level behavior. One way to deal with this problem is to compare the characteristics of the time series generated on the macro level with those of empirically observed time series. This has been another area of cooperation between the methodologically oriented and the application-oriented Initiatives.

The SFB's spectrum of research approaches has been the basis of four integrative projects, each one combining the efforts of several initiatives, where the task of the application-oriented initiatives has been the development of relevant theories, agents and modelling environments, while the methodology-oriented groups have been focusing on computing, simulation, learning, time series and data analysis.

Project **Unobserved Heterogeneity** dealt with cross-sectional heterogeneity as typically found in marketing data and longitudinal heterogeneity (structural change). Encompassing Initiatives 1, 2 and 3 a broad range of models including mixture models, non-parametric methods and clustering by ensemble methods were investigated. Also integrated models capturing both types of heterogeneity and the test of the models against artificially generated data have been pursued.

In project **Empirical Capital Market Analysis** a number of hitherto unexplained stylized facts observed in financial markets has been studied using methods from fi-

nance, statistics and agent-based computing by Initiatives 1, 2, and 6. A major topic of this effort was the study of stochastic volatility, other areas of research have been the modelling of conditional distributions for asset returns, factor models and return and volatility forecasts and the empirical analysis of the governance structure of investment funds.

Project **Agent-based Computational Economics** aimed at using the SFB's AE to study firms' strategies and industry evolution by agent-based simulation models. In particular, the effect of organizational structures for new product development, technology management, managerial incentive structures, benchmarking, product portfolio etc. has been investigated.

While the phenomena studied in this project were observed in different industries, project **Energy Markets** put an emphasis on the currently transforming energy sector. This development necessitates novel forms of forecasting, risk management, market design and strategy (capacity, cooperation, market entry). Respective methods applied here have been time-series analysis and agent-based simulation.

This reader contains work done in all four projects, organized into the sections: Modeling Consumer Behavior, Modeling Financial Markets, Agent-Based Simulation Models, and Statistical Modeling and Software Development.

Part I

Modeling Consumer Behavior

Basic Concepts and a Discrete-Time Model

Christian Buchta and Josef Mazanec

1 Purpose and Modules of the Artificial Consumer Market as a Simulation Environment

It is not the purpose of the Artificial Consumer Market (ACM) to mimic any "real" consumer population. Rather it aims at constructing an artificial environment at the marketing front end of the Artificial Firm that puts the AFs under challenge to function and survive as learning organizations. Therefore, the ACM duplicates only a selected number of *typical* properties of consumer markets that are deemed crucial for the success of marketing strategies.

The simulation environment contains two modules: (1) the Artificial Consumer Market (ACM) and (2) analytical and strategic marketing agents of the Artificial Firms (AFs) including recent methodology for conjoint analysis (Frühwirth-Schnatter and Otter, 1999) and perceptions-based market segmentation (Mazanec and Strasser, 2000; Buchta et al., 2000).

The simulation environment introduces the refinements needed to comply with contemporary consumer theory and structural equation models of buyer behavior (Howard and Sheth, 1969; Engel, Kollat and Blackwell, 1973; Howard, 1977; Mazanec, 1978; Kroeber-Riel, 1980; Bagozzi, 1986; Myers, 1996). Particularly, it distinguishes between the brand attributes (which are only observable to the AFs as binary yes/no reactions) and the underlying latent attitude dimensions. This leads to a multi-level system for the different "languages" of advertising and consumers expressing their everyday experience, the consumers' choice criteria rooted in long-term memory, and the jargon of the R&D engineers in the AF. The ACM models the brand perceptions on three levels: latent attitudinal dimensions, verbal response generating probabilities and (redundant sets of) observable indicators of the latent dimensions. The consumers' acquire product comprehension by being exposed to market communication about (modifications of) brand attributes. These bundles of perceived attributes are indicative of a set of unobservable latent attitude dimensions. Thus, the consumers preserve a condensed brand profile in a latent attitude space, which is imperfectly retrieved by the AFs owing to the consumers' limited ability to express their brand evaluations. This is a very realistic setting that puts the AFs under pressure to explore the attitude space by trial and error. The degree of ambiguity of the brand attribute indicators is systematically adjustable. It may be subject to experimentation with "would-be worlds" (Troitzsch, 1999) confronting the AFs with an either easily decodable or a rather fuzzy consumer response. Discovering the type of a "learning organization" that is more likely to survive under these challenges is an intriguing research question.

The Artificial Consumer Markets-Artificial Firm interface tackles this problem by providing a link between the latent attitudinal dimensions and the technical features, which is unknown to the AF. Both the AF's product improvement program and

market communication influences the consumers' brand perceptions, attitudes and choices. Both the product features detected during consumption and the advertising stimuli are input to the consumer's sensory, perceptual and evaluative systems. As consumers dislike to persist with an inconsistent attitudinal system they have to settle to a "compromise" post-choice attitude. Reconciling and weighting the technology-induced and the advertising-caused positions in attitude space also allows for simulating "technology-driven" vs. "market-driven" environments. Production/technology and marketing/promotion set mutual restrictions and reinforce or dampen each other. The brand perceptions and choice model makes a distinction between the consumers' (directly unobservable) abstract product comprehension ("long-term memory") and the observable consumer and advertising vocabulary. The attributes of the observational language ("short-term memory") are subject to communicative persuasion and periodically measured in consumer surveys. Advertising-induced changes in the strength of belief regarding a brand possessing a particular attribute are nonlinearly fed back into the long-term memory.

To sum up the following conceptual building blocks characterize the ACM:

- According to the tradition of product positioning theory the consumers' brand perceptions and evaluations (attitudes) are modeled as points in a latent space, which is unknown to the competing firms and can only be figured out by processing observable attribute assignments. Thus the ACM differentiates between the consumers' redundant and fuzzy manner of talking about a particular product class and the managers' and product engineers' condensed "expert" language. Brand perceptions are initialized in a segment-specific manner.

- Preferences are incorporated into the brand space as "ideal points"; unlike conventional ideal-point models, however, the ACM employs a modified unidirectional model to allow for irrelevant attitude dimensions without having to distinguish between desired and undesirable dimensions. The preferences are segment-specific and not necessarily linked to the consumer perceptions of rivalling brands.

- The consumers' "cognitive algebra" comprises compensatory as well as non-compensatory choice rules. Consumers in the ACM follow simple rules requiring very modest assumptions about the consumers' information processing and attitude formation. These rules are operative on the disaggregate level and characterize what economists may term a boundedly rational being. It is imperative that the ACM does not imply just one built-in decision mechanism but allows for a variety of rules and consumer heterogeneity in terms of decision styles.

- The ACM consumers develop pre-choice and post-choice attitudes towards the competing brands. They form consideration sets of acceptable brands based on the expectations aroused by advertising and on their personal preferences. They make random decisions in case of several brands being equally attractive and equally priced.

- Attitude change depends on confronting the brands' technology induced evaluation with the perceptual profile aroused by advertising. Consumers who purchase a brand contribute to disseminating the product comprehension and the knowledge about the brand's technological quality. The technological properties are not part of the consumer language and never experienced individually and isolated from each other. Rather the consumers experience them 'holistically' by building a technology induced attitude, which may diverge from the expectations mediated by advertising.

- Market communication happens through media advertising and through word-of-mouth. Advertising carries nontechnical persuasive information. According to what is known from communication research word-of-mouth fulfills a double function. The communicator's (opinion leader's) relay function guarantees that knowledge about the brands' technical properties gets disseminated. At the same time the communicators influence the recipients' decision making by reporting their valuing of the brands' performance. This is achieved by spreading their personal (dis)satisfaction experience.

- The (dis)satisfaction experienced after buying a brand governs the consumer's intention to repurchase, the propensity to spread word-of-mouth messages, and the persuasibility regarding future advertising. A consumer who finds his expectations fulfilled is likely to develop loyalty. In the ACM this is equivalent to keeping a brand in one's evoked set of purchasing alternatives despite one or more competing brands becoming more attractive. A disappointed consumer may (temporarily) ban the brand from his consideration set of buying alternatives; then it is disregarded irrespective of its advertising pressure. Disappointment nourishes the consumers' reactance to persuasive advertising. Personal communication is more likely to occur for more extreme (dis)satisfaction levels. Exaggerated advertising claims and unfulfilled promises thus feed dissonant information into personal communication channels and also provoke dissonance of the non-buyers receiving such messages.

- A number of sensitivity parameters governs the depth and accuracy of the consumers' information processing and cognitive effort. These parameters capture the influence of the involvement in the product class (Kroeber-Riel, 1980, p. 315). There is no separate variable for brand involvement (Mühlbacher, 1988). The involvement is consumer-specific to allow for experimental settings with different involvement segments.

2 The ACM Macro Structure

Figure 1 highlights the macro structure. It assists in describing the ACM dynamics and the data flow between the levels of latent constructs and observable indicators.

It is important to generate starting values of the ACM according to a scenario determined by the experimental design. Many experiments require an "equal opportunities" scenario, where no brand/firm benefits from some in-built competitive advantage, but

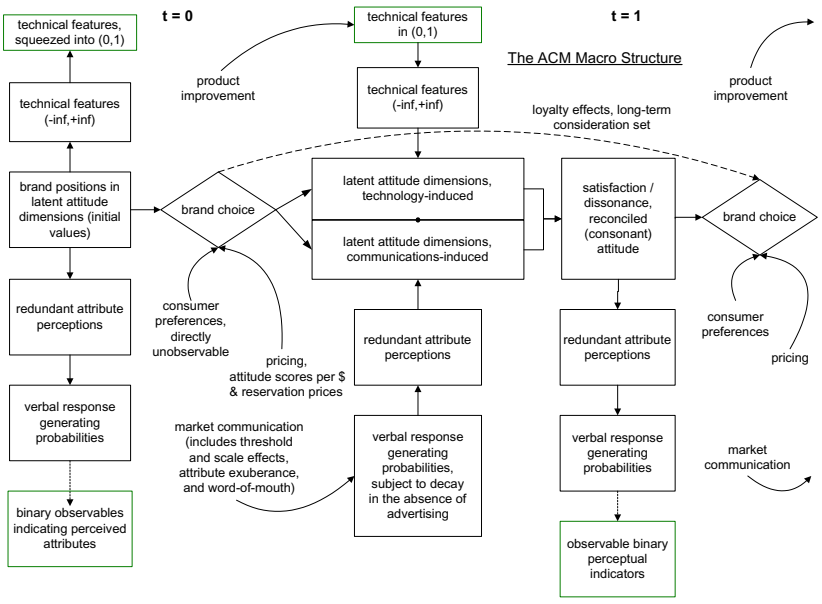

Figure 1: ACM macro structure

lives on its own analytical skills and imaginative strategies. Another requirement is the tuning of the measurement models, as the ACMs should also differ with respect to the accessibility of the latent attitude dimensions. One market may exhibit a "simple structure" in factor-analytic terminology, others may be rather obscure in terms of the perception indicators available to the AFs. Hence a "classic" factorization scheme comes to mind first. It expresses the observable product attributes as non-linear combinations of a small set of (attitudinal) factor scores. The loadings matrix introduces intercorrelations into the perceptual attributes assuming orthogonal factors and uncorrelated attitudinal dimensions. Thus, experimental variations of the empirical accessibility of the brand attitudes in the initial period may either set the values of the loadings matrix or define the desired intercorrelations of perceived attributes. Alternatively, a nonlinear mapping by means of some neural network architecture may be applied.

The ACM macro structure in Figure 1 originates from combining a factor-analytic model (which is very ordinary in product positioning theory) with a simple probabilistic independence model for generating the binary observables. As an extension, a latent-variable threshold model (Long 1997; Fahrmeir and Tutz, 1997) may be employed that allows for experimenting with group-specific or individual threshold parameters. The consequence is that the product attributes are not measured on pseudo-interval rating scales. They get squashed into probabilities, which lead to either affirmative or negative consumer response in terms of yes-no statements. The loadings

matrix determines the distinctness of the consumers' attitudinal system. By setting their values the experimenter may create a product class where the attitudinal dimensions are easy or hard to recognize by the AFs.

For the initial period it is imperative to set out with a scenario that conforms precisely with the experimental design. As mentioned above this will most frequently be a setting, where no brand or firm outperforms the other AFs because of implicit competitive advantages. The AFs product improvement and promotional spending decisions in the initial period change the brands' positions in the consumers' latent attitude space. The mass communication via media advertising uses non-expert, unprecise and emotionally loaded vocabulary. Through this language filter the AFs initiate changes of the brand positions in attitude space. If an advertising claim loses its sustained media support, the probabilities (beliefs) decay and the strengths of the perceived brand attributes decrease drastically.

Product variations and changes in the brands' technical features become known to the buyers and to those consumers receiving messages from buyers through personal communication. The consumers do not directly recognize these technical features, which are defined in production expert language. But they make a technology-based evaluation resulting in a technology induced attitude. The technical features are linked to the consumers' latent attitude space via a non-linear transformation held constant during a series of simulation cycles but unknown to the AFs. They have to figure out how technology influences the consumers' product evaluation as best they can. By sorting, filtering, and weighting all these evaluative materials ("reconciling") the consumers arrive at a post-purchase and/or post-communication attitude. Again it is reflected by the brand's position in the latent attitude space. And again, it is not directly observable, but has to be measured by verbal (or pictorial) indicators. In compliance with standard psychometric modeling the necessary strength of belief must grow exponentially for generating unity values on the measurement level with near-certainty.

3 Set Theory, Brand Choice, (Dis)satisfaction and Adaptive Preferences

In earlier versions of the simulation environment the consumers followed a cognitive algebra that allowed for three different rules of brand choice: compensatory, and non-compensatory conjunctive or disjunctive. Product perceptions and preferred attributes were involved in the consumers' utility calculus; product knowledge was subject to learning, preferences were fixed. Also the consumers did not have a memory of their brand choices made in past periods. Refinements regarding the role of preferences in the consumers' brand choice decisions are now implemented. It is straightforward to introduce variable preferences dependent on adaptive aspiration levels. While the 'learning of preferences' is still a largely unexplored notion in traditional economics (Brenner, 1999, p. 117) the Artificial Consumer Market functions more realistically as far as these perceptual and preferential dynamics are concerned. Preferences are portrayed in the latent attitude space as "ideal points". A consumer's ideal point for a product class indicates the combination of his desired levels of each attitude dimension. This is equivalent to an aspiration level that varies according to the consumer's product knowledge and experience. An unrealistically high aspiration level cannot be

maintained without continuing disappointment. A modest aspiration level easily fulfilled by an average purchase alternative is likely to rise as consumers learn to acquire better value for money.

Any combined marketing-production model benefits from the numerous studies that have been conducted in (service) quality research. Most of the empirical studies were inspired by the SERVQUAL model (Parasuraman, Zeithaml and Berry, 1985, 1988; Zeithaml and Berry, 1988). Irrespective of all the critical comments, which are rightly brought forward against the SERVQUAL concept, it has its merits as far as it triggered off a lively discussion about the construct of "perceived (service) quality". One of the lessons seems to be that a construct "perceived quality" separate from the construct of attitude toward products or services is highly superfluous (Mazanec, 1997). However, the discussion clarified the views about "transaction-specific" versus long-term attitudes and reiterated the need for focusing on attitudinal (pre and post-choice) dynamics. The conceptualization of "perceived quality" as a discrepancy between expectations and experiences (expectancy-disconfirmation approach) raised a number of criticisms, mainly from the measurement point of view. A "performance-only" concept (cf. SERVPERF as propagated by Cronin and Taylor, 1994) seems to be clearly preferable in perceived quality field research. In a simulation environment like the ACM the experimenter need not care about the consumers' ability of correctly remembering their pre-purchase expectations after acquiring consumption experience. He is in control of modeling brand expectations and performance independently of when and how often they are measured. While perceived quality may be dispensable, the (dis)satisfaction construct is not. The ACM consumers pursue the expectancy-disconfirmation paradigm (Cardozo, 1965; Oliver and DeSarbo, 1988) by deriving (dis)satisfaction from (un)fulfilled product claims. The dynamic aspects of (dis)satisfaction are consistent with equity theory (Oliver and Swan, 1989a,b) regarding the adaptation of consumers' aspiration levels to the market reality (Trommsdorff, 1998). The gradual adaptation of expectations characterizes a smoothly evolving marketplace (Johnson, Anderson and Fornell, 1995). For achieving structural breaks in a simulation run exogenous shocks leading to a disruption in the consumers' preferences are admissible in the ACM.

Actually, the concept of (dis)satisfaction plays a central role in the consumer model. It is the key factor for conducting the joint marketing and production simulation experiments. The brand positions in the (latent) attitude space are governed by the competitors' communicative and technological actions and the consumers' reactions. The AFs are put under pressure to coordinate and harmonize their market communications and product improvement decisions, as two different organizational units are in charge of these processes. The AFs may conduct a "perceived quality" or a satisfaction study where they try to measure the gap between expectations and performance. The ACM provides this information about consumer (dis)satisfaction in a very realistic manner. It is worded in the consumers' fuzzy language the same way their latent brand attitudes are reflected by a redundant set of observable indicators for perceived attributes. A composite (unidimensional) measure of (dis)satisfaction complements this set of attribute-specific indicators.

The consumers on the artificial market experience product (dis)satisfaction when

the technical/functional product features (fail to) match the advertising claims. Hence they build up loyalty, or suffer from dissonance. Consumer theory never makes generalizable propositions on a measurement level higher than ordinal. For the simulation experiments a more precise specification is required to implement relationships such as "the higher the amount of dissatisfaction the lower the probability of repurchase". It is of paramount importance, however, to realize that a theory-driven gain in precision always excels an arbitrary parameterization. For incorporating the consequences of brand (dis)satisfaction the consumer model takes recourse to elementary "set theory". The concept of consideration sets with its numerous variants originates from Howard and Sheth's concept of the "evoked set" (1969), the individual consumer's group of purchasing alternatives comprising not more than 5 to 7 product brands. In the sequel consumer research invented a variety of "sets" (see the whole zoo of awareness, inert, inapt, consideration, or choice set concepts explained by Crompton, 1992 or Goodall, 1991).

Hauser and Wernerfelt (1990) summarize results on the "consideration set phenomenon" and admit that it "is critical to the predictive ability of quantitative models" (p. 393f.). The ACM artificial consumers resemble their real counterparts in forming brand sets according to their stage in the purchase-repurchase cycle. The consumers on the ACM do not keep records of their intrinsic brand choice probabilities. Rather than bookkeeping they like to follow very simple decision rules. They build consideration sets of equally attractive buying alternatives that exceed their aspiration levels in terms of (price weighted) brand attribute dimensions. A second filter works in absolute terms, as consumers discard a brand surpassing their reservation price for the product class. Consumption experience, either self-made or mediated through word-of-mouth, changes the consumers' composition of consideration sets. Thus learning effects are responsible for two dynamic phenomena: They lead consumers (1) to (temporarily) barring a brand from their consideration sets where it does not come up to the expectations solicited by advertising; or, (2) they make consumers maintain a high-performing brand in their sets even when competing brands advertise more attractive brand profiles. Fairly elaborate learning regimes for the consumer agents in the AE such as adjusted components of Thomas Brenner's "Variation-Imitation-Decision" (VID) model (1999, pp. 71–89) have been adopted. The consumers communicate with each other about their brand usage experience and also exhibit imitative behavior by recognizing other peoples judgments of product quality.

4 The ACM Micro Structure: Tracing the Individual Consumer

The ACM microstructure is easily apprehensible on the disaggregate level of purchasing and consumption by watching an individual consumer's trajectory through these processes. Figure 2 highlights the stages in the consumer's decision making. Particularly, it indicates the feedback loops which govern the adaptive behavior of the consumers on the ACM. The loops are "intra-personal", such as aspiration level adaptation or accumulating reactance against persuasive advertising, or "inter-personal" such as spreading word-of-mouth messages to fellow consumers. In a realistic set-up of an ACM the AFs are facing the rationality restrictions imposed on the consumer's

brand choice. Satisficing behavior instead of utility maximization prevails in the brand evaluation stage and ties are broken up by random selection on the disaggregate level of the individual consumer. In the post-choice evaluation, however, the consumers are relentless in comparing expectations and actual brand performance. In this stage they also reward over-fulfillment of advertising promises. This is consistent with the global objective of the Artificial Economy project. If one aims at analyzing the AF's policies of developing a coordinated approach to corporate planning, the consumers on the ACM should be particularly sensitive to a misfit in the AF's technology-marketing policies.

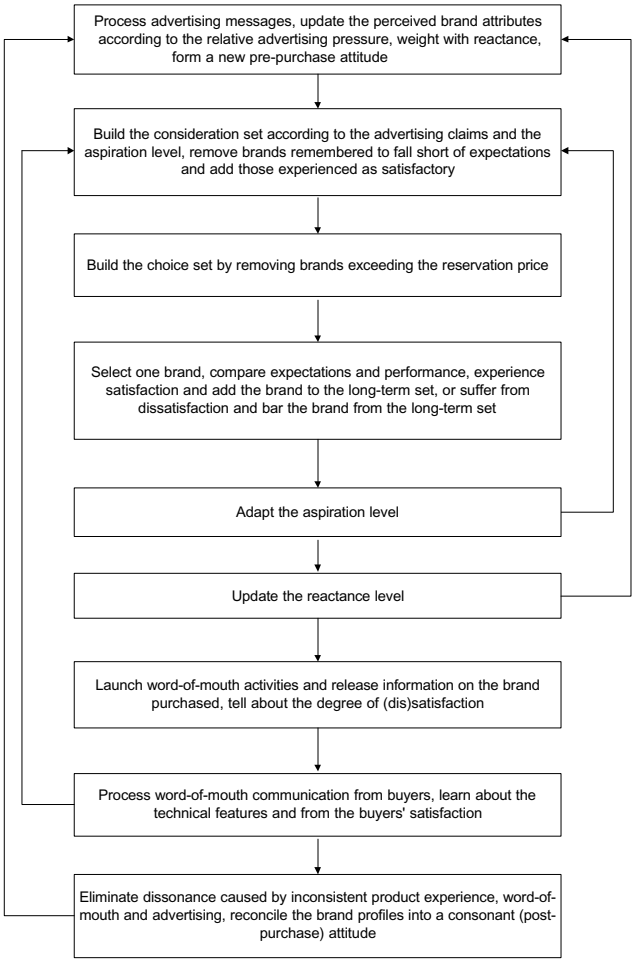

Figure 2: Communication, learning and decision-making on the ACM

It has been emphasized previously that it is the authors' ambition to conceive the ACM as parsimonious as possible and to avoid unnecessary simulation parameters. To accommodate all the empirical phenomena itemized in Figure 2 several modification are applied to the standard ideal-point model (cf. Myers, 1996; Hruschka, 1996). The consumers decide on the brands to enter their consideration sets according to a modified version of the ideal-point model, named the *Unidirectional Ideal-Point Model with threshold* (see Figures 3 and 4). The UIPM combines a spatial approach for representing attitudes with aspiration level learning.

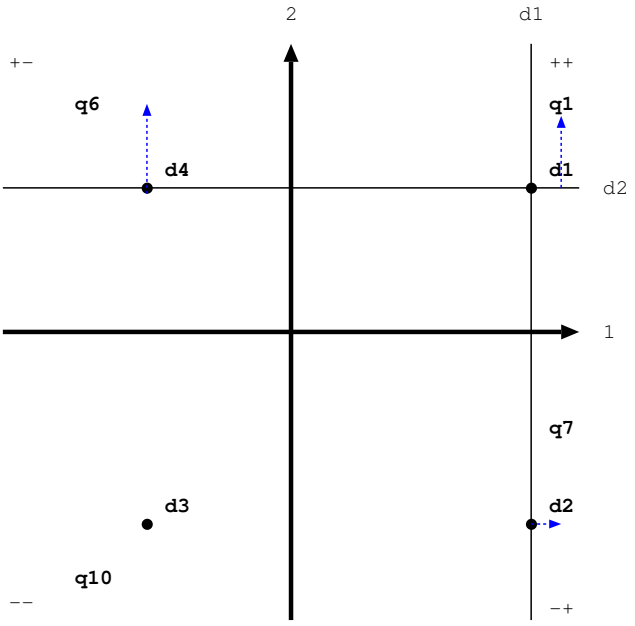

Figure 3: Unidirectional ideal-point model with threshold $d_1 = d_2 = 0$ (city-block metric)

Imagine two (price weighted) attribute dimensions $d_1$and d_2 in a latent attitude space with a city-block metric. The attitude scores increase from left to right and from bottom to top. Focus on the right upper quadrant first, indicated by $++$. The ideal point $\mathbf{d}_1$ denotes an individual or group-specific aspiration level thus introducing preferences into the brand space. Contrary to the conventional ideal-point model there are no spherical, elliptical or diamond-shaped iso-preference curves surrounding an ideal point. Over-fulfillment of the aspiration level neither increases nor decreases a brand's likelihood of entering the consideration set. This means that the consumers in the ACM are unaware of any product attributes exhibiting an inverse u-shaped utility function. They are more likely to accept a brand the more it approaches their aspiration levels "from below". Brands exceeding that level are equally welcome (unless overpriced) and induce the consumers to raise their aspirations.

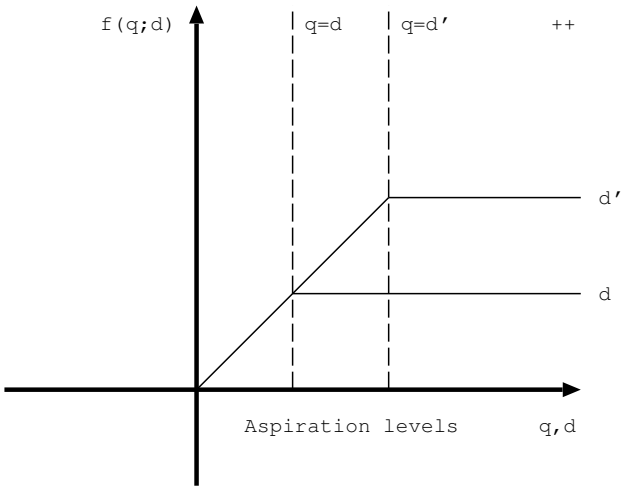

Figure 4: Moving along an attitude dimension

Figure 4 shows how the attractiveness of a brand develops moving along an attitude dimension. It stays meaningless before it crosses the irrelevance-relevance threshold and then gains more and more strength of belief until it reaches the aspiration level needed for being considered an attractive buying alternative. Unidirectionality is not an illegal simplification. It captures one of several ACM features that may be interpreted as criteria of bounded rationality. Any product attribute can be reformulated to comply with this simple cognitive algebra. (If you like your tea (beer) neither lukewarm nor boiling (frozen) you look for claims promising "adequate temperature" that will move your expectations closer to your aspiration level.) Unidirectionality is in conformity with the measurement model of binary indicators for the latent attitude dimensions. To assume a realistic usage of binary attributes, the "zero" answer should be able to capture the composite meaning of "does not fit" and "irrelevant". If a consumer rates a brand and more indicators of an attitude dimension get unity values, the brand's position shifts toward positive infinity. Somewhere in this shift the aspiration level ought to become relevant for determining the consumer's preferences where the origin of the attitude space is a natural threshold.

In principle, an adaptive aspiration level may also be introduced into the conventional, i.e. "n-directional", ideal-point model. If the attitude space exhibits a city-block metric the aspiration level does no longer collapse into the ideal point. Instead, it surrounds the ideal point in a diamond-shaped iso-preference curve. The brands inside the "diamond" area are equally eligible for consideration. This results in two separate dynamic effects: (1) the shifting of the ideal point $\mathbf{d}_t \rightarrow \mathbf{d}_{t+1}$ when expectations rise, and (2) the changing of the "satisficing threshold" when tolerance grows or shrinks.

In Figure 3 brand $\mathbf{q}_1$ will enter the consideration set of a consumer pursuing an aspiration level (ideal point) $\mathbf{d}_1$ in the $++$ quadrant. Because of the dimension relevance threshold at zero (origin of the coordinate system) the situation is different in

the other three quadrants. For a consumer with an aspiration level $\mathbf{d}_3$ neither of the two dimensions is relevant for his brand preferences. For a person characterized by $\mathbf{d}_2$ in the second and $\mathbf{d}_4$ in the fourth quadrant only one of either d_1 or d_2 influences the consideration set. Movements along the relevant dimensions (see the arrows for the most preferred brands) change the consumer's willingness to consider this brand. As one brand at least exceeds any of the aspirations consumers will learn to raise their expectations.

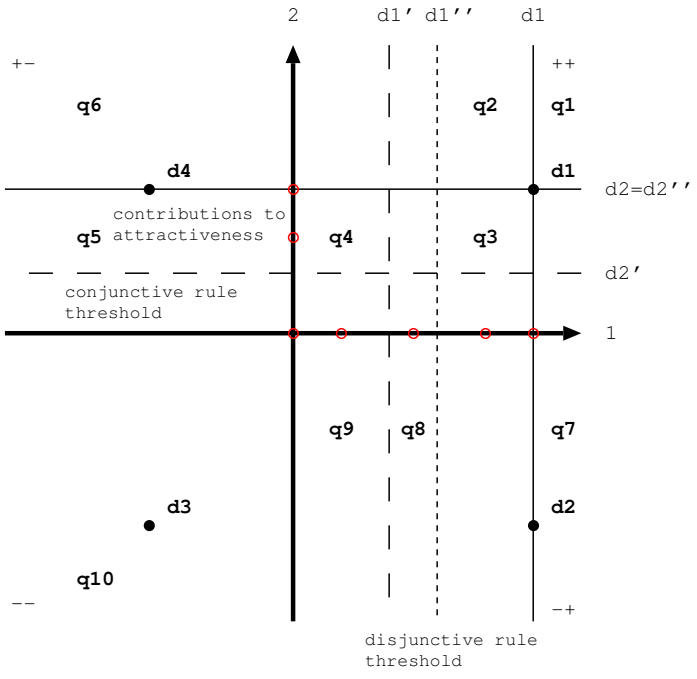

Figure 5: Noncompensatory schemes in the unidirectional ideal-point model (city-block metric)

The UIPM follows a compensatory approach. However, there is ample empirical evidence (Bettman, Luce and Payne, 1998) that suggests to relax the rather strong assumption of a fully compensatory scheme. The brand's excellent position in one attitudinal dimension may offset poor performance in another dimension once it has passed the relevance threshold. In the consideration stage, however, over-fulfillment by exceeding the aspiration level in one dimension does not further contribute to make the brand enter the set of acceptable buying alternatives or to compensate for partial failure in another dimension. A random choice occurs if the acceptable brands also charge identical prices below the consumer's reservation price. The consumer's inability or unwillingness to make meticulous judgments in the pre-choice stage is one concrete aspect of the operationalization of "bounded rationality". In the post-purchase stage the consumption experience enforces a higher level of awareness for

the brand attributes including the evaluative dimensions that may have been ignored or neglected previously. A larger amount of cognitive effort is subjectively justified. Consumer learning in the post-choice phase benefits from a richer and more reliable input than just advertising or word-of-mouth.

Noncompensatory choice rules are needed to achieve a realistic mixture of decision styles in the consumer population. Figure 5 demonstrates how a conjunctive or a disjunctive decision rule (Hruschka, 1996; Roberts and Lilien, 1993) conforms with the aspiration level concept. An additional satisfaction threshold is needed to implement these rules. Under the conjunctive regime the consumer expects an acceptable brand to offer a (modest) minimum performance in each relevant (i.e. with aspiration > 0) evaluative dimension. The disjunctive rule decider requests a (fairly high) performance in at least one relevant dimension. Thus the "ideal point" marks the consumer's preferences and sets the aspiration target; the satisfaction threshold controls the brands' entrance into the consideration set. In his post-purchase reasoning the noncompensatory consumer is likely to increase his cognitive effort. The satisficing principle is welcome to facilitate choice. Once the choice has been made much more factual knowledge assists in readjusting what may be desired (aspiration level) and what is satisfactory for becoming a new choice alternative. Both, aspiration level and satisficing threshold, are subject to learning with a nice option of convergence when the brand comprehension and the consumers' capability of discriminating among brands evolve.

The following composition of consideration sets results from the aspiration levels together with the UIPM, conjunctive and disjunctive decision styles in Figure 5:

Table 1: Composition of consideration sets by decision styles.

Ideal point	UIPM	conjunctive	disjunctive
$\mathbf{d}_1$	$\mathbf{q}_1$	$\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3$	$\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_6, \mathbf{q}_7$
$\mathbf{d}_2$	$\mathbf{q}_1, \mathbf{q}_7$	$\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_7, \mathbf{q}_8$	$\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_7$
$\mathbf{d}_3$	all	all	all
$\mathbf{d}_4$	$\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_6$	$\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_3, \mathbf{q}_4, \mathbf{q}_5, \mathbf{q}_6$	$\mathbf{q}_1, \mathbf{q}_2, \mathbf{q}_6$

5 A Formal Description of the Discrete-Time Model

In the following sections we present a formalized view of the simulation environment, on the one hand by discussing the assumptions and constructs of the ACM, and on the other, by giving proper interface definitions and an overview of the model's calibration needs. Although the latter is closely linked to the definition of market scenarios, we defer such a discussion to future work. Let us start with an overview of the "technical" assumptions of the ACM:

- In a simulation the set of consumers and the set of firms participating in the market is constant.

- The firms operate on a single product class given a fixed technological environment, in the sense that there is a bound to product improvement.
- *Market time* is thought of as synchronization points for information exchange, as well as periods of information processing. A *period* is divided into (consecutive) steps of information processing.
- The firms are assumed to act on segments of consumers. A *segment* is targeted with exactly one product (brand), and the product is available only to the consumers of a segment. Alternatively, we may assume inseparable markets where all the products are visible and available for all the consumers. A *product* is a bundle of technical (feature), advertising (claim, budget), and price information.
- The term *brand*, in the sense of the name (identifier) of a product (cf. Kotler, 1986) is meant to represents the firm's long-term concept of a product, which we use synonymously with product. We assume a fixed set of brands and that each firm's set of brands is constant for a simulation.
- A consumer responds with the choice of exactly one product, and is periodically surveyed on the perceptions and evaluations of the products on the market.
- A market-wide *reservation price* and a reference price for advertising and the production input factors must be set.

In the following discussion we require an extensive notation: consumer, product and attribute indices will be used as one block, followed by comma separated step and time indices. Following common usage, vectors (scalars) will be in lowercase bold (normal) type, matrices in uppercase bold. In a matrix context vectors are always of "column" type but for the scalar product of two vectors we use the "dot" notation. Variables that denote memory constructs will have a bar to express that they are (exponential) time averages, and variables of the interface level will be set in a different type.

6 Attitude Formation

Transformation: Initially, let us denote the following non-linear mapping from the reals to the unit interval as the *squashing function*, and assume in the context of vectors and matrices it is a function of their elements, i.e.

$$\varphi(x) := \frac{1}{1+\exp(-x)}, \quad x \in \mathbb{R}. \tag{1}$$

Note, we will also need the linear transformation $2\varphi(x) - 1$ to the interval $(-1, +1)$, which is equivalent to the hyperbolic tangent function. Further, we agree to denote the inverse function by $\varphi^{-1}(\cdot)$.

Basic Model: Let $\mathbf{q}_{ij,*,t} \in \mathbb{R}^L$, denote the ith consumer's *attitude* to product j (the position in the latent space of attitudes), in the $*$th step of attitude formation in period t, and $\mathbf{p}_{ij,*,t} \in (0,1)^K, K \geq L$ the consumer's (manifest) *perception* of the

product. On the one hand, a consumer's perception changes in response to advertising, $\mathbf{p}_{ij,0,t} \rightarrow \mathbf{p}_{ij,1,t}$ (pre/post-advertising perception), and on the other, the attitude changes in response to product experience, $\mathbf{q}_{ij,0,t} \rightarrow \mathbf{q}_{ij,1,t}$ (pre/post-purchase/word-of mouth attitude). Attitudes and perceptions are related via the (noisy) non-linear mappings:

$$\mathbf{q}_{ij,0,t} := \mathbf{A}_i \varphi^{-1}(\mathbf{p}_{ij,1,t}) + \boldsymbol{\epsilon}_{i,t}, \quad \mathbf{A}_i \in \mathbb{R}^{L \times K}, \tag{2}$$

$$\mathbf{p}_{ij,2,t} := \varphi(\mathbf{B}_i \mathbf{q}_{ij,1,t}), \quad \mathbf{B}_i \in \mathbb{R}^{K \times L}. \tag{3}$$

Thus, a product's position in attitude space is a (usually information reducing) projection of a consumer's perception, and vice versa. Note that the linear mappings are meant to differ at most between groups of consumers, as well as the independent noise components $\epsilon_{il,t} \sim N(0, \sigma_i)$ with $\sigma_i \ll 1$. Finally, attitude formation is 'embedded' in time by $\mathbf{p}_{ij,0,t+1} := \mathbf{p}_{ij,2,t}$, and we can summarize the cycle of the basic model by the following steps of information processing

$$\mathbf{p}_{ij,0,t} \rightarrow \mathbf{p}_{ij,1,t} \rightarrow \mathbf{q}_{ij,0,t} \rightarrow \mathbf{q}_{ij,1,t} \rightarrow \mathbf{p}_{ij,2,t} = \mathbf{p}_{ij,0,t+1}.$$

Let us assume $\mathbf{A}_i \mathbf{B}_i = \mathbf{E}$ (the $L \times L$ identity matrix). Obviously, this is a simplification of the unknown effect of information reduction: in fact, if we assert to know, say, $\mathbf{A}_i$ we would have to make a choice with respect to $\mathbf{B}_i$ each time the attitude changes. By the above assumption we hypothesize that the (transformed) perceptions live in a linear subspace, and if advertising tries to influence consumers 'out' of this space they 'reconcile' to the 'closest' perception in this subspace. The corresponding vector is known to be the orthogonal projection onto the subspace of $\mathbf{A}_i$[1], and following from above, this is the linear mapping $\mathbf{B}_i \mathbf{A}_i$. and, we see that the effect of advertising is just $\mathbf{B}_i \mathbf{A}_i (\varphi^{-1}(\mathbf{p}_{ij,1,t}) - \varphi^{-1}(\mathbf{p}_{ij,0,t}))$. Obviously, there is no effect if $\mathbf{p}_{ij,1,t}$ is orthogonal to the subspace. More general, the angle of this vector with its projection gives us a measure of the loss in effectiveness due to 'disorientation'. We conjecture the latter becomes more likely if $L \ll K$, and advertising does not bother to learn the consumers' perceptual redundancies.

Techitude: A product's *technical features* (attributes) $\mathbf{p}_{j,t} \in (0,1)^{\tilde{K}}$, are condensed into a position in attitude space in the same way as above. Let $\mathsf{c}_{ij,t} \in \{0,1\}$ denote if the product was chosen, and thus the information is available, assume the same type of noise as above, and let us define the position as

$$\tilde{\mathbf{q}}_{ij,0,t} := \begin{cases} \tilde{\mathbf{A}} \varphi^{-1}(\mathbf{p}_{j,t}) + \tilde{\boldsymbol{\epsilon}}_{i,t}, \quad \tilde{\mathbf{A}} \in \mathbb{R}^{L \times \tilde{K}} & : \quad \mathsf{c}_{ij,t} = 1 \\ \mathbf{q}_{ij,0,t} & : \quad \text{else} \end{cases}. \tag{4}$$

Note, the step index indicates the situation after purchase but before word-of-mouth communication. By the latter a consumer may obtain additional information (see Equation (34) in Section 10), but if he does not, his attitude is pre-defined to remain the same (see Equation (5)).

[1] From the singular value decomposition $\mathbf{A} = \mathbf{LWM}$ we see that a perception is projected onto the coordinate system of the manifest space $\mathbf{M}$, the coordinates are weighted by $\mathbf{W}$, and then are projected onto the coordinate system of the latent space $\mathbf{L}$.

Reconciliation: A consumer adjusts his current expectation (pre-purchase/post-advertising attitude) to new information on the performance of a product (either acquired directly by consumption (usage), or indirectly by word-of-mouth communication), but only on relevant attitudinal dimensions. Let us denote a consumer's current *aspiration level* on the lth dimension as $\bar{d}_{il,t} \in \mathbb{R}$, define zero to be the *threshold of relevance*, and the rate of adjustment as $\eta_{i1} \in (0, 1]$ (depending on a consumer's involvement), and let us define the change in attitude as

$$q_{ijl,1,t} := \begin{cases} \eta_{i1}\tilde{q}_{ijl,1,t} + (1-\eta_{i1})q_{ijl,0,t} & : \quad \bar{d}_{il,t} > 0 \\ q_{ijl,0,t} & : \quad \text{else} \end{cases} . \tag{5}$$

Note, now the step index indicates the situation after word-of-mouth communication.

Desire: The consumers are assumed to adapt their aspiration levels $\bar{\mathbf{d}}_{i,t} \in \mathbb{R}^L$ to their current market induced desires which are as strong as 'reasonably possible'. Further there may be (temporary) external influences to change a desire $\mathbf{d}_{i,t} \in \mathbb{R}^L$. Let $\eta_{i4} \in [0, 1]$ denote the rate of adaptation to the current desires (depending on involvement), $\mathbb{J}$ the set of products, and let us define the change in aspiration as

$$\bar{\mathbf{d}}_{i,t+1} := \eta_{i4} \max_{j \in \mathbb{J}} (\mathbf{q}_{ij,0,t}) + (1-\eta_{i4})(\bar{\mathbf{d}}_{i,t} + \mathbf{d}_{i,t}). \tag{6}$$

Note, the market induced desires are advertising biased. Thus, if we want to model a more 'consolidated' view of the market, as suggested in Section 4, we can use the 'reconciled' attitudes that include information obtained by word-of-mouth, i.e. $\tilde{\mathbf{q}}_{ij,1,t}$.

For ease of notation, let us define the *threshold indicator* function, and assume that in the context of vectors and matrices it is a function of their elements, i.e.

$$\psi(x) := \begin{cases} 1 & : \quad x > 0 \\ 0 & : \quad \text{else} \end{cases} , \quad x \in \mathbb{R}. \tag{7}$$

(Dis)satisfaction: We assume that the performance of a product is compared with the pre-purchase expectation, which leads to the overall feeling of *(dis)satisfaction*. Let us assume the intensity of this feeling depends on a consumers involvement $\eta_{i2} \in (0, 1]$, and is subject to saturation effects, i.e.

$$s_{ij,0,t} := \begin{cases} 2\varphi\left(\eta_{i2}\psi\left(\bar{\mathbf{d}}_t\right) \cdot \left(\tilde{\mathbf{q}}_{ij,0,t} - \mathbf{q}_{ij,0,t}\right)\right) - 1 & : \quad \mathsf{c}_{ij,t} = 1 \\ 0 & : \quad \text{else} \end{cases} . \tag{8}$$

Note, first if no information is available because the product was not chosen, $\mathsf{c}_{ij,t} = 0$, the feeling is 'neutral'. Second the differences in performance and expectation are noticed only on relevant attitudinal dimensions, and they are compensable. Third, we transform by the hyperbolic tangent function (see Equation (1)), because the feeling has a direction which entails specific responses. Finally, the step index indicates the situation before word-of-mouth communication.

The responses to the experience of (dis)satisfaction are threefold: first, past (current) experience can influence the current (future) choice process(es) to the effect of a different composition of the set of products considered (see Section 9), second, it determines the *reactance* to advertising which dampens the change in the perception

of a product (see Section 7), and third it determines the propensity to communicate by word-of-mouth (see Section 10).

After word-of-mouth a consumer's feeling of (dis)satisfaction adopts the current intensity and direction if the intensity exceeds the remembered level, but otherwise the memory decays. Let $\eta_{i3} \in [0,1)$ denote the persistence of the memory (depending on involvement), and let us define the change in memory as

$$\bar{s}_{ij,t+1} := \begin{cases} s_{ij,1,t} & : \quad |s_{ij,1,t}| > |\bar{s}_{ij,t}| \\ \bar{s}_{ij,t}\eta_{i3} & : \quad \text{else} \end{cases} . \tag{9}$$

Note, a consumer ignores a contradictory experience of lower intensity, but due to the memory decay he will not persist in a contradiction for long. Alternatively, asymmetric conditions for involvement dependent adjustments could be considered, e.g., $-s_{ij,1,t} > (1-\eta_{i3})|\bar{s}_{ij,t}|$ would model a sensitivity to dissatisfaction which is the higher the higher the involvement (persistence).

Reactance: The arousal of reactance is confined to the feeling of dissatisfaction but a consumer attempts to avoid contradictions, i.e.

$$r_{ij,t+1} := |\min(0, \min(s_{ij,1,t}, \bar{s}_{ij,t}))|. \tag{10}$$

Satitude: Let us assume that the experience of (dis)satisfaction with a product is also measurable in the space of perceptions — that corresponds to expectation-performance indicators based on the disconfirmation approach to perceived quality measurement (see Section 8). Thus we assume that attitudinal differences are processed (projected) in the same way as attitudes, but on irrelevant dimensions a consumer's feeling is either "neutral" or "irrelevant". Let us define the perceived (latent) (dis)satisfaction as

$$\begin{aligned} \grave{\mathbf{p}}_{ij,t} &:= 2\varphi(\mathbf{B}_i \grave{\mathbf{q}}_{ij,t}) - \mathbf{1}, \\ \grave{q}_{ijl,t} &:= \begin{cases} \tilde{q}_{ijl,1,t} - q_{ijl,0,t} & : \quad \bar{d}_{ijl,t} > 0 \\ 0 & : \quad \text{else} \end{cases} . \end{aligned} \tag{11}$$

Note, the transformation is the same as in Equation (8). For convenience we may refer to the perceived (intensity of) (dis)satisfaction as a *satisception*.

7 Dynamics of Perceptions

Let us discuss the manifest level of the model of attitude formation in this section. Technical features and perceived attributes, were introduced to live on the interval $(0,1)$. This is on the one hand a convenience for modeling technical improvement, and on the other, in the case of advertising, a necessity (see below). Further, we can postulate parsimonious measurement models, if we interpret these variables as probabilities (see the next section).

Response: In response to advertising consumers change their beliefs (perceptions), but such change is subject to saturation effects, and therefore we need bounded variables. In the following we will use the subscripts introduced in the previous section with equal meaning. Let $m_{ijk,t} \in (-1,1)$ denote the *impact* of the kth claim for

the jth product on the ith consumer, ϑ_0 the persistence of belief in the case of a zero or negative impact, and let us define the change in belief as

$$p_{ijk,1,t} := \begin{cases} p_{ijk,0,t} + (1 - p_{ijk,0,t})(1 - r_{ij,t})m_{ijk,t} & : \quad m_{ijk,t} > 0 \\ p_{ijk,0,t}\vartheta_0 & : \quad \text{else} \end{cases} . \tag{12}$$

Note, first, in case of dissatisfaction advertising is met with scepticism, i.e. the impact is dampened by reactance (see Equation (10)). Second, the impact is proportional to the current gap (unused potential) of belief $1 - p_{ijk,0,t}$, and therefore the stronger the lower the current belief. Conversely, the higher the current belief the higher the loss in credibility: $-p_{ijk,0,t}(1 - \vartheta_0)$. Finally, remember that the pre-advertising belief of the current period is just the post-word-of-mouth belief of the previous period, i.e. $p_{ijk,0,t} = p_{ijk,2,t-1}$.

Impact: We assume that the impact of advertising is decomposable and subject to saturation effects. On the one hand, let $\mathsf{b}_{j,t} \in \mathbb{R}_0^+$ be the advertising budget for product j, and $\mathfrak{n}_{j,t}$ the number of consumers addressed with the same message (size of a segment). On the other let $\mathfrak{m}_{jk,t} \in \{0, 1\}$ indicate if the kth claim is contained in the jth message. Finally, let $\mathsf{s}_{ij,t} \in \{0, 1\}$ denote if consumer i is addressed with the jth message, and remember the threshold indicator function $\psi(\cdot)$ from Equation (7). Note that replacing an index by a dot we use as a shorthand for summation. A claim's impact on consumer i depends on its *intensity* which is determined by a number of attention effects, and the responsiveness of a consumer (see Equation (14) below), i.e.

$$m_{ijk,t} := \phi \left(\mathsf{s}_{ij,t}\mathfrak{m}_{jk,t} \left(\frac{1}{\max\left(1, \sum_{j' \in \mathbb{J}} \psi\left(\mathsf{s}_{ij',t}\mathfrak{m}_{j'\cdot,t}\right)\right)} \right)^{\vartheta_1} \right.$$
$$\left. \left(\frac{1}{\max(1, \mathfrak{m}_{j\cdot,t})} \right)^{\vartheta_2} \left(\frac{\max_{k' \in \mathbb{K}}(\sum_{j' \in \mathbb{J}} \mathsf{s}_{ij',t}\mathfrak{m}_{j'k',t})}{\max(1, \mathsf{s}_{ij,t} \sum_{j' \in \mathbb{J}} \mathsf{s}_{ij',t}\mathfrak{m}_{j'k,t})} \right)^{\vartheta_{i3}} \frac{\mathsf{b}_{j,t}}{\mathfrak{n}_{j,t}^{\vartheta_4}\mathsf{b}_0}; \vartheta_{i5} \right) . \tag{13}$$

Note, first, the more messages compete for the attention of a consumer the less the effect of a message, where no claim at all (or a zero budget) does not qualify as a competing message. Second, the fewer claims a message contains the sharper its focus and consequently its effect. Third, the fewer messages that contain a claim the higher the focus on that claim. For the choice of parameters we suggest $0 \leq \vartheta_{i3} \ll \vartheta_1 < \vartheta_2 \leq 1$, where we think that modeling diversion of attention effects is mandatory. Fourth, the scale factor $\mathsf{b}_0 \in \mathbb{R}^+$ models the reference price (cost) of a contact per consumer, and it is assumed that the consumers of a segment are contacted with the same intensity (frequency). Fifth, the parameter $\vartheta_4 \in [0, 1]$, models economies of scale effects that depend on the number of consumers addressed (with the same message). Finally, the impact of a claim depends on the responsiveness of a consumer, denoted by $\vartheta_{i5} \in (0, 1]$.

Responsiveness: The following function captures the idea of a thresholded and saturable response to a claim's intensity:

$$\phi(x; \vartheta) := \vartheta - \exp(-\vartheta x), \quad x \in \mathbb{R}_0^+, \vartheta \in (0, 1]. \tag{14}$$

Note, first, that an increase in a claim's intensity, ceteris paribus by increasing the advertising budget, leads to a smaller increase in impact, $\frac{\partial^2 \phi}{\partial x} = -\vartheta^2 \exp(-\vartheta x) < 0$, but the proportions of two claims' intensities are smaller than their proportions of impacts if the latter are positive: $x_2/x_1 \leq \phi(x_2)/\phi(x_1), x_2 \geq x_1, \phi(x_1) \geq 0$. Second, the maximum impact is bounded by ϑ (saturation level), which at the same time determines the threshold of ineffective claim intensity. Note, there is a level of responsiveness, such that $\vartheta = \exp(-\vartheta) \approx 0.56714$ and a 'budget' per consumer equal the reference price (cost) has just zero impact (assume a single claim, a single product and no economies of scale effects).

We conclude this section by mentioning that a similar interpretation holds for the current level of a technical attribute (feature) $\mathrm{p}_{ij,t}$: there is an unused potential for improvement (of an existing technology) $1 - \mathrm{p}_{ij,t}$. Note, for an evolving or disruptive technology we suggest to model a continuous change or sudden structural breaks of the set of dimensions, but this is part of future work.

8 Measuring the State of a Consumer

In this section we present a general concept for measuring the internal states of the consumers. Remember, we assume that a consumer is regularly surveyed on his perception of and (dis)satisfaction with a product, and that this information is available to the firms.

Perceptions: The strength of belief in product attributes is measured on a binary scale. Let $\mathrm{x}_{ijk,t} \in \{0, 1\}$ denote the ith consumer's stated belief, i.e. the agreement to the item describing the kth attribute of the jth product, and assume the variable $p_{ijk,2,t}$ is the probability to 'agree'. Thus, belief measurements are binomially distributed random variables:

$$\Pr(\mathrm{X}_{ijk,t} = 1) := p_{ijk,2,t}. \tag{15}$$

Note, according to the interpretation we suggest below, a zero (one) measurement can be thought of as indicating a 'low' ('high') belief, but given a single measurement this is 'pointless' information.

Satisfaction/Satisception: Similarly, the intensity and direction of the overall and attribute specific (dis)satisfaction of a consumer, is measured on a five-point bipolar scale. Let $\mathrm{y}_{ij,t}, \mathrm{z}_{ijk,t} \in \{-2, -1, 0, 1, 2\}$ denote the ith consumer's stated overall (dis)satisfaction with product j, and with the product's kth attribute, respectively. Now, remembering the transformation according to Equation (1), let us return to the unit interval, i.e. $s'_{ij,1,t} := (1 + s_{ij,1,t})/2$ and $\grave{p}'_{ijk,t} := (1 + \grave{p}_{ijk,t})/2$, and assume we repeat a binary measurement with these probabilities four times and report the sum minus two as the scale values. Thus, (dis)satisfaction measurements are modeled as Bernoulli distributed random variables:

$$\Pr(\mathrm{Y}'_{ij,t} = z) := \mathrm{B}(s'_{ij,1,t}, 4), \tag{16}$$

$$\Pr(\mathrm{Z}'_{ijk,t} = z) := \mathrm{B}(\grave{p}'_{ijk,t}, 4), \tag{17}$$

where $z \in \{0, 1, \ldots, 4\}$, and the variables are transformed to $\mathrm{y}_{ij,t} := \mathrm{y}'_{ij,t} - 2$, and

$z_{ijk,t} := z'_{ijk,t} - 2$, such that the scales are "polarized" at zero. Remember, zero indicates a "neutral", as well as an "irrelevant" response (compare Equation (11)).

9 Choice of a Product

In the present section we discuss the process of consideration and choice set formation. We will assume one basic scheme, which can be varied by putting more emphasis on price or product 'quality'. The other modeling choice concerns different decision styles. Let us begin with the latter.

Utility: Under the *modified ideal point* decision rule the total attractiveness, or 'utility' of a product, is the sum of its contributions on the attitudinal dimensions, i.e.

$$u_{ij,t} := \mathbf{1} \cdot \max\left(0, \min\left(\mathbf{q}_{ij,0,t}, \bar{\mathbf{d}}_{i,t}\right)\right). \tag{18}$$

Note that a utility of zero would be ambiguous if a zero threshold of relevance could be 'fulfilled'. Although this is in fact only cosmetic, we have defined the threshold indicator function in this sense (see Equation (7)).

For a *conjunctive* decision rule we assume that the *satisfaction levels* are lower than the aspiration levels. Let $\beta_1 \in [0, 1)$ denote the lowering factor, $\mathbf{1}$ a $l \times 1$ vector of ones, and let the utility indicate if there is not a single relevant dimension where the desired level is not satisfied, i.e.

$$u_{ij,t} := \psi\left(\mathbf{1} \cdot \bar{\mathbf{d}}_{i,t}\right) - \psi\left(\psi\left(\bar{\mathbf{d}}_{i,t}\right) \cdot \left(\mathbf{1} - \psi\left(\mathbf{q}_{ij,0,t} - \beta_1 \bar{\mathbf{d}}_{i,t}\right)\right)\right). \tag{19}$$

Note, if none of the dimensions is relevant there is nothing to indicate. This is in compliance with the first definition of utilities.

For a *disjunctive* decision rule we assume, again, a lowering of the aspiration levels, $\beta_1 \ll \beta_2 \leq 1$, but the consumers further concentrate on important dimensions, i.e. the satisfaction levels are defined as

$$d^*_{il,t} := \begin{cases} \beta_2 \max_{l' \in \mathbb{L}}(\bar{d}_{il',t}) & : \quad \bar{d}_{il,t} \geq \beta_2 \max_{l' \in \mathbb{L}}(\bar{d}_{il',t}) \\ 0 & : \quad \text{else} \end{cases}. \tag{20}$$

Now, let the 'utility' indicate if there is at least one relevant and important dimension where the desired level is satisfied, i.e.

$$u_{ij,t} := \psi\left(\psi\left(\mathbf{d}^*_{i,t}\right) \cdot \psi\left(\mathbf{q}_{ij,0,t} - \mathbf{d}^*_{i,t}\right)\right). \tag{21}$$

Note again, if none of the dimensions is important then nothing is to be indicated. Further, if there are no marked differences in the aspiration levels then all relevant dimensions will be considered important and thus 'satisfiable'.

Ranking: In the context of (initial) conjoint-measurements (product development) we need a preference ordering of the products instead of choice information. Let $\mathsf{o}_{ij,t} \in \{1, 2, \ldots, |\mathbb{J}|\}$ denote the rank number the ith consumer assigns to product j, and assume he arrives at such a number by comparing the utility of a product against all the others', i.e.

$$\mathsf{o}_{ij,t} := |\mathbb{J}| - \sum_{j' \in \mathbb{J} \setminus j} \psi(u_{ij,t} - u_{ij',t}). \tag{22}$$

Note, ties in the utilities are preserved, and as a consequence the rank numbers need not be contiguous. We think, this is an appropriate assumption given the possibility of noncompensatory evaluation styles.

Choice: Let us now present the basic scheme of consideration set formation. Initially, we may assume that a firm is able to exclude a consumer from the purchase of its own competing products (brands) which it offers to a different segment. Let $\iota \in \{0, 1\}$ indicate the assumption of market *separability*, remember that $\mathsf{s}_{ij,t}$ indicates the segment membership, and let us define the set of available products as

$$\mathbb{J}_{i,t} := \{j : \mathsf{s}^{\iota}_{ij,t} = 1, j \in \mathbb{J}\}. \tag{23}$$

We assume, a consumer makes a pre-selection among the alternatives by considering only the products that are priced below the consumer's reservation price, where we denote the latter by w_{i0} and by $w_{j,t}$ the price demanded for product j, i.e.

$$\mathbb{J}_{i,0,t} := \{j : w_{j,t} \leq w_{i0}, j \in \mathbb{J}_{i,t}\}. \tag{24}$$

Note if this result is in an empty set a consumer ignores the reservation price $\mathbb{J}_{i,0,t} = \mathbb{J}_{i,t}$. Next he excludes products that are remembered as highly dissatisfying in past periods (see Equation (9)). Let $o_{ij,t} \in \{-1, 0, 1\}$ be a stochastic indicator, with $\Pr(O_{ij,t} = -1) := \max(0, -\bar{s}_{ij,t})$, and define the reduced set as

$$\mathbb{J}_{i,1,t} := \mathbb{J}_{i,0,t} \setminus \{j : o_{ij,t} = -1, j \in \mathbb{J}_{i,0,t}\}. \tag{25}$$

Note, if this results in an empty set, a consumer is assumed to ignore the feeling of dissatisfaction $\mathbb{J}_{i,1,t} = \mathbb{J}_{i,0,t}$. Alternatively, he may 'trade' dissatisfaction against the violation of reservation price, if possible. Nevertheless, if the reservation price has the implicit interpretation of a consumer's budget constraint considering overpriced products is only a last resort, as assumed above. Next, the utilities from above come into play for further set reduction, i.e.

$$\mathbb{J}_{i,2,t} := \left\{ j : u_{ij,t} \geq \max_{j' \in \mathbb{J}_{i,1,t}} (u_{ij',t}), j \in \mathbb{J}_{i,1,t} \right\}. \tag{26}$$

Then the set is 'enlarged' by products that have been (told to be) highly satisfactory in the past. Let $\Pr(O_{ij,t} = 1) := \max(0, \bar{s}_{ij,t})$, and define the enlarged set as

$$\mathbb{J}_{i,3,t} := \mathbb{J}_{i,2,t} \cup \{j : o_{ij,t} = 1, j \in \mathbb{J}_{i,0,t}\} \tag{27}$$

Note, the latter set we refer to as the *long term* consideration set because, in effect, these products do not take part in the utility based reduction step. Next, this set is reduced to the products with minimum price, where we denote by $\beta_{i,3} \in [0, 1]$ the price sensitivity of a consumer, i.e.

$$\mathbb{J}_{i,4,t} := \left\{ j : \beta_{i,3} w_{j,t} \leq \min_{j' \in \mathbb{J}_{i,3,t}} (w_{j',t}), j \in \mathbb{J}_{i,3,t} \right\}. \tag{28}$$

Finally, a consumer chooses, with probability $\frac{1}{|\mathbb{J}_{i,4,t}|}$, one among the products in the *choice set*. Remember, $\mathsf{c}_{ij,t} \in \{0, 1\}$ indicates if consumer i has chosen product j.

For illustration of the utility maximizing (satisficing) step of set formation see Figure 5 and Table 1 in Section 4.

In the case of a decision process based on price weighted utilities, the above scheme needs two modifications: first, we have to consider that the utilities of all the alternatives could be zero. Thus, let us define the price weighted utilities as follows

$$u'_{ij,t} := \begin{cases} \frac{u_{ij,t}}{w_{j,t}} & : \quad \exists j' : u_{ij',t} > 0, j' \in \mathbb{J}_{i,1,t} \\ \frac{1}{w_{j,t}} & : \quad \text{else} \end{cases} . \tag{29}$$

Second, we assume that a further reduction to minimum priced products is omitted.

10 Word-of-mouth communication

In this section we discuss the modeling of word-of-mouth activities of consumers. We think this is an important part of the model, because it will be interesting to see the difference in market response if the consumers interact more or less frequently and thus the information base is different.

First, let the probability that consumer $i \in \mathbb{I}$ contacts consumer $i' \neq i$, which is indicated by the variable $h_{ii',t} \in \{0,1\}$, depend on $\kappa_i \in \{0,1,\ldots,|\mathbb{I}|-1\}$, the average number of contacts of a consumer, i.e.

$$\Pr(H_{ii',t} = 1) := \frac{\kappa_i}{|\mathbb{I}|-1}. \tag{30}$$

Note that the probabilities are meant to differ at most between groups of consumers. Alternatively the contact structure may be fixed over some time, e.g. $\Pr(H_{ii',t} = 1) := h_{ii',t-1}, t > 0$.

Second, we assume that the propensity to communicate about a product depends on the current intensity of (dis)satisfaction. Note that ignoring the direction is a simplification as negative experience generates a greater desire to communicate than a positive one. Let $v_{ij,t}, w_{ij,t} \in \{0,1\}$ denote the stochastic *communication indicators* of the sender and the recipient, respectively, where

$$\Pr(V_{ij,t} = 1) := \begin{cases} |s_{ij,0,t}| & : \quad \mathsf{c}_{ij,t} = 1 \\ 0 & : \quad \text{else} \end{cases} , \tag{31}$$

$$\Pr(W_{ij,t} = 1) := \begin{cases} 0 & : \quad \mathsf{c}_{ij,t} = 1 \\ 1 & : \quad \text{else} \end{cases} . \tag{32}$$

According to this definition a sender has nothing to say about a product he has not consumed (chosen), and only if a recipient currently has no consumption experience, he fills the gap with the information provided by a sender. Note, as a variant we may assume that the propensity of a recipient to 'fill in' is $1 - |\bar{s}_{ij,t}|$, i.e. the more 'neutral' his feeling about a product the less prejudiced he is.

Third, the set of consumers from which consumer i receives information on the jth product is composed of the consumers he has a contact with, that communicate about the product, and from which he accepts the information, i.e.

$$\mathbb{I}_{ij,t} := \{i' : \max(h_{ii'i,t}, h_{i'i,t}) = v_{ij,t} = w_{ij,t} = \mathsf{s}^{\iota}_{ij,t} = 1, i' \in \mathbb{I} \setminus i\}. \tag{33}$$

Note, for separable markets we assume that a consumer is not interested in products that are not available to him.

Fourth, let us assume two effects of word-of-mouth communication: on the one hand the sender's (dis)satisfaction intensities become known to the recipient, and on the other, the sender's technology induced position in attitude space. Finally, let us assume a simple averaging of the information obtained from multiple senders

$$\tilde{\mathbf{q}}_{ij,1,t} := \begin{cases} \tilde{\mathbf{q}}_{ij,0,t} & : \quad \mathsf{c}_{ij,t} = 1 \vee \mathbb{I}_{ij,t} = \emptyset \\ \mathrm{avg}_{i' \in \mathbb{I}_{ij,t}}(\tilde{\mathbf{q}}_{i'j,0,t}) & : \quad \text{else} \end{cases}, \tag{34}$$

and

$$s_{ij,1,t} := \begin{cases} s_{ij,0,t} & : \quad \mathsf{c}_{ij,t} = 1 \vee \mathbb{I}_{ij,t} = \emptyset \\ \mathrm{avg}_{i' \in \mathbb{I}_{ij,t}}(s_{i'j,0,t}) & : \quad \text{else} \end{cases}. \tag{35}$$

Remember from Section 6, if a consumer does not acquire new information on a product's technical characteristics, not even by word-of-mouth, he 'fills in' with his current attitude and his current experience of (dis)satisfaction is 'neutral'. Note, obviously we assume that the recipient, does not base the feeling of (dis)satisfaction on the communicated technical characteristics of a product but takes the view of the sender(s). We think this is a realistic assumption of the reinforcement effect of personal communication.

Bibliography

The Bibliography is shared with the next chapter.

A Continuous-Time ACM Model and Experiment

Ulrike Schuster and Jürgen Wöckl

The ACM outlined in section 1 is intended to accommodate autonomous agents. These agents must be capable of adapting their behavior or even their decision rules. Periodic reanalysis and reassessment of strategy occur in discrete time. If a simulation experiment requires a large number of evaluations of market response this may be unnecessarily tedious and time-consuming. Therefore, an alternative is offered in this section.

In principle the model introduced here corresponds to the artificial consumer market (ACM). The main difference concerns the structure of time. In contrast to the discrete approach of the ACM in this model time is implemented as a continuous variable. The temporal development of the quantities in the CACM is described by differential equations instead of discrete transition functions at each discrete time step as it is implemented in the ACM. In order to derive the evolvement of a specific quantity the applied differential equations have to be integrated over time. During the simulation process also the continuous-time quantity in the CACM requires a small but still existing discretisation to enable numerical integration.

In this study a simple Euler integration method with a constant discretisation has been used to resolve the numerical integrals. The discretisation can be chosen arbitrarily with the only requirement that it should be smooth enough to provide proper results. From this point of view the discrete approach converges to the continuous one by reducing the increment for the Euler integration task.

1 Description of the Continuous Artificial Consumer Market (CACM)

The continuous model is designed to emulate the consumer behaviour concerning different brands acting in a segmented market. All firms offer the same type of product but emphasize different attributes which leads to a positioning of each firm in the product attribute space. The consumers are split up in groups of special aspiration patterns (APAT) and each consumer group has a specific ideal point which constitutes the desired features, the so called aspirations (ASP).

At the beginning of the simulation the consumer perceptions (PCEP) regarding the product features are located in the origin. Due to the firms' advertising efforts the perceptions which are related to the emphasized physical properties of the product are moving in a direction induced by the advertising claim.

In order to decide in favor of a brand the consumers consider price-weighted perceptions which are called attitudes (ATT). The brands are rated by the consumers by measuring the distance between aspirations and perceptions. In particular the distance represents an inverse measure of the utilities (UTI) of each consumer for each product.

The choice process (CHOICE) is based on this utility measure.

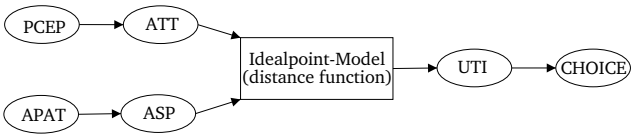

Figure 1: Application flow of the model

1.1 Dynamics of the Perceptions

Advertising impact function (aif)

The brand-specific advertising budgets affect the growth process of the consumers' perceptions concerning the position of the firms in the market. Therefore an s-shaped log-reciprocal advertising function is used (see also Hruschka, 1996, p. 214). In the following the indices i denote the aspiration groups, j the brands, k the product attributes and t the time.

$$\text{aif}(\text{budget}_j) = \exp\left\{\alpha - \frac{\beta}{\text{budget}_j}\right\}. \tag{1}$$

For the purpose of calibrating the model a special rule is appropriate. In fact the advertising impact function is adjusted in such a manner that the impact of the advertising budget is 1 at the mature market equilibrium.

$$\begin{aligned} \text{aif} &= \exp\left\{\alpha - \frac{\beta}{\text{budget}_m}\right\} = 1 \\ \Rightarrow \qquad & \alpha = \frac{\beta}{\text{budget}_m}, \end{aligned}$$

where budget_m denotes the mature advertising budget.

Differential equation of the perception dynamics

The perception dynamics are driven by the advertising budgets invested. The differential equation consists of two parts where the first describes the growth of the perceptions of the advertised attributes starting at 0 up to 1 dependent on the actual relative advertising budget. The second part describes the decay due to the forgetting of the product attributes by the consumers. The appropriate function $b(\cdot)$ is defined later.

The differential equation responsible for the temporal modification of the perceptions p of those attributes which are advertised is the following:

$$\frac{dp_{ijk}(t)}{dt} = \text{aif}(\text{budget}_j)\,(1 - p_{ijk}(t)) \; - b(t, \text{budget}_j)\, p_{ijk}(t) \Rightarrow \tag{2}$$

$$p_{ijk}(t) = \int_{\text{start}}^{t} \left(\text{aif}(\text{budget}_j)\,(1 - p_{ijk}(t)) \; - b(t, \text{budget}_j)\, p_{ijk}(t)\right) dt, \tag{3}$$

where $start$ denotes the starting time of the simulation.

Figure 2 shows the advertising impact and the solution of the differential equation for the perceptions for a special advertising budget.

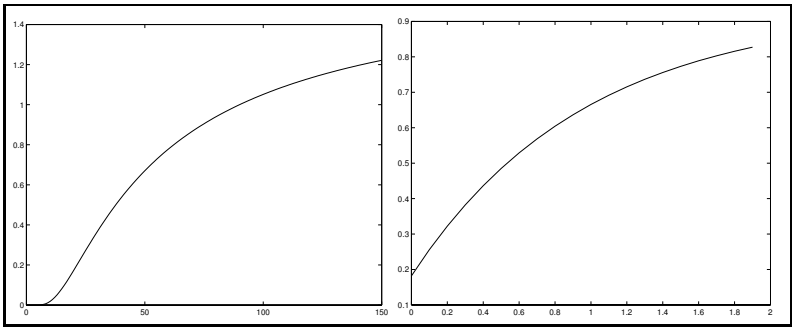

Figure 2: Schematic representation of the advertising impact function (left) and the temporal development of the perceptions for $\text{budget}_j = 90$ (right)

It can be proved that the differential equation describing the dynamics of the perceptions converges to a stationary value, namely $\lim_{t\to\infty} p_{ij}(t) = \bar{p_{ij}}$. This is valid for $\frac{dp_{ij}}{dt} = 0$:

$$\text{aif}(\text{budget}_j)(1 - p_{ijk}(t)) - b(t, \text{budget}_j)p_{ijk}(t) = 0$$

$$\Rightarrow \bar{p_{ijk}}(t) = \frac{\text{aif}(\text{budget}_j)}{\text{aif}(\text{budget}_j) + b(t, \text{budget}_j)} \,. \tag{4}$$

Calculation of attitudes

In the CACM the attitudes att are assumed to arise from the price-weighted perceptions:

$$\text{att}_{ijk}(t) = \frac{p_{ijk}(t)}{\text{price}_j^*}, \qquad \text{with} \qquad \text{price}_j^* = \frac{\text{price}_j}{\frac{1}{J}\sum_{j=1}^{J}(\text{price}_j)}.$$

Forgetting rate concerning relative budgets

The function of the forgetting rate is formulated for relative budgets. Further, it must be considered whether an attribute is advertised or not:

- non-advertised attribute:

$$b(t, \text{budget}_j) = b_0 \tag{5}$$

- advertised attribute:

$$b(t, \text{budget}_j) = \frac{1}{1 + \mathcal{F}(t, \text{budget}_j)}, \tag{6}$$

with

$$\mathcal{F}(t, \text{budget}_j) = \text{budget}_j(t) \int_{start}^{t} \frac{\text{budget}_j(\tau)}{\sum_j(\text{budget}_j(\tau))} f(t-\tau) d\tau. \tag{7}$$

The function $f(t-\tau)$ is defined by:

$$f(t-\tau) = e^{-b_0\,(t-\tau)} \tag{8}$$

$$\Rightarrow \quad \mathcal{F}(t, \text{budget}_j) = \text{budget}_j(t) \int_{start}^{t} \frac{\text{budget}_j(\tau)}{\sum_j(\text{budget}_j(\tau))} e^{-b_0\,(t-\tau)} d\tau. \tag{9}$$

The function $\mathcal{F}$ describes a mathematical convolution of former budgets with weighting function $f(t-\tau)$ which is chosen in such a way that smaller weights are imposed on past relative budgets than on actual budgets. In the actual implementation the weighting function is defined as an exponential function.

In order to calculate the actual value for the forgetting rate an Euler integration method is used where the same step-size as for the integration of the perception rates is chosen. Therefore, both integration methods are running synchronously.

Transition from discrete to continuous time

All of the above mentioned equations can be interpreted as time-discrete. Let the time intervals of the discretisation converge to zero then the model migrates to the continuous one.

$$p_{ijk}(t+1) = p_{ijk}(t) + \overbrace{[\text{aif}(\text{budget}_j)\,(1 - p_{ijk}(t)) \; - b(t, \text{budget}_j)\, p_{ijk}(t)]}^{f(p_{ijk}(t), \text{budget}_j)} \Delta t,$$

$$\begin{aligned} p_{ijk}(t+1) - p_{ijk}(t) &= f(p_{ijk}(t), \text{budget}_j)\Delta t, \\ \lim_{\Delta t \to 0} \frac{\Delta p_{ijk}(t)}{\Delta t} &= f(p_{ijk}(t), \text{budget}_j) \\ \Rightarrow \frac{dp_{ijk}(t)}{dt} &= f(p_{ijk}(t), \text{budget}_j). \end{aligned}$$

Convergence of "oblivion"

In this section the convergence of the intergral (7) is demonstrated.

The function $b(\cdot)$ corresponds to the forgetting rate and should be small but greater than zero for finite budgets. A small forgetting rate is equivalent to a stationary value of the perceptions close to 1 (equation (4)). Thus the codomain of the perception is the interval $[0, 1]$. But to reach the upper bound of the perception of 1 an infinite advertising budget would be required. In contrast the long-term or even infinite input of a finite budget does not yield the same effect. In order to provide this property of the model the function $\mathcal{F}$ should reach high values for high budgets in the former periods, but should never become infinite.

In the following the convergence of the integral over an infinite time horizon is shown. The budget of the former periods (budgetτ)) is estimated using the supremum of the function.[1]

$$\bar{b} = \sup_{j,\tau\in[0,\infty)} \left(\text{budget}_j(\tau)\right) \quad < \infty,$$

$$\lim_{t\to\infty} \int_0^t \bar{b} \cdot e^{-b_0(t-\tau)} d\tau \quad = \quad \frac{\bar{b}}{b_0} \lim_{t\to\infty} (1 - e^{-b_0 t}) \quad = \quad \frac{\bar{b}}{b_0}. \qquad (10)$$

As the factor of the relative budget $\frac{\text{budget}_j(\tau)}{\sum_j(\text{budget}_j(\tau))}$ in equation (9) may reach the maximum value of 1 this limit is valid in cases of absolute as well as relative budgets.

1.2 Ideal-Point Model

To measure the satisfaction of a consumer with a product the distance between the appropriate aspiration point and the attitude, thus the price-weighted perception, is determined by using the Euclidian norm.

Calculation of utilities

The utility of the consumers in each aspiration group i with respect to each product j can be measured with the aid of the proportional distance between the appropriate aspiration point and the attitude corresponding to brand j. The utilities uti are calculated by dividing the maximum distance by the respective one:

$$\text{uti}_{ij} = \frac{\max(\text{distance}_{ij})}{\text{distance}_{ij}},$$

thus the smaller the distance the higher the utility.

[1] As the advertising budget in each period is upper-bounded, the function budget$_j(\tau)$ can be estimated using the supremum.

Calculation of market shares

The volume of the market share MA_{ij} of brand j is calculated from the consumers of aspiration group i:

$$\text{MA}_{ij} = \frac{\text{uti}_{ij}}{\sum_i \text{uti}_{ij}}.$$

The market shares of each aspiration group i must sum up to 1.

Calculation of profits

To calculate the profits for each brand in the market the sales must be determined first:

- sales of brand j in segment i:

$$\text{sales}_{ij} = N_{C,i} \cdot \text{MA}_{ij} \cdot \text{price}_j,$$

  where $N_{C,i}$ denotes the number of consumers in segment i

- profit for brand j:

$$\text{profit}_j = \left(\sum_i N_{C,i} \cdot \text{MA}_{ij} \right) \cdot \text{price}_j - \text{budget}_j.$$

Profit serves as a target function in optimization tasks.

2 Application and Results

2.1 Experimental Market Scenario and Model Calibration

The model described above is used to explore optimal defensive strategies for a brand directly attacked by a new brand that enters a mature market. Three different firms located in three different market segments are considered. It is assumed that after half of the time period of interest a new brand enters the market and settles down in a segment yet occupied by a firm. Now this incumbent is allowed to defend itself by changing its price and advertising budget while the other two brands are assumed to show no reaction in case of a new entry because their segments aren't affected.

All brands are assumed to start at consumer perceptions of a value of 0.1 for each of the product attributes. That leads to a disadvantage for the entrant. While its consumer perceptions start at 0.1 all other incumbents' perceptions have already developed over time. Each of the three initial brands demands the same price of 3 units and advertises with identical budgets of 90 units. The entrant is hypothesized to join the market with a somewhat smaller price of 2.5 units but higher advertising expenditures of 150 units in order to compensate for lost time. Without loss of generality prices are restricted to the interval $[1, 4]$ and budgets to $[0, 200]$. As there's no boundary solution with respect to the optimization this constitutes a reasonable range.

Optimal defensive strategies under variable advertising impact functions are analyzed. For the first one (aif 1) the parameters in equation (1) are specified as $\alpha = 0.5$ and $\beta = 45$, for the second function (aif 2) $\alpha = 1$ and $\beta = 90$ are assumed, and in the third advertising impact function (aif 3) the parameters are set to $\alpha = 2$ and $\beta = 180$.

Figure 3 exhibits the shape of the advertising impact functions.

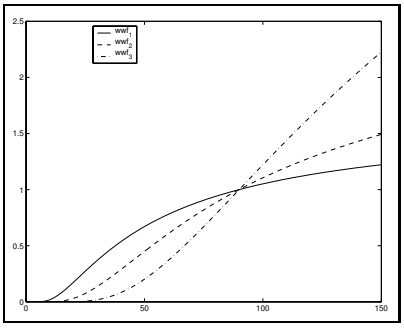

Figure 3: Advertising impact function for different parameters α and β, represented by solid (aif 1), dashed (aif 2), and dotted/dashed lines (aif 3).

Also the temporal development of the perceptions is illustrated for the incumbent as well as the entrant budget for each of the three advertising impacts (Fig. 4).

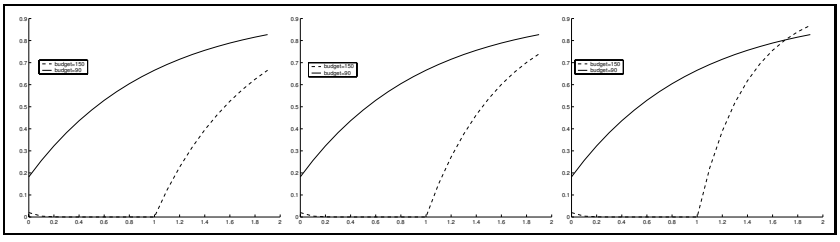

Figure 4: Comparison of the temporal development of the perceptions for incumbent's and entrant's budgets (90 [solid line] and 150 [dashed line] units respectively) under variable advertising impact functions (from left to right: aif 1, aif 2 and aif 3)

The advertising impact is responsible for the entrant's competitive strength. Under the first two advertising impact functions the entrant is unable to catch up the perceptual development of the incumbent in the past time period. In contrast with a stronger advertising impact (aif 3) the entrant manages to reach the same perceptual level as the incumbent but in a much shorter time span.

Target function

Price-budget combinations which lead to a maximum profit are considered to be optimal defensive strategies:

$$\mathrm{ZF}_j = \mathrm{profit}_j \rightarrow \max_{\mathrm{price}_j, \mathrm{budget}_j} (\mathrm{profit}_j).$$

Therefore, the profit of the attacked brand is used as target function.

2.2 Maximizing Profits under Alternative Advertising Impact Functions

Optimization of the incumbent given an entrance strategy

In this section the emphasis is put on the incumbent's reaction in case of a fixed entrance strategy.

In order to find the optimal defensive strategy a surface plot is created which shows the profits of the incumbent for different price-budget combinations (Fig. 5 and Fig. 6).

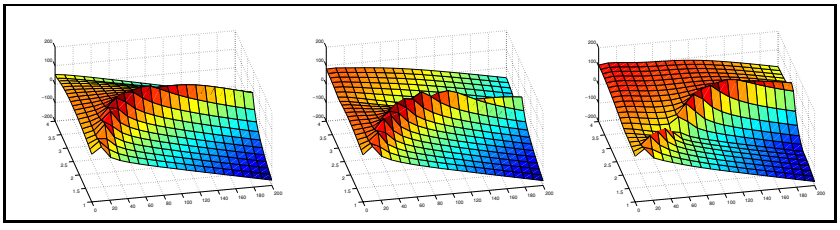

Figure 5: Surface plots of the profits of the incumbent for several price-budget combinations under a fixed entrance strategy (from left to right: aif 1, aif 2 and aif 3)

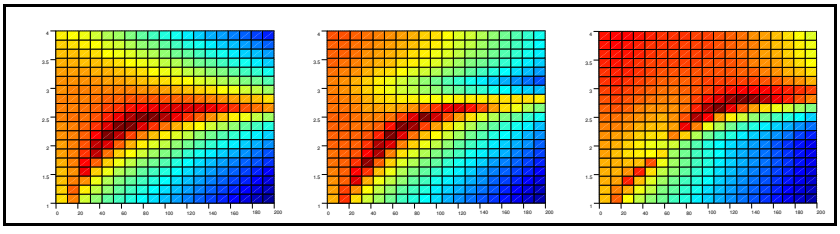

Figure 6: Upright projection of the results shown in figure 5

As the figures demonstrate there is no unique optimum. Different price-budget combinations obviously result in the same optimal profit for the incumbent.

Irrespective of the advertising impact it is advised to the incumbent to reduce its price as a reaction to a new entry. Concerning advertising expenditures it will de-

pend on the aif if the budget shall be increased or decreased. Only one of the incumbent's optimal strategies is considered. It is characterized by medium-sized budgets and prices for each of the three advertising impact functions. The specific values are presented subsequently.

In case of a weaker advertising impact (aif 1 and aif 2) the price should be reduced from 3 to about 2.5 (aif 1) or 2.4 units (aif 2) and also the budget should be decreased from 90 to approximately 80 (aif 1) or 70 units (aif 2). In case of aif 3 obviously the optimal reaction consists of a smaller price reduction (mean optimal price of 2.7 units) but here the incumbent should raise its advertising budget to 110 units.

A comparison of the results for the three advertising impact functions shows no significant difference for aif 1 and aif 2. But in contrast for aif 3 instead of lowering the budget an increase of the advertising efforts is recommended. Concerning the price reaction the difference is less distinct. Prices should still be reduced but not as drastical as under aif 1 or aif 2.

Optimization of the entrant given the optimal incumbent strategy

In the section above the entrance strategy is arbitrarily but reasonably chosen and the incumbent tries to maximize its profit by changing price and advertising expenditures.

On the other side also the entrant strategy can be optimized given the incumbent's price and budget. Therefore the mean optimal price-budget combination for the incumbent is determined in a first step. Afterwards it is assumed that the incumbent will react like this when facing a new entry. Given the fixed incumbent strategy the optimal entrance strategy can now be calculated.

Figures 7 and 8 show the profits of the entrant for different combinations of prices and budgets.

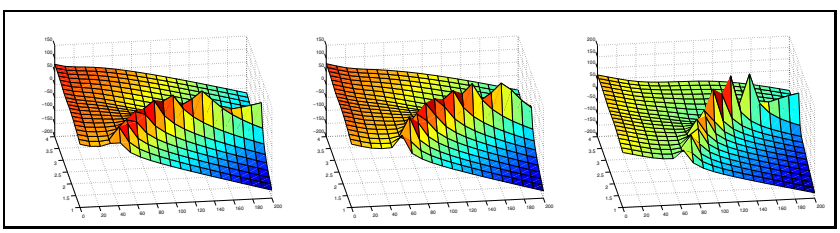

Figure 7: Surface plots of the profits of the entrant for several price-budget combinations under the optimal incumbent's strategy (from left to right: aif 1, aif 2 and aif 3)

For the entrant again no unique optimum can be specified. Because a single optimum is needed for each advertising impact mean prices and budgets are calculated from the set of optima and used as strategy recommendations. Under aif 1 optimal price and budget approximate 2.3 and 140 units respectively. With a growing advertising impact optimal entrance prices tend to increase while advertising expenditures more or less stay the same (aif 2: mean price of 2.4 units, mean budget of 140 units;

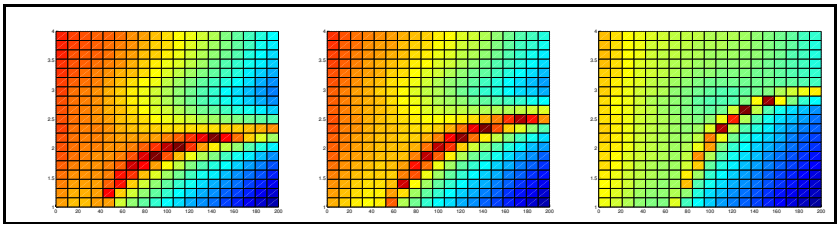

Figure 8: Upright projection of the results shown in figure 7

aif 3: mean price of 2.6 units, mean budget of 140 units).

Further research could concern the derivation of stationary strategies for both the incumbent and the entrant. This can be realized by simultaneously optimizing incumbent and entrant strategies by alternately updating optimal prices and budgets until convergence to a stable strategy is reached. Another topic of interest for future investigations is the comparison of optimal defensive strategies under a varying time of entry. A third line of research regards disaggregated markets where different distributional patterns of the consumer aspirations are studied.

Bibliography

Bagozzi, R. (1986). *Principles of Marketing Management*. Science Research Associates, Chicago.

Bettman, J., Luce, M., and Payne, J. (1998). Constructive consumer choice processes. *Journal of Consumer Research*, 25(3):187–217.

Brassel, K., Möhring, M., and Schuhmacher, E. (1997). Can agents cover all the world? In Conte, R., Hegselmann, R., and Terna, P., editors, *Simulating Social Phenomena*, pages 55–71. Springer, New York.

Brenner, T. (1999). *Modelling Learning in Economics*. Elgar, Cheltenham.

Buchta, C., Dolnicar, S., and Reutterer, T. (2000). *A Nonparametric Approach to Perceptions-Based Market Segmentation: Applications*. Springer, Vienna.

Buchta, C. and Mazanec, J. (2001). SIMSEG/ACM — A simulation environment for artificial consumer markets. Technical report, SFB 010 Working Paper Series No. 79, Vienna University of Economics and Business Administration.

Cardozo, R. (1965). An experimental study of consumer effort, expectation and satisfaction. *Journal of Marketing Research*, 2:244ff.

Crompton, J. (1992). Structure of vacation destination choice sets. *Annals of Tourism Research*, 19:420–434.

Cronin, Jr., J. and Taylor, S. (1994). SERVPERF versus SERVQUAL: Reconciling performance-based and perceptions-minus-expectations measurement of service quality. *Journal of Marketing*, 58:125–131.

Engel, J. F., Kollat, D., and Blackwell, R. (1973). *Consumer Behavior*. Holt, Rinehart, and Winston, New York, 2nd edition.

Fahrmeir, G. and Tutz, G. (1997). *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer, New York.

Frühwirth-Schnatter, S. and Otter, T. (1999). Conjoint-analysis using mixed-effect models. In Friedl, H., Berghold, A., and Kauermann, G., editors, *Statistical Modeling*, pages 181–191, Graz. Proceedings of the 14th International Workshop on Statistical Modelling.

Goodall, B. (1991). Understanding holiday choice. In Cooper, C., editor, *Progress in Tourism, Recreation and Hospitality Management*, volume 3. Belhaven, London.

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

Hauser, J. and Wernerfelt, B. (1990). An evaluation cost model of consideration sets. *Journal of Consumer Research*, 16:393–408.

Herbrich, R., Keilbach, M., Graepel, T., Bollmann-Sdorra, P., and Obermayer, K. (1999). Neural networks in economics. In Brenner, T., editor, *Computational Techniques for Modeling Learning in Economics*, pages 169–196. Kluwer, Boston.

Howard, J. (1977). *Consumer Behavior: Application of Theory*. McGraw-Hill, New York.

Howard, J. and Sheth, J. (1969). *The Theory of Buyer Behavior*. Wiley, New York.

Hruschka, H. (1996). *Marketing-Entscheidungen*. Vahlen, Munich.

Johnson, M., Anderson, E., and Fornell, C. (1995). Rational and adaptive performance expectations in a costumer satisfaction framework. *Journal of Consumer Research*, 21:695–707.

Kotler, P. (1986). *Principles of Marketing*. Prentice-Hall, Englewood Cliffs, 3rd edition.

Kroeber-Riel, W. (1980). *Konsumentenverhalten*. Vahlen, Munich, 2nd edition.

Lilien, G., Kotler, P., and Moorthy, S. (1992). *Marketing Models*. Prentice-Hall, New York.

Long, J. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Sage, Newbury Park.

Mazanec, J. (1978). *Strukturmodelle des Konsumentenverhaltens*. Orac, Vienna.

Mazanec, J. (1997). Satisfaction tracking for city tourists. In Mazanec, J., editor, *International City Tourism: Analysis and Strategy*, pages 75–100. Cassell, London.

Mazanec, J. and Strasser, H. (2000). *A Nonparametric Approach to Perceptions-Based Market Segmentation: Foundations*. Springer, Vienna.

Myers, J. (1996). *Segmentation and Positioning for Strategic Marketing Decisions*. American Marketing Association, Chicago.

Mühlbacher, H. (1988). Ein situatives Modell der Motivation zur Informationsaufnahme und -verarbeitung bei Werbekontakten. *Marketing ZFP*, 10:85–94.

Oliver, R. and DeSarbo, W. (1988). Response determinants in satisfaction judgments. *Journal of Consumer Research*, 14:495ff.

Oliver, R. and Swan, J. (1989a). Consumer Perceptions of Interpersonal Equity and Satisfaction in Transactions: A Field Survey Approach. *Journal of Marketing*, 53:21ff.

Oliver, R. and Swan, J. (1989b). Equity and disconfirmation perceptions as influences on merchant and product satisfaction. *Journal of Consumer Research*, 16:327ff.

Parasuraman, A., Zeithaml, V., and Berry, L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49:41–50.

Parasuraman, A., Zeithaml, V., and Berry, L. (1988). SERVQUAL: A multiple item scale for measuring consumer perceptions of service quality. *Journal of Retailing*, 64:12–40.

Roberts, J. and Lilien, G. (1993). Explanatory and predictive models of consumer behavior. In Eliashberg, J. and Lilien, G., editors, *Marketing, Handbooks in Operations Research and Management Science*, volume 5, pages 27–82. North-Holland, Amsterdam.

Schuster, U. and Wöckl, J. (2004a). Derivation of stationary optimal defensive strategies using a continuous market model. In *Conference Proceedings of the AMS*, volume XXVII, pages 305–311.

Schuster, U. and Wöckl, J. (2004b). Optimal defensive strategies under varying consumer disributional patterns and market maturity. In *Conference Proceedings of the AMS*, volume XXVII, pages 140–144.

Troitzsch, K. (1999). Simulation as a tool to model stochastic processes in complex systems. In Brenner, T., editor, *Computational Techniques for Modelling Learning in Economics*, pages 45–69. Kluwer, Boston.

Trommsdorff, V. (1998). *Konsumentenverhalten*. Kohlhammer, Stuttgart, 3rd edition.

Zeithaml, V. and Berry, L. (1988). Communication and control processes in the delivery of service quality. *Journal of Marketing*, 52:35–48.

Capturing Unobserved Consumer Heterogeneity Using the Bayesian Heterogeneity Model

Sylvia Frühwirth-Schnatter, Regina Tüchler, and Thomas Otter

1 Introduction

In the analysis of panel data from marketing one often is forced to deal with unobserved heterogeneity. Unobserved heterogeneity may be either cross-sectional or longitudinal. Cross-sectional heterogeneity means that important parameters such as preferences differ between the consumers whereas longitudinal heterogeneity means that important parameters such as preferences change over time.

From a methodological point of view, a broad range of methods are available to capture unobserved heterogeneity, namely non-parametric methods such as perception based market segmentation (Mazanec and Strasser, 2000; Buchta et al., 2000) and clustering by ensemble methods (Dolničar and Leisch, 2003; Dimitriadou et al., 2001), see also the chapter on ensemble methods for cluster analysis by Kurt Hornik and Friedrich Leisch in this volume.

In our own contribution we discuss a hierarchical parametric modelling approach toward unobserved heterogeneity based on mixture models, especially mixtures of random effect models. Following the seminal paper by Allenby et al. (1998), a number of authors pursued this approach, see e.g. Lenk and DeSarbo (2000), Frühwirth-Schnatter et al. (2004) and Otter et al. (2004b).

2 The General Heterogeneity Model

The data are described by a *mixture of random effects model*:

$$y_i = X_i^1\alpha + X_i^2\beta_i + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2_{\varepsilon,i} I), \tag{1}$$

where y_i is a vector of T_i observations for subject $i = 1, \ldots, N$, X_i^1 is the $T_i \times d$ design matrix for the $d \times 1$ vector of the fixed effects α and X_i^2 is the design matrix of dimension $T_i \times r$ for the $r \times 1$ random effects vector β_i. I is the identity matrix. Due to unobserved heterogeneity the random effects β_i are different for each subject i. The unknown distribution $\pi(\beta_i)$ of heterogeneity is approximated by a mixture of normal distributions $\beta_i \sim \sum_{k=1}^K \eta_k N(\beta_k^G, Q_k^G)$ with the unknown group means $\beta_1^G, \ldots, \beta_K^G$, the unknown group covariance matrices $Q_1^G, \ldots, Q_K^G$ and the unknown group probabilities $\eta = (\eta_1, \ldots, \eta_K)$.

There are two different approaches to model the error variances $\sigma^2_{\varepsilon,i}$: *homogeneous variances* where

$$\sigma^2_{\varepsilon,i} \equiv \sigma^2_\varepsilon, \qquad \forall i = 1, \ldots, N, \tag{2}$$

and consumer specific, *heterogenous variances* where

$$\sigma^2_{\varepsilon,i} = \frac{\sigma^2_\varepsilon}{\lambda_i}, \qquad \forall i = 1, \ldots, N, \tag{3}$$

with gamma distributed factors λ_i:

$$\lambda_i \sim G(\frac{\nu}{2}, \frac{\nu}{2}). \tag{4}$$

Note that marginally with respect to λ_i the variance model (3) and (4) implies the following marginal distribution for y_i:

$$y_i = X_i^1\alpha + X_i^2\beta_i + \varepsilon_i, \qquad \varepsilon_i \sim t_\nu(0, \sigma_\varepsilon^2 I). \tag{5}$$

Verbeke and Lesaffre (1996) study model (1) with the groups covariances being the same for all groups. A similar heterogeneity models is discussed in Allenby et al. (1998) however without considering fixed effects. Lenk and DeSarbo (2000) extend (1) to observations from distributions from general exponential families.

Model (1) includes many other models as a special case, especially the *aggregate model* for $K = 1$ and $Q_k^G \equiv 0$. The popular *latent class model* (LCM) is that special case where $K > 1$ and $Q_1^G \equiv \ldots \equiv Q_K^G \equiv 0$. Finally, the *random coefficient model* (RCM), which is also called hierarchical Bayes model is that special case where $K = 1$ and $Q_1^G \neq 0$.

2.1 Bayesian Estimation of the Heterogeneity Model under Heterogeneous Variances

Bayesian estimation of the heterogeneity model via MCMC methods is discussed by Allenby et al. (1998), Lenk and DeSarbo (2000) and Frühwirth-Schnatter et al. (2004). Estimation is carried out for a fixed number K of groups using Markov Chain Monte Carlo methods.

Let $y^N = (y_1, \ldots, y_N)$ denote all observations. One introduces discrete latent group indicators $S^N = (S_1, \ldots, S_N)$, with S_i taking values in $\{1, \ldots, K\}$ and thereby indicating which group consumer i belongs to, with the unknown probability distribution $Pr(S_i = k) = \eta_k$. Following the principle of data augmentation (Tanner and Wong, 1987), the parameter vector of the unknown model parameters $\phi = (\alpha, \beta_1^G, \ldots, \beta_K^G, \eta, Q_1^G, \ldots, Q_K^G, \sigma_\varepsilon^2)$ is augmented by the individual parameters $\beta^N = (\beta_1, \ldots, \beta_N)$ and the group indicators S^N. Under heterogeneous error variances the vector $\lambda^N = (\lambda_1, \ldots, \lambda_N)$ has to be added in a further data augmentation step.

A straightforward way of Bayesian estimation of the heterogeneity model via MCMC methods is Gibbs sampling from full conditional distributions. That sampler is discussed in Lenk and DeSarbo (2000) and Allenby et al. (1998). For a heterogeneity model with homogeneous error variances the parameters S^N, η, α, $(\beta_1^G, \ldots, \beta_K^G)$, β^N, $(Q_1^G, \ldots, Q_K^G)$ and σ_ε^2 are sampled in turn from the corresponding full conditional distributions. It has been demonstrated in Frühwirth-Schnatter et al. (2004) that such a full conditional Gibbs sampler is sensitive to the way model (1) is parameterized. There exist two ways to parameterize the model, depending on whether X_i^1 and X_i^2 have common columns or not. The partly marginalized Gibbs sampler suggested in Frühwirth-Schnatter et al. (2004) where the random effects β^N are integrated out

when sampling S^N, α, and $(\beta_1^G, \ldots, \beta_K^G)$ turned out to be insensitive to the parameterization.

In Frühwirth-Schnatter et al. (2004) the partly marginalized Gibbs sampler was introduced for homogeneous error variances. This sampler may easily be extended to deal with heterogeneous error variances. The MCMC sampling steps are the following.

MCMC Sampling Algorithm:

(i) Sample S^N conditional on y^N, ϕ and λ^N.

(ii) Sample η from $p(\eta|S^N)$.

(iii) Sample $\alpha, \beta_1^G, \ldots, \beta_K^G$ and β^N conditional on y^N, S^N, $Q_1^G, \ldots, Q_K^G$, σ_ε^2 and λ^N.

a) Sample α and $\beta_1^G, \ldots, \beta_K^G$ conditional on y^N, S^N, $Q_1^G, \ldots, Q_K^G$, σ_ε^2 and λ^N.

b) Sample β^N conditional on $y^N, \alpha, \beta_1^G, \ldots, \beta_K^G, S^N$, $Q_1^G, \ldots, Q_K^G$, σ_ε^2 and λ^N.

(iv) Sample $Q_1^G, \ldots, Q_K^G$ and $\sigma_{\varepsilon_i}^2$ conditional on $y^N, \alpha, \beta_1^G, \ldots, \beta_K^G, \beta^N$ and S^N.

a) Sample $Q_1^G, \ldots, Q_K^G$ conditional on $y^N, \alpha, \beta_1^G, \ldots, \beta_K^G, \beta^N$ and S^N.

b) Sample σ_ε^2 conditional on $y^N, \alpha, \beta_1^G, \ldots, \beta_K^G, \beta^N, S^N$ and λ^N.

c) Sample λ^N conditional on $y^N, \alpha, \beta_1^G, \ldots, \beta_K^G, \beta^N, S^N$ and σ_ε^2.

Details on the Sampling Steps:

The marginal heteroscedastic random effects model:

The marginal model with the random effects β^N integrated out writes as follows:

$$y_i = Z_i^* \alpha^* + \varepsilon_i^*, \qquad \varepsilon_i^* \sim N(0, V_i). \tag{6}$$

We introduce the indicators $D_i^{(k)}$, that take the value 1 iff $S_i = k$ and zero otherwise, to define the design matrix $Z_i^* = (X_i^1 \; X_i^2 D_i^{(1)} \; \ldots \; X_i^2 D_i^{(K)})$ for the parameter vector $\alpha^* = (\alpha' \; (\beta_1^G)' \; \ldots \; (\beta_K^G)')'$ and the individual model error covariances matrices $V_i = X_i^2 Q_1^G D_i^{(1)} (X_i^2)' + \ldots + X_i^2 Q_K^G D_i^{(K)} (X_i^2)' + \sigma_{\varepsilon,i}^2 I$.

(i) *Sampling the switching variable S^N:*

The indicators $S_1, \ldots, S_N$ are conditionally independent given y^N,λ^N and ϕ and we sample S_i from the discrete distribution:

$$p(S_i = k|y_i, \phi, \lambda_i) \propto p(y_i|\alpha, \beta_k^G, Q_k^G, \sigma_\varepsilon^2, \lambda_i) \cdot \eta_k,$$

where the likelihood is obtained by using the heteroscedastic model representation (6) with the random effects β^N integrated out. The likelihood is therefore normally distributed with $N(y_i; X_i^1 \alpha + X_i^2 \beta_k^G, X_i^2 Q_k^G (X_i^2)' + \sigma_{\varepsilon,i}^2 I)$.

(ii) *Sampling the weights η:*

Since η depends only on the switching variable S^N sampling η conditionally on $y^N, \phi, S^N, \lambda^N$ and β^N simplifies to sampling from the posterior $p(\eta|S^N)$. This posterior is Dirichlet distributed $D(e_{0,1} + N_1, \ldots, e_{0,K} + N_K)$, where $N_k = \#(S_i = k)$ and $D(e_{0,1}, \ldots, e_{0,K})$ is the prior Dirichlet distribution for the weights η.

(iii) *Sampling the fixed effects α, the group specific means $\beta_1^G, \ldots, \beta_K^G$ and the random effects β^N:*

The conditional posterior of $\alpha^* = (\alpha' \ (\beta_1^G)' \ldots (\beta_K^G)')'$ and $\beta^N = (\beta_1, \ldots, \beta_N)$ partitions as follows:

$$p(\alpha^*, \beta^N | y^N, \theta, \lambda^N, S^N) = \left(\prod_{i=1}^{N} p(\beta_i | \alpha^*, y_i, \theta, \lambda_i, S_i) \right) p(\alpha^* | y^N, \theta, \lambda^N, S^N),$$

where $\theta = (Q_1^G, \ldots, Q_K^G, \sigma_\varepsilon^2)$. Therefore we sample α^* from the marginal model in step (iii a) and β^N from the full conditional distribution in step (iii b).

(iii a) *Sampling α and $\beta_1^G, \ldots, \beta_K^G$:*

From the marginal heteroskedastic model (6) we see that the posterior of α^* is normally distributed:

$$p(\alpha^* | y^N, \theta, \lambda^N, S^N) \propto N(A_N^* \cdot a_N^*, A_N^*),$$

where

$$(A_N^*)^{-1} = \sum_{i=1}^{N} (Z_i^*)' V_i^{-1} Z_i^* + (A_0^*)^{-1},$$
$$a_N^* = \sum_{i=1}^{N} (Z_i^*)' V_i^{-1} y_i + (A_0^*)^{-1} a_0^*.$$

The vector a_0^* is the mean parameter and A_0^* is the covariance matrix of the prior normal distribution of α^*.

(iii b) *Sampling the random effects β^N:*

The individual parameters β_i are conditionally independent and have a normal posterior distribution:

$$p(\beta_i | y_i, \alpha, \beta_{S_i}^G, Q_{S_i}^G, \sigma_\varepsilon^2) \sim N(C_i \cdot c_i, C_i)$$

with

$$C_i^{-1} = (\sigma_\varepsilon^{-2} I)(X_i^2)' X_i^2 + (Q_{S_i}^G)^{-1},$$
$$c_i = (\sigma_\varepsilon^{-2} I)(X_i^2)'(y_i - X_i^1 \alpha) + (Q_{S_i}^G)^{-1} \beta_{S_i}^G.$$

(iv a) *Sampling the covariance matrices* $Q_1^G, \ldots, Q_K^G$*:*

The group specific covariances Q_k^G for $k = 1, \ldots, K$ and the variance of the model equation σ_ε^2 are conditionally independent given $\alpha, \beta^N, \beta_1^G, \ldots, \beta_K^G, S^N$ and y^N.

The covariance matrices are sampled for each group separately. from the inverted Wishart distribution $p(Q_k^G|\beta^N, \beta_k^G, S_i = k) \sim IW(\nu_k^Q, S_k^Q)$ with

$$\begin{aligned}\nu_k^Q &= \nu_0^Q + N_k/2,\\ S_k^Q &= S_0^Q + 1/2\left(\sum_{i=1}^N D_i^{(k)}(\beta_i - \beta_k^G)(\beta_i - \beta_k^G)'\right),\end{aligned}$$

where $D_i^{(k)} = 1$ iff $S_i = k$ and zero otherwise, $N_k = \#(S_i = k)$ and ν_0^Q and ${S_0}^Q$ are the parameters of the prior inverted Wishart distribution of Q_k^G.

(iv b) *Sampling of the error variance parameter* σ_ε^2:

Applying Bayes theorem yields the posterior inverted gamma distribution for σ_ε^2.

$$p(\sigma_\varepsilon^2|y^N, \alpha, \beta^N, \lambda^N) \propto IG(\nu_N^\varepsilon, S_N^\varepsilon),$$

where the parameters are defined as

$$\begin{aligned}\nu_N^\varepsilon &= \nu_0^\varepsilon + \frac{1}{2}\left(\sum_{i=1}^N T_i\right),\\ S_N^\varepsilon &= S_0^\varepsilon + \frac{1}{2}\left(\sum_{i=1}^N \|y_i - X_i^1\alpha - X_i^2\beta_i\|_2^2 \cdot \lambda_i\right).\end{aligned}$$

A priori σ_ε^2 follows an inverted gamma distribution with parameters ν_0^ε and S_N^ε.

(iv c) *Sampling the individual parameters* $\lambda^N = (\lambda_1, \ldots, \lambda_N)$:

The individual parameters λ_i are conditionally independent. We derive the following posterior gamma distribution for each subject specific parameter:

$$p(\lambda_i|y_i, \alpha, \beta_i, \sigma_\varepsilon^2) \propto G(\nu_i^\lambda, S_i^\lambda),$$

where

$$\begin{aligned}\nu_i^\lambda &= \frac{\nu_0^\lambda}{2} + \frac{T}{2},\\ S_i^\lambda &= \frac{\nu_0^\lambda}{2} + \frac{1}{2}\left(\|y_i - X_i^1\alpha - X_i^2\beta_i\|_2^2 \cdot 1/\sigma_\varepsilon^2\right)\end{aligned}$$

and ν_0^λ is prior parameter.

In this algorithm only steps (i) and (iii) deviate from standard full conditional Gibbs sampling. The marginal heteroscedastic model, where the individual parameters β^N are integrated out, serves to obtain S^N from $p(S^N|\phi, y^N, \lambda^N)$. The model's representation as a switching random effect model is exploited to derive a blocked Gibbs sampler to sample $\alpha, \beta_1^G, \ldots, \beta_K^G$ and β^N from $p(\beta_1^G, \ldots, \beta_K^G, \alpha, \beta^N|S^N, Q_1^G, \ldots, Q_K^G, \sigma_\varepsilon^2, y^N, \lambda^N)$.

Finally, to deal with the unidentifiability problem due to the arbitrary labeling of the groups discussed in Celeux et al. (2000), Frühwirth-Schnatter (2001) and Stephens (2000), the random permutation sampler suggested in Frühwirth-Schnatter (2001) is applied.

2.2 Bayesian Model Comparison through Model Likelihoods

Model selection based on model likelihoods follows a long tradition in Bayesian econometrics initiated by Zellner (1971) and has been applied by various authors for selecting the number of components in mixture models, see among others Chib (1995) and Frühwirth-Schnatter (2004). Previous applications to model selection problems arising for the heterogeneity model appeared in Lenk and DeSarbo (2000), Frühwirth-Schnatter and Otter (1999), Otter et al. (2002), Tüchler et al. (2002) and Otter et al. (2004b).

Different heterogeneity models $\mathcal{M}_1, \ldots, \mathcal{M}_L$ are compared through their posterior probabilities:

$$P(\mathcal{M}_l|y^N) \propto p(y^N|\mathcal{M}_l)P(\mathcal{M}_l),$$

where the model likelihood $p(y^N|\mathcal{M}_l)$ is identical with the integrated likelihood function $p(y^N|\phi)$ for model $\mathcal{M}_l$:

$$p(y^N|\mathcal{M}_l) = \int p(y^N|\phi)p(\phi)d\phi. \tag{7}$$

$\phi = (\alpha, \beta_1^G, \ldots, \beta_K^G, \eta, Q_1^G, \ldots, Q_K^G, \sigma_\varepsilon^2)$ is the vector of unknown model parameters. Note that in (7) $p(y^N|\phi)$ is the marginal likelihood, where all latent variables like S^N, β^N and, for heterogeneous error variances, λ^N are integrated out. For homogeneous variances this is the product of N multivariate normal distributions, whereas for heterogeneous errors this is a product of N multivariate t-distributions.

The computation of the model likelihood $p(y^N|\mathcal{M}_l)$ is non-trivial because it involves a high-dimensional integration. Model likelihoods have been estimated from the MCMC output using methods such as the candidate's formula (Chib, 1995), importance sampling based on mixture approximations (Frühwirth-Schnatter, 1995) and bridge sampling (Meng and Wong, 1996). For computing the model likelihoods we apply here the method of bridge sampling, which has proven to be robust against label switching and more efficient than other methods in the context of mixture models (Frühwirth-Schnatter, 2004).

3 An Illustrative Application from Conjoint Analysis

3.1 The Data

Our application comes from conjoint analysis, a procedure that is focused on obtaining the importance of certain product attributes and their significance in motivating a consumer toward purchase from a holistic appraisal of attribute combinations. Our data come from a brand - price trade off study in the mineral-water category. Each of 213 Austrian consumers evaluated their likelihood of purchasing 15 different product-profiles offering five different brands of mineral-water at different prices on 20 point rating scales. The goal of the modelling exercise is to find a model describing consumers' heterogeneous preferences towards the different brands of mineral water and their brand-price trade offs.

Applying the general heterogeneity model to these data we follow up previous work on the same data using the random coefficient model based on normal errors (Frühwirth-Schnatter and Otter, 1999), the latent class model based on normal errors (Otter et al., 2002), the general heterogeneity model based on normal errors (Tüchler et al., 2002).

3.2 The Design Matrix

Our fully parameterized design matrix consists of 15 columns corresponding to the constant, four brand contrasts (of the brands Römerquelle – *RQ*, Vöslauer – *VOE*, Juvina – *JU*, Waldquelle – *WA*), a linear and a quadratic price effect, four brand by linear price and four brand by quadratic price interaction effects, respectively. We used dummy-coding for the brands. The fifth brand Kronsteiner (*KR*) was chosen as the baseline. We subtracted the smallest price from the linear price column, and computed the quadratic price contrast from the centered linear contrast. Therefore, the constant corresponds to the purchase likelihood of Kronsteiner at the lowest price level, if quadratic price effects are not present. The investigations of these data in Otter et al. (2002) indicated that a specification with fixed brand by quadratic price interactions is preferable and is therefore chosen for the rest of this paper.

We carried out 30000 MCMC iterations and based our inference on the last 6000. The group specific means β_k^G and the fixed effects α are a priori normally distributed with $N(b_0, B_0)$ and $N(a_0, A_0)$, respectively. The prior means b_0 and a_0 are equal to the population mean of the RCM model reported in Frühwirth-Schnatter and Otter (1999) and for the information matrices we choose $A_0^{-1} = B_0^{-1} = 0.04 \cdot I$. The prior distribution of the groups covariances is an inverted Wishart distribution $IW(\nu_0^Q, S_0^Q)$. We choose $\nu_0^Q = 10$ and then derive S_0^Q from $E(Q_k^G) = (\nu_0^Q - (d + 1)/2)^{-1} S_0^Q$, where $E(Q_k^G)$ was computed by individual OLS estimation and d is the dimensionality of Q_k^G. The prior on η is the commonly used Dirichlet distribution $D(1, \ldots, 1)$. We stay noninformative about the error variances σ_ε^2 and choose the inverted gamma distribution $IG(0, 0)$.

3.3 Model Selection

We estimated various models for our data, the general heterogeneity model, varying the number of groups K, the special case of the LCM, also varying the number of groups K and the special case of the RCM. We also investigated, whether we should use homogeneous or heterogeneous variances.

Table 1 shows estimates of the logarithm of model likelihoods for all these models based on homogeneous variances, whereas Table 2 shows estimates based on heterogeneous variances.

We see that the RCM (column $Q \neq 0$, line $K = 1$) is clearly preferred to all LCMs (column $Q = 0$), but is outperformed by the general heterogeneity model (column $Q \neq 0$, line $K > 1$), regardless of the assumption made concerning the variances. The specification chosen for the variances exercises has a considerable influence on the number of optimal classes. Under the assumption of homogeneous variances the optimal *latent* class model has seventeen classes, whereas the number reduces to fourteen under heterogeneous errors. Also the general heterogeneity model has a different number of optimal classes, namely two under heterogeneous errors and three under homogeneous errors.

The optimal model out of *all* models under consideration is a general heterogeneity model with heterogeneous error variances and $K = 2$ classes.

Table 1: Bayesian model selection for the mineral water data; log of the model likelihoods based on the normal distribution (rel. std. errors in parentheses)

	$\log p(y^N \mid \text{Model})$	
K	$Q_k^G \neq 0$	$Q_k^G = 0$
1	-9222.36 (0.05)	-10077.31 (0.00)
2	-9165.66 (0.06)	-9881.49 (0.01)
3	**-9161.27 (0.06)**	-9733.98 (0.02)
4	-9165.73 (0.08)	-9669.98 (0.05)
⋮		⋮
16	-	-9464.77 (1.16)
17	-	**-9460.61 (1.19)**
18	-	-9465.79 (1.33)

The MCMC draws may be explored to indicate overfitted models with too many groups. This is illustrated in Figure 1 where we see posterior draws of the group specific parameters corresponding to the price coefficient and one of the brands obtained under assuming homogeneous error variances. The left hand side figure corresponds to the number of classes selected by the model likelihood. The right hand side figure is a plot where an additional class is added. We find the same number of simulation clusters as before, as the data do not support an additional class. The widely spread simulations overlaying the three clusters indicate that parameters are sampled from their prior because the additional class is empty on many iterations.

Table 2: Bayesian model selection for the mineral water data; log of the model likelihoods based on the t_4-distribution (rel. std. errors in parenthesis)

	$\log p(y^N \text{\|Model})$	
K	$Q_k^G \neq 0$	$Q_k^G = 0$
1	-9101.52 (0.05)	-9980.21 (0.00)
2	**-9028.81 (0.06)**	-9727.13 (0.02)
3	-9043.96 (0.06)	-9576.97 (0.03)
4	-9045.86 (0.06)	-9522.18 (0.05)
⋮	⋮	⋮
12	-	-9332.96 (0.08)
13	-	-9329.49 (0.08)
14	-	**-9326.26 (0.08)**
15	-	-9327.27 (0.08)

The preference of a model with *heterogenous* variances is also supported by Figure 8 which shows considerable differences among the individual variances $\sigma^2_{\varepsilon,i}$ for 15 randomly selected consumers by means of their posterior distribution.

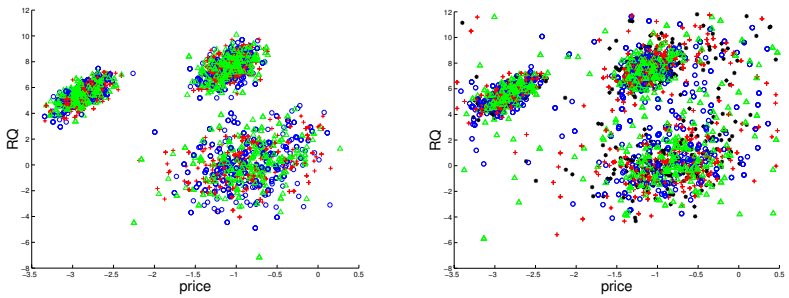

Figure 1: General heterogeneity model based on homogeneous variances, $K = 3$ (left) against $K = 4$ (right); posterior draws of group specific means obtained from random permutation sampling for the *price* against *RQ*

3.4 Model Identification for the Selected Model

Model selection pointed toward a general heterogeneity model with heterogeneous error variances with $K = 2$ classes. We proceed with estimating the group specific parameters for this model. As the model includes a discrete latent structure, we have to identify a unique labelling subspace to avoid biased estimates of the group specific parameters $\beta_1^G, .., \beta_K^G, Q_1^G, .., Q_K^G, \eta_1, .., \eta_K$ and S^N. To achieve a unique labelling we apply the method of *Permutation sampling* described in Frühwirth-Schnatter (2001).

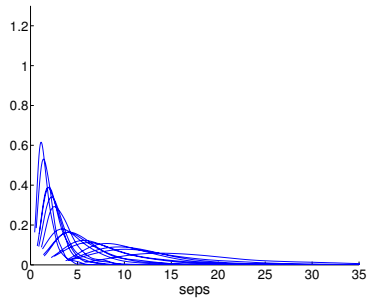

Figure 2: General heterogeneity model based on heterogenous variances with $K = 2$ classes; posterior densities of individual variances for 15 randomly selected consumers

The sampler is restricted to a unique labelling subspace by introducing a constraint $R_g : g(\beta_1^G, Q_1^G, \eta_1) < \ldots < g(\beta_K^G, Q_K^G, \eta_K)$, where g is an appropriate function of the group specifics.

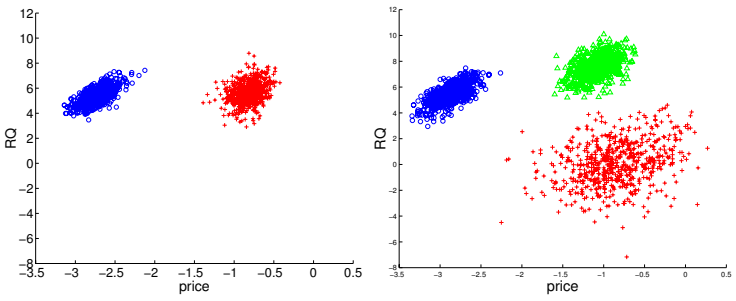

Figure 3: General heterogeneity model based on heterogenous variances with $K = 2$ (left) and based on homogeneous variances with $K = 3$ (right); posterior draws of group specific means obtained under constrained sampling for the *price* against *RQ*

We are now going to illustrate how to achieve a unique labelling for the optimal heterogeneity model. First we analyze the output of the *Random Permutation Sampler* graphically. The *Random Permutation Sampler* explores the unconstrained posterior distribution and samples from each labeling subspace with equal probability $1/K$. This can be seen in Figure 1, where the group specific mean of the *price* is plotted against the one of *RQ*. Though there is no association between individual MCMC chains and group specific parameters by definition of the *Random Permutation Sampler*—estimates from any chain integrate over between group differences—two clusters and possible constraints to separate these may be found by visual inspection. On the left-hand side of Figure 3 we see the output of the model that has been identified by separating the two groups by a restriction on the groups specific *price*

parameter.

In Table 3 we give resulting estimates for the group specific means and the group weights. We have two big groups of nearly equal size, one collecting very price sensitive consumers whereas the consumers of the other group tend to value the "high-image" brands *RQ* and *VOE*. Moreover, they are less price sensitive.

It is interesting to compare these results with the optimal heterogeneity model under homogeneous errors. On the right-hand side of Figure 3 we see the MCMC output of a model that has been identified by separating the first group from the remaining two by the constraint $price_1 < price_{2,3}$ and by dividing the second group from the third one by $RQ_2 > RQ_3$. In Table 4 we give resulting estimates for the group specific means and the group weights. We still have two big groups of nearly equal size, that have a similar meaning as the groups found under heterogeneous errors. The additional group is the smallest one and consists of consumers who are neither price sensitive nor brand conscious.

Table 3: Posterior estimates of the group specific means β_k^G and the group specific weights η_k for a heterogeneity model with heterogeneous errors and $K = 2$ (std. dev. in parentheses)

	β_k^G				β_k^G	
	$K=1$	$K=2$			$K=1$	$K=2$
$const$	14.78	12.43		$RQ \cdot p$	-0.71	-0.04
	(0.67)	(0.75)			(0.16)	(0.15)
RQ	5.44	5.65		$VOE \cdot p$	-0.85	-0.02
	(0.65)	(0.84)			(0.16)	(0.16)
VOE	5.3	5.17		$JU \cdot p$	-0.38	0.07
	(0.65)	(0.97)			(0.16)	(0.16)
JU	1.28	0.38		$WA \cdot p$	-0.58	-0.1
	(0.66)	(0.97)			(0.15)	(0.13)
WA	2.24	1.1				
	(0.68)	(0.78)				
p	-2.72	-0.82		$E(\eta_k \mid y^N)$		
	(0.15)	(0.15)				
p^2	-0.03	0			0.58	0.42
	(0.07)	(0.06)			(0.04)	(0.04)

4 Summary and Outlook

The purpose of this paper has been to illustrate a hierarchical parametric approach toward capturing unobserved heterogeneity based on mixtures of random-effects models and its practical implementation based on a fully Bayesian approach. A careful evaluation of various special cases of this general model for a conjoint study in the mineral water market demonstrated first, that strong preference heterogeneity is present, second that the hierarchical Bayes model as well as the latent class model are outper-

Table 4: Posterior estimates of the group specific means β_k^G and the group specific weights η_k for a heterogeneity model with homogeneous errors and $K = 3$ (std. dev. in parentheses)

	$E(\beta_k^G \mid y^N)$				$E(\beta_k^G \mid y^N)$		
Effect	k=1	k=2	k=3	Effect	k=1	k=2	k=3
const	14.99	12.16	13.38	RQ·p	–0.78	–0.29	0.11
	(0.79)	(0.79)	(1.49)		(0.20)	(0.22)	(0.46)
RQ	5.45	7.57	0.17	VOE·p	–0.89	–0.29	0.49
	(0.77)	(0.85)	(1.72)		(0.20)	(0.24)	(0.54)
VOE	5.23	6.91	–0.46	JU·p	–0.54	0.16	–0.18
	(0.78)	(0.94)	(2.08)		(0.21)	(0.21)	(0.51)
JU	1.83	0.02	1.76	WA·p	–0.67	–0.08	–0.38
	(0.81)	(0.94)	(1.64)		(0.18)	(0.18)	(0.38)
WA	2.35	1.09	2.66				
	(0.81)	(0.92)	(1.72)				
p	–2.87	–1.09	–0.85	$E(\eta_k \mid y^N)$			
	(0.18)	(0.18)	(0.42)				
p^2	0.01	–0.08	–0.15		0.46	0.44	0.10
	(0.09)	(0.07)	(0.18)		(0.05)	(0.05)	(0.03)

formed by the mixture approach in their ability to capture this heterogeneity, and third that also the observation variances are definitely heterogeneous.

Focus of this contribution has been entirely on capturing cross-sectional heterogeneity. Otter et al. (2004a) extend the hierarchical parametric approach based on mixtures of random-effects models in order to deal both with cross-sectional as well as longitudinal unobserved heterogeneity. For alternative work on capturing longitudinal heterogeneity we refer to the contribution of Achim Zeileis in this volume.

A certain disadvantage of a mixture of random-effects models is that representation of heterogeneity in terms of the variance-covariance matrix Q_k^G is rather highly parameterized, especially for high-dimensional preference vectors. A further interesting line of research is to find a representation of heterogeneity, that is more parsimonious than a fully unrestricted covariance matrix Q_k^G. First ideas on how to achieve a more parsimonious representation of heterogeneity based on using Bayesian variable selection ideas on the Cholesky factors of Q_1^G are found in Tüchler and Frühwirth-Schnatter (2003).

Bibliography

Allenby, G. M., Arora, N., and Ginter, J. L. (1998). On the heterogeneity of demand. *Journal of Marketing Research*, 35:384–389.

Celeux, G., Hurn, M., and Robert, C. P. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 95(451):957–970.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90:1313–1321.

Dimitriadou, E., Weingessel, A., and Hornik, K. (2001). Voting-merging: An ensemble method for clustering. In Dorffner, G., Bischof, H., and Hornik, K., editors, *Artificial Neural Networks – ICANN 2001*, volume 2130 of *LNCS*, pages 217–224. Springer, Berlin.

Dolničar, S. and Leisch, F. (2003). Winter tourist segments in Austria: Identifying stable vacation styles using bagged clustering techniques. *Journal of Travel Research*, 41(3):281–292.

Frühwirth-Schnatter, S. (1995). Bayesian model discrimination and Bayes factors for linear Gaussian state space models. *Journal of Royal Statistical Society, Series B*, 57:237–246.

Frühwirth-Schnatter, S. (2001). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96(453):194–209.

Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *The Econometrics Journal*, 7:143–167.

Frühwirth-Schnatter, S. and Otter, T. (1999). Conjoint-analysis using mixed effect models. In Friedl, H., Berghold, A., and Kauermann, G., editors, *Statistical Modelling. Proceedings of the Fourteenth International Workshop on Statistical Modelling*, pages 181–191, Graz.

Frühwirth-Schnatter, S., Tüchler, R., and Otter, T. (2004). Bayesian analysis of the heterogeneity model. *Journal of Business & Economic Statistics*, 22(1):2–15.

Lenk, P. J. and DeSarbo, W. S. (2000). Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika*, 65(1):93–119.

Meng, X.-L. and Wong, W. H. (1996). Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statistica Sinica*, 6:831–860.

Otter, T., Frühwirth-Schnatter, S., and Tüchler, R. (2004a). Unobserved preference changes in conjoint analysis. Technical report, Wirtschaftsuniversität Wien.

Otter, T., Tüchler, R., and Frühwirth-Schnatter, S. (2002). Bayesian latent class metric conjoint analysis – A case study from the Austrian mineral water market. In Opitz, O. and Schwaiger, M., editors, *Exploratory Data Analysis in Empirical Research*, Studies in Classification, Data Analysis and Knowledge Organization, pages 157–169. Springer, Berlin.

Otter, T., Tüchler, R., and Frühwirth-Schnatter, S. (2004b). Capturing consumer heterogeneity in metric conjoint analysis using Bayesian mixture models. *International Journal of Marketing Research*, 21(1):285–297.

Stephens, M. (2000). Dealing with label switching in mixture models. *Journal of Royal Statistical Society, Series B*, 62(4):795–809.

Tanner, M. and Wong, W. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82:528–540.

Tüchler, R. and Frühwirth-Schnatter, S. (2003). Bayesian parsimonious estimation of observed and unobserved heterogeneity. In Verbeeke, G., Molenberghs, G., Aerts, M., and Fieuws, S., editors, *Statistical Modelling in Society. Proceedings of the 18th International Workshop on Statistical Modelling*, pages 427–431, Leuven, Belgium.

Tüchler, R., Frühwirth-Schnatter, S., and Otter, T. (2002). The heterogeneity model and its special cases – an illustrative comparison. In Stasinopoulos, M. and Touloumi, G., editors, *Statistical Modelling in Society. Proceedings of the Seventeenth International Workshop on Statistical Modelling*, pages 637–644, Chania, Greece.

Verbeke, G. and Lesaffre, E. (1996). A linear mixed-effects model with heterogeneity in the random-effects population. *Journal of the American Statistical Association*, 91:217–221.

Zellner, A. (1971). *An Introduction to Bayesian inference in Econometric*. Wiley, New York.

Part II

Modeling Financial Markets

Non-linear Volatility Modeling in Classical and Bayesian Frameworks with Applications to Risk Management

Tatiana Miazhynskaia, Engelbert Dockner, Sylvia Frühwirth-Schnatter and Georg Dorffner

1 Introduction

Modeling and forecasting volatility of financial time series has become a popular research topic for the last several years. There are two main reasons for this development. The first is the rapid growth of financial derivatives that require volatility forecasts to calculate fair prices. The second arises from today's new global financial architecture that places more emphasis on measuring and managing financial market risks. Consequently, following the regulations of the Bank for International Settlements, many of the banks (and other financial institutions) have to measure their market risks on a regular basis. The need to improve the management of financial risks has also led to a uniform measure of risk called Value-at-Risk (VaR). The VaR of a portfolio of risky assets is the maximum potential loss of this portfolio for a given horizon and a given loss-probability. Increasing availability of financial data and rapid advances in computer technology have spurred the development of various VaR models that are applied in risk management.

To quantify the VaR of a position we need to calculate the corresponding quantile of the returns distribution. If the returns distribution satisfies specific assumptions the estimation of the VaR requires the estimation of the conditional standard deviation only. In case of normal and t-distributions the VaR of a portfolio is determined by a forecast of the standard deviation of the portfolio returns. Hence many VaR models are directly linked to modelling conditional variances.

The most famous model of this type, widely used in practice, is the GARCH model (Bollerslev, 1986) where conditional variances are governed by a linear autoregressive process of past squared returns and variances. This model captures several "stylized facts" of asset return series such as heteroskedasticity (time-dependent conditional variance), volatility clustering and excess kurtosis. Later studies (e.g., Nelson, 1991; Glosten et al., 1993; Alles and Kling, 1994; Hansen, 1994) have found that there exist additional empirical regularities that can not be described by the classical GARCH model, such as the leverage effect, negative skewness, or fat tails of conditional distributions. While traditional GARCH models only allow for constant higher order moments, Alles and Kling (1994) demonstrated that for different financial series higher order moments are time-varying.

All these insights led to the development of an enormous number of models generalizing the classical GARCH model in some directions. One possible approach is to allow for non-linear dependencies in conditional varainces. A nice review of nonlinear approaches can be found in Swanson and Franses (1999). But in spite of the wide class

of non-linear models, there is still no clear answer about the importance of non-linear dependencies in volatility modelling. As a central theme of our SFB project we have tried to address the question of "how much" non-linearity is included in financial data.

As a tool for non-linear regression we used neural network-based (NN) modelling, so called recurrent mixture density networks, describing the conditional mean and variance by multi-layer perceptrons (the same approach was applied by Schittenkopf et al., 1999, 2000; Bartlmae and Rauscher, 2000). For NN modelling, the conditional moments can be approximated with an arbitrary accuracy if the size of the neural network models is not restricted (Hornik et al., 1989).

NN modelling provides a very general framework and has become a rather popular methodology in financial modelling. To add to the literature from above, we mention more recent papers such as Boero and Cavallil (1997), Dunis and Jalilov (2002), González and Burgess (1997), Poh et al. (1998), Yao et al. (2000), where the NN approach was found to be very useful. As a semi-parametric non-linear model, neural networks have the following important advantages over the more traditional parametric models. They do not rely on restrictive parametric assumptions such as normality, stationarity, or sample-path continuity, and they are robust to specification errors plaguing parametric models. Moreover, NN models are sufficiently flexible and can easily encompass a wide range of securities and fundamental asset price dynamics.

In addition to the linearity issue of conditional variances we concentrate also on the modelling of conditional distributions. We compare three different density specifications: 1) the standard GARCH gaussian model and its non-linear generalization using a *normal distribution*; 2) the GARCH model and its non-linear neural network generalization with a *Student's t-distribution*; and 3) linear and non-linear recurrent mixture density models, which approximate the conditional distributions by a *mixture of Gaussians* (two components). All models allow for heteroskedastic data, while the model with t-distribution permits conditional leptokurtosis. But only the linear and non-linear mixture models allow the higher moments to be time varying.

To check how stable the relative performance of our models is, the empirical analysis is based on stock index return series from different financial markets. We used return series of the Dow Jones Industrial Average (USA), FTSE 100 (Great Britain) and NIKKEI 225 (Japan) over a period of more than 12 years.

The purpose of this paper is to empirically compare the performance of linear and non-linear NN models under different conditional density specifications. This leads us to the important question of how to characterize the performance of a model, i.e., which performance measure to use. We apply traditional measures but introduce model selection on the basis of the costs of value at risk forecasts as well as on the basis of Bayesian inference.

A fundamentally different approach to model selection is provided by the Bayesian methodology. Point estimates for parameters are replaced by distributions in the parameter space, which represent our knowledge or belief about the value of the parameters. The Bayesian framework allows different models to be compared using only the training data. Moreover, the Bayesian approach is more powerful with complex models than is the maximum likelihood one. In the ML-approach the problems of local optima and over-fitting sometimes lead to rather doubtful estimators and, conse-

quently, to unclear results in model performance. The principle to avoid unnecessary complexity is implicitly embodied in the Bayesian framework.

Our paper is divided into two main parts. In the first we compare models under the maximum likelihood approach. The models are evaluated with respect to their likelihood as well as with respect to their volatility forecasting performance. Due to their length, the original return series are split into several parts and each of the models is estimated separately on every part. Thus, we can not only compare the models with respect to performance results but also apply statistical tests to find out whether the differences in performance are significant. In that respect we continue the work of Schittenkopf et al. (1999) and Schittenkopf et al. (2000), comparing non-linear and non-gaussian volatility models for data from different financial markets. Moreover, we derive dynamical VaR predictions by each of our models for three different homogeneous portfolios. To evaluate the quality and accuracy of these VaR models we apply a number of standard statistical back-testing procedures. Additionally, we check the efficiency of VaR measures on the basis of economic costs resulting from VaR predictions together with corresponding capital requirements.

In the second part of the paper, we apply Bayesian analysis to our models and perform model comparison based on posterior model probabilities. The literature on Bayesian analysis applied to NN models is relatively thin. We mention the approach in MacKay (1992), based on Gaussian approximations, and the hybrid Monte Carlo algorithm applied by Neal (1996). Posterior inference in NNs is plagued by multimodality issues. Besides trivial multimodality due to relabeling, there is inherent multimodality due to non-linearity. As a consequence, there is little hope for the normal approximation with these models and we need to turn to simulation methods. First attempts to apply the hybrid MC to our models show its practical limitations because of the recurrent structure in the variance equation and consequently, rather expensive computations of the energy gradient are necessary. Thus, we mainly adopt the approach of Müller and Insua (1998) combining Metropolis-Hastings (MH) steps for simulating model parameters with Gibbs sampling of hyper-parameters.

The rest of the paper is organized as follows. In the next two sectionS we present the models and the data that are used in the empirical analysis. Sections 4 is devoted to the maximum likelihood framework, including model evaluation based on VaR applications. In the section 5 we discuss the Bayesian model selection issues. Finally, Section 6 concludes the paper.

2 Description of Models

As a benchmark we use the classical GARCH(1,1) model (Bollerslev, 1986) with conditional normal distribution and an AR(1) process for the mean equation of the returns r_t, i.e.

$$\begin{aligned} r_t &= \mu_t + e_t, \quad e_t \sim \mathbf{N}(0, \sigma_t^2), \\ \mu_t &= a_1 r_{t-1} + a_0, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 e_{t-1}^2 + \beta_1 \sigma_{t-1}^2. \end{aligned}$$

One possible extension of this GARCH model is to substitute the conditional normal distribution by a Student's t distribution with ν degrees of freedom in order to

allow for excess kurtosis.

The second direction of the extension of the classical GARCH model is to allow for non-linear dependencies in the conditional mean and in the conditional variance. As a tool for non-linear regression we use neural network-based modelling by means of a recurrent mixture density network that describes conditional mean and variance by multi-layer perceptrons (MLP) (see Bishop, 1994; Schittenkopf et al., 2000, for a detailed discussion).

In the simplest case a MLP with one input unit, one layer of hidden units and one output unit is defined by the mapping

$$\tilde{f}(x_t) = g\left(\sum_{j=1}^{H} v_j h(w_j x_t + c_j) + s x_t + b\right), \tag{1}$$

where H denotes the number of hidden units, w_j and v_j the weights of the first and second layer, s the shortcut weight, and c_j and b the bias weights of the first and second layer. In general, the activation function h of the hidden units is chosen to be bounded, non-linear, and increasing as, e.g., the hyperbolic tangent or logistic sigmoid function. The activation function of the output unit may be unrestricted, e.g. $g(x) = x$. In general, a MLP can approximate any smooth, non-linear function with arbitrary accuracy as the number of hidden units tends to infinity (Hornik et al., 1989). In such a way, the MLP can be interpreted as a non-linear autoregressive model of first order and can be applied to predict the parameters of a conditional density of the return series.

Recurrent mixture density network models RMDN(n) approximate the conditional distributions of returns by a mixture of n Gaussians:

$$\rho(r_t|I_{t-1}) = \sum_{i=1}^{n} \pi_{i,t} k(\mu_{i,t}, \sigma_{i,t}^2), \tag{2}$$

where $k(\mu_{i,t}, \sigma_{i,t}^2)$ is the Gaussian density and the parameters $\pi_{i,t}$, $\mu_{i,t}$, and $\sigma_{i,t}^2$ of the n Gaussian components are estimated by three MLPs. The MLPs estimating the priors and the centers are standard MLPs (1) with r_{t-1} as input. The MLP estimating the variances is recurrent and has the form

$$\sigma_{i,t}^2 = \sum_{j=1}^{H} v_{ij} h\left(w_j e_{t-1}^2 + \sum_{k=1}^{n} \gamma_{jk} \sigma_{k,t-1}^2 + c_j\right) + s_{i0} e_{t-1}^2 + \sum_{k=1}^{n} s_{ik} \sigma_{k,t-1}^2 + b_i. \tag{3}$$

The activation function h of the hidden units is chosen to be hyperbolic tangent.

The most appealing feature of RMDN models is the time-dependence of the higher-order moments (skewness and kurtosis), that is in contrast to the properties of GARCH and GARCH-t models.

We note that an RMDN model with one Gaussian component ($n = 1$) can be interpreted as a non-linear extension of a GARCH model.

There are two other models that must be introduced in order to analyze the influence of linear and non-linear functions and density specifications on the perfor-

mance of return series models (Schittenkopf et al., 2000). The first one is the non-linear GARCH model in the framework of a RMDN with a t-distribution replacing the weighted sum of normal densities in (2). These models will be called RMDN(n)-t models. The second one is the nonlinear GARCH model in which only linear functions are allowed. More precisely, in all three MLPs estimating the parameters of the mixture model the activation function h of the hidden units are supposed to be linear. These linear mixture models are referred to as LRMDN(n) models in the following. LRMDN(1) is again the classical GARCH model. We limited ourselves to the cases $n = 1$ and $n = 2$, mainly focusing on the non-linearity aspects.

Altogether, we concentrate on six models according to two dimensions: linearity and distribution:

type of distribution	linear	non-linear
Gaussian	**GARCH(1,1)**	**RMDN(1)**
t distribution	**GARCH(1,1)-t**	**RMDN(1)-t**
mixture of Gaussians	**LRMDN(2)**	**RMDN(2)**

3 Data Sets

In our numerical experiments we used three data sets related to different financial markets:

1. daily closing values of the Dow Jones Industrial Average (DJIA);
2. daily closing values of the FTSE 100 traded at the London Stock Exchange;
3. daily closing values of the Japanese index NIKKEI 225.

The index series were taken from public sources. The sample period for all data sets was 13 years from 1985 to 1997. All data were transformed into continuously compounded returns r_t (in percent).

In order to take care of stationarity issues and increase the reliability of the empirical analysis, all time series were divided into overlapping segments of a fixed length of 700 trading days, where the first 500 returns of each segment form a training set, the next 100 returns form a validation set and the remaining 100 returns are used for testing. The training sets are used to optimize the parameters of each model. The validation sets are used for an "early stopping" strategy to avoid overfitting for the neural network models and independent test sets are used for out-of-sample model performance evaluation. The test sets are not overlapping.

4 Maximum Likelihood Framework

A comprehensive presentation of the results in this section can also be found in the technical reports Miazhynskaia et al. (2003a,b).

4.1 Estimation of Models

We fitted GARCH(1,1), RMDN(1), GARCH(1,1)-t, RMDN(1)-t, LRMDN(2) and RMDN(2) models to each of the training sets separately. The number of optimized parameters of a particular model is 5 for the GARCH(1,1) model, 26 for the RMDN(1), 6 for the GARCH(1,1)-t, 27 for the RMDN(1)-t, 16 for the LRMDN(2), and 54 for the RMDN(2). The number of hidden units of the MLPs in the RMDN-models was chosen to be $H = 3$. The parameters of all models were optimized with respect to the average negative log likelihood of the sample

$$\mathcal{L} = -\frac{1}{N}\sum_{t=1}^{N} \log \rho(r_t|I_{t-1}),$$

where N denotes the sample size and $\rho(r_t|I_{t-1})$ is the conditional probability density function of the corresponding distribution. We refer to $\mathcal{L}$ as the *loss function* of a data set, since we will make use of values of $\mathcal{L}$ calculated for data sets which were not used to estimate the model parameters.

The optimization routine was a scaled conjugate gradient algorithm. We performed optimization of RMDN models with several parameter initializations in an attempt to approach a global optimum. For the models with t distribution, the degrees-of-freedom parameter was additionally optimized by a one-dimensional search routine.

Since the main goal of this work is out-of-sample diagnostic, i.e., comparison of model performance on a non-used data set (test set), we are interested in obtaining models with optimal generalization performance. However, all standard neural network architectures such as the fully connected multi-layer perceptron are prone to overfitting (see, e.g., Geman et al., 1992; Reed, 1993): while the network seems to become better and better, i.e., the error (in our case - the value of the loss function) on the training set decreases. In order to prevent the RMDN models from overfitting the training data, the generalization error is estimated by the performance of the model on a validation set and an "early stopping" strategy (Prechelt, 1998) is applied . More precisely, the model parameters are optimized with respect to the loss function on the training set and after each iteration the loss function on the validation set is calculated. Finally, the RMDN model on the optimization iteration t^* is selected, where

$$t^* = \arg\min_{t_0<t<T} \mathcal{L}_{\text{validation}}(t),$$

with t as an iteration number; T is the number of all iterations performed; and t_0 is the minimal iteration number chosen to avoid artefact behaviour of the loss function on the validation set in such a way that the parallel value of the loss function on the training set of a simpler (less parametrized) model is beaten.

4.2 Out-of-Sample Loss Function Performance

We investigated out-of-sample performance of the models, i.e. error values on data sets disjoint from the training data. The parameters of the models were estimated by the procedure described above using training and validation sets and then, keeping the

parameters fixed, we computed the error measures on the test sets of the corresponding segments. Favor to the out-of sample criterion for model comparison was given because in this case possible overparametrizations may be neglected.

In spite of the numerous different starting values in the optimization procedure for the neural network models, for single cases we obtained values for the loss function that were three to five times larger than the average level. Based on the smooth behaviour of the loss function for other models on the same test set, we considered such models to be "non-indicated" over-fitting cases and deleted these test sets parallel for all data sets from the analysis. After elimination we were left with 24 test sets for model evaluation.

The performance of the models on each of the test sets for the DJIA data with respect to the loss measure is summarized in Fig.1. For convenience all the results are presented with respect to the functional form of the conditional variance equation (linearity versus non-linearity) and the conditional distribution specification. We compare the performance of the Gaussian model with that having a t distribution and the mixture of Gaussians. Thus, three lines in the upper panel of the figure give the values of the relevant statistic for the linear models GARCH(1,1), GARCH(1,1)-t and LRMDN(2). The bottom panels present non-linear models RMDN(1), RMDN(1)-t and RMDN(2). Based on Fig.1, we can draw the following preliminary conclusion: in general, the differences between the models over the most test sets are negligible. On single sets (test set 5,6 and 10-12) the linear and non-linear Gaussian models show the worse results, while the models with Student's t distribution exhibit the smallest likelihood values. If we compare the upper and the lower plots in Fig.1, it is obvious that the linear models and their non-linear neural network generalizations reach equal likelihoods. Single cases, like test set 11, where the non-linear RMDN(1) model shows a loss value close to 1.6 against 1.2 for the linear GARCH(1,1), and the sets 23-24, where the non-linear mixture density RMDN(2) behaves significantly worse than its linear version, are due to the problems with the maximum likelihood estimation of the non-linear models.

In order to be statistically consistent in the model selection process, we tested the hypothesis of higher/lower errors by performing parametric and nonparametric tests. More precisely, we performed a paired t-test and a matched pairs signed rank Wilcoxon test (paired Wilcoxon test) for our loss measures. The application of the paired tests is appropriate for the following reasons: The error measures of each model vary considerably with the actual segment of the underlying return series but the differences between the error measures of different models are rather small. Therefore the differences can only be detected if a paired test which takes into account the correlations between the error measures, is applied. Additionally, for the paired t-test it is assumed that the differences are normally distributed which is not always the case and whereas for the paired Wilcoxon test it is only assumed that the distribution of the differences is symmetric. Because of this fact our conclusions are mostly based on the results of the paired Wilcoxon test.

The results of the paired test for the DJIA return series are summarized in Table 1. The column "mean" gives the mean value of the corresponding statistic over all test sets. The minimal mean value of the loss function 1.184 is reached by the

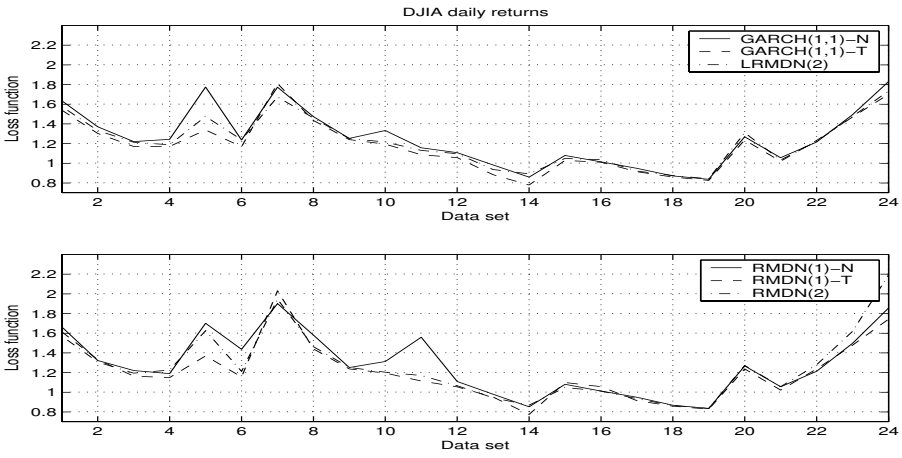

Figure 1: DJIA: The loss function values for the linear (in the upper figure) and non-linear (in the lower figure) models with different conditional distributions.

Table 1: DJIA daily returns: Loss function statistics. Mean values (second column), p-values for the paired t-tests (above the diagonal) and p-values for the paired Wilcoxon signed rank tests (below the diagonal).

Model	Mean	1	2	3	4	5	6
1: GARCH(1,1)	1.249	-	0.165	0.002	0.073	0.019	0.778
2: RMDN(1)	1.278	0.710	-	0.001	0.012	0.009	0.385
3: GARCH(1,1)-t	1.184	0.000	0.000	-	0.025	0.007	0.004
4: RMDN(1)-t	1.209	0.004	0.004	0.015	-	0.776	0.050
5: LRMDN(2)	1.214	0.008	0.010	0.000	0.103	-	0.090
6: RMDN(2)	1.255	0.265	0.230	0.000	0.032	0.391	-

GARCH(1,1)-t model. The p-values of both paired tests between this model and all other models are less than 0.025, indicating that these differences are significant, i.e. GARCH(1,1)-t significantly outperforms all other models. Its non-linear generalization RMDN(1)-t is among the best when the value of the loss function is considered, but the p-value of the Wilcoxon test for the differences between this model and the linear mixture density model LRMDN(2) is 0.103 (or even 0.776 by t-test). Hence, the performance of the RMDN(1)-t does not differ statistically much from the LRMDN(2) performance over all test sets. The third group of models consists of both Gaussian and non-linear mixture models. The performance of this group with respect to the loss values is the worst. It seems that on average linearity plays some positive role since linear models have smaller values for the loss function compared to their non-linear counterparts, but the differences are mostly not significant (the p-values

between linear and non-linear models are 0.710, 0.015 and 0.319 for the Gaussian, Student-t and the mixture of Gaussians conditional distributions, respectively).

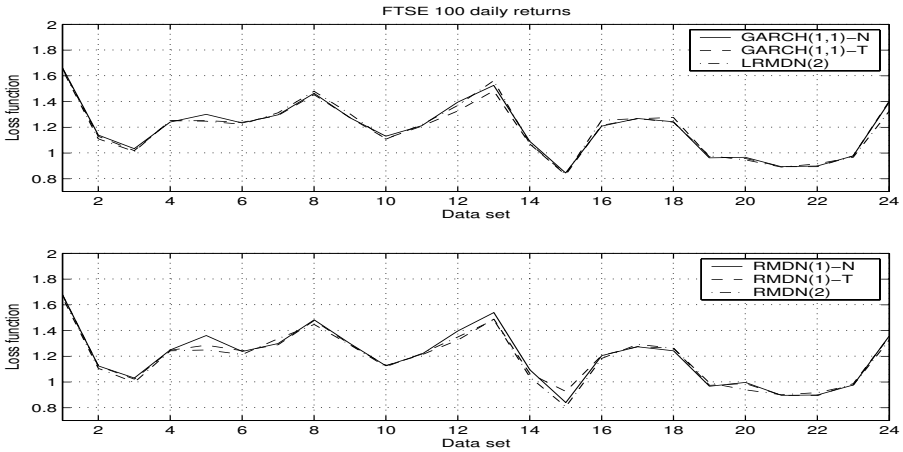

Figure 2: FTSE 100: The loss function values for the linear (in the upper panel) and non-linear (in the lower panel) models with different conditional distributions.

Table 2: FTSE 100 daily returns: Loss function statistics. Mean values (second column), p-values for the paired t-tests (above the diagonal) and p-values for the paired Wilcoxon signed rank tests (below the diagonal).

Model	Mean	1	2	3	4	5	6
1: GARCH(1,1)	1.189	-	0.227	0.028	0.389	0.457	0.612
2: RMDN(1)	1.217	0.153	-	0.125	0.830	0.190	0.214
3: GARCH(1,1)-t	1.179	0.012	0.007	-	0.250	0.111	0.083
4: RMDN(1)-t	1.215	0.831	0.059	0.048	-	0.325	0.364
5: LRMDN(2)	1.184	0.648	0.107	0.236	0.394	-	0.635
6: RMDN(2)	1.187	0.855	0.094	0.107	0.927	1.000	-

Table 2 together with Fig.2 show the results for FTSE 100 returns with respect to the loss function. The graphical plots give no preferences to any model or any class of models. With respect to the average statistics, the linear models GARCH(1,1)-t and the mixture LRMDN(2) outperform all other models, but their advantage appeared to be not significant comparing with all other models (the p-values of the Wilcoxon test between LRMDN(2) and all other models are above 0.107). Moreover, according to the paired tests, there is no statistical difference between the linear and non-linear mixture models at all (the corresponding p-value of the Wilcoxon test is 1.00). The non-linear model with Student-t shows the lowest mean value of the loss function over all test sets.

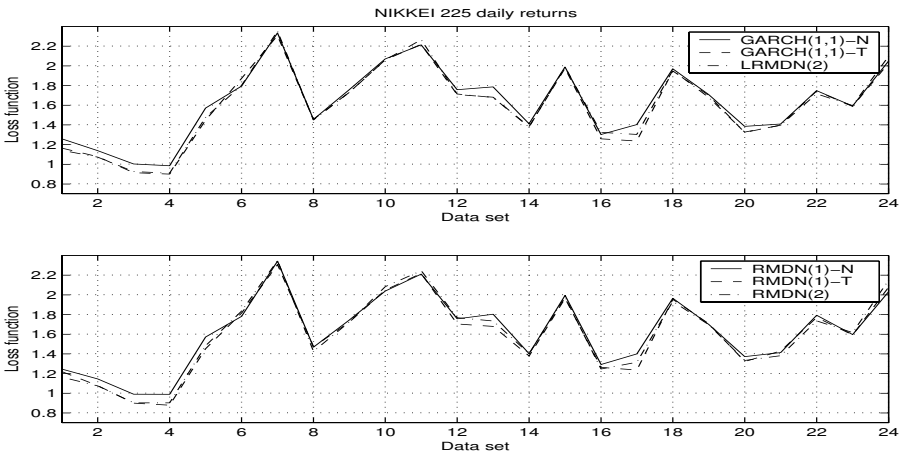

Figure 3: NIKKEI 225: The loss function values for the linear (in the upper panel) and non-linear (in the lower panel) models with different conditional distributions.

Model	Mean	1	2	3	4	5	6
1: GARCH(1,1)	1.598	-	0.643	0.000	0.001	0.002	0.011
2: RMDN(1)	1.597	0.367	-	0.001	0.002	0.004	0.022
3: GARCH(1,1)-t	1.557	0.000	0.001	-	0.531	0.207	0.058
4: RMDN(1)-t	1.559	0.002	0.003	0.840	-	0.421	0.124
5: LRMDN(2)	1.565	0.004	0.007	0.253	0.253	-	0.300
6: RMDN(2)	1.571	0.016	0.021	0.174	0.109	0.242	-

Table 3: NIKKEI 225 daily returns: Loss function statistics. Mean values (second column), p-values for the paired t-tests (above the diagonal) and p-values for the paired Wilcoxon signed rank tests (below the diagonal).

Out-of-sample diagnostics for the NIKKEI 225 series are given in Fig.3 and Table 3. As in the case with FTSE 100 data, the graphical plot of the likelihood values over all test sets in Fig.3 gives no clear preferences to any model. But the results of the paired statistical tests with respect to the loss function show a dominance of the linear and non-linear models with t-distribution. The mixture models are ranked next, but their performances relative to the models with t-distribution are not significantly worse (the corresponding p-values are more than 0.109). But both gaussian models again show the lowest efficiency with respect to all error measures.

4.3 VaR Application

After having evaluated the statistical performance of the different models we are going to apply them to a standard risk management problem.

A primary tool for financial risk assessment is the Value-at-Risk (VaR) method-

ology, where VaR is a measure of the maximum potential loss of a portfolio with a given probability over a pre-specified horizon. As soon as the probability distribution of the returns is known, a VaR can be calculated using the $(100-p)\%$ percentile r_p^* of this distribution. Hence, VaR = today's price $\cdot$ $(\exp(r_p^*)-1)$. For a more detailed discussion on VaR see, for instance, Dowd (1998) and Duffie and Pan (1997).

Our analysis proceeds in the following way: the parameters of each model are fixed within every segment and we compute a forecast of tomorrow's return distribution as well as the corresponding VaR estimates given the past data for every point in the test part of this segment. We chose $p = 1\%$ so that the significance level is 99% and compute daily VaR for an investment in the three indices. In such a way, we get a VaR series for the whole data samples. Comparing the realized losses with the VaR estimates, we determine the indicator variable θ_t as the outcome of a binomial event: either the one-day actual loss L_t is less than the VaR_t estimates (a success), or the actual exceeds the loss estimates (a failure), i.e.

$$\theta_t = \begin{cases} 1, & if\ L_t < \text{VaR}_t, \\ 0, & \text{otherwise.} \end{cases} \tag{4}$$

We assess now the quality and accuracy of the VaR predictions of our six models.

Backtesting Evaluation of VaR Forecasts

The first group of tests is based on statistical evaluation of the indicator series $\{\theta_t\}$. Such testing is often referred to as "back-testing" (Dowd, 1998).

Test 1: Basle Traffic Light. This is the back-testing framework developed by the Basle Committee on Banking Supervision. Any bank must hold regulatory capital against its market risk exposure. These capital charges are based on VaR estimates generated by the banks' own VaR models and a multiplication factor defined by supervising authorities according to the traffic light concept of the Basle Committee on Banking Supervision. According to this concept, internal banks' models are classified into three zones. The classification into green, yellow or red zones depends on how often the actual losses exceed the daily 99% VaR predictions over a period of n trading days. Based on such classification, the necessary capital reserves are assigned. The green zone implies that a multiplication factor of 3 is applied to the VaR value, yellow results in a higher (add-on) factor between 3 and 4, while red means rejection of the model.

Our back-testing period covers the sample period with $n = 2300$ that exceed the 250 days that are typically used in practice. The results of the hypothetical classifications are listed in Table 4.

All the models for FTSE 100 data are in acceptable zones. For other markets the models with t-distribution together with mixture density networks yield the most reliable risk estimates. Among the non-gaussian models, the linear mixture density network performs best. The gaussian models are mostly rejected. This test also indicates the bad performance of the RMDN(2) model on the DJIA data set.

Table 4: Classification according to the Basle rules. The column "fails" lists the number of failures in the whole sample for the corresponding index.

	DJIA		FTSE 100		NIKKEI 225	
Model	fails	zone	fails	zone	fails	zone
GARCH(1,1)	45	red	32	yellow	39	yellow
RMDN(1)	48	red	26	green	43	red
GARCH(1,1)-t	30	green	29	green	35	yellow
RMDN(1)-t	39	yellow	25	green	32	yellow
LRMDN(2)	30	green	27	green	27	green
RMDN(2)	44	red	30	green	34	yellow

Test 2: Proportion of Failures. Kupiec (1995) presents a more sophisticated approach to the analysis of exceptions based on the observation that a comparison between daily profit or loss outcomes and the corresponding VaR measures give rise to a binomial experiment. The outcomes of the binomial events θ_t in (4) are distributed as a series of draws from an independent Bernoulli distribution and the verification test is based on the proportion of failures (PF) in the sample. For more details, see Kupiec (1995).

Table 5 summarizes the performance of our models with respect to the proportion of failures test. The column denoted by "failures" lists the number of failures in the whole sample for the corresponding index. The next column shows whether the null hypothesis H0 : $p^* = 0.01$ can be rejected at the 5% significance level. The results shown in the table are consistent with the previous test.

Table 5: PF test. The largest number of failures that could be observed in the samples without rejecting the null H0 at the 5% confidence level is 32 for a sample size of 2300 points.

	DJIA		FTSE 100		NIKKEI 225	
Model	fails	H0: $p^* = 0.01$	fails	H0: $p^* = 0.01$	fails	H0: $p^* = 0.01$
GARCH(1,1)	45	rejected	32	not rejected	39	rejected
RMDN(1)	48	rejected	26	not rejected	43	rejected
GARCH(1,1)-t	30	not rejected	29	not rejected	35	rejected
RMDN(1)-t	39	rejected	25	not rejected	32	not rejected
LRMDN(2)	30	not rejected	27	not rejected	27	not rejected
RMDN(2)	44	rejected	30	not rejected	34	rejected

Economic Costs of VaR Forecasts

The common downside of the tests above is that all of them only count the number of violations of the actual loss with respect to the VaR forecast. Hence, a model with

many small violations will be rejected while a model with few large violations will be accepted. Moreover, financial institutions prefer VaR models that are not only able to pass a back-testing procedure but that provide small VaR predictions. Otherwise banks have to hold to much risk capital. Therefore, to check the efficiency of the VaR measure we developed a new test, providing a quantitative basis for the incorporation of VaR prediction into regulatory capital requirements.

Any financial institution must hold regulatory capital to cover its potential market risk exposure. We used dynamically computed daily VaR estimates generated by our VaR models and assumed for simplicity that capital reserves equal to the VaR estimates will be held for 1 day. When the actual portfolio loss L_t does not exceed the predicted loss for this day, the bank only faces opportunity costs by not being able to properly invest the capital held to satisfy the capital requirements. We compute these costs as lost interest yield $\text{VaR}_t \cdot (e^{i/250 \cdot 1} - 1)$ with some interest rate i. In the case that the portfolio loss L_t is greater than the VaR_t estimates, banks need additional capital and face capital charges as a penalty, so that the lost interest yield = $\text{VaR}_t \cdot (e^{i/250 \cdot 1} - 1) +$ Penalty, where, e.g., Penalty = $1.2 \cdot (L_t - VaR_t) \cdot e^{i/250 \cdot 1}$ to cover higher transaction costs of capital transfers.

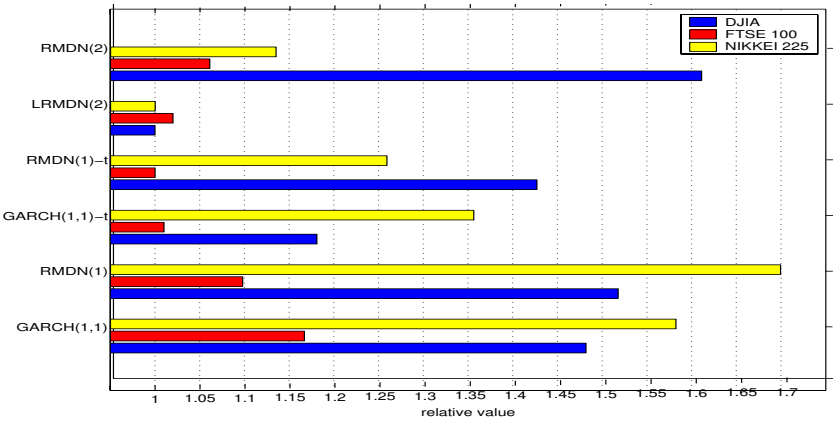

Figure 4: Lost Interest Yield

According to this strategy, we calculated the lost yield over the entire test period and then scaled it to remove the dependency on the portfolio size. These relative lost interest yields are depicted in Fig 4.

For all markets the best model is the linear mixture density network. In general, this test clearly rejects linear and non-linear gaussian models and favors the mixture density models over the models with t-distributions.

5 Bayesian Approach

In this section we consider our models in the Bayesian framework. To make the posterior simulation process easier, we simplify the model specifications. Since we con-

centrate on non-linearity issues in volatility modelling, we assume an AR(1) process for the conditional mean for all the models. Moreover, for the mixture models we use constant mixture proportions and in equation (3) the mixture variance as lagged variance is taken. We do not fix the number of hidden nodes H and investigate the "degree of non-linearity" in the data, varying $H = 1, 2, 3$. As activation functions we take the logistic for hidden nodes function and the unrestricted activation function for the output unit, assuming positivity of the parameters $(v_{ij}, s_{i0}, s_{i1}, b_i)$ to guarantee the positivity of the predicted variances.

We start the Bayesian inference with defining a prior distribution over the model parameters. The next step is to apply Markov chain Monte Carlo (MCMC) simulations to obtain posterior distributions of the parameters. And, finally, we apply a Bayesian models selection method (bridge sampling) to compute the model likelihoods and posterior model probabilities, to do model selection.

The Bayesian inference on GARCH-type models has been first implemented using importance sampling (see Kleibergen and van Dijk, 1993). More recent approaches include the Griddy-Gibbs sampler by Bauwens and Lubrano (1998), reversible jump MCMC in Vrontos et al. (2000) and the Metropolis-Hastings algorithm with some specific choice of the proposal distribution (Geweke, 1995; Kim et al., 1998; Müller and Pole, 1998; Nakatsuma, 2000).

Unlike in the mentioned literature, we adopt a hierarchical structure for the GARCH models. This is partly motivated by our intention to have a common approach for all models (so as, GARCH models are the particular case of the non-linear NN model), partly by our belief that hierarchical modeling provides a simple and general way of being weakly informative and to remain proper. For the Student-t distribution, the degrees of freedom is also a parameter to be estimated, and, following Geweke (1993), we choose an exponential prior density. To generate posterior samples for the model parameters we used the multivariate random walk Metropolis algorithm.

The literature on Bayesian analysis for NNs models is relatively thin. MacKay (1992) developed the first approaches, based mainly on Gaussian approximations to the posteriors. Bishop (1995) reviews many of these earlier attempts. Of special relevance is the work of Neal (1996). He applied a hybrid Monte Carlo algorithm, which combines the MH algorithm with methods from dynamical simulation sampling techniques. Müller and Insua (1998) use a multivariate version of MCMC and marginalize over some parameters to increase performance. Posterior inference in NNs is plagued by multimodality issues. Besides trivial multimodality due to relabeling of the hidden units, there is inherent multimodality due to non-linearity. As a consequence, there is little hope for normal approximation with these models and we need to turn to MCMC methods. First attempts to apply the hybrid MC for our models show its practical limitations because of the recurrent structure of our models and, consequently, rather expensive computation of the energy gradient. Thus, we mainly follow the Müller and Insua (1998) approach combining Gibbs sampling of hyperparameters with random walk MH for the NN parameters. We let the proposal variance depend on the current hyperparameter value.

Open questions in the NN community are the selection of the "optimal" size of the network and the identification of the hidden units. The classical model selection

methods may be misled by local modes. More recently, Müller and Insua (1998), Marrs (1998) and Holmes and Mallick (1998) have addressed the issue of selecting the number of hidden neurons from a Bayesian perspective. In particular, they apply the reversible jump MCMC algorithm of Green (1995) to feed-forward sigmoidal networks and radial basis function networks to obtain joint estimates of the number of neurons and weights. Their results indicate that it is advantageous to adopt the Bayesian framework and MCMC methods to perform model order selection. We apply the bridge sampling technique to compute the model evidences for different NN sizes and choose the "optimal" size of the network based on Bayes factors. Unlike the known literature, we select identifiability constraints based on the posterior scatter plots of the NN parameters after random permutation of the hidden nodes (the same approach as in Frühwirth-Schnatter (2001)).

Many issues about posterior multimodality and computational strategies in NN models are of relevance in the wider class of mixture models. Important references in the Bayesian literature on mixture models include Diebolt and Robert (1994), Robert (1996), Frühwirth-Schnatter (2001) and Richardson and Green (1997). We applied data augmentation for mixture models (see Diebolt and Robert, 1994) to generate the posterior samples for the mixture proportions.

At the beginning of the next section we provide a very short review of the main concepts of Bayesian model selection. For more details on the techniques presented here, see, e.g., Carlin and Louis (1996), Gilks et al. (1996), Tierney (1994), Chib and Jeliazhov (2001), Brooks (1998), Kass and Raftery (1995).

5.1 Basic Concepts and Notations

All the complex models may be viewed as the specification of a joint distribution of observables (data) which we denote by Y and unobservables (model parameters) which we denote by θ. The traditional approach to Bayesian model selection is concerned with the following situation. Suppose the observed data Y are generated by a model M_i, one of a set $\mathbf{M}$ of competing models. Each model specifies the data likelihood $f(Y|\theta_i, M_i)$ as the distribution of Y apart from an unknown parameter vector θ_i of dimension n_i. Under prior densities $\pi(\theta_i|M_i)$ the marginal distribution of Y is found by integrating out the parameters

$$p(Y \mid M_i) = \int f(Y \mid \theta_i, M_i)\, \pi(\theta_i \mid M_i)\, d\theta_i. \tag{5}$$

By analogy with the data likelihood function, the quantity $p(Y|M_i)$ is called the *model likelihood*. Typically, these probabilities will be extremely small, since any particular data set of significant size will have low probability, even under the correct model.

We assume no prior preferences between our models (prior model probabilities are equal). Then the model likelihoods yield posterior model probabilities as $p(M_i|Y) = p(Y|M_i) \,/\, \sum_{k=1}^{n} p(Y|M_k)$.

The model with greater likelihood value is declared to have better performance.

The integral (5) is analytically tractable in only certain restricted problems and sampling based methods must be used to obtain estimates of the model likelihoods

(for a review of Bayesian model selection methods see Miazhynskaia et al. (2003c)). We chose the bridge sampling technique for model likelihood computation (see, e.g., Meng and Wong (1996) and Kaufmann and Frühwirth-Schnatter (2002) for details). As an input for the bridge sampling algorithm we used the samples from parameter posterior distributions, obtained with the help of MCMC simulations.

5.2 Priors

The starting point for the Bayesian inference is a prior distribution over the model parameters. Choice of suitable priors is generally a contentious issue. One wants the priors to reflect one's believes about parameter values and at the same time to use non-informative (flat) priors that do not favor particular values of the parameters over other values. The standard choice for non-informative priors are Jeffreys priors (Jeffreys, 1961) based on the expected Fisher information in the model. But such non-informative priors are typically improper in that they do not have finite integrals. This often leads to the non-integrability of the posterior parameter distribution as well, making the Bayesian model selection (based on the normalizing constant of the posterior) questionable. Therefore, we concentrate on proper priors for all model parameters.

Because of the difficulty in interpreting the parameters for the neural network models we adopt a hierarchical prior structure (Neal, 1996) that enables us to treat the priors' parameters (hyperparameters) as random variables drawn from suitable distributions (hyperpriors). A convenient form for these hyperpriors is the inverse Gamma distribution with some fixed shape and mean parameters. To guarantee the positivity in the variance equation we work with the logarithmic transformation of the parameters (b, s_0, s_1, $v_1, \ldots, v_H$).

- $\mathbf{N}(0, 10)$ for mean parameters a_0 and a_1;
- $\mathbf{logN}(\kappa_j, \frac{1}{\tau_j}), j = \overline{1,3}$, for three linear variance parameters b, s_0, s_1, $\kappa = (-2.0, -2.0, -0.2)$;
- $\mathbf{N}(0, \frac{1}{\tau_j}), j = \overline{4,6}$, for the hidden weights (w, γ) and biases c;
- $\mathbf{logN}(0, \frac{1}{\tau_7})$ for the hidden-ouput weights v;
- hyperpriors $\tau_j \sim \mathbf{Ga}(\xi_j, \omega_j),\ j = \overline{1,7}$;
- degrees of freedom $\nu \sim \mathbf{Exp}(0.1)$;
- mixture coefficients $\eta_1, \ldots, \eta_n \sim \mathbf{Dirichlet}(1, \ldots, 1)$.

Although GARCH parameters are rather tractable we used a hierarchical prior structure for them in order to have more variability and to apply a universal approach over all models considered. After preliminary tuning, we fixed the hyperprior shape parameters at $\xi_{1:7} = 10$, $\omega_{1:3} = 1$, $\omega_{4:6} = 0.2$, $\omega_7 = 1$. The priors' centers κ were fixed reflecting our representation of GARCH parameters.

5.3 MCMC Posterior Simulation

The next step in the Bayesian procedure is the inference of the parameter vector θ and hyperparameters τ conditional on data Y via the posterior density $p(\theta \mid Y)$. Using Bayes theorem, this density takes the form $p(\theta \mid Y) = c \cdot f(Y \mid \theta)\pi(\theta)$ for some normalizing constant c, likelihood function $f(Y \mid \theta)$ and prior density $\pi(\theta)$.

For many realistic problems, evaluation of $p(\theta \mid Y)$ is analytically intractable, so numerical or asymptotic methods are necessary. In this article we adopt the MCMC sampling strategies as the tool to obtain posteriors. The idea is based on the construction of an irreducible and aperiodic Markov chain with realizations $\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(t)}, \ldots$ in the parameter space, equilibrium distribution $p(\theta|Y)$, and a transition probability $K(\theta'', \theta') = \pi(\theta^{(t+1)} = \theta'' \mid \theta^{(t)} = \theta')$, where θ' and θ'' are the realized states at time t and $t+1$, respectively. Under appropriate regularity conditions, asymptotic results guarantee that as $t \to \infty$ then $\theta^{(t)}$ tends in distribution to a random variable with density $p(\theta|Y)$, and the ergodic average of an integrable function of θ is a consistent estimator of the (posterior) mean of the function. For the underlying statistical theory of MCMC see Tierney (1994).

The best known MCMC procedures are Gibbs sampling (when we have completely specified full conditional distributions) and the Metropolis-Hastings (MH) algorithm which provides a more general framework. For an introduction to MCMC simulation methods we refer to Chib and Greenberg (1996) and Geweke (1999).

Because of the conjugate hyperpriors' form, we obtain the posterior distribution $p(\tau \mid Y, \theta) = \prod p(\tau_i \mid \theta^{(i)})$ to be a Gamma distribution with shape $\xi_i + k$ and mean $\frac{\xi_i + k}{\xi_i/\omega_i + \sum_{j=1}^{k} (\theta_j^{(i)} - \kappa_j)^2}$, where $\theta^{(i)}$ denotes the parameters directed by the hyperparameter τ_i.

Because of the autoregressive structure of the variance equation there is no property of conjugacy for all model parameters. To sample from the posterior $p(\theta|Y, \tau)$ we applied the random walk Metropolis algorithm with a Student-t distribution as a proposal for the mean parameters and the degrees of freedom and with Gaussian proposal for the variance parameters, where the variances of the proposal distributions were tuned to come close to an "optimal" acceptance rate in the range of 20-40%. After initial exploratory runs of the Markov chain we checked for correlation between the parameters. The blocking update of highly correlated parameters was implemented to increase the efficiency and improve the convergence of the Markov chain.

To sample mixture coefficients $\pi_1, \ldots, \pi_n$ we applied MH with proposal distributions constructed from the "classical" approach to mixture models (Stephens, 1997): introducing the missing component indicators $S^N = (S_1, \ldots, S_N)$ for every point from the data set $Y = (y_1, \ldots, y_N)$ of the model unknowns, we can take as proposal for the mixture coefficients $\mathbf{Dirichlet}(1+N_1, \ldots, 1+N_n)$, where $N_k = \sharp(S^N = k)$.

One of the most important issues for neural network models with $H \geq 2$ as well as for mixture models is their unindentifiability due to the invariance of relabeling the hidden units (mixture components). To cope with this problem, we applied the following strategy (Kaufmann and Frühwirth-Schnatter, 2002): during posterior simulations we performed random permutation of the hidden units' weights (components) and then by constructing scatter plots for MCMC output we checked for the possible

identification and the identification conditions.

Finally, we used the resulting posterior output as an input to the bridge sampling algorithm to compute the model likelihood. To reduce the space of unknowns, we integrated out the hyperparameters and obtained $p(\theta^{(i)}) \propto \mathbf{t}_k(\kappa_i, \omega_i, \xi_i)$, where $\mathbf{t}_k$ denotes multivariate t distribution with mean κ_i, variance $\omega_i \cdot \mathbf{E}_k$ and degrees of freedom ξ_i.

5.4 Bayesian Comparison Results

In the empirical application we used the same stock index prices, transformed in return series, as before. Contrary to the maximum likelihood approach, the data were used as a single set.

We first investigated the degree of non-linearity in the data controlled by the number of hidden nodes of the fitted non-linear models. We performed complete Bayesian analysis of the RMDN(1) model with Gaussian and t distributions for three cases $H = 1, 2, 3$. The resulting model likelihood values are presented in Table 6. The case $H = 0$ means linear GARCH models.

	normal distribution			t distribution		
H	DJIA	FTSE 100	NIKKEI 225	DJIA	FTSE 100	NIKKEI 225
0	-5776.1	-5547.0	-6810.3	-5549.9	-5512.5	-6633.1
1	-5772.9	-5549.4	-6794.5	-5546.8	-5514.9	-6629.8
2	-5757.8	-5556.8	-6788.9	-5548.7	-5517.4	-6630.5
3	-5761.6	-5559.1	-6791.3	-5551.8	-5520.3	-6633.1

Table 6: Model likelihoods (logarithm) for GARCH(1,1) and RMDN(1) models for different number of hidden nodes. The case $H = 0$ corresponds to the linear GARCH models.

It follows that in the framework of the gaussian distribution we observed clear non-linearity for DJIA and NIKKEI 225 return series. Introducing one hidden unit leads to a significant improvement in the model likelihood. The RMDN(1) model with two hidden units has the best performance. But the model with three hidden units becomes over-complex and is punished by smaller model evidence.

The optimality of the case $H = 2$ for the data sets is supported by clear identification of two hidden units in the non-linear part of the model (see scatter plots in Fig. 5). Based on this, we used $v_1 < v_2$ as the identification condition to remove the mutimodality in parameter posteriors (see plots on Fig. 6). For FTSE 100 series there is no identification for the second hidden unit (Fig. 5, the second row), which agrees with the results from Table 6, where the non-linear models (with $H \geq 1$) reached smaller model likelihoods compared to the linear GARCH model.

One hidden unit is enough to explain possible non-linearity in all data sets (Table 6, models with t-distribution) and there is no clear identification fo the second hidden unit on the scatter plots.

For mixture models LRMDN(2) and RMDN(2) we clearly identified two compo-

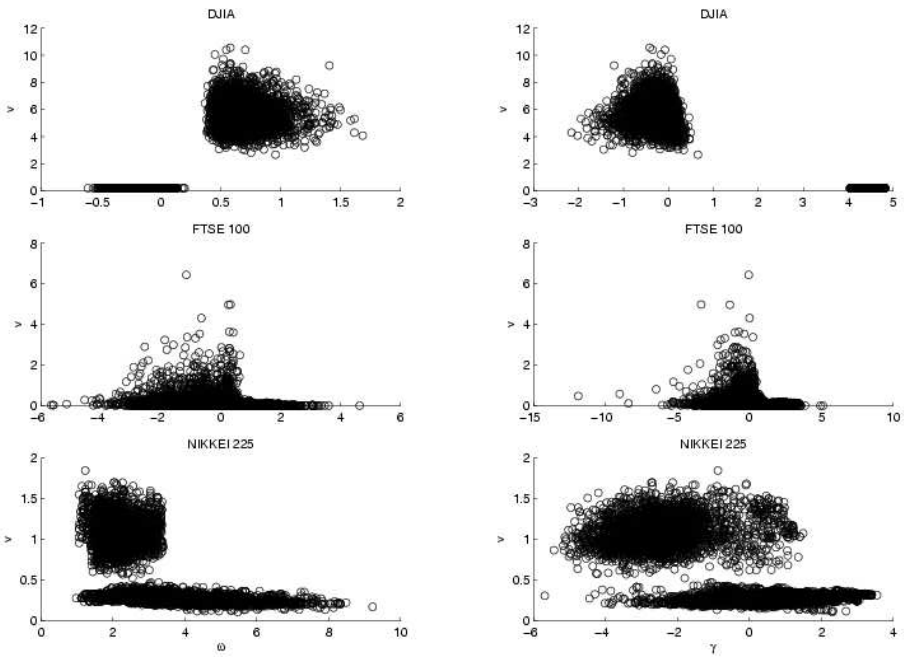

Figure 5: Scatter plot of non-linear parameters, model RMDN(1),$H = 2$.

nents based on the condition of the mixture priors $\eta_1 < \eta_2$.

To get some assessment of the robustness of the results above to the prior "informativity", we repeat the MCMC simulations varying the hyperprior means $\omega_{1:3}$ of the linear variance parameters from 0.5 (the most vague priors) to $\omega_{1:3} = 5$ (the most informative priors). The shape of the posterior density remains essentially unchanged with a slight bias to the prior mean when making the prior more informative. We calculated posterior model probabilities within these four cases of the hyperprior mean values. The results averaged over all data sets are the following:

- 0.23 for the most vague priors with $\omega = (0.5,\, 0.5,\, 0.5)$;
- 0.34 for the case $\omega = (1,\, 1,\, 1)$;
- 0.31 for $\omega = (2,\, 2,\, 2)$;
- 0.12 for the most informative priors with $\omega = (5,\, 5,\, 5)$.

In such a way, Bayesian model selection punishes unnecessary vague as well as too informative priors which do not give the best generalization.

Varying priors' width for non-linear parameters, we found a great influence of the hyperpriors on the posterior parameter distribution as well as on model likelihood values. We do not want to be very uninformative on the NN parameters and prefer to keep their range comparable to the scale of described variables (conditional variance) and be not on the limits of the sigmoid activation function. That is, we repeat the whole MC simulations for the model RMDN(1) for all data sets, taking even more

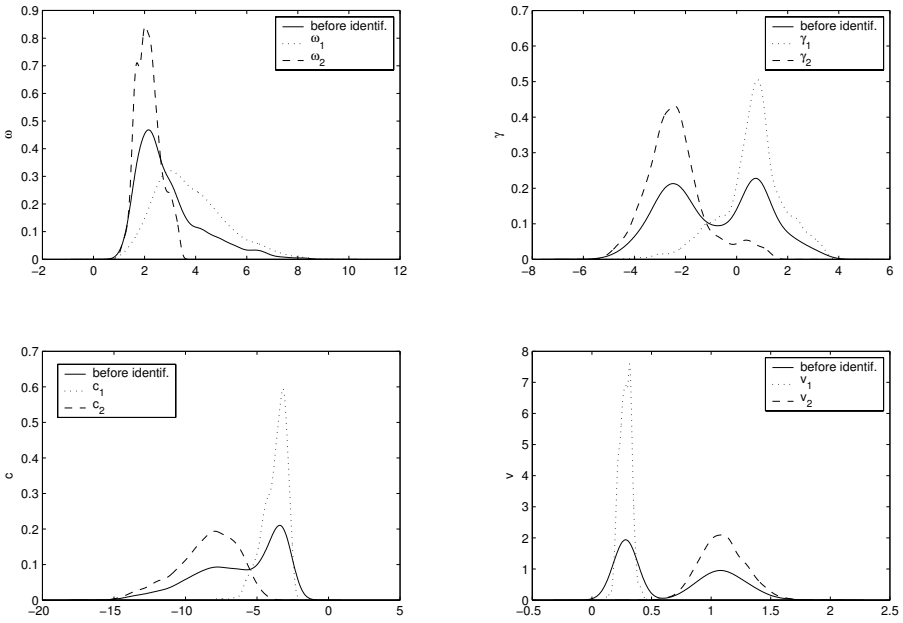

Figure 6: Posterior parameter distributions for non-linear parameters before and after identification. Data set NIKKEI 225, model RMDN(1), $H = 2$.

informative priors on the NN weights ($\omega_{4:7} = 1$) and keeping the same hyperprior means for the linear parameters ($\omega_{1:3} = 1$). In Fig. 7 we plot the posterior distributions of the non-linear parameters for the model RMDN(1) for two different hyperprior means. There is a significant influence of the hyperpriors on the posterior distributions. The posterior plots for FTSE 100 data are again symmetrical around 0. Moreover, posterior parameters' variances in this case are clearly held by the priors, drastically increasing with the priors width. The average posterior model probabilities for these cases are 0.09 for $\omega_{4:7} = 1$ and 0.91 if we use the vaguer priors.

In such a way, taking the best case within every model with respect to the hyperprior specification and the number of hidden units, we present in Table 7 the main result of this section – the model likelihood values over all data for the models considered.

Again, the conditional density specification plays the dominant role in the model performance. The gaussian distribution clearly underestimates conditional fat-tails observed in the data. Further, non-linearity issues are of much less significance. For FTSE 100 data we obtained clear preference for linear models for all types of distributions considered. For DJIA and NIKKEI 225 series we found significant non-linearity under the gaussian conditional distribution which is described best by the neural network model with two hidden units. For other density specifications non-linearity is not so pronounced, mostly hidden by the fat-tails of the distributions. Finally, there does not exist a single best model for all equity markets within the models considered.

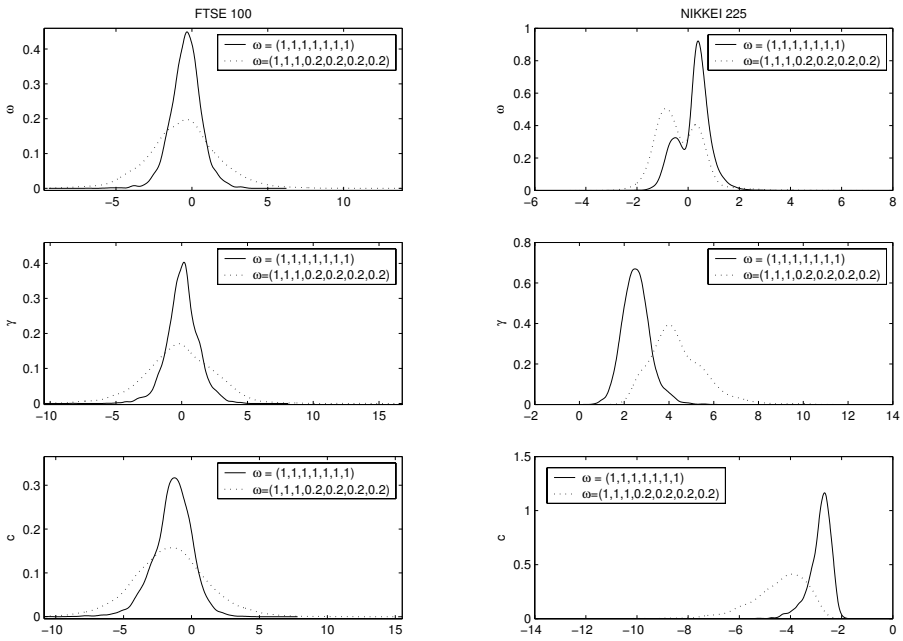

Figure 7: Posterior density plots for non-linear parameters under two hyperparameter widths. Model RMDN(1) with $H = 1$. The left panel for data set FTSE 100, the right panel for NIKKEI 225.

Table 7: Logarithm of model likelihood

Model	DJIA	FTSE 100	NIKKEI 225
GARCH(1,1)	-5776.1	-5547.0	-6810.3
RMDN(1)	-5760.1	-5549.4	-6788.9
GARCH(1,1)-t	-5549.9	-5512.5	-6633.1
RMDN(1)-t	-5546.8	-5514.9	-6629.8
LRMDN(2)	-5566.4	-5499.0	-6647.8
RMDN(2)	-5566.5	-5501.2	-6644.7

For FTSE 100 return series the best models are the mixture of gaussians with some superiority of the linear LRMDN(2). For DJIA and NIKKEI 225 data the models with t-distribution are still best with some advantage of the non-linear RMDN(1)-t model.

6 Discussion and Conclusions

We analyzed the importance of non-linearity and non-gaussian distributions in classical GARCH models. The empirical analysis was based on return series of stock in-

dices from different financial markets. We consider our models in two fundamentally different estimation frameworks: maximum likelihood and Bayesian. First, within the maximum likelihood framework the models were evaluated with respect to the likelihood performance as well as with respect to the prediction of the VaR for a portfolio position. Second, we applied full Bayesian inference to our models, including a hierarchical prior specification, MCMC implementation for the different parameter groups and model likelihood computation.

Summing up, we can derive the following conclusions:

- The conditional density specification plays the dominant role in the model performance. All statistical tests clearly confirmed the conclusion that non-gaussian models significantly dominate the gaussian ones. The gaussian distribution underestimates conditional fat-tails observed in the data.
- Non-linearity issues are of much less significance. Non-linearity found in DJIA and NIKKEI 225 data under the gaussian distribution is mostly explained by fat-tailed conditional distributions. But the Bayesian framework shows that we still have some preference for the non-linear RMDN(1)-t model for DJIA and NIKKEI series.
- The maximum likelihood framework mostly favors the linear GARCH(1,1)-t and LRMDN(2) models over all data sets considered, but the significance of its superiority differs between the markets. In the Bayesian framework, we did not find a single best model for all equity markets as well. For FTSE 100 return series the best model is a mixture of gaussians with clear superiority of linear LRMDN(2), while for DJIA and NIKKEI 225 data the models with t-distribution are the best and the non-linear version outperforms the GARCH model.

Acknowledgments

This work was funded by the Austrian Science Fund (FWF) under grant SFB#010: "Adaptive Information Systems and Modeling in Economics and Management Science". The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry for Education, Science and Culture and by the Austrian Federal Ministry for Transport, Innovation and Technology.
In the maximum likelihood framework the models were estimated using the NETLAB neural network software (can be downloaded from http://neural-server.aston.ac.uk) adapted by Christian Schittenkopf to the model specifications considered in the paper.

Bibliography

Alles, L. and Kling, J. (1994). Regularities in the variation of skewness in asset returns. *Journal of Financial Research*, 17:427–438.

Bartlmae, K. and Rauscher, F. A. (2000). Measuring DAX market risk: A neural network volatility mixture approach. Presentation at the FFM2000 Conference, London, 31 May-2 June, can be downloaded from www.gloriamundi.org/picsresources.

Bauwens, L. and Lubrano, M. (1998). Bayesian inference on GARCH models usin Gibbs sampler. *Econometrics Journal*, 1:23–46.

Bishop, C. (1994). Mixture density networks. Technical report, Neural Computing Research Group Report: NCRG/94/004, Aston University, Birmingham.

Bishop, C. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press, Oxford.

Boero, G. and Cavallil, E. (1997). Exchange rate forecasting: Neural networks versus linear econometric models. *Neural Network World*, 1:29–42.

Bollerslev, T. (1986). A generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

Brooks, S. (1998). Markov chain Monte Carlo method and its application. *The Statisician*, 47:69–100.

Carlin, B. and Louis, T. (1996). *Bayes and Empirical Bayes Methods for Data Analysis*. Chapman & Hall, London.

Chib, S. and Greenberg, E. (1996). Markov Chain Monte Carlo simulation methods in econometrics. *Econometric Theory*, 12:409–431.

Chib, S. and Jeliazhov, I. (2001). Marginal likelihood from the Metropolis-Hastings output. *Journal od American Stattistical Association*, 96(453):270–281.

Diebolt, J. and Robert, C. (1994). Estimation of finite mixture distributions through bayesian sampling. *Journal of the Royal Statistical Society, series B*, 56:363–375.

Dowd, K. (1998). *Beyond Value at Risk: the New Science of Risk Management.* John Willey & Sons, London.

Duffie, D. and Pan, J. (1997). An overview of value at risk. *Journal of Derivatives*, 4:7–49.

Dunis, C. L. and Jalilov, J. (2002). Neural network regression and alternative forecasting techniques for predicting financial variables. *Neural Network World*, 12:113–139.

Frühwirth-Schnatter, S. (2001). MCMC estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96:194–209.

Geman, S., Bienenstock, E., and Doursat, R. (1992). Neural networks and the bias/variance dilemma. *Neural Computation*, 4:1–58.

Geweke, J. (1993). Bayesian treatment of the independent student-t linear model. *Journal of Applied Econometrics*, 8:19–40.

Geweke, J. (1995). Bayesian comparison of econometric models. Working Papers 532, Federal Reserve Bank of Minneapolis.

Geweke, J. (1999). Using simulation methods for bayesian econometric models: Inference, development and communication. *Econometric Reviews*, 18:1–126.

Gilks, W., Richardson, S., and Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. Chapman & Hall, London.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance*, 48:1779–1801.

González, M. and Burgess, N. (1997). Modelling market volatilities: the neural network perspective. *The European Journal of Finance*, 3:137–157.

Green, P. (1995). Reversible jump MCMC computation and Bayesian model determination. *Biometrika*, 82:711–732.

Hansen, B. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35:705–730.

Holmes, C. and Mallick, B. (1998). Bayesian radial basis functions of variable dimension. *Neural Computation*, 10:1217–1233.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366.

Jeffreys, H. (1961). *Theory of Probability, 3rd edition*. Oxford Univiersity Press, Oxford.

Kass, R. and Raftery, A. (1995). Bayes factor. *Journal od American Statistical Association*, 90:773–792.

Kaufmann, S. and Frühwirth-Schnatter, S. (2002). Bayesian analysis of switching ARCH models. *Journal of Time Series Analysis*, 23(4):425–458.

Kim, S., Shephard, N., and Chib, S. (1998). Stochastic volatility: Likelihood inference and comparison with ARCH models. *Review of Economic Studies*, 65:361–393.

Kleibergen, F. and van Dijk, H. (1993). Non-stationarity in GARCH models: a bayesian analysis. *Journal of Applied Econometrics*, 8:41–61.

Kupiec, H. (1995). Techniques for verifying the accuracy of risk management models. *Journal of Derivatives*, 3:73–84.

MacKay, D. (1992). A practical bayesian framework for backprop networks. *Neural Computation*, 4:448–472.

Marrs, A. (1998). An application of reversible-jump MCMC to multivariate spherical gaussian mixtures. In Jordan, M., Kearns, M., and Solla, S., editors, *Advances in Neural Information Processing Systems*, volume 10, pages 577–583. MIT Press, Cambridge, Mass.

Meng, X. and Wong, W. (1996). Simulating ratios of normalizing constants via a simple identity. *Statistical Sinica*, 6:831–860.

Miazhynskaia, T., Dockner, E. J., and Dorffner, G. (2003a). On the economic costs of value at risk forecast. Technical report, SFB Adaptive Information Systems and Modelling in Economics and Management Science, Vienna. Submitted to Journal of Banking and Finance, can be downloaded from http://www.wu-wien.ac.at/am/reports.htm.

Miazhynskaia, T., Dorffner, G., and Dockner, E. J. (2003b). Non-linear versus non-gaussian volatility models in application to different financial markets. Technical report, SFB Adaptive Information Systems and Modelling in Economics and Management Science, Vienna. Can be downloaded from http://www.wu-wien.ac.at/am/reports.htm.

Miazhynskaia, T., Frühwirth-Schnatter, S., and Dorffner, G. (2003c). A comparison of bayesian model selection based on mcmc with an application to garch-type models. Technical report, SFB Adaptive Information Systems and Modelling in Economics and Management Science. Can be downloaded from http://www.wu-wien.ac.at/am/reports.htm.

Müller, P. and Insua, D. R. (1998). Issues in bayesian analysis of neural network models. *Neural Computation*, 10:571–592.

Müller, P. and Pole, A. (1998). Monte carlo posterior integration in GARCH models. *Sankhya - The Indian Journal of Statistics*, 60:127–144.

Nakatsuma, T. (2000). Bayesian analysis of ARMA-GARCH models: a Markov chain sampling approach. *Journal of Econometrics*, 95:57–69.

Neal, R. (1996). *Bayesian Learning for Neural Networks*. Springer, New York.

Nelson, D. (1991). Conditional heteroskedasticity in asset returns: a new approach. *Econometrica*, 59:347–370.

Poh, H.-L., Yao, J. T., and Jasic, T. (1998). Neural networks for the analysis and forecasting of advertising and promotion impact. *International Journal of Intelligent Systems in Accounting, Finance and Management*, 7(4):253–268.

Prechelt, L. (1998). Early stopping - but when? In Orr, G. and Müller, K.-R., editors, *Neural Networks: Tricks of the Trade*, pages 55–69. Springer, Berlin.

Reed, R. (1993). Prunung algorithm - a survey. *IEEE Transactions on Neural Networks*, 4(5):740–746.

Richardson, S. and Green, P. (1997). On Bayesian analysis of mixtures with an unknown number of components. *Journal of the Royal Statistical Society, Series B*, 59:731–792.

Robert, C. (1996). Mixtures of distributions: Inference and estimation. In Gilks, W., Richardson, S., and Spiegelhalter, D., editors, *Markov Chain Monte Carlo in Practice*, pages 441–464. Chapman & Hall.

Schittenkopf, C., Dorffner, G., and Dockner, E. J. (1999). Non-linear versus non-gaussian volatility models. Technical report, SFB Adaptive Information Systems and Modelling in Economics and Management Science, Vienna.

Schittenkopf, C., Dorffner, G., and Dockner, E. J. (2000). Forecasting time-dependent conditional densities: a seminonparametric neural network approach. *Journal of Forecasting*, 19:355–374.

Stephens, M. (1997). Bayesian methods for mixture of normal distributions. Technical report, Department of Statistics, Oxford University, England.

Swanson, N. and Franses, P. (1999). Nonlinear econometric modelling: a selective review. In Rothman, P., editor, *Nonlinear Timee Series Analysis of Economic and Financial Data*, pages 87–110. Kluwer Academic, Dordrecht.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *Annals of Statistics*, 21:1701–1762.

Vrontos, I., Dellaportas, P., and Politis, D. (2000). Full Bayesian inference for GARCH and EGARCH models. *Journal of Business & Economic Statistics*, 18(2):187–198.

Yao, J. T., Tan, C. L., and Li, Y. L. (2000). Option prices forecasting using neural networks. *International Journal of Management Science*, 28(4):455–466.

Expectation Formation and Learning in Adaptive Capital Market Models

Engelbert J. Dockner and Leopold Sögner

1 Introduction

One of the distinguishing features of financial markets are the traded contracts that require forward looking behavior of agents. Mainstream financial economics uses the framework of rational expectations to model this behavior. In a market in which agents have rational expectations and the efficient market hypothesis holds, prices are only driven by unexpected events. Hence, one can argue that any price dynamics is entirely driven by unanticipated shocks.

While the concept of rational expectations is the leading paradigm in capital market analysis, there are problems arising from its applications. One major shortcoming is the degree of sophistication in terms of expectation formation that is required for the agents and their knowledge about unconditional moments of the returns distributions. For that reason financial economists have started to look at models with boundedly rational agents where the process of expectations formation includes experience and learning.

The objective of this paper is to present some alternative examples of how the process of expectation formation can be modelled. The problem with bounded rationality is that many alternative models can be used to describe the process of expectation formation in a plausible way, i.e. there are too many degrees of freedom. Hence, it is sensible to formulate research focusses that (i) analyze under which alternative expectations hypotheses equilibrium prices in a financial market still converge to the rational expectations equilibrium and (ii) that explore if alternative theories of expectation formation and the resulting market dynamics are capable of explaining stylized facts in financial markets? We are going to explore both of these questions in this paper.

The first part of the paper investigates consistency, learning and stability in models with bounded rationality. In contrast to rational expectations models, this class of models skips the assumption that agents know both the dynamics of the economic system and its stochastic properties. The results demonstrate how different learning dynamics can result in instability or in stable but very different paths compared to the rational expectations dynamics. In terms of explaining stylized facts in financial markets models with homogenous agents are investigated (see Pötzelberger and Sögner, 2003a, 2004), and it is demonstrated that they exhibit a poor performance.

The second part of the paper describes discrete choice models, which provide an alternative approach to model heterogeneous beliefs in a financial market. These models are capable of producing time series that possess properties frequently found in asset returns, such as excess kurtosis and volatility clustering.

In financial market models with *rational expectations* all agents know the corresponding equations of the economic model and have sufficient skills to consistently

calculate the unconditional expectation. Ever since the concept was introduced by Muth (1961) it was clear that it is highly demanding since perfectly informed agents with high computational skills to solve their optimization problems are required. These agents are rational in their decision process (since the optimal strategies are derived from maximizing the agents' utility functions) and in their learning process (since the agents are able to form their expectations in a statistically correct way). As a result of interactions of rational agents at the level of markets the rational expectations equilibrium (REE) is derived. Therefore, the rational expectations equilibrium is a natural consequence of the rational agent paradigm.

As an alternative to rational expectations *bounded rationality* models have been proposed. They are based on behavioral foundations (see Sargent, 1993; Arthur et al., 1997) or discrete choice models (see Brock and Hommes, 1997, 1998). However, all models with non-rational agents allow for many degrees of freedom. For example the process of inference of new information can be modelled by means of linear rules, non-linear techniques, genetic algorithms, and even ad-hoc rules. Every forecasting technique can result in different equilibrium behavior, if equilibria of the system exist. To get some feeling for the implications of these degrees of freedom we can ask ourselves if the equilibrium paths arising from non-rational agents with simple forecasting rules do converge to any stable behavior or the corresponding rational expectations equilibrium. The goal of this work is to show that stability need not hold.

Before we proceed with our results let us define the concepts of *learning*, the *perceived law* and the *implied actual* law. Therefore, let us consider an economic system of the form $x_t = f(y^e_{t+1})$, i.e. the state variable $x_t \in \mathrm{R}^n$ is a function of the one period ahead forecast of $y_t := (y_{t,1}\ y_{t,2})' \in \mathrm{R}^m$. $y_{t,1}$ are the endogenous variables and $y_{t,2}$ are exogenous. If $y_{t,2}$ is stochastic then its distribution is $F(y_{t,2})$. The index t is the time index, where only discrete time systems are investigated in this paper. For rational expectations models $f(.)$ and the properties of the exogenous variables are assumed to be known by all agents, such that we can – at least in principle – solve for x_t. In the bounded rationality literature the function $f(.)$ is assumed to be unknown to the agents, such that predictions are needed. An agent's approximation of $f(.)$ and the dynamics of y_t – called perceived law – can be motivated by the fact that the dynamics of an economic system may be very complex such that agents prefer to work with simple forecast models and check whether the predictive power of these forecast models is sufficient and consistent with the data (see, e.g., Sargent, 1993, chapter 2; Hommes and Sorger, 1998).

Let us assume that agent i, $i = 1, \ldots, n$, believes in a functional relationship between y_t, y_{t-1} and $\tilde{\varepsilon}_t$:

$$y_t = g_i(y_{t-1}, \tilde{\varepsilon}_{i,t})\ . \tag{1}$$

Equation (1) and a specification of $\tilde{\varepsilon}_{i,t}$ is called the *perceived law* of agent i.[1] If all agents use the same perceived law, then the forecasts of the agents are homogenous.

The reader should note that we have used $\tilde{\varepsilon}_{i,t}$ in equation (1). This is because the

[1] In this article we work with $y_t = y_{i,t}$ for all agents $i = 1, \ldots, n$. I.e. all agents know the relevant variables in $x_t = f(y^e_{t+1})$. An extension to $y_t \neq y_{i,t}$ is straightforward but makes the notation much more complex.

actual noise $y_{t,2} \sim F(y_{t,2})$ can but need not agree with the agents' believes $\tilde{\varepsilon}_{t,i} \sim \tilde{F}_i(\tilde{\varepsilon}_{i,t})$, which e.g. comprises the case where agents believe in a stochastic setting despite $y_{t,2}$ being empty or purely deterministic.

The one step ahead forecast of y_t for agent i based on equation (1) and the information available when x_t will be determined, $\mathcal{F}_{t-1}$, is derived by[2]

$$y^e_{i,t+1} = \mathbb{E}^i(g(g(y_{t-1}, \tilde{\varepsilon}_{i,t}), \tilde{\varepsilon}_{i,t+1})) \,, \tag{2}$$

where $\mathbb{E}^i$ is the agent's belief of the conditional expectation. $\mathbb{E}^i$ can but need not exactly correspond to the conditional expectation operator under the perceived law consisting of $g_i(.)$ and $F_i(\tilde{\varepsilon}_{i,t})$. A composition of $f(.)$ and equation (2) results in the *implied actual law*.

In a further step learning can be introduced to an economic model. Within this paper we assume that the agents use parametric models. For learning with non-parametric models we refer the reader to Chen and White (1998). Thus, let us assume that the perceived law is a parametric model, such that $g_i(.)$ is a function of the vector of model parameters θ_i. *Learning* in this context means nothing more than a systematic update of the estimates of θ_i, denoted by $\theta_{i,t}$. A model with boundedly rational agents, where the perceived law for $g_i(.)$, $F_i(\tilde{\varepsilon}_{i,t})$ is updated systematically by some learning algorithm is called *adaptive model*. If these updates are modified model parameters $\theta_{i,t}$, we have a parametric adaptive model.

After we are equipped with these technical definitions, let us relate these definitions to a specific application. In this work we shall consider a standard capital market model. The model structure – or variations of it – has already been used in a lot of models investigating learning and capital markets (see, e.g., Arthur et al., 1997; Brock and Hommes, 1998). From an empirical point of view this model is interesting since mean-variance maximization is actually applied in finance and from a theoretical point of view it is interesting since different learning results arise from similar settings. For example with *least squares learning* rules, Bray (1982), Blume et al. (1982), Blume and Easley (1982), Marcet and Sargent (1989), Schönhofer (1997), and Routledge (1999) derived conditions for the system to converge to the rational expectations equilibrium.

This paper is organized as follows: Section 2 describes a simple capital market model. Section 3 investigates two learning schemes with homogenous agents. Section 4 discusses Hommes-Sorger consistency in these settings. Section 5 presents a short introduction to adaptive belief systems and finally Section 6 concludes.

2 A Basic Capital Market Model

Asset Demand and Market Clearing: Consider n agents which at time t invest their wealth w^i_t ($i \leq n$) in a risky asset with price p_t and in a risk-free asset paying interest r. The risky asset pays a stochastic dividend d_t in period t, where d_t has a finite second moment. The agents observe and receive their dividend payment when the price p_t is determined by the market clearing mechanism. This implies that at time t the agents

[2] As usual in economics we assume that this integral exists.

observe previous prices and dividends. If q_{t-1}^i denotes the amount of the risky asset held by agent i, then the *budget constraint* becomes $w_t^i = (1+r)w_{t-1}^i + (p_t + d_t - (1+r)p_{t-1})q_{t-1}^i$. Agents choose the quantities q_t^i by maximizing

$$\mathbb{E}^i(w_{t+1}^i) - \frac{\zeta_i}{2}\text{VAR}^{\,i}(w_{t+1}^i), \tag{3}$$

where $\mathbb{E}^i(.)$ and $\text{VAR}^{\,i}(.)$ denote the beliefs of agent i of the conditional expectation and the conditional variance of w_t^i. p_t is held fixed. The parameter ζ_i is a measure of risk-aversion of agent i. (3) is maximized for

$$q_t^i = \frac{\mathbb{E}^i(p_{t+1} + d_{t+1}) - p_t(1+r)}{\zeta_i \mathbb{V}^i(p_{t+1} + d_{t+1})}. \tag{4}$$

The market clearing price p_t is implicitly given by the equilibrium equation $S = \sum_{i=1}^n q_t^i$; where S is the constant supply of the asset. Then the market price is derived from

$$p_t = \frac{1}{1+r}\left(\sum_{i=1}^n \frac{\mathbb{E}(p_{t+1} + d_{t+1})}{\zeta_i \text{VAR}^{\,i}(p_{t+1} + d_{t+1})} - S\right) \frac{1}{1/\sum_{i=1}^n \zeta_i \text{VAR}^{\,i}(p_{t+1} + d_{t+1})}\,. \tag{5}$$

For homogenous forecasts we derive

$$p_t = \frac{1}{1+r}\left(\mathbb{E}^{(1)}(p_{t+1} + d_{t+1}) - \frac{S}{\sum_{i=1}^n 1/\zeta_i}\text{VAR}^{\,(1)}(p_{t+1} + d_{t+1})\right)\,, \tag{6}$$

where $^{(1)}$ stands for the representative forecast. Let us abbreviate the mean and the variance of d_t byμ and σ_d^2. The mean and the variance of p_t are denoted by η and σ_p^2. Furthermore we use λ for $1/(1+r)$; c for $S/\sum_{i=1}^n(1/\zeta_i)$. Then the map

$$p_t = \lambda\left(\mathbb{E}^1(p_{t+1} + d_{t+1}) - c\text{VAR}^{\,1}(p_{t+1} + d_{t+1})\right) \tag{7}$$

provides us with a map of the form $x_t = f(y_{t+1}^e)$, where $y_{t,1} = p_t$ and $y_{t,2} = d_t$.

Rational Expectations: With perfectly rational agents $\mathbb{E}^{(1)}(p_{t+1} + d_{t+1})$ is equal to the conditional expectation $\mathbb{E}(p_{t+1} + d_{t+1}|\mathcal{F}_{t-1})$; $\mathcal{F}_{t-1} = \sigma(p_s, d_s|s \leq t-1)$. $\mathbb{V}^{(1)}(p_{t+1} + d_{t+1})$ has to be equal to $\mathbb{V}(p_{t+1} + d_{t+1}|\mathcal{F}_{t-1})$. Then (7) becomes

$$p_t = \lambda\left(\mathbb{E}(p_{t+1} + d_{t+1}|\mathcal{F}_{t-1}) - c\mathbb{V}(p_{t+1} + d_{t+1}|\mathcal{F}_{t-1})\right). \tag{8}$$

Using (8) and the assumption of an iid dividend process with finite second moment results in the constant

$$p^{REE} = \frac{1}{r}(\mathbb{E}(d_t) - c\mathbb{V}(d_t)) = \frac{1}{r}(\mu - c\sigma_d^2)\,, \tag{9}$$

which is the rational expectations equilibrium (REE) for iid dividends. For first-order autoregressive dividends, the reader is referred to Tay and Linn (2001).

3 Learning and Stability for the Homogeneous Agent Model

Perceived Laws: In Sections 3 and 4 we shall use linear forecasts, where the agents' beliefs are homogeneous. Let us assume that (p_t) is an autoregressive process of order one:

$$p_t - \eta = \beta(p_{t-1} - \eta) + \varepsilon_{p,t} \ , \tag{10}$$

where the independent and identically distributed innovations $(\varepsilon_{p,t})$ have mean 0 and variance σ_p^2. η denotes the mean of p_t, while the parameter $\beta \in (-1, 1)$ refers to the first order autocorrelation of the price process. For the dividends the agents assume an iid process with mean μ and noise $\varepsilon_{d,t}$, with variance σ_d^2, i.e.

$$d_t = \mu + \varepsilon_{d,t} \ , \tag{11}$$

where the agents assume that $\varepsilon_{p,t}$ and $\varepsilon_{d,t}$ are independent. Under these assumptions the *perceived laws* are given by (10) and (11).

3.1 Sample Autocorrelation Learning

Since neither p_t nor d_t are in the agents' information set at period t the asset demand function (4) requires a two-step ahead forecast for prices and dividends. From (10) and (11) we can derive the belief $\mathbb{E}^{(1)}(p_{t+1}) = \beta^2 p_{t-1} + (1 - \beta^2)\eta$, while for the conditional variance belief Pötzelberger and Sögner (2004) use VAR $^{(1)}(p_{t+1}) = (1 - \beta^2)\sigma_p^2$. Inserting these expressions into (6) yields

$$p_t = \lambda \left(\beta^2 p_{t-1} + (1 - \beta^2)\eta + \mu - c((1 - \beta^2)\sigma_p^2 + \sigma_d^2) \right) . \tag{12}$$

To embed this model into our general definitions the state variable of interest $x_t = p_t$, $y_t = (p_t \ d_t)'$ and error terms $\tilde{\varepsilon}_t = (\tilde{\varepsilon}_{p,t} \ \varepsilon_{d,t})'$. The parameters of the forecast models (10) and (11) are $\theta = (\mu, \eta, \sigma_p^2, \sigma_d^2, \beta)'$. If learning is applied, θ is going to be updated at every period t. In this case the parameters in the implied actual law (12) have to be replaced by their corresponding estimates $\theta_t = (\mu_t, \eta_t, \sigma_{p,t}^2, \sigma_{d,t}^2, \beta_t)'$.

Pötzelberger and Sögner (2004) applied sample autocorrelation learning to derive the unknown model parameters. More precisely, θ_t is obtained from

$$\bar{p}_t := \frac{1}{t} \sum_{\iota=0}^{t-1} p_\iota = \eta_t \ , \tag{13}$$

where the sample mean of dividends $\bar{d}_t = \mu_t$ is calculated in the same way. The first order autocovariance is derived from

$$\text{COV}_t(p_t, p_{t-1}) := \frac{1}{t-1} \sum_{\iota=1}^{t-1} (p_\iota - \bar{p}_t)(p_{\iota-1} - \bar{p}_{t-1}) \ . \tag{14}$$

The sample variance of prices is derived from

$$\text{VAR}_t(p_t) := \frac{1}{t} \sum_{\iota=0}^{t-1} (p_\iota - \bar{p}_t)^2 \ , \tag{15}$$

such that the conditional variance of prices is derived from

$$\sigma_{p,t}^2 := (1 - \beta_t^2)\text{VAR}_t(p_t) \; . \tag{16}$$

The coefficient β is estimated by the first order autocorrelation coefficient:

$$\gamma_t := \frac{\text{COV}_t(p_t, p_{t-1})}{\sqrt{\text{VAR}_t(p_t)}\sqrt{\text{VAR}_{t-1}(p_{t-1})}} = \beta_t \; . \tag{17}$$

The fact that the autocorrelation coefficient stays within the interval $[-1, 1]$ by the Cauchy-Scharz inequality is essential in the proof of Proposition 1. Therefore, the results cannot be applied to a β estimated by least squares. The sample variance of dividends $\text{VAR}_t(d_t) = \sigma_{d,t}^2$ is derived from (15) by using d_t for p_t.

This ends up into the vector of estimated parameters θ_t resulting in the implied actual law with sample autocorrelation learning

$$p_t = \frac{1}{1+r}\left(\alpha_{p,t} + \beta_t^2\,(p_{t-1} - \alpha_{p,t}) + \alpha_{d,t} - c(\sigma_{p,t}^2 + \sigma_{d,t}^2)\right) \; . \tag{18}$$

For this random dynamical system Pötzelberger and Sögner (2004) show that the sequence of prices either converges to a constant real number or diverges to $-\infty$ or some given lower bound. A sufficient condition for convergence of (p_t) to z/r, where $z := \mu - c\sigma_d^2$ and $z_t := \mu_t - c\sigma_{d,t}^2$, is given in the following Proposition.

PROPOSITION 1 (Pötzelberger and Sögner, 2004) *Let the sequence (z_t) converge to z and let*

$$|p_0 - \frac{z}{r}| \leq \nu \leq \frac{r}{c} \; , \tag{19}$$

and

$$\sup_t |z_t - z| \leq r\nu - c\nu^2 \; . \tag{20}$$

Then the sequence of prices (p_t) derived from (18) converges.

In case the dividend process, (d_t), is iid Proposition 1 provides a sufficient condition to convergence to the REE (see equation (9)). The sufficient condition for convergence states that a sequence of prices converges if the initial value of the price sequence is sufficiently close to the steady-state equilibrium and a random variable derived from the dividend process is not too volatile to push the price trajectory out of the attracting region. Since these conditions are only sufficient, Pötzelberger and Sögner (2004) also provide a numerical analysis, to demonstrate that price paths can diverge if our conditions for convergence are not met. Therefore, the market price can even diverge, and the region of convergence could become very small (depending on the underlying parameters), even in a quite simple setting. A further insight of this analysis is that the support of the dividend process has to be bounded to guarantee convergence (convergence with probability one, given some distribution of the dividend process with support $supp(d_t)$). This implies that the often applied assumption of normally distributed variables in economics and finance is in conflict with almost sure convergence to the rational expectations equilibrium in this model.

3.2 Learning by Exponential Smoothing

As an alternative to sample autocorrelation learning Pötzelberger and Sögner (2003a) and Pötzelberger and Sögner (2003b) applied parameter estimation by exponential smoothing. This setting can be interpreted as a simple way to approximate bounded recall. By this algorithm prices remain stochastic. I.e. Pötzelberger and Sögner (2003b) investigate the question of *stability* and the impact of *learning* in a stochastic setup, when only limited information is used to construct estimators.

Already Muth (1961) designed the rational expectations concept to fit to stochastic equilibrium behavior as well, while by the work of Stokey and Lucas (1989) the idea of modeling stochastic equilibrium by the concepts of ergodicity and stationarity has become a familiar concept in economics. Pötzelberger and Sögner (2003b) investigate the above capital market model when θ is estimated by means of exponential smoothing with a smoothing parameter $0 \leq \nu \leq 1$. Then

$$\hat{\eta}_t = (1-\nu)\sum_{i=0}^{\infty} \nu^i p_{t-i}, \tag{21}$$

$$\hat{\mu}_t = (1-\nu)\sum_{i=0}^{\infty} \nu^i d_{t-i}, \tag{22}$$

$(\hat{\eta}_t)$ and $(\hat{\mu}_t)$ satisfy

$$\hat{\eta}_t = \nu\hat{\eta}_{t-1} + (1-\nu)p_t, \tag{23}$$

$$\hat{\mu}_t = \nu\hat{\mu}_{t-1} + (1-\nu)d_t. \tag{24}$$

The variance terms are derived from

$$\hat{s}^2_{p,t} = (1-\nu)\sum_{i=0}^{\infty} \nu^i (p_{t-i} - \hat{\eta}_t)^2 \,, \tag{25}$$

where $\hat{s}^2_{d,t} = \hat{\sigma}^2_{d,t}$ is derived equivalently; $\hat{\sigma}^2_{p,t} = (1-\hat{\beta}^4)\hat{s}^2_{p,t}$ in this setting. The autocorrelation coefficient can be calculated from

$$\hat{\beta}_{t-1} := \frac{\hat{\gamma}_{t-1}}{\hat{s}_{p,t-2}\hat{s}_{p,t-1}} \,, \tag{26}$$

where $\hat{s}_{p,t-1} = \sqrt{\hat{s}^2_{p,t-1}}$ and

$$\hat{\gamma}_t = (1-\nu)\sum_{i=0}^{\infty} \nu^i (p_{t-i} - \hat{\eta}_t)(p_{t-i-1} - \hat{\eta}_{t-1}) \,, \tag{27}$$

$\hat{\sigma}^2_{t-1} = \hat{\sigma}^2_{d,t-1} + (1-\hat{\beta}^4)\hat{s}^2_{p,t-1}$. The implied actual law of the price process with exponential smoothing is

$$p_t = \lambda(\hat{\beta}^2_{t-1}p_{t-1} + (1-\hat{\beta}^2_{t-1})\hat{\eta}_{t-1} + \hat{\mu}_{t-1} - c\hat{\sigma}^2_{t-1}) \,. \tag{28}$$

By defining a process $Y_t := (p_t, p_{t-1}, \hat{\eta}_{t-1}, \hat{\mu}_t, \hat{\sigma}^2_{d,t}, \hat{s}^2_{p,t-1}, \hat{\gamma}_t)' \in \mathrm{R}^7$ and $\varphi = \mathbb{E}(d_t)\lambda/(1-\lambda)$, a sufficient condition for ergodic prices can be derived:

PROPOSITION 2 (Pötzelberger and Sögner, 2003b) *Let* $0 < \nu, \lambda < 1$ *and let* $V > 0$ *with*

$$V \leq \frac{1}{8c}\left[-1 + \sqrt{1 + \left(\frac{1-\lambda}{\lambda}\right)^2}\right] . \tag{29}$$

Define

$$\underline{U} = \frac{1-\lambda}{8c\lambda} \tag{30}$$

$$\bar{U} = \frac{1-\lambda}{4c\lambda} \tag{31}$$

$$U = \frac{1-\lambda}{8c\lambda} + \sqrt{\left(\frac{1-\lambda}{8c\lambda}\right)^2 - \frac{V + 4cV^2}{4c}} \tag{32}$$

$$B = 2V + V^2 \tag{33}$$

$$R = 6\underline{U} + 8\underline{U}^2 + 2V + 4V^2 \tag{34}$$

$$\Gamma = \max\{\lambda\nu + (1-\nu) + 4\nu(1-\nu)(\lambda c + 1)\bar{U}, \lambda c\nu\} \tag{35}$$

$$\Lambda^+ = \frac{1}{\max\{\sqrt{\lambda(1-\nu)}, \nu\}} - 1 . \tag{36}$$

If (i) (d_t) *is* iid *with bounded support contained in* $[\mathbb{E}(d_t) - V, \mathbb{E}(d_t) + V]$*, (ii) the initial values satisfy,* p_1*,* p_0*,* $\hat{\eta}_0 \in [\varphi - U, \varphi + U]$*,* $\hat{\mu}_1 \in [\mathbb{E}(d_t) - V, \mathbb{E}(d_t) + V]$*,* $\hat{\sigma}^2_{d,1} \in [0, 4V^2]$*,* $\hat{s}^2_{p,0} \in [0, 4U^2]$ *and* $|\hat{\gamma}_t)| \leq \hat{s}^2_{p,0}$*, where* $\varphi = \mathbb{E}(d_t)\lambda/(1-\lambda)$*, (iii)* $0 \leq \Gamma < \Lambda^+$ *and (iv)* $\Gamma < \sqrt{7}\Lambda^+ - \frac{B}{R}$*, then the process* (Y_t) *is ergodic. Note that the quantities defined above satisfy* $\underline{U} > 0$*,* $B > 0$ $R > 0$*,* $\Lambda^+ > 0$ *and* $\underline{U} \leq U \leq \bar{U}$*. Furthermore, there is a* $\zeta > 0$ *such that* $\Gamma < \sqrt{7}(\Lambda^+ - \zeta) - \frac{B}{R}$*.*

The intuitive explanation of Proposition 2 can be given as follows. To ensure that prices are ergodic conditions (i) on the exogenous parameters r, S, ζ_i, (ii) on the initial values of p_t and $\hat{\theta}_t$, and (iii) on the support of the dividend process have to be fulfilled. Similarly, as with sample autocorrelation learning, simulation experiments show that convergence to a stationary limit distribution need not be attained if these conditions are not met. Last but not least Pötzelberger and Sögner (2003b) checked whether ergodic price paths match the stylized facts of excess kurtosis and volatility clustering. Simulation experiments show, that these properties cannot be derived in this model setup.

4 Consistent Expectations Equilibria

Hommes and Sorger (1998) investigated the question under which conditions the agents' model parameters can be said to be consistent with some state variables observed in a deterministic setting. In their seminal work the authors defined consistency

by "parameters of a linear model being in line with state variables generated from the implied actual law". The idea behind the consistent expectations concept is that agents try to approximate a complex system by simple linear rules. Thus, by the consistent expectations concept, the requirement that agents know the implied law of motion of the economic system has been dropped. The consistent expectations concept can result in a set of equilibria including the rational expectations equilibrium but allows for other behavior in equilibrium that has been excluded by the stringent assumption of rational expectations.

In Sögner and Mitlöhner (2002) the consistent expectations concept is adapted to a stochastic setting where the limit behavior corresponds to a fixed point. For this special application the parameters under consideration are those of the perceived laws of prices and dividends, (10) and (11), respectively. An equilibrium will be called consistent if the state variable follows the implied actual law, (12) in our case, and the parameters θ fulfill restrictions on the moments of prices and dividends.

By using the definitions of Subsection 3.1, we can derive the asymptotic sample means, variances and the covariance by $t \to \infty$. Now the parameters μ and η have to be equal to the asymptotic sample means of p_t and d_t in a consistent expectations equilibrium. Moreover, the parameters σ_p^2 and σ_d^2 have to be equal to the corresponding asymptotic sample variances. Sögner and Mitlöhner (2002) define a consistent expectation equilibrium (CEE) as follows:

DEFINITION 1 (Sögner and Mitlöhner, 2002) *A consistent expectations equilibrium is a pair $\{(p_t)_{t=0}^{\infty}; \theta\}$, where $\theta := (\eta, \mu, \beta, \sigma_d^2, \sigma_p^2)$ is a vector of the parameters. This pair has to satisfy the following conditions:*

1. *The sequence $(p_t)_{t=0}^{\infty}$ fulfills the implied law of motion.*

2.a *The asymptotic sample average of the prices $\bar{p}_{t\to\infty}$ and the asymptotic sample average of dividends $\bar{d}_{t\to\infty}$ are equal to η and μ.*

2.b *The asymptotic sample variance of the prices $VAR_{\,t\to\infty}(p^2)$ and the asymptotic sample variance of dividends $VAR_{\,t\to\infty}(d^2)$ fulfill: $VAR_{\,t\to\infty}(p^2) = \sigma_p^2/(1-\beta^2)$ and $VAR_{\,t\to\infty}(d^2) = \sigma_d^2$.*

3. *For sample autocorrelation coefficients $\gamma_{t,j}$ the following is true: $sgn(\gamma_{j,\,t\to\infty}) = sgn(\beta^j),\ j \geq 1$*

If these conditions hold Sögner and Mitlöhner (2002) derive a unique consistent expectations equilibrium for the capital market model described by equation (12). By assuming an independent identically distributed dividend process the consistent expectations equilibrium is equal to the rational expectations equilibrium, i.e. they derive the result that the rational expectations equilibrium with iid dividends of this capital market model can also be obtained by a weaker equilibrium concept, where agents do not know the actual law of motion of the system and the characteristics of the stochastic dividend process.

When considering the model investigated in Subsection 3.2, we observe that neither the REE concept nor the CEE concept can be applied to ergodic prices. However, the CEE concept can be augmented easily. First, Hommes and Sorger (2001)

extended their concept to stochastic consistent expectation equilibria (SCEE), where ergodic state variables and constant parameters – equal to some proper sample means – characterize a SCEE. We shall call this type of CEE, SCEE of type I. But also this concept cannot be applied directly. Nevertheless, if (Y_t) is ergodic, the distribution of prices exhibits an invariant measure and the expectation of the model parameters exists. Note that with exponential smoothing $\mathbb{E}(p_t) = \mathbb{E}(\eta_t)$ holds for ergodic prices (where $\mathbb{E}(.)$ is the expectation operator). More precisely, consider that the process (Y_t) satisfies $Y_t \in L^1(P)$, a natural extension to the definition of a stochastic consistent expectations equilibrium is given by:

DEFINITION 2 (Type II SCEE) *A consistent expectations equilibrium is a pair* $\{(p_t)_{t=0}^{\infty}; \hat{\theta}_t\}$, *where* $\hat{\theta}_t := (\hat{\eta}_t, \hat{\mu}_t, \hat{\beta}_t, \hat{\sigma}^2_{d,t}, \hat{s}^2_{p,t})$ *is a stochastic parameter vector. This pair fulfills the following conditions:*

1. *The process* $Z_t = (p_t\ \hat{\theta}_t)'$ *is ergodic with* $Z_t \in L^1(P)$.
2. *The model parameters satisfy:*[3] $\mathbb{E}(p_t) = \mathbb{E}(\hat{\eta}_t)$, $\mathbb{E}(d_t) = \mathbb{E}(\hat{\mu}_t)$, $\mathbb{V}(p_t) = \mathbb{E}(\hat{s}^2_{p,t})$, $\mathbb{V}(d_t) = \mathbb{E}(\hat{\sigma}_{d,t})$ *and* $\mathbb{E}[(p_t - \mathbb{E}(p_t))(p_{t-1} - \mathbb{E}(p_t))] = \mathbb{E}(\hat{\gamma}_{p,t})$.

Therefore, if (Y_t) fulfills the requirements of Proposition 2, the stochastic capital market equilibrium satisfies the properties of a type II SCEE.

5 Adaptive Belief Systems

As a an alternative model to the CEE framework we quickly present the adaptive belief system that was introduced to the literature by Brock and Hommes (1997) and Brock and Hommes (1998). The most important characteristic of these models are the way in which agents choose their forecasting rules. In general it is assumed that there are at least two different types of trades, i.e. fundamentalists and chartists. Both types use simple forecasting rules. For example the fundamentalists might use the fundamental value of the asset as forecast while the chartists use some simple trend chasing rule. The two populations of different traders do not remain constant over time but change according to a performance measure attached to each forecasting rule. If according to this performance measure the forecasting rule of the chartists is consistently higher than that of the fundamentalists, investors start to migrate into the direction of chartists. This migration of traders influences the equilibrium prices.

The starting point of the model formulation for adaptive belief systems is the asset model described in Section 2. We add to this general formulation some specific assumptions.

A1. Investors share homogenous beliefs about conditional variances i.e. $\mathbb{V}_{it}(\mathbf{p}_{t+1} + \mathbf{y}_{t+1}) \equiv \mathbb{V}_t(\mathbf{p}_{t+1} + \mathbf{y}_{t+1}), \forall i, t.$

A2. $\mathbb{E}_{it}\mathbf{y}_{t+1} = \mathbb{E}_t\mathbf{y}_{t+1}, \forall i, t$ (if, for example, $\mathbf{y}_t$ follows an AR1-process $E_t\mathbf{y}_{t+1} = a_1 + a_2\mathbf{y}_t$)).

[3]Note that $\mathbb{E}[(p_t - \mathbb{E}(p_t))(p_{t-1} - \mathbb{E}(p_{t-1}))] = \mathbb{E}[(p_t - \mathbb{E}(p_t))(p_{t-1} - \mathbb{E}(p_t))]$ for an ergodic p_t in $L^1(P)$.

A3. Beliefs about future prices are given by $\mathbb{E}_{it}\mathbf{p}_{t+1} = \mathbb{E}_t\mathbf{p}^*_{t+1} + f_i(x_{t-1}, x_{t-2}, \dots)$, $\forall h, t$, where $x_t = p_t - p^*_t$ denotes the deviation of the price from its fundamental value.[4]

On the basis of the three assumptions stated we are able to derive an equilibrium system in the deviations of the actual prices from its fundamental value,

$$Rx_t = \sum_{i=1}^{N} n_{it} f_{it}, \tag{37}$$

where n_{it} is the fraction of investors that uses the same forecasting rule i. All together there are N different forecasting rules available.

The forecasting rules are chosen on the basis of a simple discrete choice model in which we can interpret n_{it} as the probability that at time t the forecasting rule i is chosen. We specify these probabilities as

$$\begin{aligned} n_{it} &= \text{probability that forecasting rule } i \text{ is chosen at time } t & (38)\\ &= \exp(\beta U_{i,t-1})/Z_t, & (39)\\ Z_t &= \sum_i \exp(\beta U_{i,t-1}), \end{aligned}$$

where $U_{i,t-1}$ is a measure of fitness of the forecasting rule and $1/\beta$ can be interpreted as a measure of uncertainty.

In order to define a measure of fitness for the forecasting rule, one has to identify a performance concept. For that reason let us define

$$\rho_t := \mathbb{E}_t R_{t+1} = \mathbb{E}_t x_{t+1} - Rx_t = x_{t+1} - Rx_t$$

which is the excess return given rational expectations and let

$$\rho_{it} := \mathbb{E}_{it} R_{t+1} = \mathbb{E}_{it} x_{t+1} - Rx_t = f_{it} - x_{t+1} + \rho_t$$

be the belief of investor type i. Moreover for investor i we can define risk-adjusted profits as

$$\pi_{it} := \pi(\rho_t, \rho_{it}) \equiv \rho_t q(\rho_{it}) - \frac{\zeta}{2} q(\rho_{it})^2 \mathbb{V}_{it}(R_{t+1}), \tag{40}$$

where $q(.)$ is the demand for the risky asset,

$$q(\rho_{ht}) = \frac{\rho_{it}}{\zeta \mathbb{V}_{it}(R_{t+1})}$$

derived from the investors maximization problem $\max_q\{\rho_{it} q - \frac{\zeta}{2} q^2 \mathbb{V}_t(R_{t+1})\}$. If as measure of fitness past differences between utility in case of the REE forecasting rule and that of the investor i are chosen, we get

$$U_{it} = d\pi(\rho_{t-1}, \rho_{i,t-1}) + \eta U_{i,t-1}, \tag{41}$$

[4]We denote the solution to our asset pricing model which satisfies the no bubbles condition as the fundamental value p^*_t.

with

$$d\pi(\rho_{t-1}, \rho_{i,t-1}) = \pi_{i,t-1} - \pi_{t-1} = -\frac{1}{2\zeta(\mathbb{V}_{i,t-1}(R_t) + \sigma_\delta^2)}(x_t - f_{i,t-1})^2,$$

where the parameter η measures the influence of past profits on the current level of fitness.

These specifications fully characterize the adaptive belief system. Depending now on different assumptions about the forecasting rule, different variations of the model can be derived. In a model with constant expected dividend payments $\mathbb{E}(d_t) = \hat{d}$, Brock and Hommes (1998) and Brock and Hommes (1997) use two different types of traders: chartists with the forecasting rule $f_{it} = g_i xt - 1 + b_i$ where g_i and b_i are constants, and fundamentalists with the rule $f_{it} \equiv 0$. Under these assumptions it can be shown that there exists a nonlinear deterministic reduced form system in x_t that fully characterizes equilibrium behavior. Gaunersdorfer (2000) fully analyzes this system and explores a route to chaos. In particular if the intensity β is sufficiently high, chaotic price dynamics may arise in the reduced form system. Moreover as is pointed out in Gaunersdorfer and Hommes (2002) the simulated price paths from the adaptive belief system is capable of explaining some stylized facts such as volatility clustering. Hence, one can argue that the adaptive belief system embedded in a simple asset market model is an attractive alternative to existing rational expectations models.

6 Conclusions and Discussion

The early literature on least squares learning, e.g., Bray (1982), Blume et al. (1982), Blume and Easley (1982), has often been used to support rational expectations equilibria by the fact that for these settings convergence to the rational expectations equilibrium is attained under fairly mild restrictions. However, our analysis shows that the learning of variances introduces an important source of instability to this class of capital market models. All our setups share the property that due to strong distortions by the random dividend process the estimate of the variance of dividends increases. If a distortion is strong enough the quadratic terms in the estimates of the variances of prices and dividends become dominating in (12), such that (p_t) is forced to exhibit unstable behavior. Thus, the main insight of Section 3 is that learning can but need not result in stability. For a stable solution consistent equilibria can be derived.

Since the homogenous agent setups did not provide us with paths exhibiting excess kurtosis and volatility clustering, we investigated adaptive belief systems, where the fraction of forecast functions applied depends on some fitness measure.

Bibliography

Arthur, B., LeBaron, B., and Palmer, R. (1997). Time series properties of an artificial stock market. *SSRI Working Paper*, 1997(9725).

Blume, L. E., Bray, M., and Easley, D. (1982). Introduction to the stability of rational expectation equilibrium. *Journal of Economic Theory*, 26:313–317.

Blume, L. E. and Easley, D. (1982). Learning to be rational. *Journal of Economic Theory*, 26:340–351.

Bray, M. (1982). Learning, estimation, and the stability of rational expectations. *Journal of Economic Theory*, 26:318–339.

Brock, W. A. and Hommes, C. H. (1997). A rational route to randomness. *Econometrica*, 65(5):1059–1095.

Brock, W. A. and Hommes, C. H. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, 22:1235–1274.

Chen, X. and White, H. (1998). Nonparametric adaptive learnig with feedback. *Journal of Economic Theory*, 82:190–222.

Gaunersdorfer, A. (2000). Endogenous fluctuations in a simple asset pricing model with heterogeneous agents. *Journal of Economic Dynamics and Control*, 24:799–831.

Gaunersdorfer, A. and Hommes, C. (2002). A nonlinear structural model of volatility clustering. Technical report, University of Vienna.

Hommes, C. and Sorger, G. (1998). Consistent expectation equilibria. *Macroeconomic Dynamics*, 2:287–321.

Hommes, C. and Sorger, G. (2001). *Stochastic Consistent Expectation Equilibria*. Mimeo, University of Amsterdam.

Marcet, A. and Sargent, T. J. (1989). Convergence of least squares learning mechanics in self-referential linear stochastic models. *Journal of Economic Theory*, 48:337–368.

Muth, J. F. (1961). Rational expectations and the theory of price movements. *Econometrica*, 29:315–335.

Pötzelberger, K. and Sögner, L. (2003a). Equilibrium and learning in a non-stationary environment. In Neck, R., editor, *Proceedings of the 2001 IFAC Symposium on Modeling and Control of Economic Systems*, pages 191–196. Springer, Berlin.

Pötzelberger, K. and Sögner, L. (2003b). Stochastic equilibirum: Learning by exponential smoothing. *Journal of Economic Dynamics and Control*, 27(10):1743–1770.

Pötzelberger, K. and Sögner, L. (2004). Sample autocorrelation learning in a capital market model. *Journal of Economic Behavior and Organization*. To appear.

Routledge, B. (1999). Adaptive learning in financial markets. *Review of Financial Studies*, 12:1165–1202.

Sargent, T. J. (1993). *Bounded Rationality in Macroeconomics*. Clarendon Press, Oxford.

Schönhofer, M. (1997). *Konsistentes adaptives Lernen in Finanzmarkt- und Makromodellen*. Peter Lang Verlag, Fankfurt am Main.

Sögner, L. and Mitlöhner, J. (2002). Consistent expectations equilibria in an artificial stock market. *Journal of Economic Dynamics and Control*, 26(2):171–186.

Stokey, N. and Lucas, R. (1989). *Recursive methods in economic dynamics*. Harvard University Press, Cambridge, MA.

Tay, N. S. and Linn, S. C. (2001). Fuzzy inductive reasoning, expectation formation and the behavior of security prices. *Journal of Economic Dynamics and Control*, 25(3–4):321–361.

Part III

Agent-Based Simulation Models

The Artificial Economy: A Generic Simulation Environment for Heterogeneous Agents

David Meyer and Alexandros Karatzoglou

1 Introduction

Agent-based simulations are often implemented by using an object-oriented style of programming, allowing for detailed modeling of the artificial actors. In the following, we consider an agent-based economic simulation in which economic entities (firms, (groups of) consumers, investors, markets, and the like) can be thought of as interacting agents. A typical simulation combines several agents, defines their relationships, and observes their resulting interactions over time. After the simulation design has been defined (Richter and März, 2000), running a simulation usually amounts to writing a control program in one's favorite programming language, named the *Simulation Manager* (see below), that coordinates a set of previously implemented, autonomous agents.

One might wish that the agents would have standardized interfaces so that they automatically have the same bindings allowing their use in simulations as modularized components. General mechanisms for providing standardized interfaces (like CORBA) do exist, but usually require advanced programming skills. Our objective, then, is to provide an easy-to-use mechanism suitable for use in data-analytical environments like MATLAB (The Mathworks, Inc., 2003), Octave (Eaton, 2003), or R (R Development Core Team, 2003), as they offer convenient ways to analyze simulation results and are also (typically) used for implementing objects and methods. We also deal with varying parameters in controlled experiments and provide a scheduling scheme to determine the order of invocation within a single experiment (design) and the number of runs (periods) per design.

Consider the simple introductory example involving two competing firms, named `Firm "A"` and `Firm "B"`, respectively, operating in a consumer market (see Figure 1). Each firm could be modularized itself, having agents responsible for marketing, production, and finance. Market coordination and clearing may be performed by a consumer market agent, which models a (disaggregated) consumer population. In addition, a global environment, representing the common knowledge of all agents, is typically involved: this environment may be stored, e.g., in an SQL database, thus solving problems arising from simultaneous access by different agents (such as transaction control), or managed by an information broker similar to the one described in Wilson et al. (2000)—but these mechanisms are highly specific to the simulation design.

As an extension to Meyer et al. (2001) and Meyer et al. (2003), our software also offers a simple yet flexible communication system which can be used for direct information exchange between the agents, without the need of using a database. Also, there might be a need for dynamic creation of agents (when, e.g., new department or daughter firm agents shall be created) and/or communication channels (e.g., when two

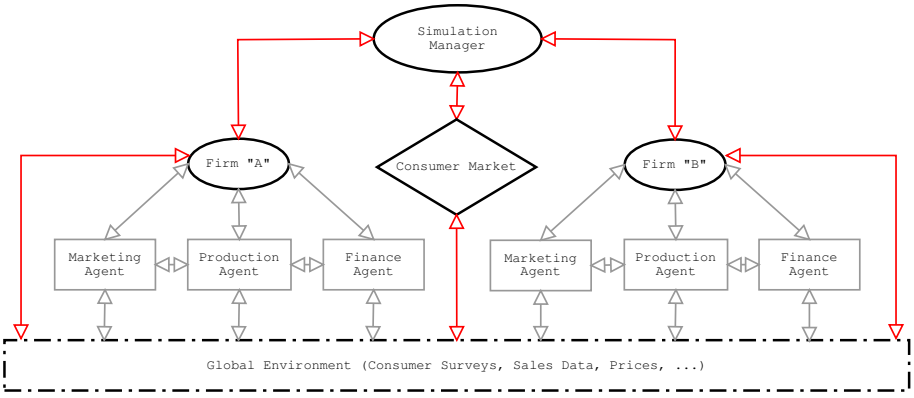

Figure 1: A simple simulation with two competing firms, each consisting of 3 agents for marketing, production and finance, respectively.

firms decide to collaborate). Both can be done at runtime during a simulation run.

The remainder of this chapter is structured as follows: First, we describe how the specification of the simulation settings is done in XML for a generic simulation manager, supporting multiple design specifications. Then, we explain the "normalization" of agent interfaces via a wrapping technique, thus allowing the simulation manager to treat all agents the same way. Section 4 deals with SIMENV's communication mechanism, and is followed by a section on mechanisms for dynamic simulation setups. The last section treats the remaining control mechanisms (such as the meta-agent, handling of random number generation, and e-Mail notifications).

2 The Simulation Manager

2.1 A Typical Simulation Cycle

A complete simulation includes several designs with (typically) different parameter settings and/or a modified set of agents. Designs can be run repeatedly. Figure 2 sketches a typical simulation for a single design. After the simulation manager and the agents have been initialized, the simulation enters the main loop: after updating the time index, all agents—grouped by phase—are run for one cycle. All agents of one phase need to complete their tasks before the next phase is entered. When the last phase is done, the next loop is entered. Upon completion of the final cycle, a cleanup is performed. This is repeated (usually with changing parameter sets) a specified number of times.

Our implementation of a generic simulation manager behaves just as described, handling "unified" agents. Because agents can be implemented in different programming languages (such as R and MATLAB) on possibly different platforms (such as Windows and Linux) depending on the user's needs or skills, the simulation manager has to be capable of operating in a technically heterogeneous environment, and

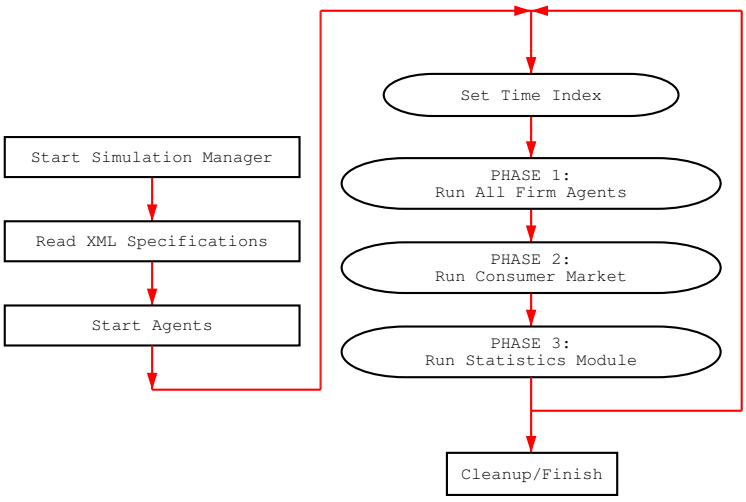

Figure 2: A typical simulation cycle

therefore is implemented in JAVA (Gosling et al., 2000), a platform independent programming language, that offers good support for network communication. Although simple in design, we consider it powerful enough to be used as a ready-made tool. It is capable of dealing with an arbitrary number of agents in different phases (e.g., a market clearing agent should only be started when all "normal" agents are done) by varying an arbitrary set of parameters through different designs. These parameters (such as market/product characteristics, initial prices, and budgets) are offered by the simulation manager to the agents at the beginning of each new design block by using a simple broadcast mechanism. Information can either be public (propagated to all agents) or private (propageted to specific agents). Usually, public information also includes technical information, like the current period (updated at the begin of each cycle), or the agent identifier (which the agent might include in its output information). The simulation components are specified in a definition file read by the simulation manager at startup.

2.2 Using XML for Simulation Settings

The SIMENV framework is based on an object-oriented approach. Conceptually, we assume the existence of agent classes with methods (functions) and attributes (parameters). A simulation setup consists of assigning one or more instances of these classes to run levels, along with a certain number of parameters. These assignments can vary from design to design. We use XML to define these settings: the Extensible Markup Language (World Wide Web Consortium, 2000). For example, the definition file for a sample simulation including firms and consumers with two designs might look as follows:

```
<?xml version = "1.0"?>
<!DOCTYPE simulation SYSTEM "simulation.dtd">

<simulation>
  <alldesigns repeat = "20" cycles = "30">
    <agent name = "consumer" level = "2" instances = "100">
      <p name = "reservprice">5</p>
      <p name = "budget">10</p>
    </agent>
  </alldesigns>
  <design name = "A">
    <agent name = "firm" level = "1" instances = "2">
      <p name = "type">mass</p>
      <p name = "budget">100</p>
    </agent>
  </design>
  <design name = "B">
    <agent name = "firm" level = "1" instances = "2">
      <p name = "type">niche</p>
      <p name = "budget">50</p>
    </agent>
  </design>
</simulation>
```

The tags are described in the following:

- The first two lines form the XML header, which is common to all XML files; the specific structure is defined in the "`simulation.dtd`"-file, indicated in the second line.
- The document starts with the `<simulation>` root tag. This tag has several parameters, which are described in Section 6 on control structures. `<simulation>` may contain an arbitrary number of
- `<design>` tags with the parameters:
  - `name` for identification in log files,
  - `repeat` for design replications, and
  - `cycles` for the number of periods.

  For convenience, parts common to all designs can be put into an (optional) `<alldesigns>` section, as has been done for the `consumer` agents. Each `<design>` tag may contain an arbitrary number of
- `<agent>` tags with the attributes `name`, (number of) `instances`, and `level` (the phase, in which the agent is scheduled to run). Parameters common to all agents can be put into an optional `<allagents>` section. For each `<agent>` tag, an arbitrary number of

- <p> tags specify the parameters for this particular agent, the name attribute this time indicating the parameter name. Each agent "inherits" the parameters from the corresponding <agent> section in the <alldesigns> section, if any, as well as all parameters from a possibly existing <allagents> section.

If the same parameters exist both in general sections (<alldesigns> or <allagents>) and the <design> sections, the more specific parameters overrule the more general ones. By using this rather general framework, one is able to specify whole design plans in a flexible way. Simple design plans (like the full-factorial plan) are usually created in an automated way, but the structure also enables more elaborated, fractional plans: if one is not interested in the influence of all possible factor combinations, it is possible to reduce the number of parameter combinations by following certain design rules (see, e.g., Dey and Mukerjee, 1999), thus substantially reducing the simulation time needed. So far, we described how parameters are specified in simulations. Now, we move to the agents' side to see how the methods are defined.

3 Agent Specification

One of the basic motivations for SIMENV was the need for integrating agents implemented in (possibly heterogeneous) high-level programming environments. To make such agents "simulation-aware", we need

1. a translator which accepts generic method calls from the simulation manager and passes the corresponding method call to the agent, and
2. an interface definition which describes the corresponding translation mappings.

3.1 Wrapping Agents

First, we describe the program acting as an intermediary, translating the simulation manager's JAVA calls to the native method calls in the agent's programming environment. This program "wraps" the agent and exports through JAVA a standardized interface (we refer to this program as the *wrapper*). The translation of the agents' interface is stored in an XML-based interface definition format, which (mainly) defines one-to-one correspondences to the JAVA interface calls.

The whole concept is illustrated in Figure 3: a typical simulation takes several agents, defines their relationships, and observes the resulting interactions between the objects over time. The agents interact with the environment and subsequently with each other. The whole simulation is coordinated by a central agent that starts and synchronizes the simulation components (agents). The central coordinating agent (simulation manager) makes the JAVA calls to the wrapper, and the latter—initially having parsed the XML interface definition of the agent—translates the call and executes the interpreter command as if it were typed at the command prompt. Note that SIMENV currently targets interpreter-based environments only, redirecting their standard input and standard output devices. Compiled code (from, e.g., C or Fortran programs) can

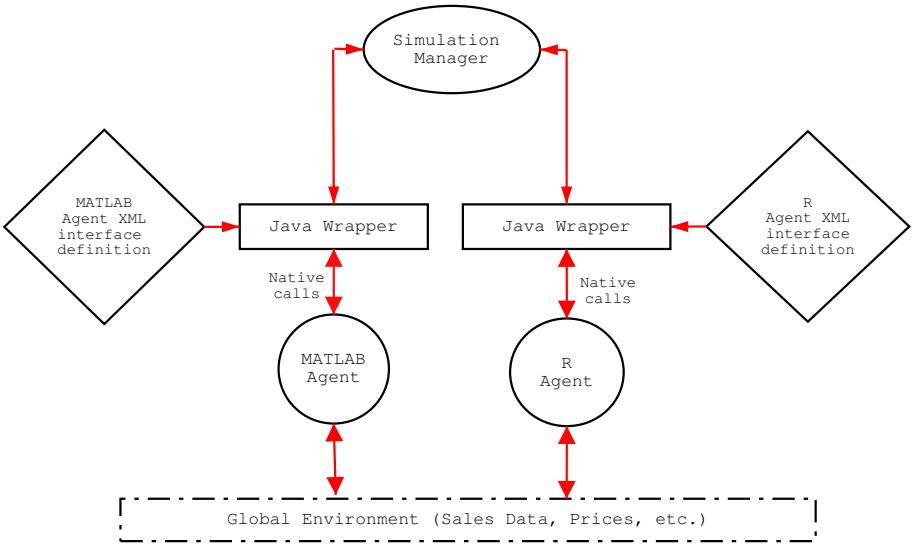

Figure 3: A simple simulation with two agents

be integrated using a command shell as "interpreter", which subsequently is used to call (execute) binary programs instead of making function calls.

3.2 How Agents Are Controlled during Simulations

Before we can use XML to define agent interfaces, we have to look at an agent's simulation "life" to derive the functionality to be handled by the wrapper. It can be summarized in the following steps:

1. Start of the interpreter (MATLAB, R, ...).
2. Loading of the agent's source code.
3. Setting of some variables by the manager (like the data base name).
4. Initializing of variables, opening of a database connection, etc.
5. *Action loop (executed several times):*
   (a) setting of periodic information (like the time index) by the simulation manager,
   (b) execution of task function,
   (c) possible retrieving of results by the simulation manager.
6. Cleaning up (saving of results, closing of database connections), and finally

7. Quitting from the interpreter.

From this "life-cycle", we derive the specification for an appropriate interface.

3.3 Using XML for Defining Agent Interfaces

The agent's interface is reduced to six main methods: `start`, `boot`, `init`, `action`, `finish` and `stop`, corresponding to the main steps just mentioned, and two helping methods: `setattr` and `getattr`, for information passing. In addition, we require a `printdone`-method, along with the definition of a `donestring`, both needed for communication control: each command string sent to the interpreter is immediately followed by the command defined by `printdone`, which should print a specified OK-message. If this string is detected by the wrapper, an OK-signal is sent to the simulation manager which subsequently can assume that the command has been executed completely and that the agent is ready for more commands.

The interface specified in the example XML file below defines a simple R agent:

```
<?xml version = "1.0"?>
<!DOCTYPE wrapper SYSTEM "wrapper.dtd">

<wrapper>
  <start>R --quiet --vanilla</start>
  <boot>source(" Ragent.R" )</boot>
  <init>init()</init>
  <action>action()</action>
  <finish>finish()</finish>
  <stop>q()</stop>

  <setattr>assign(``<name/>'',<value/>)</setattr>
  <getattr>print(<name/>)</getattr>

  <printdone>printdone()</printdone>
  <donestring>OK</donestring>
</wrapper>
```

The tags are described in the following:

- The `<start>` tag defines the start-command (in 'quiet' mode), executed as a shell command.
- The `<boot>` tag (optional) typically encloses a command for loading files into the interpreter.
- `<init>`, `<finish>` and `<action>` specify user-defined functions (for initialization, cleanup and the main `action` method, respectively); `<init>` and `<finish>` are optional.
- The `<setattr>` tag contains a method call that sets the various attributes of the agent (e.g., `TIME` or `NAID`). It takes 2 parameters: `<attrname/>` and `<value/>`,

which are empty tags and play the role of placeholders. They are replaced by real values provided by the simulation manager before the command is executed. It should be called as many times as necessary to set all agent parameters. Note that the implementation could, of course, be simplified by an ordinary variable assignment like `<name/> = <value/>` as in the example, but the use of a separate function allows the mapping from the generic variable names to the specific variable names of the implementation, thus preserving the agent's name space.

- `<getattr>` defines the complementary function to `<setattr>`: the implementation should simply print the requested value; it is parsed and returned to the client.

- `<printdone>`, as mentioned above, should simply print a defined OK-message enabling flow control. This message must be specified in the `<donemessage>` tag. Note that for this method, one should implement an extra function and not, for example, simply use a `print` statement.

In addition to the parameters specified in the "`simulation.dtd`", the simulation manager offers some internal information to all agents (using the specification in `<setattr>`). Agents of course are free to choose to use or not to use this information. This is another reason why we recommend implementing explicit functions for the parameter handling, allowing to keep the name space clean and to filter unused information. Currently, the public information set includes: `TIME` (the current time index), `SEED` (the current seed to be used for random number generation), `NAID` (the agents' internal unique identifier), and `ANAME` (the agents' full name). In addition, `NRID` (the current replication), `NDID` (the unique internal design identifier), and `DNAME` (the optional design name) are passed to the special `meta` agent (see Section 6). The `<wrapper>` tag has an optional parameter: `separator`, which can be used to replace the dot separator (`"."`) by another character, as the dot is not allowed in variable names in all programming environments.

On the other hand, the simulation manager may also retrieve some information from the agent (as specified in `<getattr>`): currently, only three variables—`CTRL`, `CTRL.TARGET`, and `CTRL.PARAMETERS`—are scanned. These "control" variables are used to pass commands to the simulation manager and are described in Section 5 on dynamic settings.

4 Communication Structures

Evidently, agents must be enabled to interact, for it is the outcome of this interplay which is of interest in agent-based simulations. In addition, there are simulation setups explicitly focused on the study of communication structures and cooperation. One possibility is the use of a database, modeling the agents' global environment. But databases clearly have two main disadvantages: the simulation performance significantly decreases, while on the other hand the implementation complexity increases. The user (scientist) is forced to take care of database design to avoid redundancies,

and to account for transaction control in the agents' program code to avoid concurrency problems. Therefore, we offer an alternative facility for information exchange, namely the specification of direct communication channels from one agent to another or to a group of agents. These channels can subsequently be used for the transmission of character-string messages. Note that this is not a too severe restriction as character strings can encode data objects of great complexity when one uses a structuring meta-data language like XML. For example, Meyer et al. (2004) have designed an XML-based format for statistical data allowing, e.g., for multidimensional arrays and recursive, tree-like list structures, which should suffice for most applications.

We distinguish three types of channels: "one-to-one", "one-to-group", and "one-to-all" (or "broadcast"). "one-to-one" relates one instance of a class to another instance of (possibly the same) a class. A "one-to-group" relation targets all instances of a class (again, possibly the own class). "broadcast" informations obviously are passed to all instances in the simulation.

Collection and delivery of mails is done in one step after the agents' `init` phases (allowing dynamic initializations like random-generated start scenarios) and after all `action` calls of one level. The Simulation Manager collects mails by applying the `getattr` call of the sender agents on each registered communication variable. At the target side, delivery is done by setting a variable with a unique name to the message string using the `setattr` method of the target agent When the target agent does not exist, the message is ignored.

5 Dynamic Settings

Some kind of simulations, in particular in the context of evolutionary research and network industries, necessitate a dynamic setup, that is, agents and/or communication channels are created and discarded during the simulation. These dynamics are handled by the SIMENV framework using special control variables at the agent side: `CTRL`, `CTRL.TARGET`, and `CTRL.PARAMETERS`, which can be used by agents to alter the initial setting defined in the XML design file. Currently, four `CTRL` commands are handled: `"start"` and `"stop"` for the instantiation of new agents, and `"commAdd"` and `"commRemove"` for the construction and destruction of communication channels. `CTRL` variables are scanned and possible commands are executed right before message exchange takes place.

6 Control Issues

In addition to the basic functionality so far described, the simulation framework offers several control facilities. First, a "meta" agent can be defined, which differs from other agent in several ways:

- It is only created once at the beginning of the simulation, that is, "survives" the beginning of new replications and designs unlike the other agents which are restarted at these occasions.

- It has full information on the simulation schedule, that is, in addition to `TIME` also gets `NDID`/`DNAME` (design number/design name) and `NRID` (replication number).
- The `init`, `finish`, and `action` methods are replaced by several other methods, allowing the `meta` agent to perform tasks other agents cannot: `<preSim>` and `<postSim>` are called before/after a simulation is started/stopped, `<preDesign>` and `<postDesign>` before and after designs, `<preRepeat>` and `<postRepeat>` before and after replications, and `<preRun>` and `<postRun>` at the beginning and at the end of every period. Typical applications for these methods are database management (initialization, cleanup between designs) and logging.
- The `meta` agent is passive: it can receive messages (e.g., for logging purposes), but is not able to send messages or to start agents, as it is not expected to influence the simulation itself.

Further, the `<simulation>` tag allows the specification of additional parameters, such as:

- `seed` for the control of the agents' random number generators,
- `mailserver`, `mailto`, and `mailfrom` for sending optional status emails (e.g., in case of abnormal termination of a simulation),
- `timeout` for detecting non-terminating agents (due, e.g., to programming errors or dead locks)

7 Summary

In this work we introduced SIMENV, a generic simulation framework suitable for agent-based simulations featuring the support of heterogeneous agents, hierarchical scheduling, and flexible specification of design parameters. One key aspect of this framework is the design specification: we use a format based on the Extensible Markup Language (XML), that is simple-structured yet still enables the design of flexible models, with the possibility of varying both agent population and parameterization. Further, the tool allows the definition of communication channels to single or group of agents, and handles the information exchange. Also, both (groups of) agents and communications channels can be added and removed at runtime by the agents, thus allowing dynamic settings with a agent population and/or communication structures varying during the simulation time. A further issue in agent-based simulations, especially when ready-made components are used, is the heterogeneity arising from both the agents' implementations and the underlying platforms: for this, we presented a wrapper technique for mapping the functionality of agents living in an interpreter-based environment to a standardized JAVA interface, thus facilitating the task for any control mechanism (like a simulation manager) because it has to handle only one set of commands for all agents involved. Again, this mapping is made by an XML-based definition format.

Bibliography

Dey, A. and Mukerjee, R. (1999). *Fractional Factorial Plans*. Wiley, New York.

Eaton, J. W. (2003). Octave software version 2.0.17, `http://www.octave.org/`.

Gosling, J., Joy, B., Steele, G. L., and Bracha, G. (2000). *The Java Language Specification*. Addison-Wesley, Boston, second edition.

Meyer, D., Buchta, C., Karatzoglou, A., Leisch, F., and Hornik, K. (2003). A simulation framework for heterogeneous agents. *Computational Economics*, 22(2):285–301.

Meyer, D., Karatzoglou, A., Buchta, C., Leisch, F., and Hornik, K. (2001). Running agent-based simulations. Working Paper 80, SFB "Adaptive Information Systems and Modeling in Economics and Management Science".

Meyer, D., Leisch, F., Hothorn, T., and Hornik, K. (2004). StatDataML: An XML format for statistical data. *Computational Statistics*, 19(3):493–509.

R Development Core Team (2003). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.

Richter, H. and März, L. (2000). Towards a standard process: The use of UML for designing simulation models. In *Proceedings of the 2000 Winter Simulation Conference*, pages 394–398.

The Mathworks, Inc. (2003). MATLAB software: Release 13. Natick, MA: The Mathworks, Inc., `http://www.mathworks.com/`.

Wilson, L. F., Burroughs, D., Sucharitaves, J., and Kumar, A. (2000). An agent-based framework for linking distributed simulations. In *Proceedings of the 2000 Winter Simulation Conference*, pages 1713–1721.

World Wide Web Consortium (2000). *Extensible Markup Language (XML), 1.0 (2nd Edition)*. Recommendation 6-October-2000. Edited by Tim Bray (Textuality and Netscape), Jean Paoli (Microsoft), C. M. Sperberg-McQueen (University of Illinois at Chicago and Text Encoding Initiative), and Eve Maler (Sun Microsystems, Inc.). Reference: `http://www.w3.org/TR/2000/REC-xml-20001006`.

Disruptive Technologies: the Threat and its Defense

Christian Buchta, David Meyer, Andreas Mild, Alexander Pfister, and Alfred Taudes

1 Introduction

Based on extensive long-term studies of the disk drive and other industries, Christensen (1997) introduced the concept of "disruptive technology". According to Christensen, initially such a technology is employed in a novel market segment, and, when judged according to the features most relevant to the incumbents' current customers, is inferior to the technology used by the incumbents in the established market segment. Nevertheless, over time the firms using the disruptive technology are able to successfully invade the established market segment from the lower end of the market and industry leadership changes. Christensen's finding provides empirical support to the resource-based and organizational learning perspective of the theory of the firm, whereas other approaches in general predict advantages for incumbents due to learning by doing, economies of scale and scope, network economies of scale, etc. (see, e.g., Klepper and Simons, 1997; Rumelt, 1981; Mas-Colell et al., 1995).

Table 1 provides an example of a disruptive technology: 5.25 inch disk drives were used in the early eighties' desktop computers and, initially, were inferior to the 8 inch drives used in minicomputers in terms of capacity, access time and cost/MB – the features most relevant to a minicomputer user. However, by 1986 industry leadership changed from CDC, the leading 8 inch vendor, to the new entrant Seagate, and most of the firms that were producing 8 inch drives vanished (see Christensen, 1993, p. 543). Christensen also demonstrates that it is the incumbents who are leading in "sustaining technologies", i.e. innovations that follow the current trajectory of technological improvement, and are trying to find new technical solutions to tackle the flattening of the current technology's S-curve. Thus, technological (in)competency cannot explain the failure of industry leaders, but this is rather done by factors rooted in the way new

Table 1: Disruptive Technology 5.25 Inch Drives (Christensen 1993, p.15)

Feature	8 Inch Drives (Minicomputer)	5.25 Inch Drives (Desktop Computer)
capacity (MB)	60	10
peripheral volume (inch^3)	566	150
weight (pounds)	21	6
access time (ms)	30	160
cost/MB ($)	50	200
unit cost ($)	3000	2000

product development projects are valued. Empirical evidence suggests the following causes for disruption:

Market Segment Overlap: Disruption can only occur if consumers of different segments have basically the same needs with different feature weights, though. As shown in Table 1, lower system price can compensate for inferior product features. Learning by entrant firms must be faster than the adaptation of the customers' needs, allowing them to follow the new, disruptive trajectory of improvement to catch up with the incumbents from below (Christensen and Bower, 1996).

Incentives: If an incumbent considers switching to the trajectory of a new disruptive technology early, it has to deal with the fact that important current customers are given up for highly insecure new markets. Initially, these are too small to support the growth rate of the incumbent's current organization and—given the current organizational design—offer lower margins (Christensen and Bower, 1996).

Organizational Inertia: An organizational design is adapted to the needs of the firm's customers (Hauser and Clausing, 1988) and frames the way the environment is seen and how problems are solved. This makes radical change hard and time-consuming. Also, an integrated firm is conflict-ridden and hard to manage if the degree of commonality (economies of scope) is low. Henderson and Clark (1990), for instance, show that incumbents often fail when confronted with architectural innovations rather than with the introduction of new components as the internal distribution of labor and communication channels have to change. Frequently, disruptive technologies entail new architectures based on standard, off-the-shelf components (Christensen, 1997, see). Similarly, Tushman and Anderson show that in the minicomputer and airline industries, competence-destroying innovations were made by new firms while competence-enhancing ones were made by incumbents (Tushman and Anderson, 1986).

Given these empirical findings, Christensen suggests that disruptive technologies can best be tackled by continuous monitoring of potentially overlapping market segments, long-term projections of technological trajectories and, to provide the appropriate learning environment, the setup of a completely separated, independent new organization in the market segment where disruption is expected to originate.

Several authors have developed formal models to study disruption: Adner (2002) formulates a market-driven model to analyze market conditions under which disruption occurs. Adner introduces the concepts of preference symmetry and preference overlap to characterize the relationship between preferences of different market segments. Using an agent-based computer simulation with myopic firms, he identifies different competitive regimes: convergence, isolation, and disruption. Focusing on the market conditions under which these regimes arise, Adner uses a simplified technological model: firms can move freely to reach any position within a certain distance, i.e., there are no predefined technological trajectories in his model (for a similar "history-friendly" model of the computer industry, see Malerba et al., 1999).

Nault and Vandenbosch (2000) identify conditions under which an entrant is able to outperform an incumbent in a rational, game-theoretic setting. In their view, disruptive technologies lead to a next-generation product with a greater market response and, therefore, higher cash flows. They define capability advantages as lower launching costs for the next-generation product. Under the condition that the entrant has a capability advantage in a disruptive technology, it is able to outperform the incumbent even though both technologies are available to both firms at the same time and both players are perfectly rational.

This article endeavors to add another important aspect to the explanation of the emergence of disruption and its defense: using rational, myopically optimizing firms, we study the influence of organizational inertia and technological efficiency on the emergence of competition between an incumbent and an entrant using a new technology. This new technology is characterized by its efficiency and is also available to the incumbent. While technological efficiency determines the speed of improvement offered by a technology per se, organizational inertia determines the speed at which an organization can be adapted so as to actually reach a desired product position. We thus endogenize the cost differences exogenous to the Nault & Vandenbosch model and characterize each technology via a simplified S-curve model. Using an agent-based simulation, we study the effect of technological efficiency under various market conditions and organizational structures and identify four competitive scenarios: entrant failure, diverse and duopolistic competition, and disruption. These competitive scenarios show robustness for all parameter combinations other than technological efficiency and organizational inertia. We then study realistic ways of defending industry leadership. We increase rationality of firm agents by increasing the planning horizon and allowing the setup of a daughter company in the case of a perceived threat of leadership loss. Respective simulations show that simple forecasting techniques allow the incumbent to pre-detect a threatening entrant product, to create a new startup firm intercepting the entrant, and to defend leadership as a group of firms, at the price of lower profits caused by more intense competition, though.

The remainder of this article is organized as follows: in the next section, we present our model of technologies, describe the market structure and consumers' behavior and define the firms' decision-making process (agent design). On this basis, the third section presents the structure of the agent-based simulation and the experimental design. In the results section, we look at the outcome of our experiments. Then, we discuss model extensions dealing with defensive incumbent strategies. In the final section, we draw conclusions and discuss the managerial implications of our findings.

2 Model

Our model consists of 3 components: technology, market and a firm's decision. The technology part connects product performance (features) to a firm's investment, i.e. the movement of the product position in the feature space as a function of the investment of the firm. The market describes the consumers' choice, their preferences and market dynamics. In the firm's decision part, we describe the firm's objective function and decision-making process. In the following, let i, j, k, l denote indices of consumers,

firms, technologies, and features, respectively.

Technology

A technology α_k is a vector that specifies a linear trajectory of possible product positions in a two-dimensional feature space that are reachable through investments in product development over time:

$$\alpha_k = \lambda_k(\sin \delta_k, \cos \delta_k), \tag{1}$$

where $\delta_k \in (0, \pi/2)$ describes the direction (feature mix), and $\lambda_k > 0$ the efficiency of the technology, i.e. the larger λ_k, the higher the feature levels of a product for a given investment sum.

There are two technologies available: at first, only α_1 used by the incumbent is available. α_2 is the (potentially) disruptive technology and the only choice available to the entrant. By the time of entry τ, the incumbent firm is free to choose either of the two technologies. Let us denote technology choice by index variables $c_{j,t} \in \{0, 1, 2\}$, where $c_{1,t} = 1$, $t < \tau$ for the incumbent and $c_{2,t} = 2$, $t \geq \tau$ for the entrant. A zero choice indicates absence of a firm from the market, as in the initial period of a simulation ($t = 0$).

The total investment in the current technology of a firm, $E_{j,t}$, is the sum of investments $e_{j,t}$ over time. In the basic version of the model, we assume that the incumbent has to give up its former technology and forfeit its prior investments, if it decides to switch to the disruptive technology:

$$E_{j,t} = \begin{cases} E_{j,t-1} + e_{j,t} & \text{if } c_{j,t} = c_{j,t-1} \\ e_{j,t} & \text{otherwise,} \end{cases} \tag{2}$$

where we assume that in the initial period of a simulation $E_{j,0} = 0$.

A firm's product position, a vector with components $x_{j1,t}, x_{j2,t}$, is defined as the firm's effective investment multiplied by the technology chosen:

$$\mathbf{x}_{j,t} = \ln(1 + E_{j,t})\alpha_{c_{j,t}}. \tag{3}$$

This means that, using a logarithmic transformation of the total investment, we suggest a simplified S-curve model where successive investments in a technology show decreasing returns to scale.

A firm's total cost consists of two components: fixed cost and investment cost. Regarding the fixed cost, we assume a factor $\gamma > 0$ on total investments. Through the investment cost, we model organizational inertia by introducing a factor $\kappa \geq 1$ that scales the investments of a firm without inertia:

$$C_{j,t} = \gamma E_{j,t} + \kappa^{e_{j,t}} e_{j,t}. \tag{4}$$

Thus, while a level of $\kappa = 1$ describes a situation where a firm is faced with linear increases in cost for linear improvements of its technological position in a single period, $\kappa > 1$ punishes fast technological progress by exponentially increasing investment

cost. This implies reaching a specific level of investment within more periods is less costly for inert organizations. However, as a consequence of the simplified S-curve model, a linear increase in a firm's position leads to an exponential increase in total cost, even if the firm is not inert. Therefore, cost acts as a bound on investments.

Market

Following Adner (2002), we assume that the behavior of a consumer is guided by a Cobb-Douglas utility function with two arguments: product performance $y_{ij,t} \geq 0$ and price $p_{j,t} > 0$ with the parameter $\beta > 0$ balancing the importance of product performance versus price:

$$u_{ij,t} = y_{ij,t}^{(1-\beta)}(1/p_{j,t})^{\beta} \tag{5}$$

Product performance depends on the feature levels $x_{jl,t}$, performance thresholds $d_{il,t} > 0$, and the relative preferences for the features $\eta \geq 0$, again in the form of a Cobb-Douglas function:

$$y_{ij,t} = \begin{cases} 1 + (x_{j1,t} - d_{i1,t})^{\eta}(x_{j2,t} - d_{i2,t})^{1-\eta} & \text{if } x_{jl,t} > d_{il,t}, l \in \{1,2\} \\ 0 & \text{otherwise} \end{cases} . \tag{6}$$

We assume that a consumer considers a product for choice only if its utility exceeds an overall utility threshold $u > 0$, i.e. $u_{ij,t} > u$, and chooses one unit of the product with maximum utility (denoted by $s_{i,t} \in \{1,2\}$). Ties are broken with equal probability, and in order to avoid artificial results, we assume that consumers are also indifferent between products with too small a difference in utilities. From the definition of the utility function it follows that the choice set is empty if the available products do not satisfy the performance and implicit price thresholds.

Parameters η, β and u describe general market conditions and are thus assumed equal for all consumers. Consumer heterogeneity is introduced by a distribution of $(d_{i1,t}, d_{i2,t})$.

We study both time-constant and adaptive consumer thresholds. Using time-invariant preferences, consumers are not influenced in their preferences by technological progress, i.e. $d_{i,t} = d_{i,0}$. In the case of adaptive consumer behavior, which we indicate by the switch variable $\zeta \in \{0,1\}$, the minimal performance thresholds are adapted according to direction and rate of improvement of the product purchased, that is such that, with $\rho_t(x_I) = \frac{x_{I,t}}{x_{I,t-1}}$ for arbitrary index set I,

$$\rho_{t+1}(d_{il}) = \begin{cases} \rho_t(x_{c_{i,t}l}) & \text{if } \zeta,\ x_{c_{i,t}l,t-1} > 0 \\ 1 & \text{otherwise} \end{cases} . \tag{7}$$

holds. This means that if the features of a product increase by, say, 10%, the buyers of this product also increase their minimal performance requirements by the same percentage in the next period. In case the product is just launched, consumers do not change their requirements as there was no improvement.

Firm's Decision

Besides technology choice, in each period of time a firm has to decide on a proper level of investment and price. We assume the firms to be: 1) well-informed (they know the consumers' utility functions and their competitors' past actions), 2. rational (they make optimal best response decisions), and 3. myopic (they have a one-period forecast horizon).

The equations from the preceding paragraphs can be reformulated so as to express a consumer's reservation price for a product as a function of a firm's investment and price, given the consumer's current preference and the utility of the competitor's product. By reservation price we mean the maximum price a consumer is willing to pay for a product, which is all we need to know in order to define a demand function. For ease of presentation, let $\hat{D}_{c_{j,t},t}$ denote the demand forecast of firm j using technology $c_{j,t}$ in period t, based on the information about the market up to period $t-1$. Then we can summarize the profit maximization problem of a firm as follows:

$$\begin{aligned} \hat{\pi}_{j,t} \quad &= \quad p_{j,t}\hat{D}_{c_{j,t},t} - C_{j,t} \rightarrow \max_{c_{j,t},\, e_{j,t},\, p_{j,t}} \\ &\text{s.t.} \quad c_{1,t} = 1 \ \text{ if } \ t < \tau, \\ &\qquad\ c_{2,t} = 0 \ \text{ if } \ t < \tau, \\ &\qquad\ c_{2,t} = 2 \ \text{ if } \ t \geq \tau, \\ &\qquad\ e_{j,t} \leq F_{j,t}. \end{aligned} \tag{8}$$

By $F_{j,t}$ we denote a firm's current funds, that is cumulated profits plus initial funds. Although the constraint on investments implies that we do not consider the possibility of external funding, we can always relax this constraint by a proper choice of $F_{j,0}$. Further, we assume that a firm leaves the market if it does not expect a positive profit or if it runs out of funds.

3 Simulation Setup and Experimental Design

Based on the definitions given in the previous section, the emergence of different competitive scenarios is studied using the following simulation scheme:

The first step is to initialize the population of consumers and firms. Next, the incumbent enters the market with technology α_1. For the first three periods, the incumbent can act as a monopolist, and in the fourth period the entrant joins the market with the new technology $\alpha_2 \neq \alpha_1$, which from this time on ($\tau = 4$) is also available to the incumbent. The firms calculate their profit-maximizing strategies (including the option to leave the market) according to Equation 8, and consumers then make their utility-maximizing choices (including the option not to buy) according to Equations 5 and 6. In the case of adaptive preferences, the buyers further adapt their thresholds according to Equation 7. Finally, the market outcome is evaluated in terms of market shares and profits.

The parameters held constant are as follows: the market consists of 100 consumers where a consumer's thresholds of acceptable product performance are drawn from a

uniform distribution over the rectangle $(0, 3) \times (0, 3)$. Note that we hold the distribution constant across different simulations. For the overall utility threshold of the consumers, we assume a level that scales the reservation prices at zero surplus performance properly: we decided to use $u^{-1/\beta} = 3$ (approximately the mean of the component sums). For parameter η we choose a level of 0.5, meaning that for a consumer, both dimensions are equally valuable. Thus, we do not model segments of relative preference as in Adner (2002), but a potentially competitive market that is segmentable by the firms' choices of technology, investments, and price.

With regard to the firms, we assume initial funds of 1000 monetary units and a fixed cost factor $\gamma = 0.2$, which ensures unconstraint investments and proper scaling with reservation prices (so that initially the incumbent can make a profit). We further assume a considerable bias of the incumbent technology in favor of feature two ($\delta = \pi/10$) and a balanced entrant technology ($\delta = \pi/4$). That is, given the same level of total investment and equal efficiency, the entrant technology outperforms the incumbent's with respect to the second feature but is inferior in the first. Thus, the new technology fulfills Christensen's criterion of potentially disruptive technologies (Christensen, 1997).

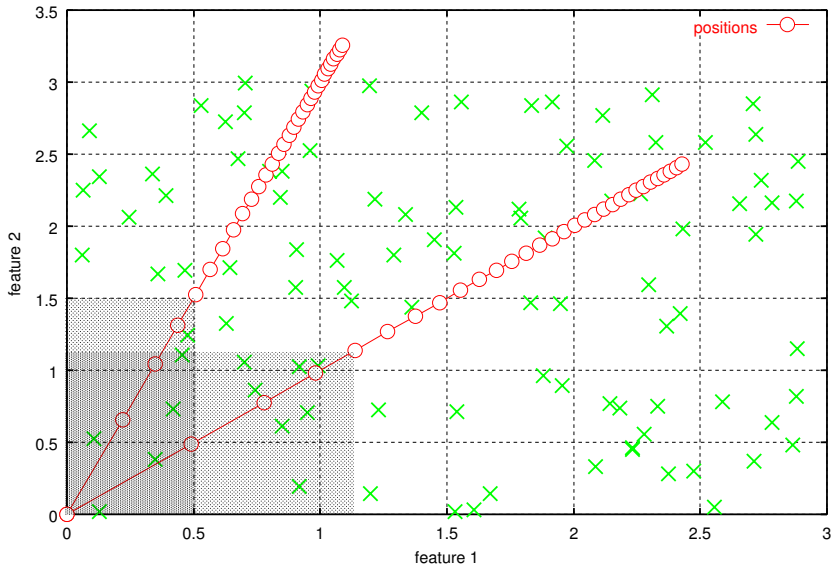

Figure 1: An illustration of the market space

Figure 1 illustrates the key features of the market so far defined. The consumers' performance thresholds are drawn as crosses. The lines mark the technological trajectories for the incumbent (close to the vertical axis) and the entrant technology (45°), respectively. Points on these lines depict product positions corresponding to linearly increasing levels of total investment $(0, 1, 2, \ldots)$ given equal technological efficiency. Thus, the market volume grows quadratically in the inner of the rectangle as the firms

develop their products over time. The shaded areas illustrate that the utilization of the entrant technology implies a better market coverage, i.e., for equal levels of effective investments the number of potential buyers of the product based on this technology is always greater than for the other one (allowing for variations in the distribution of thresholds). Further, the ratio of exclusive to competitive market coverage is clearly in favor of the entrant technology.

We study the influence of specific model parameters on the competition between an incumbent and an entrant firm. Four competitive regimes can be distinguished according to technology choice and market shares:

Entrant Failure: The incumbent sticks to the initial technology but the entrant fails to capture a reasonable share of the market ($\leq 30\%$), or even does not enter the market.

Diverse Competition: The incumbent sticks to the initial technology, and the entrant can equal the incumbent in terms of market share ($\approx 50\%$).

Disruption: The incumbent sticks to the initial technology but the entrant is able to outperform the incumbent, i.e., the entrant gains a considerable share of the market ($\geq 70\%$), or even may force the incumbent out of the market.

Duopolistic Competition: The incumbent switches to the entrant technology and thus competes with the entrant on a similar product. Therefore, we expect the market shares to be rather identical ($\approx 50\%$).

Table 2: Design Factors

Factor	Levels
λ_2	0.4, 0.6, ..., 1.8
κ	1.0, 1.1, ..., 1.3
β	0.5, 0.7
ζ	0,1

Table 2 shows a full factorial design of the model parameters we consider relevant for market outcome. As we conjectured that relative technological efficiency and organizational inertia are key determinants of the market outcome, we decided to search these parameters with a reasonably high resolution while economizing on the levels of price sensitivity. Thus, with $\lambda_1 = 1$ the range of λ_2 includes entrant technologies that are inferior, equal, and superior in terms of the incumbent's technological efficiency. Especially, we expect a considerable influence on the decision to switch and, thus, on the market outcome. With respect to organizational inertia κ, we analyze levels between 1.0 (no inertia) and 1.3 (high inertia). Note that in the present setting, differentials in inertia are meaningless for the incumbent's technology choice at $t = \tau$, because information on the entrant product is not available by that time. By variation of β, we study the effect of high (0.5) and low (0.7) price elasticity, modeling the market's receptiveness to innovation. Further, we compare markets

with consumers adapting their performance thresholds ($\zeta = 1$) to markets with static consumers ($\zeta = 0$). We expect adaptation to act in favor of the incumbent because initially, as a monopolist, it is able to thin out the low end of the market and thus could block out the entrant.

4 Results

The model was implemented in the mathematical language Octave (Eaton, 2003), and the results were analyzed using the R software for statistical computation and graphics (R Development Core Team, 2003). The source code of the implementation is available upon request from the authors. A total simulation time of 20 periods proved sufficient to get a clear picture of the market outcome. Further, since our model is rather deterministic (random product choices should be rare except in duopolistic competition where they act as stabilizers on market share), the simulation was run repeatedly mainly in order to determine a proper calibration of the random search: using 1000 steps and a restriction on the upper search range provided stable results.

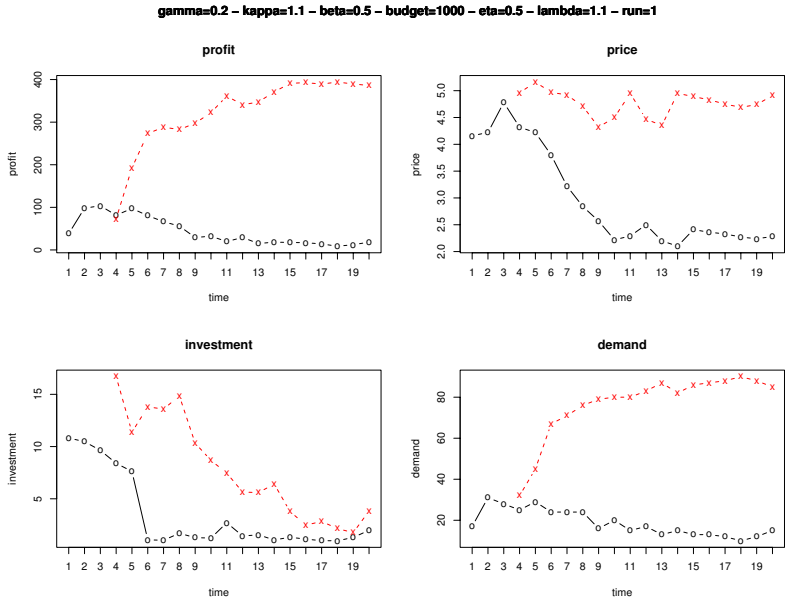

Figure 2: An illustration of a scenario

Figure 2 shows the outcome of a scenario with parameter combination $\lambda_2 = 1.1$, $\kappa = 1.1$, $\beta = 0.5$, and $\zeta = 1$: the incumbent's results are shown as solid lines, the entrant's are presented using dashed lines. Utilization of the initial (incumbent) technology is indicated by circles whereas lines marked with crosses indicate the use of the new (entrant) technology. The upper left graph shows profit over time, i.e., the success of a firm's actions. It can be seen that the entrant outperforms the incumbent

from the fifth period on, since the incumbent does not switch to the new technology. In the upper right and lower right diagrams, we see that this outperformance results from both a higher unit price and a higher number of units sold. These higher unit prices can be demanded because of higher product performance resulting, in turn, from higher investments (see the lower left diagram). The gap in investments also results in differences in market coverage, and therefore the entrant has a larger number of (exclusive) buyers (see Figure 1).

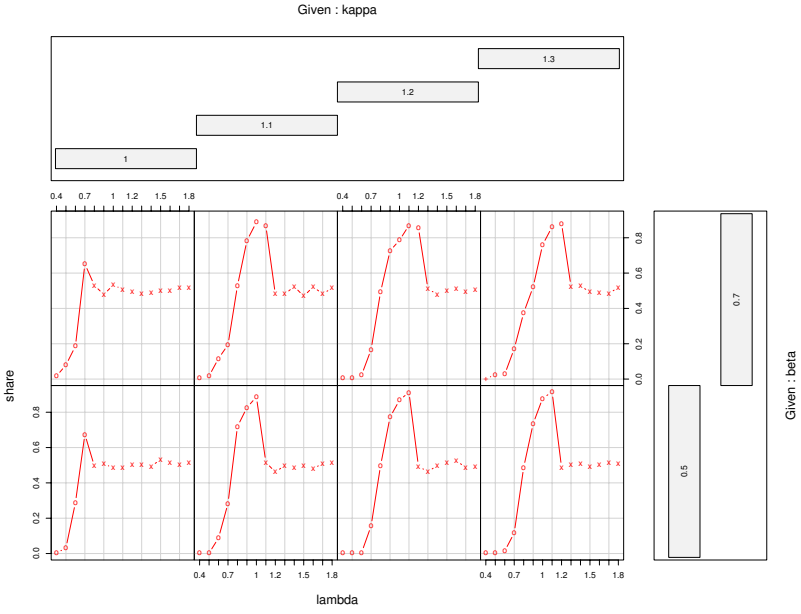

Figure 3: Results for static scenarios

Figure 3 shows aggregate results for all parameter combinations with static performance thresholds ζ: we plot the entrant's average market share (vertical axis) for different levels of efficiency of the entrant's technology (horizontal axis). The average market share (in terms of profit) is defined as the mean of the shares from periods 11 to 20, i.e. when the market has already stabilized. Points marked with an 'x' ('o') indicate a (no) switch of the incumbent to the entrant technology and points marked with a '+' the failure of the entrant to enter the market (in the beginning of scenario $\kappa = 1.3$, $\beta = 0.7$). The subplots represent the results for different combinations of the remaining design factors: from left to right, the level of organizational inertia κ increases, from top to bottom the level of price elasticity β decreases.

First, we notice that the entrant is never able to outperform the incumbent if no organizational inertia exists ($\kappa = 1.0$), i.e., there exists no level of λ_2 where disruption occurs. The reason for this is the following: if the new technology is efficient enough, the incumbent switches (without exception in $t = 4$) and duopolistic competition emerges, which is characterized by rather balanced market shares. In case the

market is targeted by different technologies, entrant failure or diverse competition is the outcome: the more inferior the entrant's technology is, the smaller is its market share. This is due to the fact that low investments result in a higher market coverage of the incumbent (see Figure 1).

All scenarios with $\kappa > 1$ show a different pattern as compared to the scenarios with $\kappa = 1$: now, we observe a range of efficiency in between diverse and duopolistic competition, where the incumbent does not consider it profitable to switch technology and subsequently looses a significant share of the market to the entrant, i.e., where disruption occurs. Obviously, the disruptive range does not considerably depend on whether consumers adapt their performance thresholds or not. In the case of adaptation, closer inspection reveals that—although the incumbent is able to maintain exclusive coverage of a small part of the market—he cannot catch up with the entrant, because the incumbent technology does not follow the main direction of the market and, therefore, the entrant's market is almost exclusive. Conversely, in the case of static consumer thresholds and disruption, the whole incumbent market is competitive whereas the entrant's one is by far more exclusive. Further, in the long run, the entrant captures part of the incumbent market since both firms lack the incentive to offer a distinguishable product to these consumers (see Figure 1).

Table 3: Summary of results

ζ	β	κ	λ				t
			failure ($\leq 30\%$)	diverse ($\approx 50\%$)	disruption ($\geq 70\%$)	duopolistic ($\approx 50\%$)	switch
0	0.5	1.0	0.4–0.6 (0)	0.7–0.7	-	0.8–1.8	-
		1.1	0.4–0.7 (0)	-	0.8–1.0	1.1–1.8	4
		1.2	0.4–0.7 (0)	0.8–0.8	0.9–1.1	1.2–1.8	4
		1.3	0.4–0.7 (0)	0.8–0.8	0.9–1.1	1.2–1.8	4
0	0.7	1.0	0.4–0.6 (0)	0.7–0.7	-	0.8–1.8	-
		1.1	0.4–0.7 (0)	0.8–0.8	0.9–1.1	1.2–1.8	4
		1.2	0.4–0.7 (0)	0.8–0.8	0.9–1.2	1.3–1.8	4
		1.3	0.4–0.7 (1)	0.8–0.8	0.9–1.2	1.3–1.8	4
1	0.5	1.0	0.4–0.6 (0)	0.7–0.7	-	0.8–1.8	-
		1.1	0.4–0.7 (1)	0.8–0.8	0.9–1.0	1.1–1.8	4
		1.2	0.4–0.8 (2)	-	0.9–1.1	1.2–1.8	4
		1.3	0.4–0.8 (3)	0.9–0.9	1.0–1.1	1.2–1.8	4
1	0.7	1.0	0.4–0.6 (0)	0.7–0.7	-	0.8–1.8	-
		1.1	0.4–0.7 (0)	0.8–0.8	0.9–1.1	1.2–1.8	4
		1.2	0.4–0.7 (1)	0.8–0.8	0.9–1.2	1.3–1.8	4
		1.3	0.4–0.7 (1)	0.8–0.8	0.9–1.2	1.3–1.8	4

Table 3 gives a summary of the ranges in the market outcome (distinguished by share and technology choice) for λ, switching times, and the number of no-entry fail-

ures (in parentheses). It further shows important aspects of our model: first, the breakpoints for switching do not increase with higher levels of inertia. Therefore, we are inclined to conclude that in our model disruption is mainly a result of myopic decision making. To understand this, notice that in the absence of organizational inertia investments are concentrated on the time of entry, which is rather similar to making a single, long-term decision, whereas with increasing inertia investments become more and more distributed over time. Now, as the firms have only a one-period horizon, they loose more and more their sense of long-term optimality. To be precise, the long-term levels of total investment are the lower the higher the level of organizational inertia, and that is—besides disruption—clearly suboptimal.

Another important aspect of the present model is that the occurrence of disruption does not depend on possible differences in organizational inertia because we observed that switching takes place when there is no competitive information available, i.e. on the time of entry. Thus, even if the entrant is assumed to be less inert than the incumbent our results hold, only the range of efficiencies with disruptive market outcomes increases. Our experiments for this setup have shown that in the case of duopolistic competition, the market shares of the entrant are slightly higher if price sensitivity is low, since the entrant can demand higher premium prices. Further, among the adaptive scenarios, there are cases of no entry as well as cases where the incumbent leaves the market (in $t = 18$), and thus the entrant's market share goes up. Clearly, low price sensitivity and a high differential in inertia is not in favor of the incumbent.

5 Defending Disruption

In the previous setting, we have detected conditions under which an incumbent firm may fail in detecting a new, disruptive technology because it underestimates its efficiency. In addition, even when the incumbent becomes aware of the peculiar situation, its 'defending' strategy—switching to the entrant technology—is often not practical due to risk considerations not incorporated in our framework: a real incumbent firm is unlikely to sink all its former investments, resulting in giving up its leading position in the high-end of the market, with the additional risk of failure due to inappropriate organization and cost structures.

As we have learnt so from our basic model, incumbant failure is mainly caused by too short a planning horizon. It thus seems promising to increase rationality in this direction and to allow internal differentiation in the sense that a group of firms can pursue different technological trajectories. This resembles Christensen's suggestion based on empirical evidence, who adivses incumbant managers facing threat from disruptive technology the following:

1. Try to predict the technological path of the entrant product in order to assess its competitive threat potential.

2. When a potential competitor is detected, do not try to change your firm, but create a new, efficient entrant instead and accept possible cannibalization effects.

We now explain how these suggestions are operationalized in our artificial environment, and present the results of "corresponding" experiments.

5.1 Model Extensions

Better Forecast of the Entrant's Product Position

First, we allow the incumbent to make a better forecast of the entrant position. In the basic model, the incumbent's estimate of the future entrant position was just its current position. We now replace this crude guess by a model-based approach, assuming that each dimension of the entrant technology follows an exponential model, that is, given the entrant's product $\boldsymbol{x}_{E,t} = (f_{1,t}, f_{2,t})$, we assume:

$$f_{i,t} = a_i t^{b_i},\ i = 1, 2 \tag{9}$$

which is a simple linear regression model accounting for decreasing positional gains on the technology path, that is, assumes a simplified S-curve model. We need at least two observations to estimate the two parameters a_i and b_i. Note that this is not the actual mechanism implemented in our base model: the incumbent does not know the exact characteristics of the entrant's product. Using this model, the incumbent is able to make a prediction of the entrant's future position. To keep things simple, we use the average of the entrant product's last two prices as an estimate for the future price. For its own product, the incumbent assumes an investment rate increase of 10%, which is what approximately happened in the base simulations (in real life situations, the incumbent simply uses the figures from its investment plan. The aim here is a conservative, worst-case estimation of the future situation). Now, as the demand function is supposed to be known, the incumbent can forecast the optimal price for its product, and also the future profits and market shares. This allows the incumbent to assess the entrant product for any future period.

Cloning of the Entrant Firm

In our simulation, the incumbent considers the entrant technology as perilous in period t if its market share in period $t + 3$ drops to under 50%. But instead of switching to the new technology, we assume the incumbent has the ability to create a new firm similar to the entrant—which will be called 'clone' in the following—with the same technology, but with 80% of the incumbent's budget. (We choose 80% because we want to explore an experimental setting in which the clone's investments in the first period are maximized. When the clone has normal—optimizing—behavior, the budget size effectively does not matter, because only a small part of it is used.) The role of this 'cloned' firm is to catch up with the entrant's position and thus to participate in the better product performance and the new market segment. On the other hand, the incumbent can no longer choose to switch to the new technology on its own.

5.2 Experiments and Results

In order to see whether the new defense mechanism is effective, we run the simulation within the parameter ranges of κ, entrant λ, and incumbent budget associated with disruptive outcomes in the basic model. Alternatively, we test the assumption of a more 'aggressive' clone, investing its whole endowment in the first period to make up the initial technological disadvantage. We stop the simulations after 30 periods (there is no considerable change in the figures in later periods). The experimental setup thus follows the full-factorial design defined by the factors in Table 4.

Table 4: Experimental Design Factors

factor	levels
incumbent budget	100 1000 10000
entrant λ	0.6 0.8 1.0 1.2
κ	1.1 1.2 1.3
aggressive?	YES NO

In all cases where the entrant has the potential to defeat the incumbent (which is not the case for most settings with $\lambda = 0.6$), the pursuing firm catches up with the entrant and finally gets half of the market. Hence, the incumbent-clone group survives in all settings. However, the consolidated profits are lower if the pursuing firm is created than if it is not. This is due to the more pronounced price competition, initiated by the clone which tries to reach the entrant firm, and aggravated by the incumbent firm lowering its price in response to the advent of the entrant firm. The price level is also higher in the basic model when the incumbent switches to the entrant technology, resulting in a duopolistic competition which is well known to have a higher equilibrium price than settings with full competition (Cournot game). This also explains why the profits of incumbent and clone combined are lower than the cumulated profits of the entrant at the end of the simulation. Finally, neither λ nor the incumbent starting budget are of great influence, except the trivial effects that the overall cumulated profits increase with λ (i.e., the product's efficiency), and the cumulated profits are higher in the case of huge incumbent starting budgets.

As to the final market shares, the picture has more nuances: here, the value of λ is most influential, whereas κ and incumbent starting budget are not. For $\lambda = 0.6$, the entrant—most of the times—is not menacing: no clone is created, and the incumbent keeps the whole market. For $\lambda \geq 0.8$, however, the incumbent's assessment of the future situation leads to the creation of a clone. For λ = 0.8, the incumbent still stays in the leading position, but for $\lambda \geq 1$, it vanishes from the market at the end of the simulation (!) and the market becomes a duopoly with entrant and clone firm. Interestingly, the results for settings with aggressive investing behavior do not show an advantage for the clone: despite the faster catching-up, it is not able to defeat the entrant whose budget is already important enough to survive periods without (or with low) profits. On the contrary, due to the exponentially increasing costs, the clone makes a huge loss in the first period which it can never recover in future periods.

6 Conclusions

We have analyzed the influence of organizational inertia and technological efficiency on the emergence of competition between an established firm and an entrant and studied simple and effective means of defending industry leadership. We have assumed that the firms maximize their profit expectation for the next period based on full information of the needs of the entire consumer market and the competitor's current product position and price, and that the incumbent has the choice to switch to the new entrant technology. Technologies are modeled as linear trajectories of possible product positions in a two-dimensional feature space. A simplified S-curve model describes the relationship between a firm's investments and its technological progress, which comes at increased fixed and investment cost. The firms are faced with a highly competitive market of compensatory, utility-maximizing consumers with differing minimum performance requirements. We have studied the influence of differentials in technological efficiency on the entrants' success under different market conditions and levels of organizational inertia.

Using an agent-based computer simulation, we have shown that the entrant is never able to outperform the incumbent if organizational inertia does not exist. This is an interesting finding as we expect organizational inertia to be higher for larger companies and/or complex industries: reducing an organization's complexity is, therefore, advisable to large companies that are faced with potential entrants. This is consistent with Christensen's suggestion that firms should not pursue the development of potentially disruptive technologies within their existing organization but ought better outsource this task to a new company.

Furthermore, we have found that outperformance of the incumbent firm depends on a specific range of the entrant's (relative) technological efficiency. If the new (disruptive) technology's efficiency is too low, the entrant is not able to reach a satisfactory product performance and thus is unable to capture a significant share of the market. On the other hand, if the efficiency is very high, it is more attractive to the incumbent to switch to the new technology than to continue with its initial one. The result is a duopolistic market where price competition between similar products prevails. Finally, we have found that differentials in organizational inertia expose the incumbent to an increased risk of early failure.

Both results regarding technological efficiency and organizational inertia are rather independent of the demand structure. In contrast to Adner, we therefore conclude that the phenomenon of disruption does not necessarily occur as a result of changes in consumer preferences, but that technological and organizational aspects seem to be more important.

Finally, experiments with an extended model have shown that the use of even simple forecasting techniques, applied to the positions of the entrant technology, allow the detection of threatening competitors. The creation of a new firm similar to the entrant assures the survival of the consolidated firm group, but leads to lower profits due to intense competition, and may cause severe cannibalization effects: when the incumbent has a technology which is less efficient than the entrant's, it vanishes from the market. The message of this finding is clear and has already been applied by leading

high-tech firms (see, e.g., Brown and Eisenhardt, 1998): an incumbant under threat by disruptive technologies does not have to be overly innovative himself. Rather technology management is important in the sense that the development of entrants has to be closely watched and that a homogeneous, centrally controlled firm structure has to be given up. Managerial advice is thus, that the firm should organize as a patch work of small, independent units pursueing different technologies independently, also competing with each other. Strategy in such a framwork resembles to the close monitoring of technological developments in other market segments, the forecasting of technological positions to detect threats, and the making of appropriate portfolio decisions—that is, setup of new units, also including the acquisition of successfull entrants. However, while survivial can be secured in this way, it is the consumers who benefit from more intense competition and lower prices.

Bibliography

Adner, R. (2002). When are technologies disruptive? A demand-based view of the emergence of disruption. *Strategic Management Journal*, 23:667–688.

Brown, S. L. and Eisenhardt, K. M. (1998). *Competing on the Edge: Strategy as Structured Chaos*. Harvard Business School Press, Boston, Mass.

Christensen, C. M. (1993). The rigid disk drive industry: History of commercial and technological turbulence. *Business History Review*, 67:531–588.

Christensen, C. M. (1997). *The Innovator's Dilemma*. Harvard Business School Press, Boston, Mass.

Christensen, C. M. and Bower, J. L. (1996). Customer power, strategic investment, and the failure of leading firms. *Strategic Management Journal*, 17:197–218.

Eaton, J. W. (2003). Octave software version 2.0.17, `http://www.octave.org/`.

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

Henderson, R. M. and Clark, K. B. (1990). Architectural innovation: The reconfiguration of existing product technologies and the failure of established firms. *Administrative Science Quarterly*, 35:9–30.

Klepper, S. and Simons, K. L. (1997). *Technological Extinctions of Industrial Firms: An Inquiry into their Nature and Causes*. Oxford University Press, Oxford.

Malerba, F., Nelson, R., Orsenigo, L., and Winter, S. (1999). History-friendly models of industry evolution: The computer industry. *Industrial and Corporate Change*, 8(1):3–40.

Mas-Colell, A., Whinston, M., and Green, J. R. (1995). *Microeconomic Theory*. Oxford University Press, New York.

Nault, B. R. and Vandenbosch, M. B. (2000). Research report: Disruptive technologies – explaining entry in next generation information technology markets. *Information Systems Research*, 11:304–319.

R Development Core Team (2003). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.

Rumelt, R. P. (1981). Towards a strategic theory of the firm. In Boyden, R., editor, *Competitive Strategic Advantage*, pages 556–570. Prentice Hall, Englewood Cliffs.

Tushman, M. L. and Anderson, P. (1986). Technological discontinuities and organizational environments. *Administrative Science Quarterly*, 31:439–465.

Agent-Based Simulation of Power Markets

Thomas Steinberger and Lucas Zinner

1 Introduction

Simulations using artificial agents to model the dynamics of new developing markets have already been proved to be a useful tool to analyze these changing markets (Bunn and Oliveira, 2001; Tesfatsion, 1997, see, e.g.,). The development of a framework which includes the various components of a market such as producers, consumers, networks or different clearing mechanisms is aimed to gain insights into the results depending on the different strategies of the main players in these markets. The modular framework enables us to study certain phenomena by changing parts and leaving others unchanged. We are especially interested in the producers part in this market and therefore made rigorous simplification with respect to consumers and for instance power lines compared to real world scenarios.

The simulation environment is based on the MASS – MATLAB Aided Simulation System developed by Scherrer (2001), and is intended to be the base of a tool which should provide useful information in the decision making process. In a first step we aim in generating close to real world time series concerning the prices over time as well as usage rates of power plants using demand curves as an external parameter. By changing parameters such as fuel prices or price sensitivity of consumers we try to estimate chances in a developing market.

The simulation depends mainly on the interaction of different agents representing consumers, producers, markets and regions. All of them are acting with respect to their utility function. Each of these agents is implemented as a MATLAB program. The communication is based on a message system which is coordinating the data of the different agents. During the clearing process three kind of products are traded, namely certain volumes of constant load offers for a period of 24 hours, 16 hours and 7 hour called Base, Peak and Superpeak.

The consumers communicate a demand in the form of 3 quantities. This is exogenously specified on the basis of a load curve. The demand is (weakly) price flexible thereby.

The production agents are equipped with a certain portfolio of power plants and an associated cost structure and put day ahead trading offers for the three groups of loads, namely Base, Peak and Superpeak, in the form of price-quantity pairs. The producer puts its offer knowing the expected demand as well as prices and quantities of its competitors and/or results of the previous day. In each simulation step the budget of the producer is updated according to its costs and sold quantities.

The inquired quantity is supplied at each time. In case the offers are not sufficient, a so-called generator of last load resort (GLR) ensures the covering of the demand. The price setting of the GLR is likewise exogenously given. His capacities are quasi unlimited and thus his prices form upper bounds for the prices of the production agents.

2 Market agent

The market program (the market) provides on the one hand the completion of the trade in all regions and is responsible on the other hand for the organization of the simulation itself. In each simulation step it passes on the mean of the expected demand of the region for each of the three groups of loads to the producers. After receipt of all offers from the producers, namely pairs of price and quantity for each group of loads, a supply curve is formed. By intersecting this with the demand curve the size of the market and the price for each of the 3 groups of loads is determined.

The market conveys price and volume per group of loads with the appropriate quantity requirements of the individual producers to the consumer. The consumer order the corresponding quantities from the producers themselves.

The market can be cleared either as pool or OTC. In the pool model only one cost per group of loads and region for all producers and consumer is determined by the intersection of supply and demand curve. Within the OTC trading framework only the overall traded volume is determined by the market. Each producer receives its required price, whereas the consumer price is calculated as weighted means of the producers prices.

In the available version only one market is implemented, which manages the clearing for everyone in the regions. The transportation net is only very rudimentarily implemented. Within a region there are neither restrictions of capacity nor costs for transportation. Between the regions the interconnecting capacities are externally given. Costs of the transportation are likewise not implemented at present. An extension of the simulation plans connections of several markets, which are coupled exogenously to a certain degree.

One simulation step from the market agent point of view:

The market requests the consumers of all regions in a ‘call for offer’ message to specify their demand.
After receipt of all demand curves the market calculates a mean demand for each region and sends these and the request to put an offer in a ‘call for offer’ message to all producers. If all offers arrived, the aggregated supply curves for each region are computed and the market clearing is accomplished. In an ’offer’message the consumers of each region receive the offers as well as the market price and the size of the market.
As soon as the trade was completed by the consumers, the market receives a ‘close’ message and the next trading day is opened, as described above.

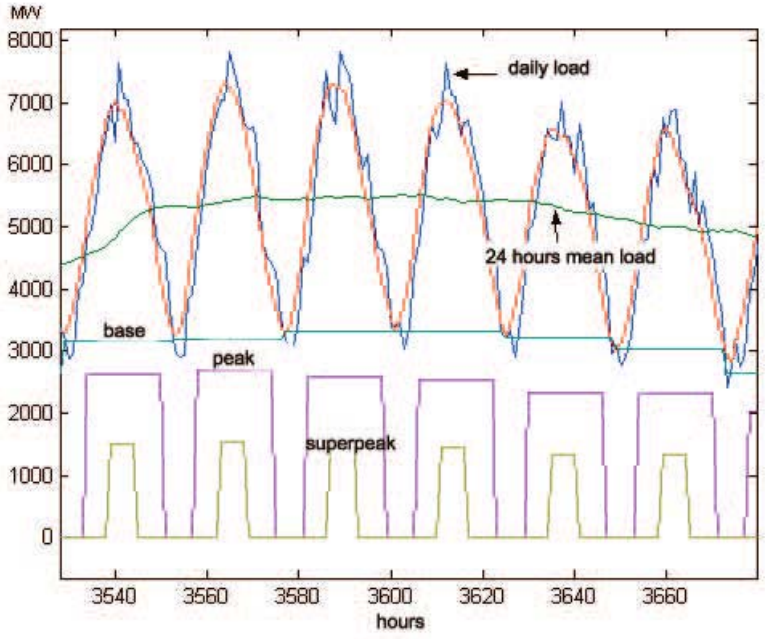

Figure 1: Typical load curve for 6 days in the summer, beginning on Monday

3 The Aggregated Demand–The Consumer

For each region an aggregated demand is given, i.e. the consumer side is not particularly modelled in the sense of a strategic behavior. According to its demand the consumer buys electricity beginning with the cheapest offer of a certain producer ascending until its need is covered.

The demand is characterized by a diurnal, a weekly and a yearly variation (see Figure 1). For instance the aggregated demand of a simulation of the Austrian electricity market is set at the initialization by approximately 53 TWh per year with maximum daily peaks in the winter of approximately 10 GW and in the summer of approximately 7.5 GW. The minimal demand ranges between approximately 3 GW respectively 2 GW. Additionally an annual increase in the consumption of 2% is assumed.

Since the strong diurnal variation of the load plays an essential role for energy supply, it is not

possible, even for long-term simulations, to restrict oneself averaged daily loads. On the other hand an hourly trade would slow down the simulation due to the computing intensity. An simulation over 10 years would need about 87600 simulation steps. Therefore 3 groups of loads are derived from the daily load curve, namely Base, Peak and Superpeak, which are traded together.

Base denotes the load, which is constantly demanded for an entire day of 24 hours. Peak denotes the load, which is constantly demanded additionally to Base during a period of for example 16 hours, whereby the duration is a selectable parameter.
Superpeak denotes the load, which additionally to base and peak is constantly demanded during a short, likewise selectable, period of for example 7 hours (see Figure 1).

The daily demanded energy corresponds to the sum of the demanded energy of the 3 groups of loads. Always the whole demand is satisfied, since a so-called generator of last resort (GLR) supplies arbitrary energy quantities at a price specified at the beginning of the simulation.

At the end of a simulation step the consumer submits the estimated demand of its region for Base, Peak, and Superpeak for the next day to the market agent. From this a linear demand curve is formed, by assigning these loads to the prices of the GLR where for the price 0 an for instance 10% increased demand is assumed. This short term price elasticity is determined with its own parameter.

The estimated demand, which the consumer submits for the three groups of loads, does not have to agree with the actual demand for two reasons. First of all the demand depends to a certain extend on the actual price of the respective electricity product obtained in the market clearing process. Secondly the estimated demand is affected by certain noise. The order of magnitude of this error is likewise adjustable with a parameter. Thus we have modelled the uncertainty of the estimation of the day ahead load without paying attention to the complicated details of balancing load and supply in the network. The risk, which lies in an over or an underestimation of the load is carried by the producers.

A simulation step from the view of the consumer:

After receipt of the ‘call for offer’ message the consumer computes the demand curves (Figure 1) for Base, Peak and Superpeak for its region and sends these in a ‘demand’ message to the market.

In the ‘deal’ message the consumer receives all necessary information to complete the trade with the respective producers. First the actual volume is calculated for all three groups of loads by the size of the market and the given prediction error. Thereupon each producer receives a ‘deal’ message with the volumes and the prices for the three groups of loads. The price corresponds thereby either to the market clearing price (pool) or the one offered by the producer (OTC). The volumes correspond to the volumes offered by the producers for all except the marginal one. If there are several marginal supplies, the inquired quantity of electricity is partitioned according to their offers.
When trading is completed, the consumer sends a ‘close’ message to the market.

4 Modeling of the Producers

Each producer is implemented as an own MATLAB program and is assigned to one region. Within this region there is no restriction concerning the interconnection capacities. To supply electricity into other regions, restrictions of capacity have to be considered. During the initialization of the simulation the portfolio of power plants of the individual producers is specified. From an external cost table (see Table 1) the producer calculates his costs for generation, i.e. his variable costs and his fixed costs. These data were raised from generally accessible sources and examined for plausibility (see Schneider, 1998).

The restriction on 7 types of power stations is strong simplifying. Furthermore we do not consider the costs of heating up and cooling down power stations or the no-load operation costs. In addition it is supposed that the costs of a type of power station are the same for all producers. The variable costs of thermal power stations are calculated in each simulation step using exogenously given fuel prices.

In Table 1 the installed capacity of all types of power stations is indicated. For hydro power stations the actual capacity depends strongly on the water level. We implement this technical restriction by defining an availability-variable (in per cent of installed capacity). Further technical restrictions are given for storage power stations. Here we give a minimum and a maximum value for the energy per day, which must resp. can be delivered. These values are indicated in per cent of the installed energy (installed capacity*24). The minimum energy must be delivered to prevent overfilling of the reservoir and the maximum energy may not be exceeded to prevent emptying the reservoir of the storage power plant. At present the minimum and maximum energy for storage power stations are externally given and no detailed modelling of the water level is made. It is supposed that all power stations are always available (no repair and maintenance times).

The determination of an offer by the producer:

All prices and loads are discretized. The prices are divided into 20 steps between the smallest variable costs of the producer and the prices of the GLR. The prices of the GLR are given and can be different for each group of loads. The capacities are set for each producer in 5%–steps between 0 and the installed capacity.

The production agent receives information from the market-agent in each simulation step containing the demand in the three groups of loads Base, Peak and Superpeak. The producer is requested by the market for an offer (i.e. a pair of price and quantity) for the 3 groups of loads.

Two different price setting strategies are implemented, Marginal Pricing and Best Response Pricing.

Marginal Pricing: The price is determined by the production costs of the respective

Table 1: Data used to calculate generation costs

Type of Power Plant	new Hydro	old Hydro	old Storage	Nuclear	Hard coal-fired	Brown coal-fired	Gas/ Steam	Gas turbine
Capital outlays (Mill. Euro)	839.9	66.6	330.8	2 060.4	1 152.1	960.1	314.7	65.6
Amortisation period	30	30	30	30	30	30	30	30
Planning horizon	20	20	20	20	20	20	20	20
Interest rate (%)	6	6	6	6	6	6	6	6
Capacity (MWel)	172	280	360	1530	950	900	800	250
Initialized Fuel price (Euro/MW)	0	0	0	5.931	5.931	5.11	18.2	18.2
Efficiency	1	1	1	0.36	0.49	0.46	0.57	0.39
Fixed operating costs per year (Mill. Euro)	35.5	35.5	15.4	61.2	52.6	27.5	10.3	2.6
Variable operating costs (Euro/MWel)	5.38	5.38	0.5	1.7	4.6	3.3	0.9	1.31
Annual rate of price increase (%)	1	1	1	1	1	1	1	1

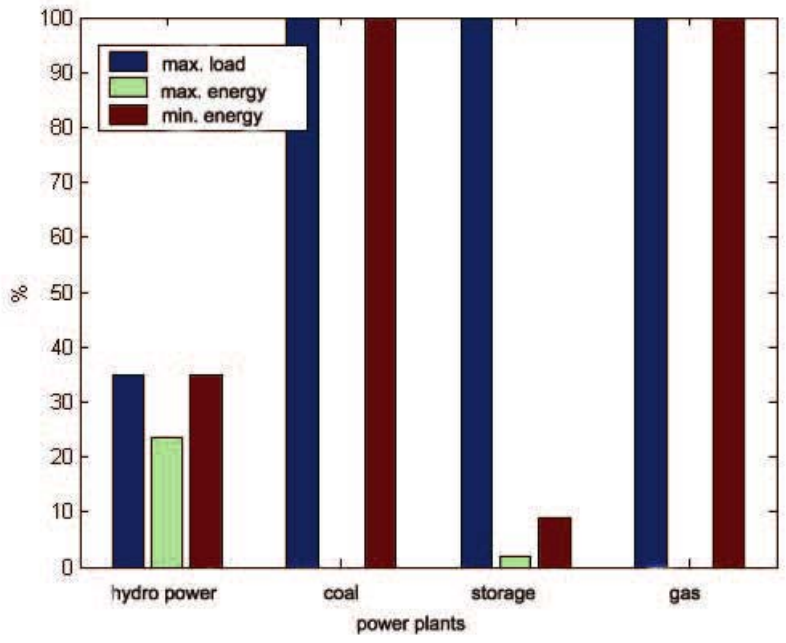

Figure 2: Technical constraints (maximum load, maximum energy, and minimum energy) for different types of power stations for a producer per day

energy quantity plus an additional charge, which is selectable as parameters and can be interpreted as amount of coverage for short or long-term planning.

Best Response Pricing: For putting his offer the producer uses the following knowledge:

- his price-quantity pairs of the previous day,
- the estimated demand in the 3 groups of loads
- his calculated production costs and capacities
- the prices and quantities of the other producers of the day before.

With this knowledge he performs a local optimization, by computing his expected profit, if he in- resp. decreases the price resp. the quantities by a unit under the assumption that the other producers behave exactly the same as on the day before. The power stations are optimally used under the constraints maximum load, minimum energy and maximum energy (Figure 2).

Table 2: Initial parameter for the aggregated load curve of the Austrian electricity market

Annual Load	Annual Growth	fr, sa, su against rest of week in %	Winter min/max in GW	Summer min/max in GW	Base, Peak, Superpeak in hours
53TWh	2%	90, 70, 50	~ 3 / $\sim 9,5$	$\sim 2,5$ / ~ 7	24, 16, 7

Remarks:

- The local optimization was selected, in order to prevent too strong short term fluctuations in the offer behavior.
- In each simulation step the agent computes 729 alternatives and selects from these that one, which maximizes his profit in the next time step under his constraints. To reduce the computation time we accept a small error in computing the production costs. The production costs of the three groups of loads are not independent, but for optimization we neglect these dependencies and use the loads of the day before.
- The agent tries to only maximize his profit for the subsequent day, i.e. the cost function is exclusively aligned to the daily profit and not to possibly different use (e.g. market power, customer connection).
- The lower bound for the prices of the production agent is his production costs. It is not possible to dump consciously.

A simulation step from the view of a producer:

- After receiving the 'call for offer' message the producer calculates the optimal price-quantity pairs for all three groups of loads as described in the paragraph above. The quantity of 0 is also possible and stands for no offer. These pairs are communicated to the market in an "offer"-message.
- In a "deal" message from the consumers all information is contained which the producer needs to compute his new budget and his best response offers for the next day.

5 Simulation of the Austrian Electricity Market

As a first application and as a test for the made simplifications a simulation of the Austrian electricity market was implemented. It should be also clarified whether the simulation system is already powerful enough, to reproduce important stylized facts. The demand side of the Austrian electricity market was initialized thereby as in Table 2.

The production side was simplified extremely. There are only two producers, one with very high portion of hydro power, corresponding to the VerbundAG, and a second producer that sums up all other Austrian power producers. OTC was selected for market clearing and the offers are put by means of Best Response Pricing.

The following stylized facts should be reproduced by the simulation:

Table 3: Parameters of the producers in the simulation of the Austrian electricity market

Producer	Power plant					Total capacity
P1	13 Hydro old	5 Storage old	1 Brown coal	0 Gas/Steam	3 GT	6981 MW
P2	6 Hydro old	1 Storage old	5 Brown coal	5 Gas/Steam	0 GT	10.472 MW

1. Baseload has to be cheaper than Peakload and this again has to be cheaper than Superpeak.
2. The prices for the individual products are approximately 20 Euro/MWh for Base, approximately 30 Euro/MWh for Peak and up to 50 Euro/MWh for Superpeak.
3. In long-term simulations a general price increase should be obtained because of removing over-capacities.
4. Likewise due to the worse relationship between installed capacity and consumption the price should be tendentious higher in winter than in summer.

These four price-referred stylized facts are particularly meaningful, since the prices depend on very many factors and additionally are affected by a strategic optimization of the production agents.

5. The usage rate of the individual types of power station should agree with usage rates at the Austrian market.

In the following figures (Figures 3 and 4) it is to be recognized that the above-mentioned points 1–4 are met by the simulation environment. In Figure 3 the consumer prices for the three groups of loads are shown for a period of 20 years. The consumer price of a group of loads is, under the assumption of OTC trade, the sum of the prices of the producer weighted with the quantity of electricity, bought at the respective producer. An annual growth of demand of 2% and an unchanged installed capacity is supposed, so that the over-capacities are reduced in course of time. It can be seen clearly that point 1 of the stylized facts is fulfilled, i.e. Baseload is cheaper than Peakload and this again is cheaper than Superpeak. In addition the long-term price increase (point 3) of all three electricity products is apparent. The yearly variation in the prices with higher prices in the winter and lower prices in the summer (point 4) is particularly clear in the second half of the simulation run, i.e. in the second 10 years. In the first years the price series are less regular probably because of Best-response-dynamic. In general, the price level for all three groups of loads is somewhat too low, particularly in the first half of the simulation run. The Base price reaches hardly 20 Euro/MWh and also Peak shows lower prices than 30 Euro/MWh. Likewise Superpeak hardly obtains prices over 40 Euro/MWh. The reasons for this are not totally clear. One reason could be that the actually available amount of energy is higher in the simulation than in reality (due to the simplifications, we described in chapter 3) on the

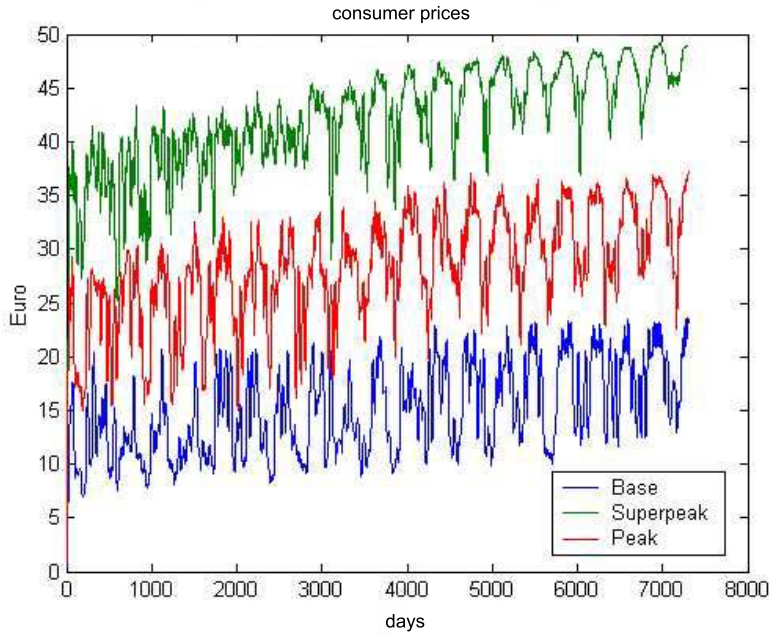

Figure 3: The consumer prices for the three groups of loads for a period of 20 years

other hand it is possible that the production costs were set too low, or that the price setting mechanism is responsible.

In Figure 4 the usage rate of the different types of power plants is shown. It can be seen that the entire energy provided by the hydro power stations is sold. The usage rate of the hydro power stations is somewhat too high, it lies in the yearly cut at approximately 5000 full load hours in reality. This is because our data report the actual water level of the rivers only very rudimentarily. The small usage rate of the storage power stations reflects their meaning as a supplier of peakload again. The usage rate of the coal power stations agrees in the order of magnitude with the reality. In principle a more exact modelling of the power plant portfolio of the individual producers should improve simulation results, since electricity market is strongly determined by technical basics.

Figure 5 indicates exemplarily, of which kind the statements could be, which one can expect from a simulation like this. The development of the budgets of the two producers is shown under the assumption that the demand increases annually by 2% and no capacities are added. In the first picture a constant fuel price is supposed, in the second picture gas price increases by 2%, while the other fuel prices remain constant and in the third picture an annual increase by 2% for all fuel prices is supposed. Thereby the development is surprising, when gas price increases and the other fuel prices stay constant. The better performance of P2 in this scenario might have to be attributed to the fact that P1 must fall back toward the end of the simulation to fuel-intensive small

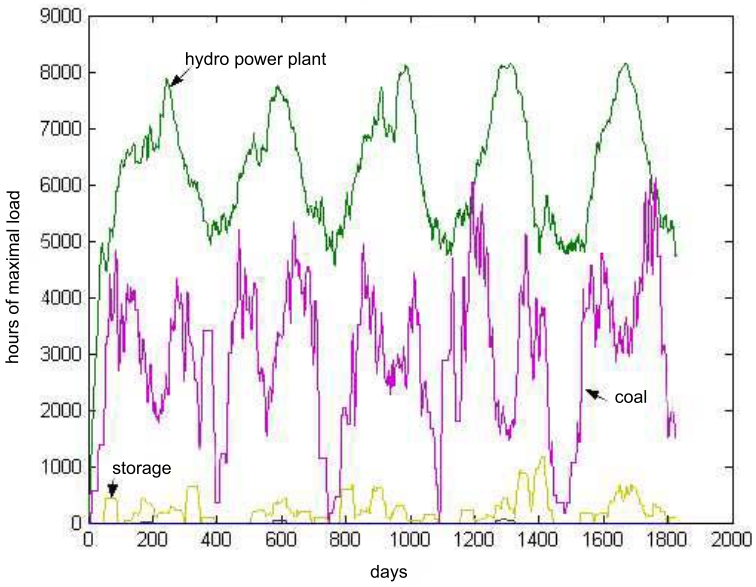

Figure 4: The usage rate of the different types of power plants

gas turbines.

6 Conclusion and Outlook

The simulation environment implemented at present can be seen only as a first step. Accordingly extensions and improvements are possible into almost all directions. It is the more amazing that with this simple means a satisfying picture of some parts of the electricity market is possible. This might be due to the underlying structure of the simulation environment. The decision to deal three electricity products daily is well justified and seems to be successful. Likewise the simple modelling of the demand side is not a large disadvantage. In the following the most important extensions and improvements are described briefly.

An implementation of accurate information about the used power stations seems particularly important. On the one hand exact data for the cost structure are important, on the other hand the technical side should be modelled in more detail. Of particular importance is the problem of heating up and cooling down thermal power units. In addition a more exact view of storage and pump-storage power stations is important. Maintenance and repair times and costs should be modelled. The transmission network between the individual regions is implemented only very rudimentarily. Only transmission capacities are restricted, but for example no transmission costs are implemented yet.

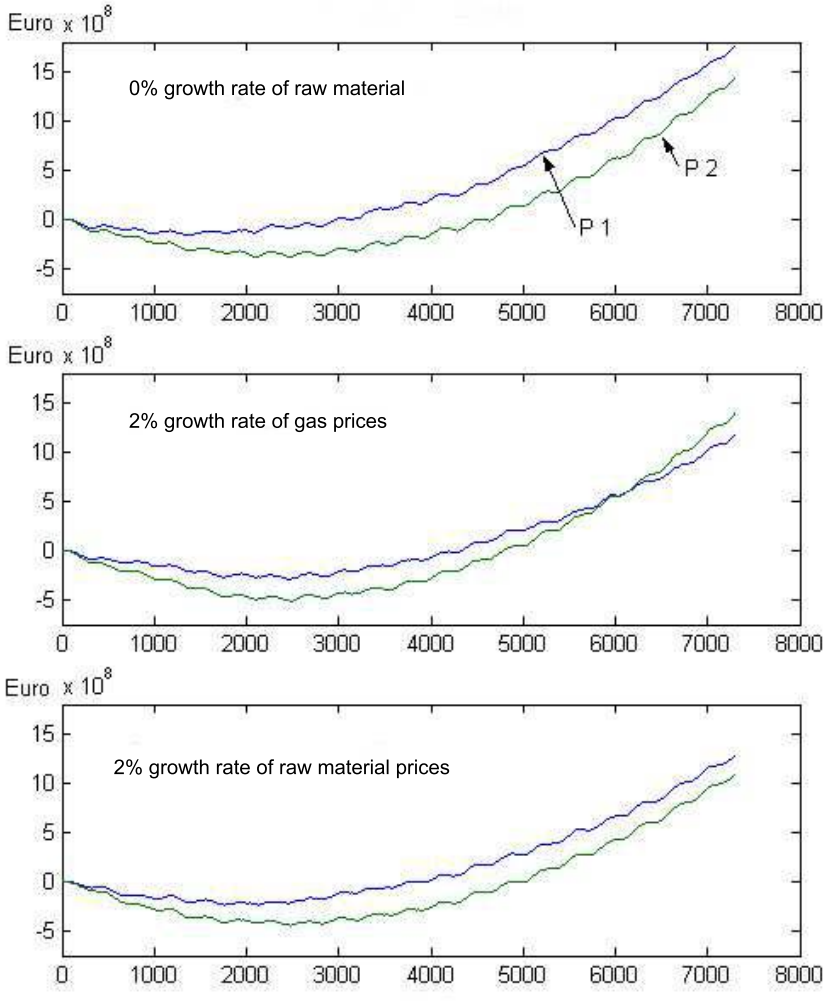

Figure 5: The development of the budgets of two competing producers under different assumption of fuel prices and with annually by 2% increasing demand

A scientifically founded analysis of simulation results can not be limited on "visual inspection" of individual trajectories. Testable hypotheses must be set up and examined at the results or other statistically founded methods have to be used.

The two methods of the price setting used so far are improvable. Marginally pricing appears too rigid and best response pricing produces unnatural fluctuations in the price time series. Some attempts with Machine Learning methods (e.g. Q-learning algorithm) were not successful. Although the theoretically necessary assumptions are not fulfilled, these methods are used extensively (Harp et al., 2000). Maybe agents with given, however adaptive strategies are working better.

Bibliography

Bunn, D. W. and Oliveira, F. S. (2001). Agent-based simulation: An application to the new electricity trading arrangements of england and wales. *IEEE-TEC on Evolutionary Computation, Special Issue: Agent Based Computational Economics*, 5(5):493–503.

Harp, S. A., Brignone, S., Wollenberg, B., and Samad, T. (2000). SEPIA: a simulator for electric power industry agents. *IEEE Control Systems Magazine*, 20(4):53–69.

Scherrer, W. (2001). MASS – MATLAB Aided Simulation Systems. Inst. of Econometrics, OR and System Theory, TU Vienna, Internal Report.

Schneider, L. (1998). *Stromgestehungskosten von Großkraftwerken*. Öko-Institut, Werkstattreihe 112.

Tesfatsion, L. (1997). How economists can get alive. In Arthur, W., Durlauf, S., and Lane, D., editors, *The Economy as an Evolving Complex System*, volume II. Addison-Wesley, Reading, Mass.

A Simulation Model of Coupled Consumer and Financial Markets

Brian Sallans, Alexander Pfister, Alexandros Karatzoglou and Georg Dorffner

1 Introduction

The study of economic phenomena involves not just the domain of economics, but also dynamical systems theory, game theory, the theory of adaptive learning systems, psychology and many others. Beginning with the seminal work of Herbert Simon (1982), there has been a realization that classical economic theory, based on rational equilibria, is limited. In reality, economic agents are bounded both in their knowledge and in their computational abilities. Recent work in simulation-based computational economics has sought to implement boundedly rational economic actors as learning agents, and to study the implications on the resultant economic systems. See Tesfatsion (2002) for a review of agent-based computational economics.

In this chapter we describe and study a discrete-time agent-based economic model which incorporates three types of boundedly rational agents: Production firms, Consumers, and Financial traders. These three agents operate in two coupled markets: a consumer market and a financial equities market. In the consumer market, production firms offer goods for sale, and customers purchase the good. The financial equities market consists of stock traders who can buy and sell shares in the production firms. The two markets are coupled through the production firms, which try to increase shareholder value. They might do this by increasing profits, or by taking actions which directly boost their stock price. Each firm explicitly implements a boundedly-rational agent which learns from experience, and has limited knowledge and computational power.

The other simulation work described in this collection generally takes a more detailed look at a single agent type or market. Models of consumers (Baier and Mazanec, 1999; Buchta and Mazanec, 2001), financial traders (Gaunersdorfer, 2000; Pfister, 2003), and production firms (Natter et al., 2001; Dawid et al., 2002) have been studied previously. The focus is on a single type of actor (firm, consumer or trader), with the other actors modeled as exogenous inputs or simple random processes. These models are appropriate for the study of intra-market phenomena like market segmentation, stock market volatility clustering or strategic decision making.

In contrast, this work takes an integrative approach, by simplifying and linking simulation models of consumers, stock markets and firms. While the individual agents are simpler than others being studied, the interaction between agent types and markets gives the model rich and interesting dynamics. Specifically, we build on the work of Steiglitz et al. (1995), Arthur et al. (1997), Brock and Hommes (1998), Gaunersdorfer (2000), Dangl et al. (2001) and Pfister (2003) in financial market modeling; Baier and Mazanec (1999) and Buchta and Mazanec (2001) in consumer modeling; and Tesauro (1999) and Natter et al. (2001) in production firm modeling. The goal is to simplify

and integrate these previous models, while still retaining their empirical behavior. In addition, the integrated model can address new phenomena that can not be investigated in separate models.

The purpose of integration is to allow us to study the mutual influence of the two markets. In particular, because the firms learn to act based on feedback from the financial market, we can examine the influence of the financial market on production firm behavior. Given rational expectations of firms and financial traders, one might expect that it does not matter whether a firm bases its actions on its own estimate of future performance, or on its stock price (the shareholders estimate of future performance). This would be the case if firms and traders were fully rational, since both would have the same estimate of the value of a firm and its actions. However, when both are only boundedly rational, then their estimators might be in disagreement. In this case the financial market could have a positive or negative influence on firm performance. The type and degree of influence will depend on how firms and stock traders estimate future performance. Further, if more sophisticated compensation schemes are used (such as stock options), then both stock market value and volatility can have an influence on firm behavior.

Before we use the model to investigate inter-market effects, we have to satisfy ourselves that it behaves in a reasonable way. We validate our computational model by comparing its output to known "stylized facts" in consumer and financial markets. Stylized facts are robust empirical effects that have been identified in a number of examples of real markets. Successful reproduction of empirical phenomena suggests that the dynamical properties of the model are similar to those of the real markets that it tries to emulate. We can then use the model to better understand the underlying dynamics and causes behind the observed effects. For example, by looking at what model parameter settings encourage particular behavior, we can get some insight into the underlying mechanism which causes it.

We also describe a novel validation technique based on Markov chain Monte Carlo (MCMC) sampling. By using the technique, we can investigate how model parameters influence model behavior, even for large parameter spaces. We can explicitly investigate how different model parameters are correlated, and under what conditions the model reproduces empirical "stylized facts". This new validation and exploration technique is widely applicable to agent-based simulation models, and is an important contribution of this work.

In this chapter we begin by giving an overview of the results that have come from this work. We then describe the integrated markets model (IMM) itself. After describing the model in detail, we explain the model exploration and validation technique based on MCMC sampling. We describe a number of stylized facts, and show simulation results from the integrated markets model. Using MCMC exploration, we show that the dynamics of competition in the consumer market are an important part of the overall dynamics in the financial market. Similarly, the dynamics of the financial market have an impact on the learning abilities of firms in the consumer market.

After specifying and validating the model, we give two examples of interactions between the consumer market and the financial market. First, we introduce an additional stock trader which bases its behavior on product position rather than profits. We

show that a minority of these traders in the market can have a significant influence on the firm's behavior.

The second example explores managerial compensation. We begin by showing that the integrated markets model replicates both known empirical behaviors from the compensation literature, and theoretical behaviors from optimal contract theory. We then use the model to generate new hypotheses about the usefulness of stock option compensation in a competitive market, based on their influence on exploration and risk-taking.

2 Overview of Results

Our implementation and use of the IMM has given rise to a number of interesting results in the areas of market behavior, feedback effects, model validation and managerial compensation. We give an overview of the most interesting results here. The details and supporting evidence are the subject of the remainder of this chapter.

2.1 Integration and Stochasticity

One of the fundamental results of this work is simply that such an integrated model is possible. In previous work on consumer and financial markets, agent types from outside of the market in question were replaced by simple stand-ins such as random processes. In the integrated model, there are no such random processes. All of the decisions made in the model are deterministic, based on agents maximizing their utility within the bounds of their knowledge.

In our model, a constant random influx of new information to the stock market is replaced by highly predictable fundamental information with occasional unpredictable shocks. Despite this, the stock market simulation has the same degree of unpredictability as other simulated markets. This is due to a combination of heterogenous trading strategies in the market, and the difficulty in predicting the timing and direction of these occasional shocks. The simulated stock market is also qualitatively similar to real stock markets in terms of volatility clustering and volume-volatility correlations.

An interesting effect can be seen in the tradeoff between firm learning and autocorrelations in the market returns. Fundamentalist traders trade based on profits of the firms. If the firms do very well or very poorly, profits are easy to predict, and the returns of the stock market becomes more autocorrelated. There is a regime in which the performance of the firms is variable, causing a decrease in the autocorrelation of market returns. In particular, the parameter controlling the length of the firm's memory trades off against the parameter controlling what kind of information (profits or stock price) is used to make decisions. When firm memory is short, or more difficult (stock) information is used by the firms to make decisions, firm performance is uniformly bad. When firm memory is long, and reliable (profit) information is used, firm performance is uniformly good. Memory lengths and reliance on stock information trade off against one another when the goal is to have decorrelated stock returns.

2.2 Bounded Rationality and Information Usage

The firms successfully make use of both profit- and stock-based information to make decisions and learn over time. In the consumer market, we see reasonable behavior including price competition and market segmentation. The managers learn to maximize their compensation over time.

Despite this, we can see the effect of bounded rationality, both in the consumer and the financial markets. The clearest example is the volatility of stock markets. Because of the inclusion of both fundamentalist and technical traders, the market price does not converge to the equilibrium price, but rather oscillates around it.

A more interesting example can be seen in the behavior of the firms. In the model, because the stock price is constructed from firm profits, both the profits and the stock price contain the same underlying information about firm performance. Given an unlimited memory and knowledge, a firm should therefore have the same performance whether its source of information is stock price or profits. Instead, we see that the firm performs best with a mixture of both sources of information. Further, some stock market structures are better than others from the point of view of firm performance. Specifically, there was an intermediate value for the influence of fundamentalist stock traders which gave the best firm performance. Because of the bounded rationality of the firms, the form of the information they receive, not just the content, is important.

2.3 Validation

Model complexity is always an issue. It is especially pressing in our case, because of the number of different agent types present in the integrated model. Even with each agent-type only requiring a few parameters, the total number of parameters quickly adds up. This makes it difficult to assess model sensitivity to parameters, and to find reasonable parameter settings.

In order to combat this problem, we introduce a novel model validation and exploration technique based on a Markov chain Monte Carlo sampling algorithm. This method samples model parameters according to how well the model reproduces behavior based on empirical "stylized facts" from real markets. Using this technique, we can explicitly see how a parameter influences model behavior, and how parameters interact to cause model effects.

During validation, we found that high kurtosis of marginal stock returns, and volatility clustering are very robust features of our artificial market. All reasonable parameter combinations produced these features. The one exception was that the stock market simulation was sensitive to the number of fundamentalists in the market. In a market with 20% fundamentalists, the returns look Gaussian. If the proportion of fundamentalists drops below 10% the stock price collapses. The heterogeneity of the market traders is necessary to maintain market liquidity and trading volume, and compensates for the lack of arbitrary randomness in our model.

We also found that our artificial stock market exhibited a positive correlation between volatility and trading volume, as seen in real markets. This occurs because when the stock price is volatile, more traders want to adjust their portfolios, and volume

increases. Similarly, when the market has high volume, more trades can be made, allowing for larger adjustments to the stock price.

2.4 Fundamental Value and Stock Price Inflation

The goal of building an integrated markets model is to use it to study feedback and interaction effects between the consumer and financial markets. We have already mentioned several effects where the behavior in one market influences the stylized facts seen in the other market.

We have also investigated the ability of the stock market to integrate multiple information sources, and to influence the performance of the firms. One form of feedback occurs when financial traders base their valuation judgments on some aspect of the firm's performance in the consumer market. Previous stock-market models have shown evidence of short-term stock market "bubbles" caused by technical trading. Our simulations show that traders with alternative and conflicting "fundamental" valuations can explain periods of sustained stock price inflation, which outlast short term dynamic fluctuations.

We found that when managers are rewarded based on stock performance, they will extract from the stock price information about these conflicting ideas of fundamental value. Moreover, they will consistently cater to the valuation theory over which they have the most direct control. In our case, we show that firms will develop "trendy" products which elicit investor excitement, even when this sacrifices profits.

As a byproduct, we also show that the stock market can integrate conflicting ideas of what constitutes fundamental "value", and that these separate ideas can then be extracted from the stock signal by the firm.

2.5 Managerial Compensation

Optimal contracts are a major subject of theoretical study. Unfortunately, it is very difficult to confirm theoretical contract results with empirical evidence, because of difficulties in gathering and analyzing the empirical data. We show that the integrated model can be used as an alternative to empirical tests to check theoretical predictions . The simulation model is more realistic than theoretical models, in that it incorporates bounded knowledge and learning, and operates over many repeated trials. However, it is more accessible than empirical studies, because we have access to the "ground truth". We also use the model to generate new hypotheses, to be explained by theory and tested empirically.

In particular, we use the model to test the predictions of principal-agency theory with respect to the effect of stock options on firm risk taking. The simulation model tests these predictions under conditions that are quite far removed from the theoretical models: The firms are boundedly rational and have no prior knowledge of consumer or stock market preferences. This indicates that the effect of stock options is quite robust, and does not overly rely on the assumptions of the theoretical models.

Although the effect of stock options is as predicted by principal-agency theory, the mechanism in our model is quite different. In theoretical models, managers know

the probable outcomes of their actions and the risks that they take. They are willing to gamble, because stock options insulate them from negative consequences. This results in a higher expected return, although with higher risk. In our simulation model, managers must experiment to discover the outcomes of their actions. They are willing to experiment because the stock options insulate them from failed experiments. Experimentation results in finding more effective strategies, and a higher return. This alternative mechanism is not suggested by theoretical models.

The theoretical mechanism requires a priori knowledge, while the simulated mechanism will only occur when managers must learn consumer and stock market preferences. We also show that learning and knowledge acquisition are influenced by the timing of stock option grants. We find that options are most effective when they are introduced in response to a need to learn new behavior, rather than being included as a standard part of a compensation contract.

3 The Integrated Markets Model

In this section we give a brief overview of the integrated markets model. The reader is directed to Sallans et al. (2003) for a detailed description of the model and validation results. For completeness, the model parameters and equations are included as Appendix A.

The model consists of two markets: a consumer market and a financial equities market. The consumer market simulates the manufacture of a product by *production firms*, and the purchase of the product by *consumers*. The financial market simulates trading of shares. The shares are traded by *financial traders*. The two markets are coupled: The financial traders buy and sell shares in the production firms, and the managers of firms may be concerned with their share price. The traders can use the performance of a firm in the consumer market in order to make trading decisions. Similarly, the production firms can potentially use positioning in product space and pricing to influence the decisions of financial traders (see Figure 1).

The simulator runs in discrete time steps. Simulation steps consist of the following operations:

1. Consumers make purchase decisions.
2. Firms receive an income based on their sales and their position in product space.
3. Financial traders make buy/hold/sell decisions. Share prices are set and the market is cleared.
4. Every N_p steps, production firms update their products or pricing policies based on performance in previous iterations.

We describe the details of the markets, and how they interact, in the following sections.

3.1 The Consumer Market

The consumer market consists of firms which manufacture products, and consumers who purchase them. The model is meant to simulate production and purchase of non-durable goods, which the consumers will re-purchase at regular intervals. The product

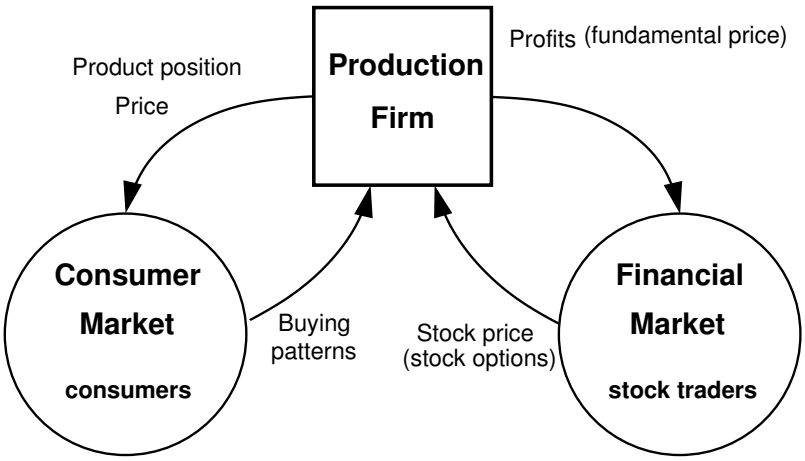

Figure 1: The integrated markets model. Consumers purchase products, and financial traders trade shares. Production firms link the consumer and financial markets, by selling products to consumers and offering their shares in the financial market.

space is represented as a two-dimensional simplex, with product features represented as real numbers in the range [0,1]. Each firm manufactures a single product, represented by a point in this two-dimensional space. Consumers have fixed preferences about what kind of product they would like to purchase. Consumer preferences are also represented in the two-dimensional product feature space. There is no distinction between product features and consumer perceptions of those features.

Firms

The production firms are adaptive learning agents. They adapt to consumer preferences and changing market conditions via a reinforcement learning algorithm (Sutton and Barto, 1998). In each iteration of the simulation the firms must examine market conditions and their own performance in the previous iteration, and then modify their product or pricing.

A boundedly rational agent can be subject to several kinds of limitations. Our model explicitly implements limits on knowledge, and representational and computational power. These limits manifest themselves in the firm's representation of its environment and its knowledge of its competitors.

The firms do not have complete information about the environment in which they operate. In particular, they do not have direct access to consumer preferences. They must infer what the consumers want by observing what they purchase. Purchase information is summarized by performing "k-means" clustering on consumer purchases. K-means is a common clustering technique used in consumer market research. The current state information consists of the positions of the cluster centers in feature space, along with additional state information such as whether or not the previous

action was profitable or boosted stock price, and where the competitors products are located.

This information gives a summary of the environment at the current time step. Firms make decisions based on a finite history of states of some length. This limited history window represents an additional explicit limit on the firm's knowledge.

In each iteration the firms can take one of several actions. The actions include taking a random action, doing nothing, raising or lowering product price, or moving the product in feature space. The random action was included to allow the firm to explicitly choose to take a "risky" exploratory action.

A firm's manager seeks to modify its behavior so as to maximize an external reward signal. This reward signal can be viewed as the managers compensation for its actions. The influence of this reward signal on the firm's behavior will be the focus of our investigation, and is described in section 6.1.

Given the reward signal, the firm learns to make decisions using a reinforcement learning algorithm (Bertsekas and Tsitsiklis, 1996; Sutton and Barto, 1998). Given the reward signal at each time step, the learning agent attempts to act so as to maximize the total (discounted) reward received over the course of the task. The discounting indicates how "impatient" the manager is to receive her reward. It can also be related to the interest rate for a low-risk investment or the rate of inflation.

Consumers

Consumers are defined by their product preference. Each consumer agent is initialized with a random preference in product feature space. During each iteration of the simulation, a consumer must make a product purchase decision. For each available product, the consumer computes a measure of "dissatisfaction" with the product. Dissatisfaction is a function of product price and the distance between the product and the consumer's preferred product. Given dissatisfaction ratings for all products, each consumer selects from this set the product with the lowest dissatisfaction rating.

3.2 The Financial Market

The financial market model is a standard capital market model (see, e.g., Arthur et al., 1997; Brock and Hommes, 1998; Dangl et al., 2001). Myopic investors maximize their next period's utility subject to a budget restriction. At each time step agents invest their wealth in a risky asset (a stock or index of stocks) and in bonds, which are assumed to be risk free. The risk free asset is perfectly elastically supplied and earns a risk free and constant interest rate. Investors are allowed to change their portfolio in every time step.

As in Brock and Hommes (1998); Levy and Levy (1996); Chiarella and He (2001, 2002) the demand functions are derived from a Walrasian scenario. This means that each agent is viewed as a price taker (see Brock and Hommes, 1997; Grossman, 1989).

As in many other heterogeneous agent models we assume that two kinds of investors exist: Fundamentalists and chartists. The two types of investors differ in how they form expectations of future prices. Additionally investors have different time

horizons which are modeled via the time length agents look back into the past.

Fundamentalists determine their price expectations according to a model based on fundamental information, which in our model are past dividends. They calculate a fair price and expect that the current price will gradually move towards it at some fixed rate. A fundamentalist assumes that the fair price is a linear function of past dividends. Chartists use the history of the stock prices in order to form their expectations. They assume that the future price change per period equals the average price change during the previous periods.

In our model we want to focus on the formation of expectations about prices and not on the formation of expectations about variances. Therefor we assume homogeneous and time independent expectations about the variance.

The market uses a sealed-bid auction, where the clearance mechanism chooses the price at which trading volume is maximized. The first step is to construct supply and demand curves based on the transaction requests. Then, a price is found which maximized the volume of shares traded. Note that there may be a range of prices that would maximize volume. We select the maximum price in this range. If there are buy orders but no sellers then the share price is set to the maximum bid. If there are only sell orders then the price is set to the minimum ask. If there are no orders in a time period, then the price remains unchanged. Each trader specializes in a single firm, and only buys or sells shares in this firm. Each trader is initialized with a supply of shares in its firm of interest.

Let us have a look at the timing of the events within the financial model. The first step is the formation of expectations. Based on past prices and dividends an investor forms his/her expectation about the distribution of the next period's price and dividend. The trading agent is then able to determine the demand function, which is submitted to the stock market via limit buy orders and limit sell orders. After the orders of all agents are submitted the stock market calculates this period's equilibrium price At the end of the period the current dividend is announced and becomes public information.

4 Model Validation

One goal of constructing agent-based economic models is to gain some insight into the mechanisms that cause observed market behaviors. Agent-based economic models offer a kind of economic laboratory, in which parameters can be changed, and the results observed. Useful models will reproduce known market behaviors for reasonable parameter settings. Knowing the behavior of the model in different parameter regimes is therefore important both for validating that a model is reasonable, and using the model to understand economic phenomena. However, in complicated models with many parameters, it may be difficult to discover relationships between model parameters, and find regions in parameter space where the model has interesting behavior.

We will validate our model by confirming that it can indeed reproduce empirically observed market behaviors, or "stylized facts". In this section we propose a novel algorithm for exploring the relationship between model parameters and stylized facts. The algorithm is based on Markov chain Monte Carlo (MCMC) sampling. We describe a number of empirical phenomena that have been observed in consumer and financial

markets, and give corresponding simulation results. We show that a number of stylized facts within the two markets can be reproduced by our model under reasonable parameter settings. We further show that the behavior of each of the markets is dependent on the dynamics of the other market. In other words, the integrated model is not simply two separate models joined together. The behavior of each market is intimately tied to the parameters and dynamics of the other market. We explore the mechanisms behind some stylized facts by examining correlations between model parameters. More details about the model and its validation can be found in Sallans et al. (2003).

4.1 Model Parameters

Although it has been our intention to keep the model simple, the firm's learning algorithm and trader's decision rules have tuning parameters. Parameter values must be selected before a simulation can be run. These parameters have been introduced in earlier sections describing each of the agents in the model. Using preliminary simulations, some of the parameters were found to have a large influence on the outcome of the simulation, and others were found to be relatively unimportant. All parameters are summarized in Table 1. The "value" column indicates the value used for simulations (see section 4.3). The values of parameters in the first group (above the double line) were found using the Markov chain simulation technique described in the next section. Those in the second group were found to be relatively unimportant. These values were set based on initial trial simulations, and held fixed for all simulation runs.

Table 1: Parameters for Integrated Markets Simulator

Parameter	Description	Range	Value
α_ϕ	strength of profitability reinforcement	$[0, 1]$	0.47
α_p	strength of stock price reinforcement	$[0, 1]$	0.53
N	Number of cluster centers	$\mathbb{N}$	2
ν	product update rate	$\mathbb{R} \geq 0$	0.03
γ	reinforcement learning discount factor	$[0, 1]$	0.83
H_s	History window length for firms	$\mathbb{N}$	3
N_f	Proportion of fundamentalists	$[0, 1]$	0.57
N_c	Proportion of chartists	$[0, 1]$	0.43
α_f	Fundamentalist price update rate	$[0, 1]$	0.18
α_n	Chartist price update rate	$[0, 1]$	0.36
K	Number of bins	$\mathbb{N}$	10
N_p	product update frequency	$\mathbb{N}$	8
S_f	base salary	$\mathbb{R} \geq 0$	0
λ	reinforcement learning rate	$\mathbb{R} \geq 0$	0.1
ϵ	reinforcement learning temperature	$[0, 10]$	$5 \rightarrow 0.2$
α_c	Consumer feature/price tradeoff	$[0, 1]$	0.5
MAXDIS_i	Maximum dissatisfaction for consumer i	$[0, 1]$	0.8
f	inverse fair dividend yield	$\mathbb{R}$	50

We would like to understand the effect of parameters on model behavior. We could "grid" the space of parameters, and then run a large number of repetitions of the simulator, one for each grid point. However, this approach would very quickly become infeasible with more than a few parameters. Ten parameters, each taking on one of ten values, would require 10^{10} runs to cover the grid. Many of these parameter combinations will not be of particular interest.

Instead we would like a way to focus computational power on areas of parameter space that are "interesting". We will define as interesting areas where a stylized fact is well-reproduced. To this end, we will adapt Markov chain Monte Carlo sampling to do a "directed" random walk through parameter space.

4.2 The Metropolis Algorithm

Consider the problem of evaluating the expected value of some multivariate function with respect to a probability distribution or density. In some cases (such as linear functions and Gaussian distributions) expectations can be computed analytically. In many cases this is not possible. Monte Carlo algorithms allow for the approximate evaluation of expectations in more difficult circumstances. In the following, bold face will denote a vector and subscripts will denote elements of a vector or set: $\mathbf{x} = \langle x_1, ..., x_J \rangle$. Given a set of multivariate samples $\{\mathbf{x}_1, ..., \mathbf{x}_N\}$ from a distribution $P(\mathbf{x})$, we can approximate the expected value of a function $f(\mathbf{x})$ as follows:

$$E[f(\mathbf{x})]_{P(\mathbf{x})} \approx \frac{1}{N} \sum_{i=1}^{N} f(\mathbf{x}_i) \tag{1}$$

Before this approximation can be employed, we need a set of samples $\{\mathbf{x}_1, ..., \mathbf{x}_N\} \sim P(\mathbf{x})$. In many cases we do not have a closed-form distribution from which samples can be drawn. The Metropolis algorithm (Metropolis et al., 1953) is a method for drawing a set of samples from a distribution $P(\mathbf{x})$. Further, we need not have access to $P(\mathbf{x})$, but only need an unnormalized energy function $\Phi(\mathbf{x})$, where:

$$P(\mathbf{x}) = \frac{\exp\{-\Phi(\mathbf{x})\}}{\sum_{\mathbf{x}'} \exp\{-\Phi(\mathbf{x}')\}} \tag{2}$$

Given an initial point $\mathbf{x}_0$, the i^{th} step of the Metropolis algorithm operates as follows:

1. Select a dimension k. Select a proposed sample $\mathbf{x}$ from a *proposal distribution* $\Pr_k(\mathbf{x}; \mathbf{x}_{i-1})$. The proposal distribution can be a function of the previous point, and leaves all of the elements of $\mathbf{x}_{i-1}$ unchanged except for the k^{th} element.
2. Set $\mathbf{x}_i \leftarrow \mathbf{x}$ with probability $\min\{1, \exp\{-(\Phi(\mathbf{x}) - \Phi(\mathbf{x}_{i-1}))\}\}$. This is called *accepting* the proposed sample. Set $\mathbf{x}_i \leftarrow \mathbf{x}_{i-1}$ otherwise (*rejecting* the proposed sample).

Note that when a proposal is rejected, the old point is added to the sample in its place. In the algorithm as described above, the proposal distributions should be symmetric. That is, $\forall k \forall \mathbf{x} \forall \mathbf{x}' \ \Pr_k(\mathbf{x}; \mathbf{x}') = \Pr_k(\mathbf{x}'; \mathbf{x})$.

In the limit, the sequence of samples will converge to a unique stationary distribution with marginal distribution $P(\mathbf{x})$. Thus the set of samples can be used for the approximation in Eq.(1). In practice, the speed of convergence of the chain to the stationary distribution will depend on the dimensionality of $\mathbf{x}$, the energy function of interest and the proposal distribution. Assessing convergence can be problematic. If a non-convergent set of samples is used, then the estimate will be biased. The algorithm can also be extended to include non-symmetric proposal distributions.

4.3 Markov Chain Model Exploration

In our application, we do not want to evaluate expectations of a function. Instead, we want to find settings for model parameters that reproduce stylized facts. The Metropolis sampler has the following property: Samples are more likely to be drawn from low-energy areas.

Given a stylized fact, we can define an energy function such that low energy corresponds to good reproduction of the fact. Then, we implement a Metropolis sampler using this energy function. In the limit, parameter samples are drawn according to the normalized probability distribution defined by the energy function. In practice, we will not generate Markov chains which are sufficiently long to reach the equilibrium distribution. But even without theoretical guarantees on the distribution of sampled parameters, the sampler can find good model parameter settings, and reveal interesting correlations between model parameters. The Metropolis sampler acts as a "directed" random walk through parameter space, avoiding high energy areas.

We have constructed energy functions for several stylized facts including: learning-by-doing in the consumer market, low autocorrelations in stock returns, high kurtosis in marginal returns, and volatility clustering. The sampler operated over the parameters in the first group of Table 1. We used symmetric Gaussian proposal distributions over real-valued parameters, and uniform distributions over discrete parameters. It was assumed that the energy function took on a value of $+\infty$ wherever parameters fell outside of their valid range, ensuring that such values would be rejected by the sampler. One thousand samples were drawn using the Metropolis sampler. While this is too short to allow for convergence, we can still examine the sample set to identify regions where stylized facts are well-reproduced, and look for significant correlations between parameters.

As it turns out, two of the four Markov chain experiments were uninteresting. These were the runs trying to achieve high kurtosis in the stock market returns, and getting high autocorrelations in the absolute stock returns. The simulated stock market had both of these features for almost all parameter values, and there were no interesting correlations or relationships between parameters for these energy functions. The only parameters of interest were the proportion of fundamentalists and chartists. If the number of chartists fell below 20%, the returns looked Gaussian. This suggests that high kurtosis and volatility clustering are very robust features of the artificial stock market, and are driven by the interaction between fundamentalists and chartists.

In the sections below, we show the results for two energy functions: The "learning-by-doing" effect, and low-autocorrelations in the stock market returns.

Learning by Doing

The "learning-by-doing" effect (Argote, 1999) encapsulates the idea that firms gain knowledge and optimize their behavior over the course of performing a task. Empirically, costs go down, and efficiency and profits go up as a function of the number of units of a particular product produced. Our model explicitly includes learning by doing in the production firm. As the firm produces its product, it learns what sells in the marketplace and at what price. This results in an increase in profits over time. Note that this is very different from models which include a "learning" component in populations of agents, implemented as an evolutionary algorithm. Our individual firms learn over the course of the task.

We investigated which parameter settings influence the learning-by-doing effect using our adapted Metropolis algorithm. The energy function was the negative profits:

$$E = \frac{1}{Z_p} \sum_i \sum_{t=2}^{T} -(\phi_{i,t}) \tag{3}$$

where i indexes the firms, and Z_p is simply a scaling factor designed to bring the energies into a reasonable range (set to 10000 for our simulations).

We found that the learning effect was quite robust to parameter settings. In general, firms learned to perform well in the market place for almost all parameter settings (Figure 2).

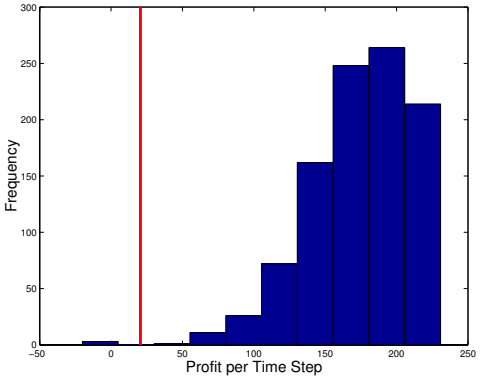

Figure 2: The "Learning by Doing" effect was robust across almost all parameter settings. The bar graph shows the per-time-step profits of the firms sampled by the Metropolis algorithm. The vertical line shows mean profits achieved by a randomly-behaving firm. The learning firms do better than a randomly acting firm for nearly all parameter settings.

There was a significant negative correlation between the proportion of fundamentalist traders N_f in the simulation, and the adaption rate α_f of the fundamentalists.[1]

Note what this implies: Two parameters of the stock market are correlated when trying to maximize a quantity from the consumer market (profits). This suggests that the feedback mechanism from the stock market to the production firms (via stock price) is having an influence on the behavior of the firms in the consumer market,

[1] Significance was measured in the following way: First, the sequence of parameter values was subsampled such that autocorrelations were insignificant. Given this independent sample, the correlations between parameters could be measured, and significance levels found.

and that some intermediate behavior of the financial market is optimal from the point of view of firm learning. This may be because of an exploration/exploitation tradeoff: A certain amount of noise or uncertainty in the financial market could help the firms avoid shallow local minima and prompt them to find products and prices that are clearly winning in the consumer market. Too much noise can inhibit learning. The proportion and adaptation rate of the fundamentalists influences the volatility of the financial market, and therefore the noise in a firm's reward signal.

Low Predictability and Volatility Clustering

A fundamental feature of financial markets is that they are not easily predictable. The perfect market hypothesis claims that new information is immediately factored into prices, so that the price at any given time reflects all prior knowledge. Under this assumption, it is in principle impossible to predict market movements. In practice, it has been found that many financial return series have insignificant autocorrelations. Unlike most artificial stock markets, our model does not include any extrinsic noise (such as randomized trading strategies (Gaunersdorfer, 2000; Raberto et al., 2001), or a randomized dividend process (Arthur et al., 1997)). Interestingly, the autocorrelations are nevertheless very low. This is due to a combination of heterogenous trading strategies in the market, and the difficulty in predicting shocks in the consumer market.

Unlike price movements, price volatility is highly autocorrelated. Empirically, market volatility is known to come in "clusters". That is, periods of high or low volatility tend to follow one another. This is the basis of conditional heteroskodacity models, which predict volatility in the next time step based on volatility in previous time steps. In our model technical traders will tend to adjust their market position during large price movements. This will in turn cause greater price movements. Similarly, when the price is near the fundamental price, fundamentalists are satisfied and hold their stock. This in turn stabilizes prices, and causes the chartists to hold their stock as well.

We investigated which parameter settings lead to low autocorrelations in the returns of the artificial stock market. The energy function used was the squared error between the actual autocorrelations in the returns, and an idealized set of autocorrelations:

$$E = \sum_{i=1}^{A} (v_i - v_i^*)^2 \tag{4}$$

where v_i denotes the autocorrelation at lag i, and v_i^* is the idealized autocorrelation. We used $A = 5$, and $v^* = \{-0.05, 0.0, 0.0, 0.0, 0.0\}$. That is, a slight negative autocorrelation at the first lag, and zero autocorrelation thereafter.

After sampling with this energy function, we found significant correlations between some sampled production firm parameters. This is particularly interesting, because it indicates that the statistical properties of the stock returns are substantially affected by the dynamics in the consumer market. Specifically, there is a significant negative correlation between the firm's "history depth" parameter H_s, and the weighting placed by the firm on profits α_ϕ (at the 95% confidence level). That is, in order to get low autocorrelations in the stock returns, it is best to have either a short history

depth, or to place the most weight on improving the stock price (at the expense of profits).

This is likely related to how hard it is for the firms to learn to do well in the market. Recall, the sampler is trying to find parameter values for which the stock returns have low autocorrelations. That is, the sampler prefers stock prices that are unpredictable. If the firms do very well or very poorly, then their fundamental price is predictable, and the stock returns have higher autocorrelation. There is a regime in which firms have variable performance. The amount of information available to firms (the history length H_s) and the kind of information available (profits or stock price) appear to trade-off in determining firm performance.

This suggests an alternative to previous models that required a continuous influx of random information, or the continuous use of randomized decisions. In our model, fundamental information is predictable for long periods of time, interspersed with occasional unpredictable shocks. When the shocks are absent, return series from the market become autocorrelated. These occasional but unpredictable shocks drive the market dynamics, and are sufficient to decorrelate the market returns.

4.4 Ideal Parameters

We identified a set of parameter settings for which all of the stylized facts were well reproduced (see Figure 3 and column "Value" in Table 1). We did this by intersecting the histograms of parameter values from the MCMC simulation runs, and finding common parameter settings. Since nearly all parameter settings gave good kurtosis and volatility clustering behavior, these have been omitted from the figure for clarity. After identifying a set of parameter settings for which all of the stylized facts were well reproduced we ran 20 repetitions of the simulation at these ideal parameter settings. The simulation consisted of two competing firms, 50 stock traders, and 200 consumers.

The "ideal" parameter values are reasonable. There are no parameters which must take on extreme or unlikely values in order to get good simulation behavior.

The following sections show stylized facts reproduced by simulation runs at the ideal parameter settings.

The Learning Effect

Figure 4 shows simulated profits as a function of time, across the 20 simulation runs at the parameter settings specified in Table 1. Median profits increase as a function of time, indicating that firms learn to identify good product positions and prices. The increase is significant at the 5% level, as tested with a Wilcoxon signed rank test.

Autocorrelations of Returns

Figure 5 shows autocorrelations of returns and absolute returns for the artificial market. The autocorrelations were computed for the last 2000 periods of each runs, and averaged over 20 runs. For these plots, p_t is stock price at time t, and returns at time t are defined as $\mathrm{ret}_t = \log(p_t/p_{t-1})$. There are small negative autocorrelations in the

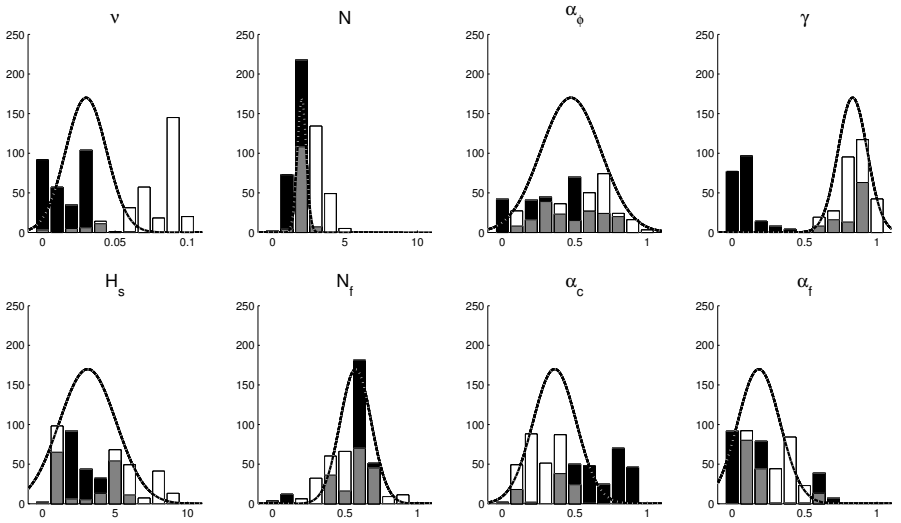

Figure 3: Histograms of parameter values from Markov chain Monte Carlo sampling. The plot for each parameter shows three histograms: black for the "learning-by-doing" energy function (Section 4.3), white for the low-autocorrelations energy function (Section 4.3), and gray for the intersection of the other two. Each histograms includes the top 30% of samples from the MCMC sampler, ranked by the negative energy. The curve shows a Gaussian fit to this intersection. The "ideal" parameters were taken to be the means of these best-fit Gaussians.

first few lags, followed by zero autocorrelations. The kurtosis of the market returns was quite high at 57.8. The error bars show 95% confidence bounds.

Fundamental Price

In financial markets, it is generally assumed that share price oscillates around a "fundamental" fair value, or fundamental price. This price can be related to dividends, cash flow or profits made by the firm. Empirically, it has been shown that models of the fundamental price can account for some of the variance in share price (Shiller, 1981; Fama and French, 1988; Kim and Koveos, 1994). Computational models of stock markets have typically assumed either a static fundamental price, or a simple time-varying price such as a first-order autoregressive process (Arthur et al., 1997; Gaunersdorfer, 2000). Because our model includes a consumer market, our fundamentalist traders construct a fundamental price based on the actual past profits of the firm.

Figure 6 shows a simulated stock price and the associated fundamental price, as calculated by the fundamentalist traders, from a sample run. The simulation used the parameter settings from Table 1. The fundamental price is the equilibrium price in a market of only fundamentalist traders. The fundamental price shown was rescaled and translated to compensate for the adaptation rate of the fundamentalists.

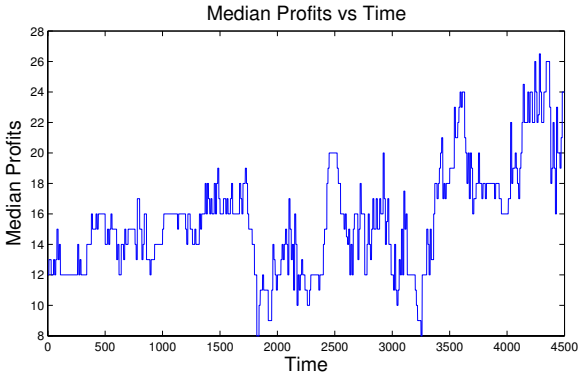

Figure 4: "Learning by Doing" in the consumer market. The plot shows median profits as a function of time, across 20 simulation runs. The longer a firm stays in the market, the higher its profits.

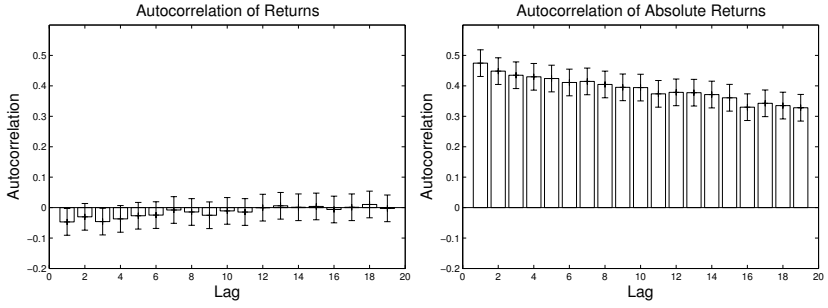

Figure 5: Autocorrelations of log returns and absolute log returns in the artificial stock market.

This sequence shows several aspects of the artificial stock market. First, the stock price roughly reflects the underlying fundamental price. The price differential is due to the number of stocks held by the traders initially (in our case 120). Second, the stock price oscillates at a higher frequency than the underlying fundamental price. Despite this, fundamental price information is incorporated slowly, due to the adaption rate α_f less than 1.0. Large stock price changes lag behind similar changes in the fundamental price. Third, large changes in fundamental price lead to high volatility in the stock price. Fourth, the stock price tends to over- or under-shoot and then oscillate after a large change.

For this run, the proportion of fundamentalists was quite high ($N_f = 0.57$). It is interesting that, under our model, decreasing the proportion of fundamentalists tends to also decrease the kurtosis of the returns. In a market with only 20% fundamentalists, the returns look Gaussian. If the proportion of fundamentalists drops below 10%, the stock price collapses. The heterogeneity of the market traders is necessary to maintain market liquidity and trading volume.

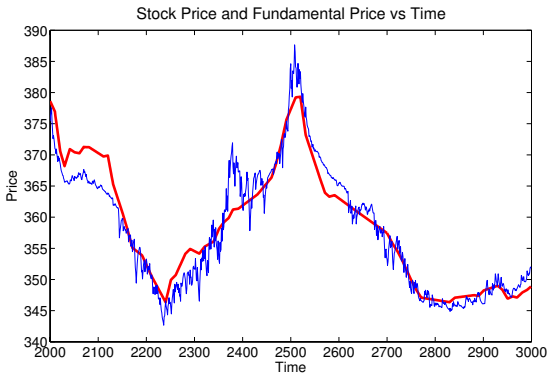

Figure 6: Fundamental price (thick line) and stock price (thin line) from a section of a single run of the integrated markets model. The fundamental price has been translated and rescaled to compensate for the adaptation rate of the fundamentalist traders.

If the fundamental price remains static over a long period of time, then the share price tends to decay in a deterministic way to the fundamental price. The variation in fundamental price due to the dynamics in the consumer market is an integral part of the stock returns in our model.

Volatility and Trading Volume

There is a known positive correlation between volatility and trading volume in financial markets (Karpoff, 1987). That is, periods of high volatility are also those of high trading volume.

Our integrated model exhibits the same behavior. There is a correlation between volatility and trading volume. High volume and high volatility are interrelated, and each can significantly predict the other, although the effect of high volatility on trading volume is longer lasting. Figure 7 shows average cross-correlations and 95% confidence intervals for stocks from the 20 runs of the simulator, with parameters set as in Table 1.

4.5 Discussion

In this section we have described an integrated model consisting of three agent types: production firms, consumers and financial traders. The agents operate in two coupled markets: a consumer market and a financial market. The model builds on previous work by simplifying and integrating previous models of consumers, firms and traders. We have found that for a particular reasonable setting of the parameters, a large number of stylized facts can be reproduced simultaneously in the two markets. We have also indicated in which parameter regimes the model does not perform well with respect to different stylized facts.

We have shown that it is possible to incorporate a profit signal from a competitive

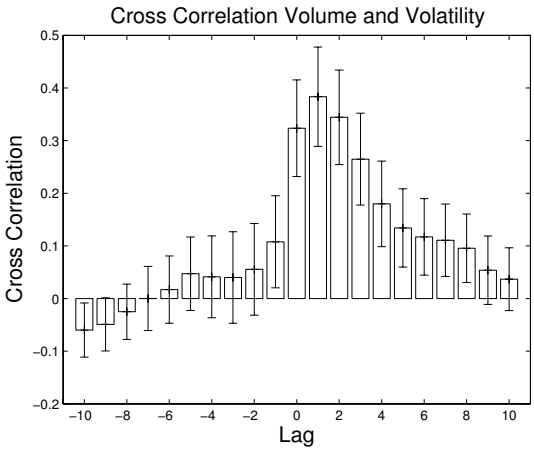

Figure 7: Cross correlation between trading volume and absolute returns. The figure was generated by averaging 45-day periods of volume and absolute returns for 40 stocks (20 runs, 2 firms per run). Cross correlations were measured for each stock. The plot shows mean cross correlations and 95% confidence intervals. The plot shows that volatility and trading volume are interrelated, with each being a significant predictor of the other, although the effect of volatility on trading volume is longer-lasting.

consumer market endogenous to the model itself. This endogenous profit signal provides some of the low-frequency and large-scale variability seen in the financial market model. The fundamental information is correlated with stock market returns. Our model demonstrates that a market that follows deterministic fundamental information can still have low autocorrelations, despite the fact that the fundamental information itself is highly autocorrelated. In fact, the shocks provided by changes in fundamentals are necessary to achieving low autocorrelations in stock market returns.

We have introduced a new model validation technique based on Markov chain Monte Carlo sampling, and used the new technique to investigate under which model parameter regimes the model exhibits realistic behaviors. We have shown that this technique can highlight interesting correlations between model parameters and offer insights into the mechanisms underlying the behavior of the model. We feel that this technique has wide applicability to other agent-based models, and is an important contribution of this chapter.

We have demonstrated that the combined model is more than just the sum of its parts. The behavior of each of the markets is substantially influenced by the dynamics of the other market. Firm performance in the consumer market is significantly affected by how the firm estimates future performance. Firms operate best given a mixture of performance-based and stock-based pay. This offers an example of the influence of the form, and not just the content, of information in a boundedly rational system. Similarly, the statistical properties of the stock market are best for intermediate values of firm parameters.

In the next sections, we use the integrated model to investigate inter-market stylized facts that are beyond the reach of individual models. These include product hype in the financial market and managerial compensation schemes.

5 Share Price Inflation and Product Hype

In the previous section we have seen that the two markets can have a significant impact on one another. In this section we explore the extent to which the financial market can reconcile different views of "fundamental value", and the extent to which these different views influence the behavior of the firm. In particular, we would like to see if conflicting views of fundamental value can account for long-term price inflation. In order to do this we introduce a new type of trader. More details can be found in (Sallans et al., 2002).

A stock market "bubble" is said to occur when the price of a stock rises significantly above its fundamental price. Previous agent-based market models have shown bubble formation simply as a consequence of the dynamics of the market. That is, a bubble can form because of random fluctuations, and then be inflated due to trend-following traders. However, these type of "dynamic" bubbles are relatively short term. Technical trading can not sustain the price inflation when the "fundamental" traders begin to sell their shares.

In real markets, we have recently seen bubbles which lasted over several years. That is, valuation of shares in certain markets was significantly above their "fundamental" value, in terms of the profitability or cash flow of the firm, over a prolonged period. However, these shares were only overvalued according to a profit-based valuation. This leaves open the possibility that traders were acting rationally, but on alternative definitions of "fundamental value". In this section we experiment with traders who have conflicting definitions of fundamental value, and see to what extent they can cause and sustain stock price inflation above fundamental value.

5.1 Hypist Traders

Hypists are meant to simulate traders who base their trading decisions on what they think will be "the next big thing". Like the other traders, hypists base their buying and selling decisions on predicted price movements. However, the Hypist bases its buying decision solely on the position of the firm's product in product space (see Figure 8).

Like the consumers, each hypist is initialized with a fixed preferred point in product space. If a firm moves its product closer to this preferred point, the hypist assumes that the stock price will increase. Otherwise, the hypist assumes that the stock price will decrease. The distance is measured as Euclidean distance in product feature space. If the price is predicted to increase in the future, it tries to buy, and bids $p_t(1+\mathrm{margin})$. Otherwise, it tries to sell with an offer of $p_t(1-\mathrm{margin})$.

One might ask if this is a realistic model of the way some financial traders behave. We would argue that the recent "dot-com" phenomenon demonstrates exactly this kind of effect. For example, Subramani and Walden (2001) have shown evidence for positive cumulative abnormal returns to shareholders following immediately after

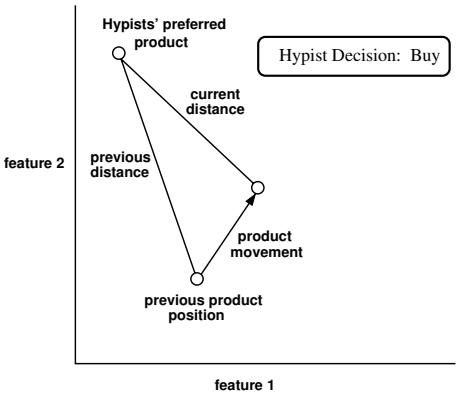

Figure 8: Hypist decision making. If the product moves closer to the hypist's preferred product, it will buy. If it moves away, it will sell.

"e-commerce" announcements by publicly-traded firms during the period October 1, 1998 to December 31, 1998. We would argue that some of these e-commerce investors were not simply trend-followers, but believed that producing e-commerce products added to the fundamental underlying value of the company.

5.2 Simulation Results

To investigate the effect of alternative valuations on price inflation and firm behavior, we ran three sets of simulation. In the first set the production firms ignored their stock price when making decisions ($\alpha_{\text{sp}} = 0$). In the second, the financial market did influence reward ($\alpha_{\text{sp}} = \alpha_{\text{as}} = 0.5$), but there were no hypist traders in the market. In the third set of simulations, $\alpha_{\text{sp}} = \alpha_{\text{as}} = 0.5$, and there were hypists in the market. Each set of simulations consisted of twenty repetitions with different random initializations of consumer and hypist preferences.

Each repetition contained 2 firms, 50 traders and 200 consumers. When hypists were present, hypist product preferences were Gaussian distributed around a mean of $[0.1\ 0.9]$.

The mean profitability when firms ignored the stock market (the "no market influence" condition) was 2459.8. With no hypists in the marketplace it fell to 1861.7. With hypists in the marketplace, the mean profitability dropped further to 1461.9. Both drops are statistically significant (at the 1% level estimated with a Wilcoxon signed rank test).

This result suggests that the stock-price based reward is distracting the firm from focusing on improving profitability. However, it does not tell us whether or not the firm was ignoring profits in order to explicitly boost stock price, or if the stock information simply added noise to the decision process. We would also like to know if "hypist" product placement information was integrated into the stock price and used by firms.

We can confirm that product placement information is incorporated into the stock

price by hypists and used by firms by examining product movement decisions made during the simulations. We computed the angle between the actual direction of product movement, and the movement direction that would have been preferred by hypists. Without hypists the mean angle is 90.8. With hypists it is 87.5. This difference is significant (at the 1% level, estimated with a Wilcoxon signed rank test).

Figure 9 shows the placement of products in the last 2000 iterations of the simulation. Large circles indicate clusters of consumer preferences. The small circle at $[0.1\ 0.9]$ shows the mean hypist product preference.

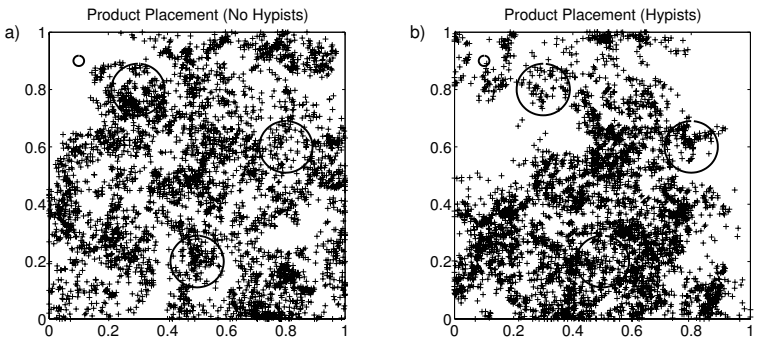

Figure 9: Product placement and movement during the last 2000 iterations of the simulations. Panels (a) and (b) show product placement relative to customer clusters and hypist preferences in a market without and with hypists. The three large circles indicate clusters of consumer preferences. The small circle at $[0.1\ 0.9]$ indicates the average hypist product preference.

When hypists are present in the market, they do not dominate product movement. The firms still try to select products that are profitable in the consumer market. However when hypists are present, many more products are located near their preferred product. That is, the firms are decoding the hypist preferences from their influence on the stock price, and attempting to satisfy their definition of "value".

So far we have seen that the stock market can successfully integrate two different views of "fundamental price", and the firms are able to extract this information from the stock price and act on it. Implicitly this shows that the firm's performance and product positioning also influenced its stock price, since stock price is the only feedback mechanism that could influence product placement. We can also look explicitly at the stock price as a function of having or not having hypists in the marketplace.

Figure 10 shows the share price of the two firms, averaged across firm and averaged across the 20 trials. First notice that the average stock price rises in all cases. This is consistent with the fact that average profitability rises over the course of the simulation. When there is no market influence, profits are higher, and stock price is similarly higher. In this case, fundamentalist traders are driving a sustained increase in the stock price.

When the profits are lower, but there are only "profit" fundamentalists and chartists in the market, the stock price is depressed. Even though the firms have an interest in

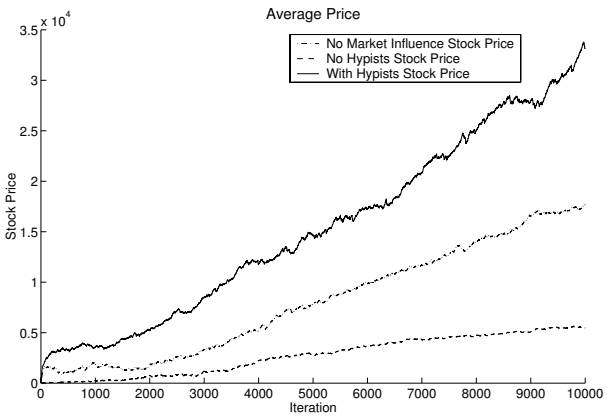

Figure 10: Average stock price with no market influence (middle line), and with (top line) and without (bottom line) hypists. In general, the market value of firms rises over the course of the simulation. However, the rate of growth in firm value is dramatically greater when hypists are present in the financial market.

boosting their stock price, they can not sustain more than a short-term stock price increase in the face of the "profit" based fundamentalists. There can be short term "bubbles" due to market dynamics, but no sustained stock price increase.

When hypists are present in the market the stock price rises much faster than in either of the other two cases, and shows a sustained increase. This indicates that the firms are explicitly sacrificing profits to boost their stock price. They can afford to do this, because the hypist traders sustain a high stock valuation despite a lack of profits. Because the firms can explicitly control product position, and only implicitly control profitability, they prefer to cater to the hypist traders rather than the fundamentalists. There are an equal number of hypists and fundamentalist in the market, but the firms sacrifice profits in order to adjust their product positioning.

5.3 Discussion

In this section we have experimented with the idea of alternative fundamental valuations as a source of sustained stock price inflation. We have shown that the artificial stock market can integrate multiple "fundamental price" signals, and that the firms can extract and use this information. Although the hypists constitute a minority of traders, they can sustain a high stock price valuation, even in the face of lower profits. Moreover, the firms prefer to cater to view of "fundamental" value that they can directly control, rather than the one that can only be increased indirectly.

These results suggest an alternative mechanism for stock bubble formation. Instead of depending on short term market dynamics, we suggest that alternative views of "fundamental price" can account for long-term price inflation. These results also suggest that boundedly rational firms will be very tempted to "game" the market, by identifying and catering to definitions of fundamental value which they can directly

control.

6 Managerial Compensation

Designing compensation for top-level managers is an important aspect of firm governance. Understanding compensation contracts is therefore of great general interest. Specifically, we focus on the role of stock options as part of a compensation contract.

Theoretical models of compensation address the problem of how to craft contracts which maximize firm value. Principal–agency models have achieved some prominence as a model of contracts, including those between managers and owners of a company (Holmström, 1979). The key feature of principal–agency models is that the principal is limited in terms of the observability of the agent's actions, effort, results, or preferences. The principal therefore delegates decision making to the agent. Alternatively, compensation models can focus on how contracts or instruments are valued by managers as opposed to unrestricted traders (see for example Hall and Murphy (2002)).

Typically, the manager in a theoretical model of compensation is assumed to be a fully rational economic actor. It has complete knowledge and understands the consequences of its actions, at least probabilistically. It can therefore assess the riskiness of alternative actions, and weigh this against possible gains in its compensation. This assumption results in analytic models for which the contract can be found which optimizes the gain by the firm's owners (Bushman and Indjejikian, 1993; Baiman and Verrecchia, 1995; Choe, 1998).

One limitation of principal–agency models is there restriction to rational agents. Empirical work seeks to address this limitation by studying the influence of compensation in actual firms. However, real firms are complicated. It is difficult to control for all variables in an empirical study. Empirical studies of compensation have led to conflicting and inconclusive results (Murphy, 1999).

Computational economic models bridge the gap between theoretical and empirical economics. On one hand, a computational model can be used to test the predictions of theory under conditions which are too complex to be addressed analytically. On the other hand, computational models can be used to give insight into complex systems and suggest new hypotheses to be tested in empirical studies. Computational models offer an environment which is complex but controlled, where all assumptions are explicitly encoded in the model.

In this section we study management compensation using the IMM. The managers of the firms in the IMM try to optimize their own compensation. Depending on their contract, they might do this by increasing profits, or by taking actions which directly boost the value of their stock-based compensation. Thus, as in principal-agency models, the problem of moral hazard still exists in the IMM. However, each manager explicitly implements a boundedly-rational agent which learns from experience, and itself has limited knowledge and computational power. We supplement the classical principal-agency framework by considering issue of limited knowledge, learning from experience, exploration versus exploitation of existing knowledge, and asymmetry between different sources of information for the decision-making agent.

Under the IMM, the role of the compensation contract is somewhat different than under a principal-agency framework. The manager does not start with intimate knowledge of the consumer or financial market. Rather, it must learn what the markets want through experience. The manager receives feedback in terms of profits and movements in stock price. These two measures give the manager two different views of firm performance. Because these two estimators are generated by two different populations of boundedly-rational agents (consumers and stock traders respectively), they do not necessarily agree.

In the next section, we use the IMM to test some predictions of empirical studies of compensation and theoretical models of contracts. Then we generate new hypotheses that can be tested empirically. In the following section section we describe the different compensation schemes examined, and simulation results.

6.1 Compensation in the Integrated Markets Model

Owners of a company delegate authority to a manager. The goal of the owners is to encourage the manager to increase the (risk-adjusted) value of their company. The goal of the manager is to maximize its compensation.

In the IMM, each manager seeks to modify its behavior so as to maximize an external payment signal. This payment takes the form of a fixed cash salary, a variable amount based on the firm's profitability, and a variable amount due to change in the value of the firm's stock.

A manager in the IMM is rewarded once in every time period, using a combination of cash and and stock-based bonuses. Specifically, the manager's compensation is based on a profit-based cash bonus, a stock grant, and a stock option.

The profit-based bonus is proportional to the profits of the firm, the stock grant bonus is proportional to the change in stock price, and the value of the stock option bonus is the change in value of the stock option held by the manager. The value of the stock option is computed using the Black–Scholes formula. It is dependent on the current stock price; the strike price; the risk-free interest rate; the volatility of the underlying stock; and the time periods until the option vests.

6.2 Risk Aversion

In order to assess the worth and riskiness of an action, managers estimate two quantities. First, they estimate the expected discounted payment they will receive after taking an action given the current world state. Second, they estimate the variance of this discounted payment, again using stochastic dynamic programming. Given this estimate of expected value and variance of value, the manager selects which action to perform based on its risk-adjusted expected discounted value. The risk penalty associated with an action is the above-average variance of the payoff associated with that action.

The reader should note that both the value and risk used by the manager are estimates, based on past experience. Unlike many analytic models, we do not assume that the manager has a priori perfect knowledge of value or risk. In fact, the quality of the estimates will be influenced by the actions taken by the manager, which in turn are

influenced by the estimates.

6.3 Simulation Results

In this section we present simulation results from the IMM. First we test some predictions of theoretical compensation models and empirical studies. We then propose new hypotheses based on simulation experiments.

Comparison to Empirical Studies

There have been a number of empirical studies of compensation, testing whether or not different compensation contracts improve firm performance. Overall, these studies have been somewhat inconclusive (Murphy, 1999). In contrast, studies of when and options are used, and their effect on market volatility have been more conclusive.

Simulations with the IMM have successfully reproduced empirical stylized facts from the compensation literature. First, it is known that stock-option grants are correlated with an increase in stock market volatility (Rajgopal and Shevlin, 2002). By manipulating the proportion of profit- and stock-based bonus, we can observe the effect on the financial market. Figure 11 shows the volatility of the simulated stock market as the proportion of stock-based bonus is increased. There is a significant trend, indicating that as the proportion of stock-based pay increases, the volatility of the market also increases.

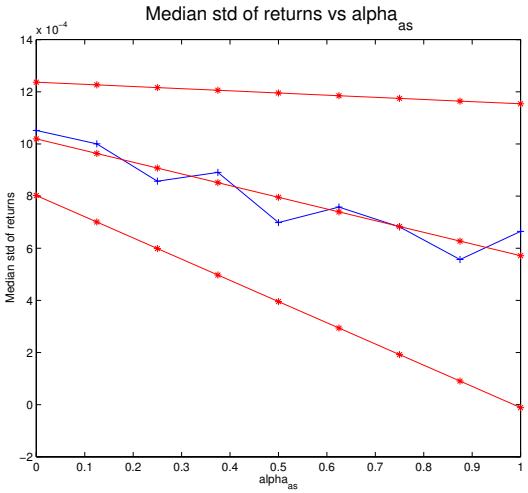

Figure 11: Stock price volatility as a function of proportion of profit-based pay ($\mathrm{alpha}_{\mathrm{as}}$). As the proportion of profit-based pay increases, the volatility of the stock price decreases. The outer lines show linear fits at the 95% confidence level.

Second, it is known that fewer stock options are granted under highly volatile market conditions (Core and Guay, 1999). We have simulated different volatility conditions by changing the mixture of different trader types in the stock market. We have

found that the optimal stock-option grant is indeed lower under more volatile market conditions. Figure 12 shows the profits obtained by the firms under two different market conditions. The first curve is for 30% chartists, and the second curve is for 70% chartists. The latter leads to a significant increase in market volatility. While the overall profits go down in the more volatile market, the peak of the curve shifts to the left, indicating that under more volatile market conditions, lower stock-based pay is optimal.

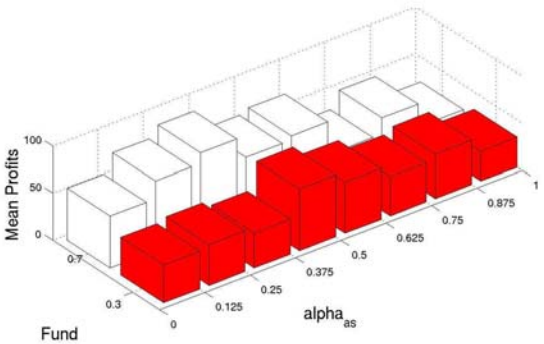

Figure 12: Firm profits under two market conditions. The "Fund" axis shows the proportion of fundamentalist traders. The axis $\mathrm{alpha}_{\mathrm{as}}$ shows the proportion of compensation coming from profit-based pay. We test two conditions, a low-volatility market (70% fundamentalists) and a high-volatility market (30% fundamentalists). In the more volatile market, the optimal pay package has more profit-based pay and less stock-based pay.

Risk Aversion and Options

Because they limit down-side risk, stock options become more valuable in volatile conditions. This is reflected in option pricing models such as Black–Scholes (Black and Scholes, 1973). Because of this, it has been theorized that executive stock options should reduce risk aversion. That is, a manager's risky actions will result in stock price volatility, increasing the value of the options. Empirical studies suggest that in some industries, stock options can lead to more risk-taking behavior (Rajgopal and Shevlin, 2002). However, there are other conflicting and inconclusive results. To add to the confusion, it is not entirely clear how executive options should be valued (Hall and Murphy, 2002), or if this mechanism is even considered when awarding options. See Murphy (1999) for a review of executive compensation.

We simulated the use of options with risk-neutral and risk-averse managers. We also used two different market conditions: one where the firms can quickly modify

their products, and one where the products can only be modified slowly. For each market condition we did three types of simulations: One with risk-neutral managers, one with risk-averse managers, and one with risk-averse managers with stock options. The results are shown in Figure 13. For the options, the option duration was 250 periods; and the options were granted slightly out of the money at 1.05 times the current stock price.

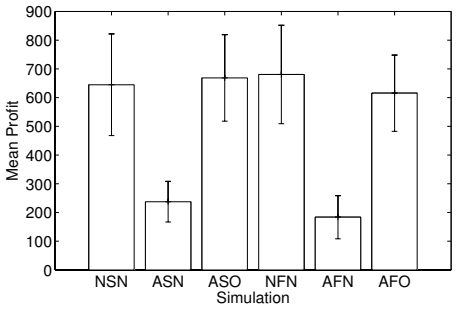

Figure 13: The effect of stock options on risk aversion. Mean and standard error of per-time-step profits are shown for six conditions: N??, risk neutral; A??, risk averse; ?S?, slow product movement; ?F?, fast product movement; ??N, no stock options; ??O, stock options. The risk-averse managers cause mean profits to drop, and the variance of the profits to be reduced. In the slow market case, options boost expected profits and profit volatility. In the fast market conditions, the options also boost profits, but not as much as in the slow market case.

In both market conditions, the risk-averse managers have much worse performance, and lower variance in their achieved profits. The effect of risk aversion is reduced by adding stock options to the manager's compensation. In the case of the slow-market condition, the average profits achieved by the firm are equivalent to the risk-neutral case. The performance with stock options in the fast-market condition is slightly inferior to the performance in the slow-market condition ($p = 0.04$).

The simulated stock options have the effect predicted by principal–agent theory. The behavior of the risk-averse manager leads to lower but more stable profits on average. The use of stock options boosts both expected profits and profit volatility.

Our purpose for doing these simulations is twofold. First, we would like to "sanity-check" our model. According to theory, awarding stock options should increase risk-taking. We would like to confirm that our managers act in the expected way. Second, theoretical models which predict this behavior make strict assumptions about the rationality of managers. Empirical studies are inconclusive. To our knowledge, this is the first attempt to validate the predictions of principle-agency theory in a boundedly-rational, learning system. The IMM tests how robust this prediction is to limitations on rationality imposed by learning, limited memory and imperfect estimation of reward and risk. The simulation results show that this behavior does indeed occur in systems with these limitations. This suggests that the risk-enhancing effect of options is quite robust.

It is interesting to note that, although the effect of stock options appears to be robust, the mechanism by which this effect occurs is quite different under our model than under theoretical models. In theory, managers know the probable outcomes of their actions. Given options, they are willing to gamble on risky outcomes, because the options insulate them from negative outcomes. In our simulation model, the managers must learn the outcomes of their actions. Given options, they are willing to try risky experiments to acquire new knowledge. This distinction between gambling with known risks, and experimenting to acquire new knowledge, suggests the simulation experiments of the next section.

Market Competition and Options

In this section we model a scenario which is difficult to address with an analytic model: How stock options influence the performance of a new competitor entering a market dominated by an incumbent firm. The incumbent has the benefit of prior experience in the market. This could also be seen as a disadvantage: The competitor does not have to overcome old habits. In this scenario, learning market preferences, and reacquiring knowledge are crucial to firm performance.

Hypothesis 1: The incumbent will have an inherent advantage because of its prior knowledge of the market.

Hypothesis 2: Options will help the incumbent, because they will promote experimentation.

Hypothesis 3: Options will help the entrant for the same reason.

We modeled three scenarios: The incumbent and entrant have no stock options, the entrant receives stock options with the incumbent having none, and both the entrant and the incumbent having stock options (see Figure 14). In the last scenario, the incumbent receives its options when the entrant arrives. All managers are risk averse, and all simulations run for 5000 iterations. The entrant enters the market at iteration 2000.

Given no options, the incumbent on average does better than the new entrant. This supports hypothesis 1. The incumbent firm does better on average than the new entrant (significant at the 1% level according to a t test). The results also support hypothesis 2. When the entrant is granted options, it does as well as the incumbent. The results of hypothesis 3 are mixed. When options are granted to the incumbent at the beginning of the simulation, it does no better than when it had options (the difference is not significant at the 5% level, according to a t test). However, when options are granted only after the new entrant appears, it does better on average than without options (significant at the 5% level, according to a t test).

This suggests that encouraging risk-taking is not enough, but rather the incumbent manager needs to be encouraged to experiment specifically after the new threat appears. This is followed by a period of new experimentation and learning, which results in the boosted profits. Figure 15 shows average profits versus time for the incumbent firm for both stock options conditions. Initially, there is no difference in performance

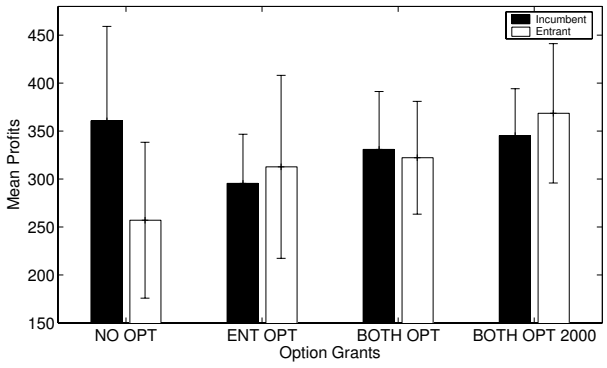

Figure 14: Results of a new entrant in the market. The bar graphs show the average per-time-step profits and standard errors for four scenarios: no stock options (NO OPT); an entrant firm with options (ENT OPT); both with options (BOTH OPT); and both with options, where the incumbent firm has options granted at the time the new entrant arrives (time $t = 2000$) (BOTH OPT 2000). The incumbent is black, and the new entrant is white. The new entrant was always granted stock options from time $t = 1$. The averages were taken over the time from the arrival of the new entrant (time $t = 2000$) to the end of the simulation (time $t = 5000$). Options granted to the entrant help it to compete with the incumbent. Options helped the incumbent when they were granted at the time that the new entrant arrived.

between the two. It is only after a period of experimentation and learning that the performance difference is seen.

We can therefore make a new prediction to be tested using empirical data: While options will help a new entrant be competitive in a new market, they will be most effective in helping the incumbent when they are granted after the new entrant appears. This suggests that re-examination of compensation is particularly important when a firm is facing new competition. Encouraging risk-taking at this time can be particularly helpful.

6.4 Discussion

In this section we have presented a computational economics model of managerial compensation. We have simulated risk-averse managers with and without stock-option compensation, and shown that the computational model confirms the predictions of principal agency theory. In particular, stock options encourage risk taking in otherwise risk-averse managers, and can boost overall profits.

This work shows that these effects are quite robust, occurring in our model in the presence of learning and incomplete knowledge. However, we also show that alternative mechanisms could explain the effect of stock options. Under our simulation model, it is not gambling on known risks which causes a boost to expected profit. Rather, profits are increased because of a willingness to experiment with new strate-

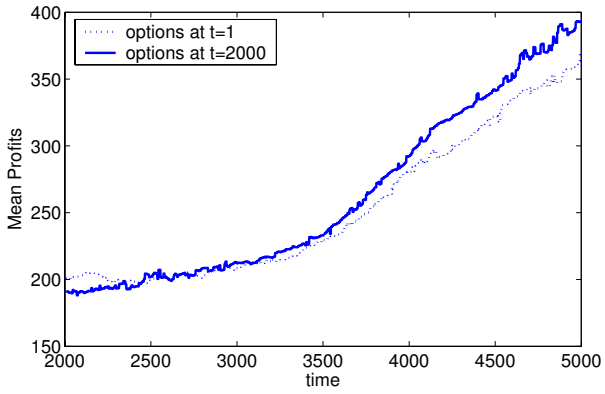

Figure 15: Effect of learning on the granting of stock options. The dotted line shows the average profits of the incumbent when options are granted from the start of the simulation. The solid line shows the average profits of the incumbent when options are granted only starting at time $t = 2000$, when the new entrant appears. After a period of learning and exploration in the new environment, the second scenario produces greater profits.

gies and learn whether or not they are effective. That is, under our model, the fact that the managers begin with incomplete knowledge is crucial to understanding the effect of options.

Prompted by this alternative mechanism, we simulated the scenario of a new entrant appearing to challenge an incumbent firm. We show that stock options can boost the competitiveness of the entrant, and also help the incumbent to fight off competition. In the latter case, the options are most effective when they are introduced as a response to the new competition. This is because the options encourage experimentation and learning in the new competitive environment. Based on these results, we have suggested new empirical studies to test the influence of stock options on knowledge acquisition.

7 Conclusions

In this chapter we have presented an integrated markets model. The model incorporates consumers, stock traders and firms. We have shown that the integrated model can reproduce a number of empirical stylized facts from both the consumer and financial markets.

Studies with the integrated model have shown that the two markets can have a large impact on one another. Product development can be influenced by the preferences of consumers, the behavior of stock traders, and the compensation given to managers. The properties of the stock market are influenced by the responsiveness of the firms, as well as the behavior of stock traders. The model has suggested alternative mechanisms for the low autocorrelations of stock return series; for prolonged periods of stock price

inflation; and for the effectiveness of stock option compensation.

We believe that our current work demonstrates the merits of simulating models of multiple markets. Feedback effects are of paramount importance in many economic situations. Gathering empirical data on these effects is a difficult challenge, because of the need to collect simultaneous data from multiple sources. Through computational simulations, we can begin to study cross-market phenomena. Using the integrated markets model as a tool, we can explore the mechanisms underlying known cross-market phenomena, and suggest new hypotheses to be tested empirically.

Acknowledgements

The authors would like to acknowledge helpful discussions with Christian Buchta, Andreas Mild, Tatiana Miazhynskaia, Martin Natter, and Rudolf Vetschera. This work was funded by the Austrian Science Fund (FWF) under grant SFB#010: "Adaptive Information Systems and Modeling in Economics and Management Science". The Austrian Research Institute for Artificial Intelligence is supported by the Austrian Federal Ministry for Education, Science and Culture and by the Austrian Federal Ministry for Transport, Innovation and Technology.

Bibliography

Argote, L. (1999). *Organizational Learning: Creating, Retaining and Transferring Knowledge*. Kluwer Academic, Dordrecht.

Arthur, W. B., Holland, J., LeBaron, B., Palmer, R., and Tayler, P. (1997). *The Economy as an Evolving Complex System II*, chapter Asset pricing under endogenous expectations in an artificial stock market, pages 15–44. Addison-Wesley, Reading, Mass.

Baier, T. and Mazanec, J. (1999). The SIMSEG project: A simulation environment for market segmentation and positioning strategies. Technical report, SFB Adaptive Information Systems and Modeling in Economics and Management Science, Vienna University of Economics and Business Administration.

Baiman, S. and Verrecchia, R. (1995). Earnings and price-based compensation contracts in the presence of discretionary trading and incomplete contracting. *Journal of Accounting and Economics*, 20:93–121.

Bertsekas, D. P. and Tsitsiklis, J. N. (1996). *Neuro-Dynamic Programming*. Athena Scientific, Belmont, Mass.

Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654.

Brock, W. and Hommes, C. (1997). A rational route to randomness. *Econometrica*, 65:1059–1095.

Brock, W. and Hommes, C. (1998). Heterogeneous beliefs and routes to chaos in a simple asset pricing model. *Journal of Economic Dynamics and Control*, 22:1235–1274.

Buchta, C. and Mazanec, J. (2001). SIMSEG/ACM: A simulation environment for artificial consumer markets. Technical report, SFB Adaptive Information Systems and Modeling in Economics and Management Science, Vienna University of Economics and Business Administration.

Bushman, R. and Indjejikian, R. (1993). Accounting income, stock price and managerial compensation. *Journal of Accounting and Economics*, 16:3–24.

Chiarella, C. and He, X. (2001). Asset pricing and wealth dynamics under heterogeneous expectations. *Quantitative Finance*, 1:509–526.

Chiarella, C. and He, X. (2002). Heterogeneous beliefs, risk and learning in a simple asset pricing model. *Computational Economics*, 19:95–132.

Choe, C. (1998). A mechanism design approach to an optimal contract under ex ante and ex post private information. *Review of Economic Design*, 3(3):237–255.

Core, J. and Guay, W. (1999). The use of equity grants to manage optimal equity incentive levels. *Journal of Accounting and Economics*, 28:151–184.

Dangl, T., Dockner, E., Gaunersdorfer, A., Pfister, A., Soegner, A., and Strobl, G. (2001). Adaptive Erwartungsbildung und Finanzmarktdynamik. *Zeitschrift für betriebswirtschaftliche Forschung*, 53:339–365.

Dawid, H., Dörner, K., Dorffner, G., Fent, T., Feuerstein, M., Hartl, R., Mild, A., Natter, M., Reimann, M., and Taudes, A., editors (2002). *Quantitative Models of Learning Organizations*, New York. Springer, Vienna.

Fama, E. and French, K. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–26.

Gaunersdorfer, A. (2000). Adaptive beliefs and the volatility of asset prices. Technical report, SFB Adaptive Information Systems and Modeling in Economics and Management Science, Vienna University of Economics and Business Administration.

Grossman, S. (1989). *The Informational Role of Prices*. MIT Press, Cambridge, MA.

Hall, B. J. and Murphy, K. J. (2002). Stock options for undiversified executives. *Journal of Accounting and Economics*, 33:3–42.

Holmström, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, 10:74–91.

Karpoff, J. (1987). The relationship between price changes and trading volume: A survey. *Journal of Financial and Quantitative Analysis*, 22:109–126.

Kim, M. and Koveos, P. (1994). Cross-country analysis of the price-earnings ratio. *Journal of Multinational Financial Management*, 4(3/4):117–127.

Levy, M. and Levy, H. (1996). The danger of assuming homogeneous expectations. *Financial Analysists Journal*, 52(3):65–70.

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21:1087–1092.

Murphy, K. J. (1999). *Handbook of Labor Economics*, volume 3, chapter Executive Compensation. North Holland, Amsterdam.

Natter, M., Mild, A., Feurstein, M., Dorffner, G., and Taudes, A. (2001). The effect of incentive schemes and organizational arrangements on the new product development process. *Management Science*, 47:1029–1045.

Pfister, A. (2003). Heterogeneous trade intervals in an agent based financial market. Technical report, SFB Adaptive Information Systems and Modeling in Economics and Management Science, Vienna University of Economics and Business Administration.

Raberto, M., Cincotti, S., Focardi, S., and Marchesi, M. (2001). Agent-based simulation of a financial market. *Physica A*, 299(1–2):320–328.

Rajgopal, S. and Shevlin, T. J. (2002). Empirical evidence on the relation between stock option compensation and risk taking. *Journal of Accounting and Economics*, 33(2):145–171.

Sallans, B., Dorffner, G., and Karatzoglou, A. (2002). Feedback effects in interacting markets. In Urban, C., editor, *Proceedings of the Third Workshop on Agent-Based Simulation*, pages 126–131. SCS-European Publishing House, Ghent, Belgium.

Sallans, B., Pfister, A., Karatzoglou, A., and Dorffner, G. (2003). Simulation and validation of an integrated markets model. *Journal of Artificial Societies and Social Simulation*, 6(4). `http://jasss.soc.surrey.ac.uk/6/4/2.html`.

Shiller, R. (1981). Do stock prices move too much to be justified by subsequent changes in dividends? *The American Economic Review*, 71:421–436.

Simon, H. A. (1982). *Models of Bounded Rationality, Vol 2: Behavioral Economics and Business Organization*. MIT Press, Cambridge, Mass.

Steiglitz, K., Honig, M., and Cohen, L. (1995). A computational market model based on individual action. In Clearwater, S., editor, *Market-Based Control: A Paradigm for Distributed Resource Allocation*. World Scientic, Hong Kong.

Subramani, M. R. and Walden, E. (2001). The impact of e-commerce announcements on the market value of firms. *Information Systems Research*, 12(2):135–154.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, Mass.

Tesauro, G. (1999). Pricing in agent economies using neural networks and multi-agent Q-learning. In *Proceedings of the International Joint Conference on Artifial Intelligence 1999, Workshop on Agents Learning About, From, and With other Agents*, Stockholm.

Tesfatsion, L. (2002). Agent-based computational economics: Growing economies from the bottom up. *Artificial Life*, 8(1):55–82.

Product Diversification in an Artificial Strategy Environment

Roland Bauer, Albert Schwingenschlögl, and Rudolf Vetschera

1 Introduction

The relationship between research on corporate strategy and economic theory has always been a delicate one (Besanko et al., 1996; Khanna et al., 2000). On one hand, in neoclassical economic theory, most problems and phenomena that characterize corporate strategy do not exist. At the focus of corporate strategy, there is the quest for sustainable, above average profit. This quest is a direct contradiction to the neoclassical equilibrium, in which only the most efficient firms survive, and even those firms have zero profit.

The fact that neoclassical economics renders strategy more or less obsolete has directed researchers on corporate strategy to focus their attention, and thus base their strategy recommendations, on the differences between the assumptions of economic theory and economic reality.

Apart from the bounded rationality of actors (Schoemaker, 1990; Amit and Schoemaker, 1993; Greve, 1998), one obvious source of such differences is the lack of perfect markets. Consequently, much of the literature on strategy can be interpreted as a search for market imperfections. This is clearly evident in the literature inspired by Porter's five competitive forces (Porter, 1998a), which describes imperfections of product markets and leads to the conclusion that a firm (or, more in line with Porter's arguments, an industry) is able to achieve sustained profits only if its product markets are imperfect. Similarly, the resource based view of strategy (Barney, 1986; Peteraf, 1993; Wernerfelt, 1984) can be seen as a quest for imperfections in factor markets.

However, in their quest to exploit the differences between the assumptions of economic theory and existing markets, strategy research with a few notable exceptions (e.g., Karnani, 1984; Khanna et al., 2000; Vining and Meredith, 2000; Bruggeman and Nuallain, 2000) has abandoned much of the analytical rigor that characterizes economic models. This development is quite natural, since most of the market imperfections that play a dominant role in strategy result from the dynamics and the complexity of the systems involved, which defy most of the analytical tools available today.

While complex, dynamic systems cannot easily be analyzed with analytical models, they still might be amenable to quantitative studies using simulation models, especially agent-based models. Such models are increasingly being used to analyze complex, dynamic systems in economics (Holland and Miller, 1991; Judd, 1997; Tesfatsion, 2000) or organization theory (Carley, 1995), which share many common traits with strategy. Thus simulation models based on agent technology might also be useful tools for analyzing corporate strategy.

Several benefits can be expected from this approach. First of all, while most of

the strategy literature commonly refers to economic concepts like productivity, demand, or preferences, most of these concepts and even more their relationships are only ambiguously defined. Incorporating them into simulation models requires a precise definition of the concepts and their interactions. An explicit, formal definition can be subject to rigorous analysis and criticism by other researchers, leading to a constant refinement of models.

The ambiguity of concepts and verbal methods of analysis leads traditional strategy research to consider only relationships between a limited number of variables, thus ignoring the very complexity that invalidates many of the assumptions of economic theory. Consequently, traditional strategy literature tends to overestimate the generalizability of results. By representing complex relationships in simulation models, we expect to be able to gain insight into the particular settings in which our results are applicable, and those settings where they are not.

By providing a precise, formal definition of concepts computer simulation models might also form an excellent docking point for empirical research. While there has always been a tendency in the strategy literature to perform empirical studies on hypotheses derived from theory, this empirical research often has led to inconclusive results. The precise definition of concepts embodied in a formal model could help to identify subtile differences in the operationalization of concepts, which are one cause of these mixed results.

In the present paper, we use an agent-based simulation model to study one central question of strategy research, the impact of diversification on corporate performance under different conditions of the competitive environment. The research goals of this paper are therefore twofold:

- From a *methodological* point of view, the aim of this paper is to study the applicability of agent-based modeling to strategy research. We want to study if agent-based models can replicate common results of strategy research, which were obtained using other methods, and possibly improve upon them.

- This potential improvement of existing results can lead to *substantive* results. Specifically, we expect that agent-based models, which require a precise and formal specification of the concepts involved, will allow us to delineate more precisely the conditions under which results are actually applicable.

The remaining part of this paper is structured as follows: in section two, we introduce the concept of diversification and review empirical evidence which leads us to the research hypotheses studied in this paper. Section three introduces the Artificial Strategy Environment used for the simulation experiments. Section four presents the experimental setup and the results. Section five concludes the paper by summarizing its main results.

2 Diversification Strategies

The composition of the portfolio of business units is one of the most important strategic questions at the corporate level. There has been a long debate in the strategy lit-

erature on the benefits of diversification into a broad range of different businesses vs. following a more focused, core-competence oriented strategy. While some authors (Keats and Hitt, 1988; Wiggins and Ruefli, 2002) found only weak empirical evidence for benefits of diversification, others argued that diversification is indeed beneficial provided that a firm diversifies into products that react differently to business cycles (Amit and Livnat, 1989), or products that use similar resources (Markides and Williamson, 1996), and when possible synergies are adequately managed (Hill, 1995). Others argued that rather than following a pure diversification or core competence strategy, firms should aim for a modest amount of diversification to find an optimum balance between reducing competition by differentiation and maintaining legitimacy by similarity to other firms (Deephouse, 1999).

The dichotomy between diversification and focused strategies was extended by Miles and Snow (Miles et al., 1978), who added the frequency of new product introductions as an additional parameter and thus distinguished four basic strategies:

- The *prospector* strategy has a strong focus on diversification and innovation. Firms following this strategy are characterized by high rates of innovation and introduce highly diversified products.
- The *analyzer* strategy is also characterized by high innovation rates, but unlike in the prospector strategy, new products are more similar and build on common core competencies.
- The *defender* strategy combines a core competence focus with low innovation rates. Firms using this strategy focus on continuous development and refinement of existing core products.
- The fourth strategy combines low innovation rates with a high degree of diversification between products. It was labelled *reactor* by Miles and Snow, who attributed the wide diversification to a lack of clear focus rather than a conscious strategic choice.

Miles and Snow further argued that these strategies would fit to different environments. A prospector strategy should be most appropriate in a dynamic, growing environment, while a defender strategy would be more suitable in a stable environment, and the analyzer strategy takes a middle position.

By considering the strategy type and its fit to the environment as the main factors influencing a firm's performance, we make two important assumptions:

Firstly, we consider "diversification" as a one-dimensional concept, which can be measured using the dispersion of a firm‘s products in feature space. In the empirical literature, a distinction is often made between the degree (extent) and the type (related vs. unrelated) of diversification. However, there is strong empirical evidence that these concepts are highly related (Palich et al., 2000; Hoskisson et al., 1993; Keats and Hitt, 1988) and thus we consider product similarity as an adequate indicator of diversification.

Secondly, by focusing on diversification and thus on current strategy, we deliberately ignore some historical facts, which might also influence a firm's performance.

One central argument of the resource based view of strategy (Wernerfelt, 1984; Amit and Livnat, 1989) is that the initial endowment of a firm with resources and the way these resources are managed over time have a considerable impact on performance. In order to isolate the impact of diversification, we consider only firms with identical initial resource endowments. However, firms in our model can follow different paths over time according to their strategy and operational decisions.

There is already considerable empirical evidence on the relationship between strategy, environment and performance, on which we can base our research hypotheses. On the one hand there are some studies which directly address the relationship between diversification and performance (Chatterjee, 1991; Bettis, 1981; Varadarajan, 1986; Amit and Livnat, 1989; Grant et al., 1988). On the other hand, there is empirical research on configurational theories that often refers to the typology of Miles and Snow (Miles et al., 1978) and therefore focuses on the match between firm strategy and its external environment (Lengnick-Hall and Wolff, 1999; Doty and Glick, 1993; Burnes, 1997; Hambrick, 1983; Segev, 1989).

Studies in the former area often produce inconclusive results. Researchers have identified three possible relationships between diversification and performance:

- In the *linear model*, an increase in diversification is assumed to (unconditionally) increase performance.

- The *u-shaped model* assumes an optimal level of diversification beyond which further diversification will reduce performance. This optimum level of diversification is often associated with the maximum of related diversification.

- In the *bounded model*, diversification ceases to have a positive impact on performance after a certain extent, but does not have a negative impact.

Similar differences exist with empirical research based on configurational theories. Doty and Glick (1993) criticize the lack of systematic empirical research on the theory of Mintzberg (1979) of five ideal types of organization which is the most prominent configurational theory. In contrast the typology of Miles and Snow seems to be moderately supported by empirical research (Doty and Glick, 1993); for a review see Zahra and Pearce (1990). Additionally, Segev (1989) suggests a wide congruence of Miles and Snow‘s strategy types with Porters Cost Leadership and Differentiation strategy.

Therefore, we base our analysis mainly on the typology of Miles and Snow, which we modify to more precisely operationalize the strategy parameters. Since Miles and Snow do not assign the reactor strategy to any type of environment and because empirical findings show that this strategy is of little relevance in practice (Slater and Olson, 2001), we only consider three distinct types of strategies:

Table 1: Strategy types

Frequency of new product introduction	Range of new products	
	Wide	Narrow
Often	Diversifier	Variator
Rarely	—	Core Competence

- diversifier,
- variator, and
- core competence firm.

Diversifying firms, which are similar to the prospector strategy of Miles and Snow, have a very high probability for developing new products and they try to launch very different products to ensure the coverage of the complete feature space as far as possible. Variators correspond to the analyzer strategy of Miles and Snow. They have the same rate of innovation, but their products are less dispersed in feature space. Core competence focused firms are exactly the opposite of diversifiers. They focus only on a small sector of the feature space and hardly ever launch a new product.

Core competence firms are the only strategy in this framework with a low innovation rate. The fourth possible combination, a low innovation rate combined with a high degree of diversification, can hardly be considered as a relevant strategy and therefore is not analyzed. Table 1 summarizes the key characteristics of the three strategy types.

Comparing the core competence strategy to the high innovation strategies, its likely advantages are lower costs for production and product development. On the other hand, a core competence focus causes firms to compete in few closely related markets and therefore they have a higher operating risk (Amit and Livnat, 1989) as markets may break away instantly due to disruption or a change in consumer preferences (D'Aveni, 1999). Thus, like in the typology of Miles and Snow, core competence firms should perform better in a stable environment, while the other two types should fit better into a more dynamic environment.To test the "fit" between strategies and environments, a performance measure is needed. In this study, we use the average operating cash flow of firms. Based on the predictions for Miles and Snow's typology, we formulate:

Hypothesis 1: In a more dynamic environment, firms following a diversification strategy will have a higher aggregated cash flow than firms following a core competence strategy. In a more static environment, this relationship will be reversed. Firms that follow a variator strategy will always perform in between the two other types.

The main advantage of a core competence strategy in a stable environment is the possibility to obtain lower costs, which can be passed on to the consumers. We therefore expect:

Hypothesis 2: Firms following a core competence strategy will have lowest unit costs and charge lowest prices followed by firms pursuing a variator and then a diver-

sification strategy.

It should be noted that hypothesis 2 is formulated without referring to different types of environment. However, lower costs will be of advantage only in stable environments, while in more rapidly changing environments, the greater flexibility of diversifiers should pay off. Their broader product portfolio should also enable them to stabilize profits even in turbulent environments. Therefore, we formulate:

Hypothesis 3: In a more dynamic environment, firms following a diversifier strategy will have a lower variation of their cash flows over time than firms following a core competence strategy, firms following a variator strategy will take a middle position.

3 The Artificial Strategy Environment

The artificial strategy environment used for our simulations provides a tool for analyzing the impact of different strategies in an environment characterized by imperfect markets and bounded rationality of actors. Each firm is represented by one agent, which makes the operational and strategic decisions for that firm. The agents are embedded in an environment, which represents both the internal technological and cost related as well as the external, market-related conditions under which the agent has to operate.

3.1 Internal Factors

Products and costs

Following earlier research (Natter et al., 2001; Krishnan and Ulrich, 2001), we characterize products by n-dimensional feature vectors. The similarity of products can thus be measured by their distance in feature space. This distance can be considered as a measure of product diversification similar to the spread in industry classification codes that is often used in the empirical literature (Amit and Livnat, 1989).

We denote a firm's product portfolio at time t by S_t. Each product $k \in S_t$ is characterized by its feature vector $f_k = (f_{k,1}, f_{k,2}, \ldots, f_{k,N})$, where N is a constant. The number of units of product k produced in period t is $x_{k,t}$.

It is possible to represent product innovations in this framework as long as these innovations concern features already contained in the vector. This does not restrict the generality of the model, since at the beginning of the simulation some attributes can be zero for all products.

One important aspect, which the cost function needs to capture is the effect of synergies when related products are manufactured using the same core competencies. To model this effect, we introduce the firm's current "focus of knowledge" E_t. E_t is a point in the feature space, its location is determined over time as

$$E_t = \lambda E_{t-1} + (1 - \lambda) F_t \tag{1}$$

where

$$F_t = \left(\frac{\sum_{k \in S_t} x_{k,s} \cdot f_{k,1}}{\sum_{k \in S_t} x_{k,s}}, \ldots, \frac{\sum_{k \in S_t} x_{k,s} \cdot f_{k,N}}{\sum_{k \in S_t} x_{k,s}} \right) \tag{2}$$

is the weighted average of features of all products manufactured in the current period. Parameter λ represents the relative importance of past vs. current experience.

Each unit of product k enters the firm's cost function with a weight of

$$c_{k,t} = 1 + \gamma_{k,t} \cdot \sqrt{\sum_{n} (f_{k,n} - E_{n,t})^2} \tag{3}$$

where $\gamma_{k,t}$ is a scaling parameter. Thus a product located exactly at the firm's "focus of knowledge" is a standard product, which is used as a reference value for manufacturing costs. The more a product differs from this (hypothetical) standard product, the higher are its unit costs.

To allow for (dis-)economies of scale, we specify the firm's total costs in period t similar to Karnani (1984) as :

$$K_t = \beta_t \left(\sum_{k \in S_t} x_{k,t} \cdot c_{k,t} \right)^{\alpha} \tag{4}$$

where β_t is a parameter denoting the overall efficiency of the firm. It should be noted that for $\alpha \neq 1$, cost function (4) makes it impossible to allocate costs correctly to products. Thus any production decision which the agent makes will necessarily be based on approximate unit costs.

In addition to these variable costs, we also consider fixed costs at two levels: fixed costs of individual products Kp_k are incurred whenever product k is manufactured at all, and corporate fixed costs Kc are always incurred. The value of Kp_k is determined as a random number when product k is introduced and remains constant over time. Kc is generated at the beginning of the simulation and is also constant.

Learning and investment effects

Over time, the cost position of a firm will change due to learning and productivity enhancing investments. In our model, this corresponds to a change in parameter β_t, which is varied over time according to the following equation:

$$\beta_t = \beta_0 \cdot \left(\sum_{\tau<t} \sum_{k \in S_\tau} \frac{x_{k,\tau}}{1 + \sqrt{\sum_n (f_{k,n} - E_{n,\tau})^2}} \right)^{\beta_1} \cdot \frac{1}{1 + \ln\left(1 + \sum_{\tau<t} I_\tau\right)} \tag{5}$$

The first term in equation (5) represents learning effects and the second the effects from productivity enhancing investments. Learning is modeled as a standard learning curve (Belkaoui, 1986) with a learning rate of $r = 2^{\beta_1}$ where $\beta_1 < 0$. By weighting the amount manufactured of each product by its distance to the focus of knowledge, we take into account that products which are dissimilar to other products also contribute less to organizational learning.

The second factor represents the effects of productivity enhancing investments I_τ in periods $\tau < t$, for which we assume decreasing returns to scale.

Short run production and marketing decisions

In each time period, the agent has to make short run decisions on the quantity and price of each product. Products can not be stored between periods.

In its planning process, the agent takes the following information into account:

- the plan of the previous period,
- a demand forecast for the next period,
- cost information,
- the total production capacity.

The plans of the previous period are stored in the agent's memory. The actual amount sold, on which the forecast is based, is determined by the market model. We do not assume that the agent has perfect information about the cost function (4), but uses a standard cost accounting system, which allocates the observable total costs to the products. The production capacity at the beginning of the simulation is exogenously given. During the simulation, the agent can increase this capacity by investments.

The production and sales plans of an agent are developed in two stages. At the first stage, the agent considers each product individually and determines a target volume and a target price for each product. Both target volume and target price are dynamically adjusted using data of the previous period, forecasts for the demand of the current period and simple heuristics. At the second stage, all products are considered simultaneously to take into account capacity restrictions. Production capacity is allocated to products based on their relative contribution margins and their target volumes.

3.2 External Factors

Product lifecycles

The external environment consists of different groups of consumers (markets). Firms do not perform a segmentation of consumers, they manufacture products with specific features and offer them to all consumers in the same way. Each product can be sold on different markets and it is possible that several products of one firm compete against each other in the same market. Products are commodities which every customer (who is part of the relevant market) purchases exactly once in every period (Adner and Levinthal, 2001).

Markets are modeled at the aggregate level, not at the level of individual consumers. Each market is characterized by a point in feature space, which represents the ideal product for a group of consumers. These ideal points are not known to firms and move in feature space. Both the probability and maximum distance of a movement are exogenously given parameters of the simulation. By changing these parameters, the experimenter can deliberately expose firms to different environmental scenarios. During the simulation, new markets are created, while existing markets may decline and eventually vanish.

A market follows a product lifecycle as it is commonly assumed in the corporate strategy literature (Porter, 1998b; Onkvisit and Shaw, 1989). We denote the maximum size of market m by g_m and the relative size at time t with respect to its maximum size by z_t.

Product lifecycles are usually defined in terms of four phases: an introduction phase characterized by slow growth, a growth phase in which sales rapidly increase, a maturity phase in which sales stabilize at a high level and a phase of decline, in which the market deteriorates.

The first three phases thus form an s-shaped curve. These phases are modeled by a differential equation of the relative size of the market as:

$$z_t' = \delta_0 \cdot z_t + \delta_1 \cdot z_t \cdot (1 - z_t) \tag{6}$$

where δ_0 represents the adoption rate of innovators and δ_1 the adoption rate among imitating users of the product (Bass, 1969).

While the parameters δ_0 and δ_1 have a convenient interpretation in terms of the diffusion process, they cannot be related directly to the duration of the product lifecycle. Solving equation (6) under the starting condition $z_0 = 0$, we obtain the time path for z_t as

$$z_t = \frac{\delta_0 \left(e^{(\delta_0+\delta_1)t} - 1\right)}{\delta_0 e^{(\delta_0+\delta_1)t} - \delta_1} \tag{7}$$

Using equation (7), the point in time at which the market reaches a certain fraction q of its ultimate size g_m can be determined and this relationship is used to calibrate the product lifecycle in the simulation.

For the last phase of the product lifecycle, a progressive decline in a product's market size is generated by the differential equation

$$z_t' = -\delta_2(1 - z_t) \tag{8}$$

Market shares

The market share submodel determines how many units of its products each firm sells on each market. The total demand observed by a firm is the aggregated demand for a product on all markets.

Corporate strategy is a meaningful concept only if there are market imperfections and markets are not in equilibrium. We therefore assume that it is possible to sell similar products at different prices in one market, although the number of items sold will be influenced by their price. The quantity of each product is also influenced by the fit of the product's features to the market's ideal point.

Consumers follow a satisficing strategy. They perform only a limited, random search among suppliers and make their purchase at the first supplier who meets their aspirations with respect to price and product features.

The aggregate demand function of market m is specified as:

$$D(p) = s_t \cdot (ln(\overline{p}) - ln(p)) \tag{9}$$

where $\overline{p}$ is a limit price and $s_t = g \cdot z_t$ represents the current size of the market.

The price asked by firm i is denoted by p_i. Without loss of generality, we assume that firms are numbered in ascending order of prices, i.e. $p_1 < p_2 < \ldots < p_I$. There are

$$v_i = D(p_i) - D(p_{i+1}) \tag{10}$$

customers who have a reservation price between p_i and p_{i+1}. These customers would buy the product from any of the firms $1, \ldots, i$. Assuming that these customers are randomly split among those firms, each firm can sell its product to

$$N_i = \frac{D(p_i) - D(p_{i+1})}{i} \tag{11}$$

customers in this segment. Thus firm i can sell at most

$$y_i = \sum_{j \geq i} \frac{D(p_j) - D(p_{j+1})}{j} \tag{12}$$

units. The actual number of product sold by firm i is then given by $min(x_i, y_i)$, since a firm cannot sell more products than it has produced.

To take into account different product features, one could consider a mismatch between a product's features and the market's ideal point as an opportunity cost to customers (Prietula and Watson, 2000). We use a different approach, which is more consistent with the demand structure defined above. Denote the Euclidean distance of the feature vector of product k to the ideal point of the market by d_k. We then define a coefficient g_k for product k as

$$g_k = \max\left(1 - \frac{d_k}{d_{\max}}; 0\right) \tag{13}$$

where $d_{\max}$ is a model parameter which represents the maximum distance a customer will accept. Features at this distance correspond to a "functionality threshold" (Adner and Levinthal, 2001) for the product. Customers in each segment are split in proportion to g_k. Thus the number of potential buyers of product k in segment i is given by

$$N_i \cdot \frac{g_k}{\sum_{\kappa \leq i} g_\kappa} \tag{14}$$

Substitution between markets

Whenever a new market is created, it is randomly assigned to one of two types:

- *Independent markets*, which consist entirely of new customers and do not directly affect the size or structure of existing markets.

- *Disruptive markets*, which reduce the customer base of existing markets.

A disruptive market will divert a fraction of an existing market's potential size g_m towards itself. In that case, an existing market is randomly selected and a randomly determined fraction of that market's total size is transferred to the new market. It should be noted that this diversion takes place at the level of potential, not actual customers. Thus the actual loss of sales will depend on the stage of the affected market as described by variable z_t. By increasing the probability of disruptive markets, the experimenter can create a more turbulent environment for the firms.

3.3 Cash Flow and Investment

From a firm's sales revenues and costs, the cash flow can be determined. Free cash flows are used by the firm to implement its strategy. Specifically, they can be used to

- introduce new products,
- extend production capacity, or
- improve the efficiency of production.

The frequency of new product introductions is determined by the strategy of an agent.The following algorithms are used in the model by the three strategy types to position products.

For the initial product portfolio, core competence focused firms and variators launch their products between the two markets that have the smallest distance in feature space. So they have a chance to compete in at least two markets with their narrow product portfolio. Diversifiers determine the extreme consumer preferences that exist in the markets and spread their products evenly between these coordinates.

In positioning new products during the simulation, variators take the past product performance into account. Features of new products are positioned in the direction of the two most successful current products. This way they adapt their product portfolios to consumer preferences without increasing dispersion.

Core competence focused firms use the same algorithm for positioning a new product, but have a much lower innovation rate. Diversifiers deliberately try to increase dispersion and calculate the direction that allows them to move as far as possible away from their current knowledge focus.

When firms do not develop new products, they can use their free cash flows to invest into additional capacity or productivity improvements. For these investment decisions, firms calculate the impact of their investment alternatives on the total cash flow for the next period and select the investment which will lead to the highest revenues.

4 Simulation Experiments and Results

To test the hypotheses formulated above, the four environmental settings shown in Table 2 were analyzed. In each experiment, 6 firms compete against each other, where two firms each follow the diversifier, variator and core competence strategies. The firms retain their strategies throughout the experiments.

Table 2: Types of experiments

Nr.	*Market Life Cycles*	*Shift in Consumer Preferences*	*Market Size*
1	long	slow	small
2	long	slow	large
3	short	fast	small
4	short	fast	large

For each setting, 100 experiments of 250 time periods were run. To take into account start-up effects of the model, the first 20 periods of each run were discarded from the following analyses.

The simulation results indicate that the different strategies obtain the expected distribution of products in feature space. Figure 1 shows the distribution of product features obtained for the diversifier and core competence strategies at the end of one representative experiment. Figure 2 shows that diversifiers shift their knowledge focus much faster and thus can react better to changing customer requirements than core competence firms.

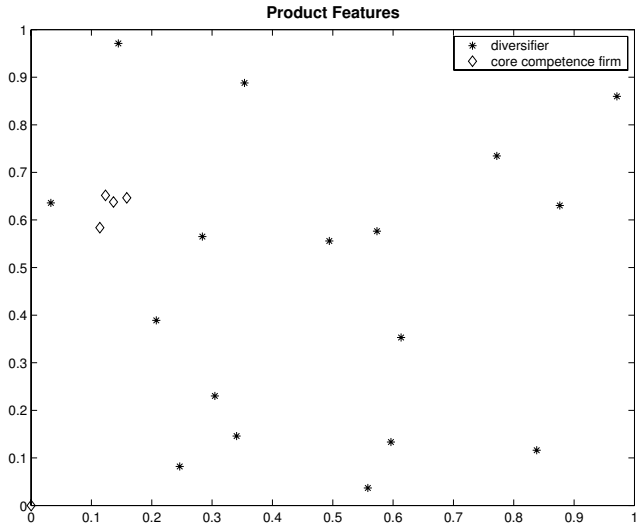

Figure 1: Product features of different firm types

Hypotheses 1 referred to the performance of the different strategy types. Figure 3 shows box plots of the cumulated cash flows of the different strategy types in the four environments in the 100 experiments. The figure clearly indicates that there are performance differences between firms, and the ranking of strategy types is dependent on the type of environment.

In the more stable environments, core competence firms perform better, while in

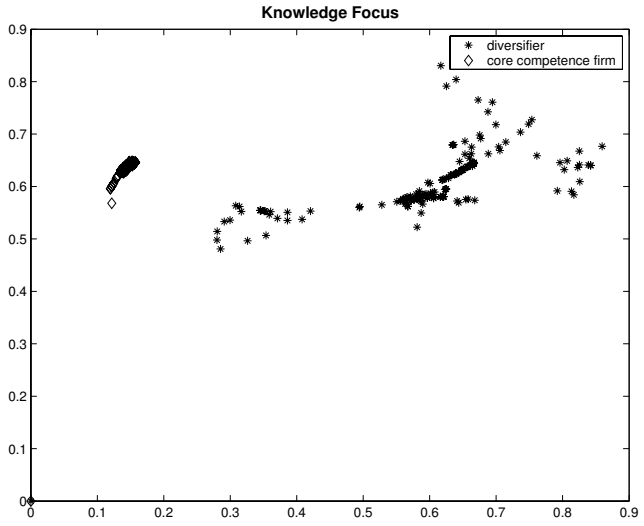

Figure 2: Movement of knowledge focus over time

the other two environments diversifiers are superior. Our results also indicate that this effect is moderated by market size. The performance advantage of core competence firms in stable environments is more accentuated when the total market size is small. On the other hand, market size has less influence on the advantages of diversifiers in dynamic environments.

To test the statistical significance of these results, a nonparametric Wilcoxon signed rank test was used since the observed average cash flows do not fulfill normality assumptions. The results shown in Table 3 confirm that the observed differences are statistically significant.

As expected, performance of the variator strategy is in between that of the diversifier and the core competence strategies, except for the stable environments in experiments 1 and 2.

If one compares the competitive situation in this environment setting with the other experiments, this outcome is not very surprising. In the dynamic environments of experiments 3 and 4, the variator clearly outperforms the core competence focused firm because of its ability to adapt the product portfolio to changing consumer preferences. In experiments 1 and 2, these advantages do not exist due to stable market conditions. Moreover, the variator launches its initial products the same way as the core competence focused firm. This causes the variator to compete against the core competence focused firm in terms of prices and increases the disadvantage of its weak cost position. In experiment 2, capacity restrictions prevent core competence firms from capturing the entire market, which leaves more opportunities for variators.

Hypothesis 2 predicted that core competence firms would achieve lower costs, leading to lower prices for their products. Figures 4 and 5 clearly confirm this hy-

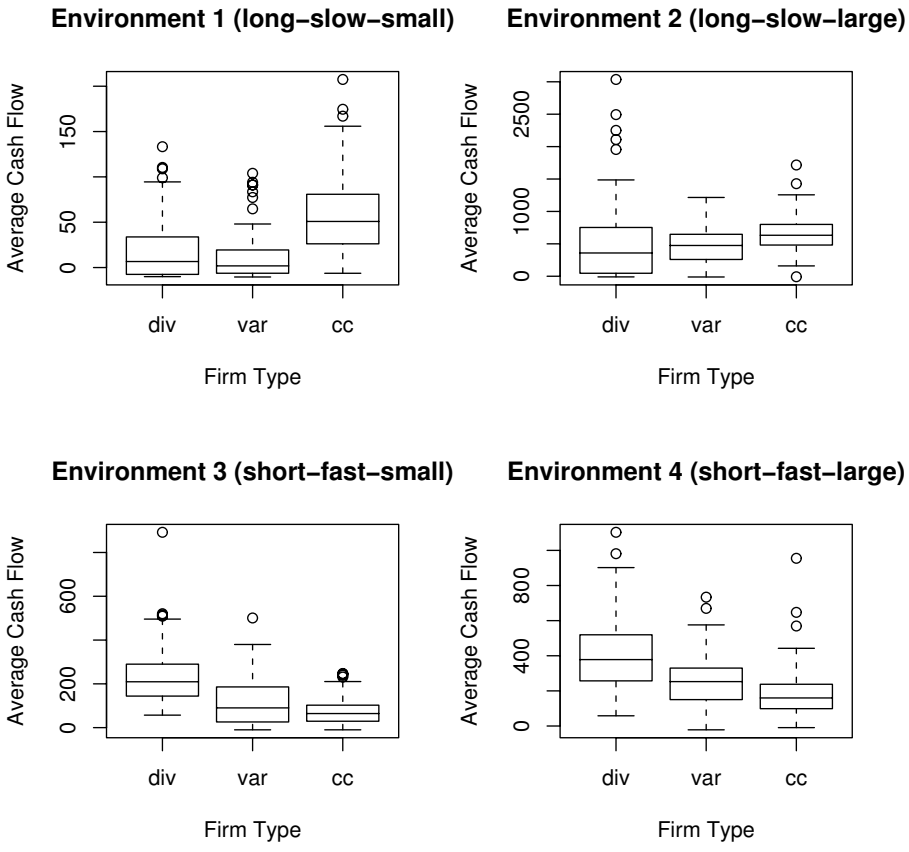

Figure 3: Average Cash Flow of different strategy types in different environments (100 runs, average over periods 21–250)

Table 3: Wilcoxon tests for cash flow differences

Exp	*Div > Var*	*Div > CC*	*Var > CC*	*Median*
1	$V+ = 3010$ $V- = 2040$ $p = 0.048$	$V+ = 743$ $V- = 4307$ $p = 1$	$V+ = 159$ $V- = 4891$ $p = 1$	$m_{div} = 6.626$ $m_{var} = 1.821$ $m_{cc} = 50.829$
2	$V+ = 2407$ $V- = 2643$ $p = 0.658$	$V+ = 1529$ $V- = 3521$ $p = 0.9997$	$V+ = 1102$ $V- = 3948$ $p = 1$	$m_{div} = 359.120$ $m_{var} = 474.856$ $m_{cc} = 632.308$
3	$V+ = 4327$ $V- = 723$ $p < 0.001$	$V+ = 4873$ $V- = 177$ $p < 0.001$	$V+ = 3505$ $V- = 1445$ $p < 0.001$	$m_{div} = 209.4594$ $m_{var} = 90.16545$ $m_{cc} = 64.3703$
4	$V+ = 4275$ $V- = 775$ $p =< 0.001$	$V+ = 4612$ $V- = 438$ $p < 0.001$	$V+ = 3634$ $V- = 1416$ $p < 0.001$	$m_{div} = 378.2$ $m_{var} = 252.773$ $m_{cc} = 159.7932$

pothesis. Core competence focused firms have the lowest unit costs and charge lowest prices, followed by variators and diversifiers.

This cost leadership strategy of core competence firms seems to be less successful in dynamic environments. But a sensitivity analysis performed on our parameter settings indicates that this is not always the case. The relative performance of the different strategy types in a dynamic environment is strongly influenced by the reservation prices of consumers.

Figure 6 illustrates the effects of low reservation prices on environment 3, i.e. a small market with rapid innovation cycles and fast movement of consumer preferences. In the standard parameter setting for environment 3, the reservation prices are uniformly distributed between 3 and 100 whereas in the additional experiment they vary only between 3 and 15.

Increasing the price sensitivity of consumers ("Price-Competition" in Figure 6) leads to a reversal in the order of firm performance. This effect can be explained by the higher costs of variators and diversifiers. In this setting only core competence firms manage to fulfill the demand of consumers for low-cost products.

The relationship between reservation price and performance is illustrated in Figure 7. For low reservation prices, core competence firms are able to adjust very well to the environment and increase their cash flows rapidly with increasing reservation prices. But after a certain threshold is reached and consumers become less concerned with prices, their performance ceases to improve. On the other hand, diversifiers and variators exhibit a more regular relationship between reservation prices and performance.

In accordance with the standard argument of strategy literature, hypothesis 3 predicted that diversifiers would have more stable cash flows than less diversified firms.

However, Figure 8 indicates exactly the opposite situation: diversifiers exhibit the highest variance of cash flows, followed by variators, and core competence firms have the most stable cash flows.

But a more detailed analysis of results reveals that it is misleading to consider vari-

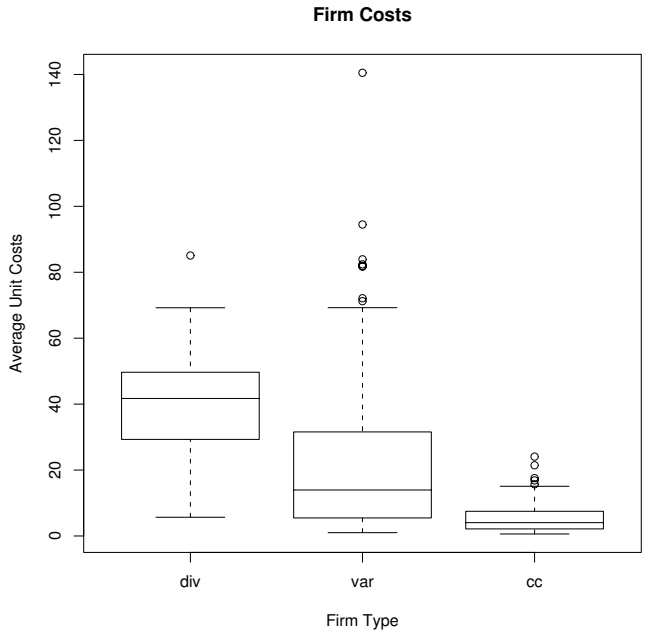

Figure 4: Costs by strategy type

ations in cash flow as a risk that should be avoided. Figure 9 shows the development of cash flows over time for the three strategy types in a representative run. While the core competence firms indeed exhibit the lowest variation in cash flows, their cash flows are also low compared to diversifiers. Core competence firms in this run tend to focus on a small and stable niche market, while diversifiers are able to quickly exploit any new opportunities emerging in the dynamic environment. This leads to more frequent changes in their cash flows, but often these changes are increases, not decreases. Thus, a high variance in this case is an indicator of opportunities, not risks.

5 Conclusions and Further Research

In this paper, we have used an agent-based computational model to study a classical question of strategy research, the impact of diversification strategies on firm performance. Starting from the typology of Miles and Snow, we defined three prototypical diversification strategies and studied their performance in different environments.

The results from this exercise on one hand show the viability of agent-based models as an instrument of strategy research. Our simulations to a significant extent confirmed existing results of the strategy literature. This correspondence of results can be seen as a validation of our model.

But the results go beyond those of traditional strategic analysis and thus indicate

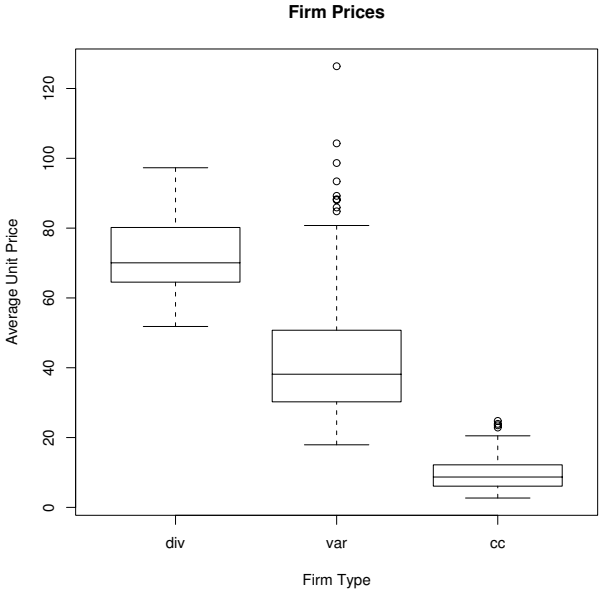

Figure 5: Prices by strategy type

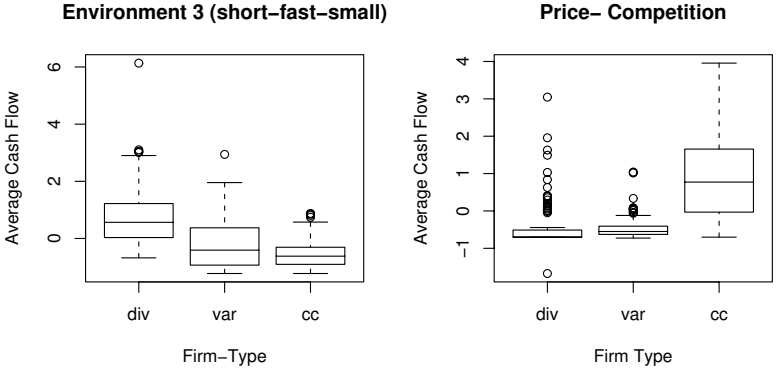

Figure 6: Price Competition (Z-Values of Average Cash Flows)

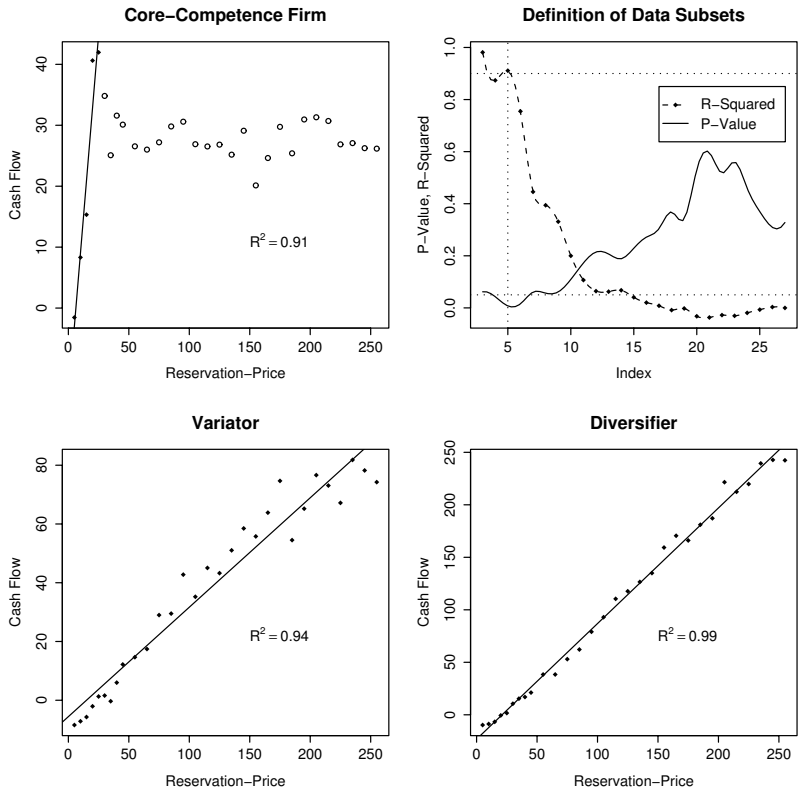

Figure 7: Price-Performance Development

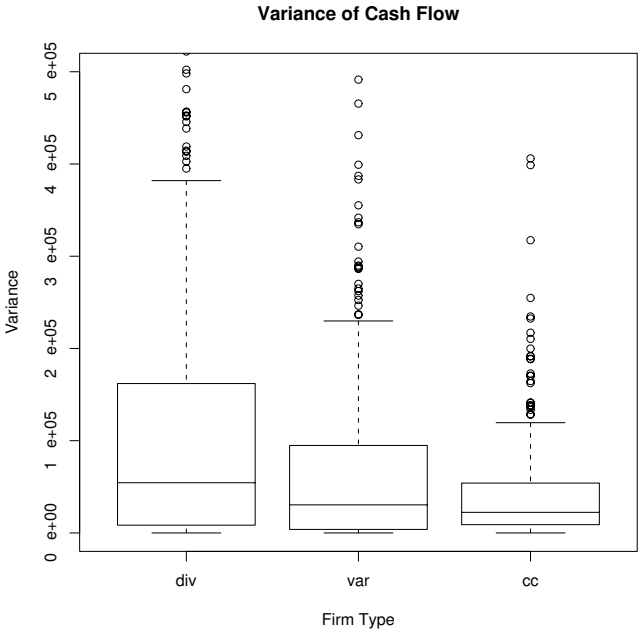

Figure 8: Cash Flow Variances

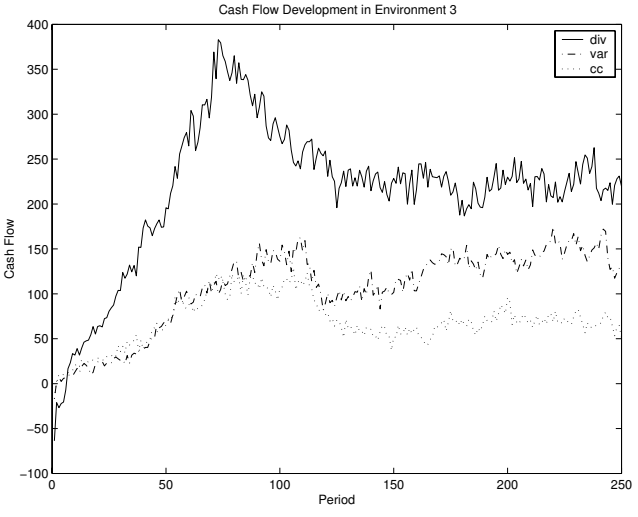

Figure 9: Performance Development in Environment 3

the potential of agent-based models to obtain new results in this field. We have identified market size as a moderating variable, which mitigates the performance advantage of core competence strategies in stable environments. Perhaps even more interesting is the result that the commonly assumed performance advantage of highly diversified firms in dynamic environments might cease to exist if the market is characterized by strong price competition. This effect could explain the current wave of divestitures one can observe in the economy.

The main purpose of this simulation exercise was to serve as a feasiblity study for the use of agent-based models in strategy research. Thus, more and richer applications of this methodology can be expected in the future. The present study has identified several promising areas.

It has indicated that results taken for granted in the existing strategy literature are in fact dependent on specific parameter settings and might be weakened or even eradicated in different settings. Thus one important research task is to more precisely delineate parameter ranges for which given results hold. This will require a systematic analysis of parameter space as well as the inclusion of additional dimensions (like the number of competing firms) in the model. These extensions can lead to more substantial contributions of agent-based modeling to research on corporate strategy.

Bibliography

Adner, R. and Levinthal, D. A. (2001). Demand heterogeneity and technology evolution: Implications for product process innovations. *Management Science*, 47:611–628.

Amit, R. and Livnat, J. (1989). Efficient corporate diversification: Methods and implications. *Management Science*, 35:879–896.

Amit, R. and Schoemaker, P. J. (1993). Strategic assets and organizational rent. *Strategic Management Journal*, 14:33–46.

Barney, J. B. (1986). Strategic factor markets: expectations, luck, and business strategy. *Management Science*, 32:1231–1241.

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

Belkaoui, A. (1986). *The Learning Curve: A Management Accounting Tool.* Quorum Books, Westport, London.

Besanko, D., Danove, D., and Shanley, M. (1996). *The Economics of Strategy*. Wiley, New York.

Bettis, R. (1981). Performance differences in related and unrelated diversified firms. *Strategic Management Journal*, (4):379–393.

Bruggeman, J. and Nuallain, B. O. (2000). A niche width model of optimal specialization. *Computational and Mathematical Organization Theory*, 6:161–170.

Burnes, B. (1997). Organizational choice and organizational change. *Management Decision*, 35:753–759.

Carley, K. (1995). Computational and mathematical organization theory: Perspective and directions. *Computational and Mathematical Organization Theory*, 1:39–56.

Chatterjee, S. (1991). The link between resources and type of diversification: Theory and evidence. *Strategic Management Journal*, 12:33–48.

D'Aveni, R. A. (1999). Strategic supremacy through disruption and dominance. *Sloan Management Review*, 40:127–134.

Deephouse, D. L. (1999). To be different, or to be the same? It's a question (and theory) of strategic balance. *Strategic Management Journal*, 20:147–166.

Doty, J. and Glick, W. (1993). Fit, equifinality, and organizational effectiveness: A test. *Academy of Management Journal*, 36:1196–1250.

Grant, R., Jammine, A., and Thomas, H. (1988). Diversity, diversification, and profitability among British manufacturing companies. *Academy of Management Journal*, 31:333–346.

Greve, H. R. (1998). Managerial cognition and the mimetic adoption of market positions: What you see is what you do. *Strategic Management Journal*, 19:967–998.

Hambrick, D. (1983). Some tests of the effectiveness and functional attributes of Miles and Snow's strategic types. *Academy of Management Journal*, 26:5–26.

Hill, C. W. (1995). Diversification and economic performance: Bringing structure and corporate management back into the picture. In Rumelt, R. P., Schendel, D. E., and Teece, D. J., editors, *Fundamental Issues in Strategy*, pages 297–321. Harvard Business School Press, Boston, Mass.

Holland, J. H. and Miller, J. H. (1991). Artificial adaptive agents in economic theory. *American Economic Review Papers and Proceedings*, 81:365–370.

Hoskisson, R., Hitt, M., Johnson, R., and Moesel, D. (1993). Construct validity of an objective (entropy) categorical measure of diversification strategy. *Strategic Management Journal*, 14:215–235.

Judd, K. (1997). Computational economics and economic theory: Substitutes or complements? *Journal of Economic Dynamics and Control*, 21:907–942.

Karnani, A. (1984). Generic competitive strategies – an analytical approach. *Strategic Management Journal*, 5:367–380.

Keats, B. W. and Hitt, M. A. (1988). A causal model of linkages among environmental dimensions, macro organizational characteristics, and performance. *Academy of Management Journal*, 31:570–596.

Khanna, T., Gulati, R., and Nohria, N. (2000). The economic modelling of strategy process: 'clean models' and 'dirty hands'. *Strategic Management Journal*, 21:781–790.

Krishnan, V. and Ulrich, K. T. (2001). Product development decisions: A review of the literature. *Management Science*, 47:1–21.

Lengnick-Hall, C. and Wolff, J. (1999). Similarities and contradictions in the core logic of three strategy research streams. *Strategic Management Journal*, 20:1109–1132.

Markides, C. and Williamson, P. (1996). Corporate diversification and organizational structure: A resource-based view. *Academy of Management Journal*, 39:340–367.

Miles, E., Snow, C., Meyer, A., and Coleman, H. (1978). Organizational strategy, structure, and process. *Academy of Management Review*, 3:546–562.

Mintzberg, H. (1979). *The Structuring of Organizations*. Englewood Cliffs, NJ: Prentice-Hall.

Natter, M., Mild, A., Feurstein, M., Dorffner, G., and Taudes, A. (2001). The effect of incentive schemes and organizational arrangements on the new product development process. *Management Science*, 47:1029–1045.

Onkvisit, S. and Shaw, J. (1989). *Product Life Cycles and Product Management*. Quorum Books, Westport, London.

Palich, L., Cardinal, L., and Chet, M. (2000). Curvilinearity in the diversification-performance linkage: An examination of over three decades of research. *Strategic Managment Journal*, 21:155–174.

Peteraf, M. A. (1993). The cornerstones of competitive advantage: A resource-based view. *Strategic Management Journal*, 14:179–191.

Porter, M. (1998a). *Competitive Advantage: Creating and Sustaining Superior Performance*. Free Press, New York.

Porter, M. (1998b). *Competitive Strategy: Techniques for Analyzing Industries and Competitors*. Free Press, New York.

Prietula, M. J. and Watson, H. S. (2000). Extending the Cyert-March duopoly model: organizational and economic insights. *Organization Science*, 11:565–585.

Schoemaker, P. J. (1990). Strategy, complexity and economic rent. *Management Science*, 36:1178–1192.

Segev, E. (1989). A systematic comparative analysis and synthesis of two business-level strategic typologies. *Strategic Management Journal*, 10:487–505.

Slater, S. and Olson, E. (2001). Marketing‘s contribution to the implementation of business strategy: An empirical analysis. *Strategic Management Journal*, 22:1055–1067.

Tesfatsion, L. (2000). Agent-based computational economics: A brief guide to the literature. Report, Iowa State University.

Varadarajan, P. (1986). Product diversity and firm performance: An empirical investigation. *Journal of Marketing*, 50:43–57.

Vining, A. and Meredith, L. (2000). Metachoice for strategic analysis. *European Management Journal*, 18:605–618.

Wernerfelt, B. (1984). A resource-based view of the firm. *Strategic Management Journal*, 5:171–180.

Wiggins, R. R. and Ruefli, T. W. (2002). Sustained competitive advantage: Temporal dynamics and the incidence and persistence of superior economic performance. *Organization Science*, 13:82–105.

Zahra, S. and Pearce, J. (1990). Research evidence on the Miles-Snow typology. *Journal of Management*, 16:751–768.

Part IV

Statistical Modeling and Software Development

Parameter Estimation and Forecasting under Asymmetric Loss

Thomas Steinberger and Lucas Zinner

1 Introduction

In this paper we assume that the process generating observed data is approximately described by an element of some parametric class of models. We are interested in forecasting and hence seek statistics, such that the distribution is near the true parameter, assuming the model is correct. But the main goal is not to find the best estimator for the model parameters, but to give a forecast which minimizes the risk under certain loss functions.

Granger (1993) stated that "asymptotically, if we believe that a particular criterion ... should be used to evaluate forecasts then it should also be used at the estimation stage of the modelling process". This is also one of the starting points of Weiss (1995) where various aspects of Grangers suggestion are discussed, especially methods are given to estimate the parameters and to produce forecasts using general cost functions. Although it seems convincing that knowledge of the cost function should be incorporated in the estimation procedure of the model parameters there are by best knowledge of the authors no rigorous proofs for this suggestion, meaning that there is no result which states explicitly that in some sense estimating the model parameters as well as evaluating the forecasts with the same criterion is superior to what we call plug-in procedure, meaning that one estimates the model parameter under the minimal variance criterion and then use the special loss function to evaluate the forecast.

There is quite a large amount of literature considering the question of optimal prediction under certain asymmetric criteria starting with the paper of Granger (1969). In Christoffersen and Diebold (1996, 1997) the optimal prediction problem under general loss structures, especially under the linlin criterion, are studied and the optimal predictor is characterized. All these papers have in common that the parameter models are assumed to be known.

It is a well known fact that the use of the least square error criterion implies that the conditional expectation is the optimal predictor in the sense that it minimizes the expected cost conditional on the information set. In the regression model under normality assumption due to the Gauss–Markov Theorem the optimal predictor for the least square error criterion is given by estimating the parameter via OLS for instance and then insert this parameter into the forecast equation. The natural question arising is, if this procedure remains optimal in case the risk should be minimized under a linlin criterion. It is a result by Koenker and G. Bassett (1978a) that in the non-Gaussian case the regression quantile is sometimes superior to the least square error criterion for the parameter estimation in the sense that it gives an asymptotically unbiased estimator which converges in distribution to a Gaussian random variable with smaller covariance. But it is not at all clear if the regression quantile minimizes the risk.

2 Concept

We consider the static system

$$y_t = g(x_t, \beta) + \varepsilon_t, \qquad t = 1, 2, \ldots, T, \tag{1}$$

where $g(\cdot)$ is a function of the exogenous variable x_t and the parameter β with bounded Jacobian with respect to β. For convenience of presentation we assume that $g(\cdot)$ contains an additive constant, meaning $g(x_t, \beta) = \beta_0 + \tilde{g}(x_t, \beta_1)$.

The disturbance ε_t is assumed to be intertemporally independently distributed with mean zero and covariance Σ. Furthermore we assume that the distribution of ε_t denoted by F is absolute continuous and has positive at least continuous density f with bounded first derivative a.s.

In addition to the assumption above we assume that the distribution is invariant under linear transformations, meaning that

$x \sim F(\mu, \sigma)$ implies that $\frac{x-\mu}{\sigma} \sim F(0, 1)$ where we denote the later by Φ with according density ϕ.

Recall that the linlin loss function L is defined by

$$L(y - x) = \begin{array}{ll} a(x - y) & x > y \\ b(y - x) & x \leq y \end{array}$$

with $a, b > 0$ and set $\frac{a}{a+b} = \theta$.

We define the risk $\mathcal{R}$ to be the expected loss, namely $\mathcal{R}(x) = \mathbb{E}(L(y_{T+1}, x))$. Note that $\mathcal{R}(\cdot)$ is strictly convex due to the assumption on the density f.

Given the true parameter of the model the optimal predictor is the unique solution of the minimization problem

$$\min_{x \in \mathbb{R}} \mathbb{E}(L(y_{T+1}, x)),$$

in the linlin case this best forecast is given by $F^{-1}(\theta)$, which is a straightforward consequence of the first order condition. We shall denote the optimal forecast by y_{opt} which means that

$$\frac{d}{dx}\mathcal{R}(x)\Big|_{x=y_{\text{opt}}} = 0.$$

Note that

$$y_{\text{opt}} = \mathbb{E}(y_{T+1}) + \sigma\Phi^{-1}(\theta) = g(x_t, \beta) + \sigma\Phi^{-1}(\theta) \tag{2}$$

where Φ is the $F(0, 1)$ c.d.f.

In many application the main problem is in the construction of *ex ante* predictions, usually the model parameters are unknown and have to be estimated. The purpose of the paper is to compare different forecasts which are known to converge to y_{opt} meaning that the following asymptotic expansion holds

$$x - y_{\text{opt}} = \frac{a_x}{\sqrt{T}} + O_P(T^{-1})$$

where a_x is a random variable with mean zero and variance Σ_x. In order to do this we take a second order Taylor series expansion of the risk around y_{opt} as $T \to \infty$. For a random variable x we get

$$\mathcal{R}(x) = \mathcal{R}(y_{\text{opt}}) + \frac{1}{2}\mathcal{R}''(y_{\text{opt}})(x - y_{\text{opt}})^2 + O_P(x - y_{\text{opt}})^3$$

since $\mathcal{R}'(y_{\text{opt}}) = 0$. Furthermore $\mathcal{R}''(x) = (a + b)f(x)$ and therefore

$$\mathcal{R}(x) = \mathcal{R}(y_{\text{opt}}) + \frac{a+b}{2} f(y_{\text{opt}})(x - y_{\text{opt}})^2 + O_P(x - y_{\text{opt}})^3$$

Though the risk turns out to be the sum of the risk taking the optimal predictor which cannot be omitted plus terms coming from parameter estimation of the model. Due to the assumption that x converges to y_{opt} for $T \to \infty$ the crucial rule in the asymptotic is played by the second term in the Taylor expansion.

Then we get

$$\mathcal{R}(x) - \mathcal{R}(y_{\text{opt}}) = \frac{a+b}{2T} f(y_{\text{opt}})\, a_x^2 + O_P(T^{-3/2}).$$

Though if $T \to \infty$ we gain that asymptotically

$$T\mathbb{E}\big(\mathcal{R}(x) - \mathcal{R}(y_{\text{opt}})\big) \;\approx\; \frac{a+b}{2} f(y_{\text{opt}})\Sigma_x \tag{3}$$

The size of the term on the right hand side above is our benchmark for different forecasts.

The two forecast we want to compare are briefly described as follows. Set $\mu = \mathbb{E}(y_{T+1})$. First, since $y_{\text{opt}} = \mu + \sigma\Phi^{-1}(\theta)$ we estimate μ respectively β and σ by standard methods, for instance OLS or Maximum Likelihood, to get $\hat{\beta}$ resp. $\hat{\sigma}$ and finally set $\hat{y} = \hat{y}_{T+1} = g(x_{T+1}, \hat{\beta}) + \hat{\sigma}\Phi^{-1}(\theta)$. This is what we call plug-in procedure. More precisely we assume that the following asymptotic expansion holds for the parameter estimators $\hat{\sigma}$ and $\hat{\beta}$, namely

$$\begin{pmatrix} \hat{\sigma} \\ \hat{\beta} \end{pmatrix} = \begin{pmatrix} \sigma \\ \beta \end{pmatrix} = + \frac{1}{\sqrt{T}} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} + O_P(T^{-1}) \tag{4}$$

where

$$\begin{pmatrix} a_0 \\ a_1 \end{pmatrix} \sim N(0, \Omega) \quad \text{with} \quad \Omega = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}.$$

Furthermore we expand the model function g around the true parameter β,

$$g(x_{T+1}, b) = g(x_{T+1}, \beta) + G(b - \beta) + O(b - \beta)^2$$

where $G' = \nabla_\beta g(x_{T+1}, \beta)$. Hence using the Mann–Wald Theorem

$$\hat{y} - y_{\text{opt}} = \frac{1}{\sqrt{T}} G a_1 + \frac{1}{\sqrt{T}} \Phi^{-1}(\theta) a_0 + O_P(\frac{1}{T}).$$

Inserting this into (3) we get

$$\mathbb{E}\big(\mathcal{R}(x) - \mathcal{R}(y_{\text{opt}})\big) = \\ = \frac{a+b}{2T} f(y_{\text{opt}}) \Big(\Phi^{-1}(\theta)\Omega_{11} + 2\Phi^{-1}(\theta) G\Omega_{12} + G\Omega_{22}G' \Big) + O_P(\frac{1}{T^{3/2}}).$$

Secondly we define $\tilde{\beta} = (\tilde{\beta}_0, \tilde{\beta}_1)'$ to be the minimizer of $\sum_{t=1}^{T} L(y_t - u_0 - \tilde{g}(x_t, u_1))$ over all u_0, u_1 in the parameter space and define the forecast $\tilde{y} = g(x_{T+1}, \tilde{\beta})$. Again we assume that

$$\tilde{\beta} = \begin{pmatrix} \tilde{\beta}_0 \\ \tilde{\beta}_1 \end{pmatrix} = \begin{pmatrix} \beta_0 + \sigma\Phi^{-1}(\theta) \\ \beta_1 \end{pmatrix} + \frac{1}{\sqrt{T}} \begin{pmatrix} b_0 \\ b_1 \end{pmatrix} + O_P(T^{-1})$$

where

$$\begin{pmatrix} b_0 \\ b_1 \end{pmatrix} \sim N\Big(0, \begin{matrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{matrix}\Big).$$

Hence

$$\tilde{y} - y_{\text{opt}} = \frac{1}{\sqrt{T}} G b_1 + \frac{1}{\sqrt{T}} \Phi^{-1}(\theta) b_0 + O_P(\frac{1}{T}).$$

and our benchmark inequality becomes

$$\mathbb{E}\big(\mathcal{R}(x) - \mathcal{R}(y_{\text{opt}})\big) = \\ = \frac{a+b}{2T} f(y_{\text{opt}}) \Big(\Phi^{-1}(\theta)\Sigma_{11} + 2G\Sigma_{12} + G\Sigma_{22}G' \Big) + O_P(\frac{1}{T^{3/2}}).$$

Proposition 2.1 *Assume that the asymptotic expansion* (4) *holds and that the estimators $\hat{\sigma}$ and $\hat{\beta}$ are efficient in the sense that Ω equals the Fisher Information matrix, then the plug-in forecast $\hat{y}_{T+1} = g(x_{T+1}, \hat{\beta}) + \hat{\sigma}\Phi^{-1}(\theta)$. is best possible with respect to our criterion.*

Proof. The proof is straightforward, since under the assumption above $\hat{y}_{T+1} = g(x_{T+1}, \hat{\beta}) + \hat{\sigma}\Phi^{-1}(\theta)$ attains the Cramer–Rao Lower Bound, see Caines (1988), p.301 for details. □

3 Location estimator

We consider the model

$$y_t = \mu + \varepsilon_t \quad \text{where } \varepsilon_t \sim F(0, \sigma).$$

We start with the OLS plug-in procedure and estimate μ and σ by

$$\hat{\mu} = \frac{1}{T}\sum_{t=1}^{T} y_t \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{T}\sum_{t=1}^{T} (y_t - \hat{\mu})^2.$$

Then it is a well know fact that

$$\begin{aligned}\hat{\mu} &= \mu + \tfrac{a_1}{\sqrt{T}} + O_P(T^{-1}) \quad \text{with } a_1 \sim N(0,\sigma^2)\\ \hat{\sigma}^2 &= \sigma^2 + \tfrac{b_1}{\sqrt{T}} + O_P(T^{-1}) \quad \text{with } b_1 \sim N(0,\sigma_4-\sigma^2) \text{ and hence}\\ \hat{\sigma} &= \sigma + \tfrac{c_1}{\sqrt{T}} + O_P(T^{-1}) \quad \text{with } c_1 \sim N(0,\Sigma)\end{aligned}$$

where $\frac{\sigma_4-\sigma^2}{4\sigma^2} = \Sigma$ and $\sigma_4 = \mathbb{E}(\varepsilon^4)$. Though the OLS plug-in forecast is given by $\hat{y} = \hat{\mu} + \hat{\sigma}\Phi^{-1}(\theta)$. Furthermore note that

$$f(y_{\text{opt}}) = f(F^{-1}(\theta)) = \frac{1}{\sigma}\,\phi(\Phi^{-1}(\theta)).$$

Inserting this into (3) we get

$$\begin{aligned}\lim_{T\to\infty} T\mathbb{E}\big(\mathcal{R}(\hat{y}) - \mathcal{R}(y_{\text{opt}})\big)&\\ = \lim_{T\to\infty} \tfrac{a+b}{2\sigma}\phi\big(\Phi^{-1}(\theta)\big)\mathbb{E}\big(\hat{\mu}-\mu+(\hat{\sigma}-\sigma)\Phi^{-1}(\theta)\big)^2&\\ = \lim_{T\to\infty} \tfrac{a+b}{2\sigma}\phi\big(\Phi^{-1}(\theta)\big)\Big(\mathbb{E}\big(\hat{\mu}-\mu\big)^2 + \Phi^{-1}(\theta)^2\mathbb{E}(\hat{\sigma}-\sigma)^2\Big)&\\ = \tfrac{a+b}{2\sigma}\phi\big(\Phi^{-1}(\theta)\big)\big(\sigma^2+\Phi^{-1}(\theta)^2\Sigma\big).&\end{aligned}$$

Here we used the fact that $\hat{\sigma}$ and $\hat{\mu}$ are asymptotically independent.

Secondly, denote by $\tilde{\mu}$ the minimizer of $\sum_{t=1}^{T} L(y_t - m)$ over all $m \in \mathbb{R}$ and define the forecast $\tilde{y} = \tilde{\mu}$. It is well known, see for instance Koenker and G. Bassett (1978b), p. 42, that $\sqrt{T}(\tilde{\mu} - y_{\text{opt}}) \sim N(0,\Omega)$ where

$$\Omega = \frac{ab}{(a+b)^2}\frac{\sigma^2}{\phi\big(\Phi^{-1}(\theta)\big)^2}.$$

Inserting this into (3) we get

$$\begin{aligned}\lim_{T\to\infty} T\mathbb{E}\big(\mathcal{R}(\tilde{y}) - \mathcal{R}(y_{\text{opt}})\big) &= \lim_{T\to\infty} \tfrac{a+b}{2\sigma}\phi\big(\Phi^{-1}(\theta)\big)\mathbb{E}\big(\tilde{y}-y_{\text{opt}}\big)^2\\ &= \tfrac{ab}{2(a+b)}\frac{\sigma}{\phi\big(\Phi^{-1}(\theta)\big)}.\end{aligned}$$

Hence the OLS plug-in procedure is superior if

$$\frac{a+b}{2\sigma}\phi\big(\Phi^{-1}(\theta)\big)\big(\sigma^2+\Phi^{-1}(\theta)^2\Sigma\big) < \frac{ab}{2(a+b)}\frac{\sigma}{\phi\big(\Phi^{-1}(\theta)\big)}. \tag{5}$$

It is straightforward to check that in the Gaussian case $\Sigma = \sigma^2/2$. Hence the inequality above can be rewritten as

$$\phi\big(\Phi^{-1}(\theta)\big)^2\big(2+\Phi^{-1}(\theta)^2\big) < \frac{2ab}{(a+b)^2} = 2\theta(1-\theta) \tag{6}$$

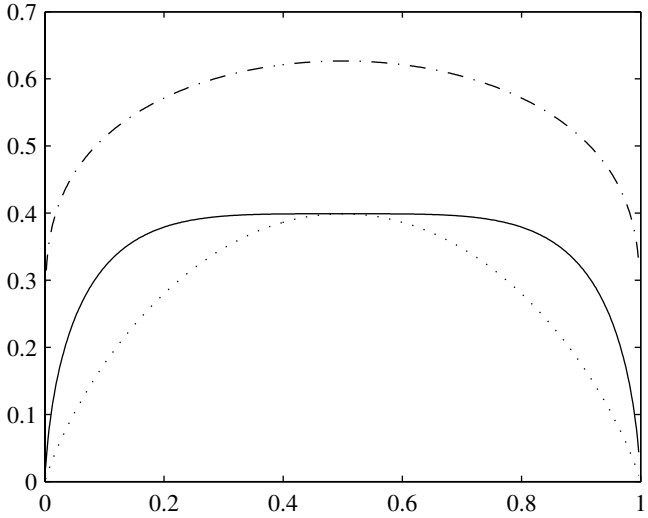

Figure 1: $\frac{1}{\sigma^2}\,\phi\big(\Phi^{-1}(\theta)\big)\mathbb{E}\big(y - y_{\text{opt}}\big)^2$ for $0 < \theta < 1$ where y is either $\hat{y}$ or $\tilde{y}$. The doted line indicates the case $y = \hat{y}$ where σ is known, the solid line $y = \hat{y}$ where σ has to be estimated as well, finally the dashdoted line gives $y = \tilde{y}$.

Figure 1 indicates that in the Gaussian case the OLS plug-in procedure is superior to direct forecast.

But we show that inequality (5) fails in the double exponential case for certain θ, more precisely when the density is defined by

$$f(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2}} \exp\Big(-\frac{\sqrt{2}|x - \mu|}{\sigma}\Big). \tag{7}$$

A straightforward calculation yields that $\Sigma = 5\sigma^2/4$ Assuming without loss of generality that $\theta < 1/2$ gives $\phi\big(\Phi^{-1}(\theta)\big) = \theta\sqrt{2}$. Now (5) becomes

$$2\theta^2\big(1 + 5\Phi^{-1}(\theta)^2/4\big) < \theta(1 - \theta) \tag{8}$$

Of course the left hand side of (8) is bounded from below by $2\theta^2$. Hence the inversed inequality in (8) holds at least if $1/3 \leq \theta \leq 2/3$. Let us finally note that (8) remains true for θ close to zero or 1. Furthermore if σ is known, the second term in the bracket on the left hand side of (8) vanishes and inequality holds for $\theta = 1/3$ and $\theta = 2/3$.

The reason why in the non-Gaussian case the direct forecast $\tilde{y}$ is superior to the OLS estimator $\hat{y}$ is fact that here the OLS estimator for μ and σ are non–optimal with respect to their asymptotic behavior. Instead, knowing the distribution family given by (7) one could estimate μ and σ by Maximum Likelihood, namely

$$\hat{\mu}_L = \operatorname*{argmin}_{m \in \mathbb{R}} \sum_{t=1}^{T} |y_t - m| \;=\; \text{median}$$

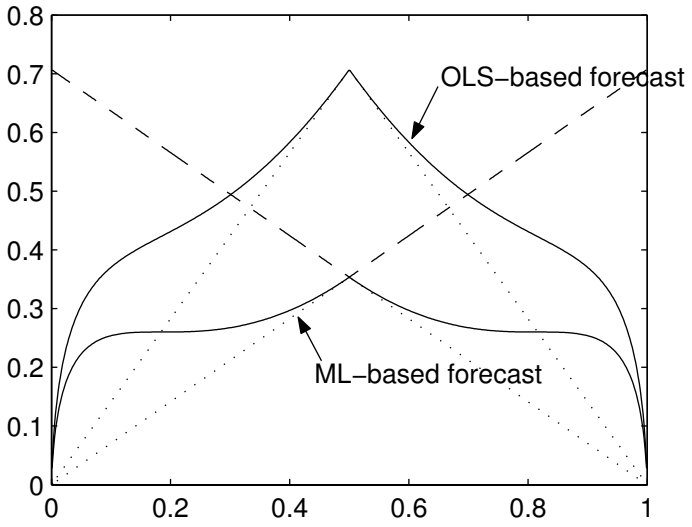

Figure 2: $\frac{1}{\sigma^2}\,\phi\big(\Phi^{-1}(\theta)\big)\mathbb{E}\big(y-y_{\text{opt}}\big)^2$ for $0<\theta<1$ for the double exponential case where y is one of the forecasts considered. Especially, the doted line and the solid line indicates the case $y=\hat{y}$ resp $y=\hat{y}_L$ where σ is either known or not, finally the dashed line gives $y=\tilde{y}$, which is seen to be superior to the OLS case for certain θ.

and

$$\hat{\sigma}_L=\frac{\sqrt{2}}{T}\sum_{t=1}^{T}|y_t-\hat{\mu}_L|.$$

Then $(\mu-\hat{\mu}_L,\sigma-\hat{\sigma}_L)\sim N(0,\Omega_L)$ where

$$\Omega_L=\left(\mathbb{E}\Big(-\frac{\partial^2\log f(x,\mu,\sigma)}{\partial\mu^i\partial\sigma^j}\Big)_{i+j=2}^{i,j}\right)^{-1}=\begin{pmatrix}\frac{\sigma^2}{2} & 0\\ 0 & \sigma^2\end{pmatrix}.$$

We note that partial derivatives of $\log f(x,\mu,\sigma)$ exist a.e. with respect to F. It is straightforward to see that the crucial benchmark-inequality (6) with $\hat{\mu}_L$ instead of $\hat{\mu}$ resp $\hat{y}_L=\hat{\mu}_L+\hat{\sigma}_L\Phi^{-1}(\theta)$ instead of $\hat{y}$ becomes

$$2\theta^2\big(1/2+\,\Phi^{-1}(\theta)^2\big)\;<\;\theta(1-\theta) \tag{9}$$

or equivalently since $\Phi^{-1}(\theta)=\frac{1}{\sqrt{2}}\log(2\theta)$ for $\theta\le 1/2$

$$\theta\left(1+\log^2(2\theta)\right)<1-\theta. \tag{10}$$

Of course, if $\theta=\frac{1}{2}$ equality holds and furthermore using ML estimators is superior to OLS. If $\theta<\frac{1}{2}$ the right hand side of (10) is strictly decreasing whereas the left hand

side of (10) is increasing. Hence the ML estimation plug-in procedure is superior to direct estimation as indicated by figure 2. Independently of θ the ML plug-in procedure is superior to OLS which can be seen by comparing the left hand side expressions in (8) and (9) also indicated by figure 2.

In fact the ML-estimation plug-in procedure is best possible. To see this recall that we have to minimize Σ_x in (3) where

$$\Sigma_x \geq (1, \Phi^{-1}(\theta))I^{-1}(\mu,\sigma)(1, \Phi^{-1}(\theta))'$$

where $I(\mu,\sigma)$ denotes the information matrix and the right hand side gives the Cramer-Rao lower bound for any estimator $x = u(y_1,\ldots,y_T)$ of $\mu + \sigma\Phi^{-1}(\theta)$, see Caines (1988), p. 301 Hence in any case where the ML–estimator achieves the Cramer-Rao lower bound the ML–estimation plug-in procedure is best possible. This is true in the Gaussian as well as in the double exponential case.

4 Linear Regression

As a second example for our method we consider a regression model, namely

$$y_t = \beta_1 + \beta_2 x_t + \varepsilon_t \quad \text{where } \varepsilon_t \sim F(0,\sigma).$$

We assume that

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}(x_t - \bar{x})^2 > 0 \qquad \text{with } \bar{x} = \frac{1}{T}\sum x_t. \tag{11}$$

We start with the OLS plug-in procedure. Therefore we estimate

$$\begin{aligned}
\hat{\beta}_1 &= \frac{1}{T}\sum_{t=1}^{T}(y_t - \hat{\beta}_2 x_t) \\
\hat{\beta}_2 &= \frac{\sum_{t=1}^{T}(y_t - \bar{y})(x_t - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2} \\
\hat{\sigma}^2 &= \frac{1}{T-2}\sum_{t=1}^{T}\hat{\varepsilon}_t^2 = \frac{1}{T-2}\sum_{t=1}^{T}(y_t - \hat{\beta}_1 - \hat{\beta}_2 x_t)^2
\end{aligned}$$

where $\bar{y} = \frac{1}{T}\sum y_t$. It is well known that

$$\mathrm{Var}(\beta_1 - \hat{\beta}_1) = \frac{\sigma^2 \sum x_t^2}{T\sum(x_t - \bar{x})^2}, \qquad \mathrm{Cov}(\beta_1 - \hat{\beta}_1, \beta_2 - \hat{\beta}_2) = \frac{-\sigma^2\bar{x}}{\sum(x_t - \bar{x})^2},$$

$$\mathrm{Var}(\beta_2 - \hat{\beta}_2) = \frac{\sigma^2}{\sum(x_t - \bar{x})^2} \quad \text{and} \qquad \mathrm{Var}(\hat{\sigma}^2) = \frac{2\sigma^4}{T}.$$

We set $\hat{\mu} = \hat{\beta}_1 + \hat{\beta}_2 x_{T+1}$. A straightforward calculation gives

$$\mathbb{E}(\sqrt{T}(\mu - \hat{\mu}))^2 = \frac{\sigma^2 \sum_{t=1}^{T}(x_t - x_{T+1})^2}{\sum_{t=1}^{T}(x_t - \bar{x})^2}$$

We set

$$\lim_{T\to\infty} \frac{\sum_{t=1}^{T}(x_t - x_{T+1})^2}{\sum_{t=1}^{T}(x_t - \bar{x})^2} = V_x$$

and note that $\hat{\mu}$ and $\hat{\sigma}$ are asymptotically independent. Thus

$$\hat{\mu} = \mu + \frac{a_\mu}{\sqrt{T}} + O_P(T^{-1}) \quad \text{with } a_\mu \sim N(0, \sigma^2 V_x)$$

and assuming normality of the errors

$$\hat{\sigma} = \sigma + \frac{a_\sigma}{\sqrt{T}} + O_P(T^{-1}) \quad \text{with } a_\sigma \sim N(0, \sigma^2/2).$$

Though the OLS plug-in forecast is given by $\hat{y} = \hat{\mu} + \hat{\sigma}\Phi^{-1}(\theta)$. Inserting this into (3) we get

$$\lim_{T\to\infty} T\mathbb{E}\big(\mathcal{R}(\hat{y}) - \mathcal{R}(y_{\text{opt}})\big) = \frac{(a+b)\,\sigma}{4}\phi\big(\Phi^{-1}(\theta)\big)\,\big(2V_x + \Phi^{-1}(\theta)^2\big).$$

Secondly, denote by $\tilde{\beta}_1, \tilde{\beta}_2$ the minimizer of $\sum_{t=1}^{T} L(y_t - b_1 - b_2 x_t)$ over all $(b_1, b_2) \in \mathbb{R}^2$ and define the forecast $\tilde{y}_{T+1} = \tilde{\beta}_1 + \tilde{\beta}_2 x_{T+1}$. Since by assumption (11) the matrix

$$Q = \lim_{T\to\infty} \frac{1}{T} \begin{pmatrix} T & \sum x_t \\ \sum x_t & \sum x_t^2 \end{pmatrix}.$$

is positive definite the result of Koenker and G. Bassett (1978b), p. 43, applies, namely that

$$\sqrt{T}\begin{pmatrix} \tilde{\beta}_1 - \beta_1 - \sigma\Phi^{-1}(\theta) \\ \tilde{\beta}_2 - \beta_2 \end{pmatrix} \sim N(0, \Omega \otimes Q^{-1})$$

where

$$\Omega = \frac{ab}{(a+b)^2}\frac{\sigma^2}{\phi\big(\Phi^{-1}(\theta)\big)^2}.$$

Furthermore since

$$\begin{pmatrix} T & \sum x_t \\ \sum x_t & \sum x_t^2 \end{pmatrix}^{-1} = \begin{pmatrix} \frac{\sum x_t^2}{T\sum(x_t-\bar{x})^2} & \frac{-\bar{x}}{\sum(x_t-\bar{x})^2} \\ \frac{-\bar{x}}{\sum(x_t-\bar{x})^2} & \frac{1}{\sum(x_t-\bar{x})^2} \end{pmatrix}$$

we get

$$\lim_{T\to\infty} T\mathbb{E}(\tilde{y}_{T+1} - y_{\text{opt}})^2 = V_x\Omega$$

Inserting this into (3) we get

$$\lim_{T\to\infty} T\mathbb{E}\big(\mathcal{R}(\tilde{y}) - \mathcal{R}(y_{\text{opt}})\big) = \frac{ab}{2(a+b)}\frac{\sigma V_x}{\phi\big(\Phi^{-1}(\theta)\big)}.$$

Again the plug-in procedure is superior if

$$\frac{(a+b)\,\sigma}{4}\phi\big(\Phi^{-1}(\theta)\big)\,\big(2V_x + \Phi^{-1}(\theta)^2\big) < \frac{ab}{2(a+b)}\frac{\sigma V_x}{\phi\big(\Phi^{-1}(\theta)\big)}.$$

or equivalently if

$$\phi\big(\Phi^{-1}(\theta)\big)^2\big(2V_x + \Phi^{-1}(\theta)^2\big) < 2\theta(1-\theta)V_x. \tag{12}$$

Completely analog to section 3 above we can state that in the double exponential case for certain θ the corresponding inequality fails, more precisely the corresponding inversed inequality holds at least if $\theta \in [1/3, 2/3]$, but remains true for θ close to zero or 1. Again if σ is known, the second term in the bracket on the left hand side of (12) vanishes and the interval where (12) fails is sharp.

The main result of Koenker and G. Bassett (1978a) implies that for distributions for which the median is superior to the mean as an estimator of location, the LAE estimator is preferable to the OLS estimator in the general model. In case of the double exponential distribution the Maximum Likelihood estimator for the parameters is exactly the LAE and in fact one gains estimators $\beta_{1\,L}$ and $\beta_{2\,L}$ for β_1 and β_2 respectively, more precisely

$$(\beta_{1\,L}, \beta_{2\,L}) \;=\; \underset{(b_1,b_2)\in\mathbb{R}^2}{\operatorname{argmin}} \sum_{t=1}^{T} |y_t - b_1 - x_t b_2|$$

and

$$\hat{\sigma}_L = \frac{\sqrt{2}}{T}\sum_{t=1}^{T} |y_t - \beta_{1\,L} - x_t\beta_{2\,L}|.$$

with

$$\sqrt{T}\begin{pmatrix} \beta_{1\,L} - \beta_1 \\ \beta_{2\,L} - \beta_2 \end{pmatrix} \sim N(0, \Omega_L \otimes Q^{-1}) \qquad \text{with } \Omega_L = \frac{\sigma^2}{4\,\phi\big(\Phi^{-1}(1/2)\big)^2}$$

which results in

$$\lim_{T\to\infty} T\mathbb{E}\big(\mathcal{R}(\hat{y}) - \mathcal{R}(y_{\text{opt}})\big) = \frac{(a+b)\,\sigma}{4}\phi\big(\Phi^{-1}(\theta)\big)\,\big(V_x + 2\,\Phi^{-1}(\theta)^2\big).$$

The benchmark inequality becomes

$$\theta\,\big(V_x + \log^2(2\theta)\big) < (1-\theta)\,V_x$$

and the same argument following inequality (10) shows that Maximum Likelihood plug-in procedure is superior to direct forecast.

Bibliography

Caines, P. E. (1988). *Linear Stochastic Systems*. Wiley, New York.

Christoffersen, P. and Diebold, F. (1996). Further results on forecasting and model selection under asymmetric loss. *Journal of Applied Econometrics*, 11:561–572.

Christoffersen, P. F. and Diebold, F. (1997). Optimal prediction under asymmetric loss. *Econometric Theory*, 13:808–817.

Granger, C. W. J. (1969). Prediction with a generalized cost of error function. *Operational Research Quaterly*, 20:199–207.

Granger, C. W. J. (1993). On the limitations of comparing mean squared forecast errors, comment. *Journal of Forecasting*, 12:651–652.

Koenker, R. and G. Bassett, J. (1978a). Asymptotic theory of least absolute error regression. *Journal of the American Statistical Association*, 73:618–622.

Koenker, R. and G. Bassett, J. (1978b). Regression quantiles. *Econometrica*, 46:33–50.

Weiss, A. A. (1995). Estimating time series models using the relevant cost function. *Journal of Applied Econometrics*, 11:1962–1968.

Identification of multivariate state-space systems

Thomas Ribarits and Manfred Deistler

1 Introduction

Apart from traditional identifiability analysis, questions of parameterization do not attain much attention in econometrics. In the linear dynamic case a major reason for this fact seems to be that mainly AR(X) models are used, where parameterization problems are simple. However, for ARMA(X) and state-space models, parameterization issues are important for the properties of model selection and estimation procedures. Identification of such models may still cause problems and the idea is that such problems can be mitigated by 'intelligent' parameterizations. Another issue connected with parameterization is the curse of dimensionality in the multivariate case.

The problem of parameterization is concerned with the relation between transfer functions and parameters. To be more precise, we consider a parameter space $T \subseteq \mathbb{R}^d$, a set of transfer functions U and a surjective mapping $\pi : T \to U$, attaching transfer functions to parameters. The first questions in this context are identifiability, i.e. injectivity of π, and for the identifiable case continuity and differentiability of the inverse mapping $\psi : U \to T$; see Deistler (2001) for details.

The paper is organized as follows: In Section 2, the main advantages and disadvantages arising from the use of ARX, ARMAX and state-space models are considered. It will be argued that some of the disadvantages of ARMAX and state-space modelling can be mitigated by the choice of appropriate parameterizations. Section 3 presents parameterizations for state-space models, including very recent parameterization approaches. Finally, Section 4 contains possible directions for future research activities.

2 ARX, ARMAX and State-Space Systems

We consider a linear dynamic relation of the form

$$y_t = \sum_{j=0}^{\infty} L_j u_{t-j} + \sum_{j=0}^{\infty} K_j \varepsilon_{t-j} \quad L_j \in \mathbb{R}^{s\times m}, K_j \in \mathbb{R}^{s\times s} \tag{1}$$

where y_t denotes the s-dimensional observed output, u_t denotes the m-dimensional observed input and the ε_t form the white noise innovations process with $\Sigma = \mathbb{E}\varepsilon_t\varepsilon_t'$; note that $K_0 = I_s$. Furthermore, $(u_t|t \in \mathbb{Z})$ is assumed to be uncorrelated with $(\varepsilon_t|t \in \mathbb{Z})$. Thus, $\sum_{j=0}^{\infty} L_j u_{t-j}$ is the best linear approximation of y_t by $(u_t|t \in \mathbb{Z})$.

Relation (1) can more conveniently be written as

$$y_t = l(z)u_t + k(z)\varepsilon_t \tag{2}$$

where $l(z) = \sum_{j=0}^{\infty} L_j z^j$ is the transfer function from u_t to y_t and $k(z) =$

$\sum_{j=0}^{\infty} K_j z^j$ is the transfer function from ε_t to y_t. Note that z denotes a complex variable as well as the backward shift operator.

In most cases it is assumed that $l(z)$ and $k(z)$ are rational, and in this case they can be represented by a 'common denominator matrix' $a(z)$ as $l(z) = a^{-1}(z)d(z)$ and $k(z) = a^{-1}(z)b(z)$, where $a(z) = \sum_{j=0}^{p} A_j z^j$, $d(z) = \sum_{j=0}^{r} D_j z^j$ and $b(z) = \sum_{j=0}^{q} B_j z^j$ are polynomial matrices of suitable dimensions. Equation (2) can then be rewritten in ARMAX representation

$$a(z)y_t = d(z)u_t + b(z)\varepsilon_t \tag{3}$$

It is very common to choose $b(z) = I_s$ in (3), resulting in an ARX model of the form

$$a(z)y_t = d(z)u_t + \varepsilon_t \tag{4}$$

An alternative approach is to rewrite (2) is in terms of a state-space representation. Starting from (3), the following can easily be shown: Assume w.r.o.g. that $a(0) = A_0 = I_s$ and put $x_t = (y_{t-1}, \ldots, y_{t-p}, u_{t-1}, \ldots, u_{t-r}, \varepsilon_{t-1}, \ldots, \varepsilon_{t-q})'$. Then (3) can immediately be written in terms of the (in general) nonminimal state-space system

$$\begin{aligned} x_{t+1} &= Ax_t + Bu_t + K\varepsilon_t \\ y_t &= Cx_t + Du_t + \varepsilon_t \end{aligned} \tag{5}$$

where the matrices A, B, C, D and K are composed of the blocks $-A_1, \ldots, -A_p$, $D_0, \ldots, D_r$, $B_0, \ldots, B_q$, I_s, I_m and many zero matrices. This nonminimal state-space system can always be reduced to a minimal one, where x_t has minimal dimension, n say, among all state-space representations of (2). From now on we will thus assume that x_t in (5) denotes the n-dimensional state and that $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{s\times n}$, $D \in \mathbb{R}^{s\times m}$ and $K \in \mathbb{R}^{n\times s}$ are parameter matrices.

In the sequel we will always assume that (l, k) satisfies the stability condition, i.e. (l, k) has no pole for $|z| \leq 1$. In addition, we assume that k satisfies the minimum phase condition, i.e. k has no zero for $|z| < 1$; for an appropriate definition of poles and zeros of rational matrices see e.g. chapter 2 in Hannan and Deistler (1988). We always assume that $\Sigma > 0$.

Let U_A be the set of all rational and causal $s \times (m + s)$ transfer functions (l, k) where (l, k) is stable, $k(0) = I$ and k satisfies the minimum phase condition. Identification in this context is mostly considered in a *semi-nonparametric framework* meaning that in a first step a suitable model subclass of the (infinite dimensional) model class U_A is determined by a data driven model selection procedure. Each of these subclasses is described by a finite dimensional parameter space. In a second step, the real valued parameters corresponding to the subclass are estimated by 'parametric' procedures.

Despite the fact that identification of linear dynamic systems is a quite mature subject by now, there are still some practical problems in applying identification routines for ARMAX and state-space models. As a consequence, in many applications ARX modelling is still preferred. The reasons for this fact are the following:

1. In ARX modelling the structure of the parameter spaces is much simpler than in the case of ARMAX or state-space models. The $ARX(p, r)$ model class, where p and r denote the prescribed *maximum* degrees of $a(z)$ and $d(z)$, respectively, can be parameterized by simply considering all entries in the coefficient matrices of $a(z)$ and $d(z)$ as free parameters, if no stability assumption has been imposed. By imposing the stability assumption $\det a(z) \neq 0$, $|z| \leq 1$, we obtain an open subset of this parameter space. Note that in particular every point in the parameter space is identifiable. For state-space or ARMAX models, however, the parameterization problem is much more involved and the parameter spaces may have a quite complicated structure. Additionally, commencing e.g. from an identifiable parameterization of a state-space model of state dimension n, models of state dimension $\bar{n} < n$ are in general not identified.

2. ARX modelling may be preferable because MLEs coincide with least squares estimates. Hence, the MLE is given explicitly and least squares formulas are robustly implemented in any standard piece of software. For state-space or ARMAX models, however, in general the MLE has to be determined by using more involved and less reliable iterative numerical optimization algorithms. In contrast to the ARX case, a major additional problem is that the likelihood function may have many local optima and that conventional optimization algorithms are prone to converge to these local optima if the starting values are chosen inappropriately.

On the other hand, the more general class of state-space or ARMAX models clearly allows for a more flexible and parsimonious parameterization, resulting in less parameters in a number of applications. Note again that every rational and causal transfer function can be described by an ARMAX or by a state-space system, and in this sense both approaches are equivalent.

Given the greater flexibility of ARMAX and state-space systems, the improvement of identification procedures for these model types is still an important task, in particular with respect to the issues raised in the list above. A key issue for this improvement is the choice of appropriate parameterizations, which should ideally be of a simple structure (1) and lead to numerically well conditioned estimation procedures (2).

In the sequel, we restrict ourselves to parameterization problems for state-space models.

3 Parameterizations of State-Space Systems

In the sequel we assume that a particular model subclass of U_A has already been determined by means of some model selection procedure (such as information criteria or test procedures). Of course, there are many ways in which one can break U_A into subclasses, and we will consider the particular case where the subclasses are $M(n) \subset U_A$, denoting the subset of all transfer functions of order n. It is well known that each transfer function of order n can be represented in terms of a minimal state-space realization of the form (5), where the state dimension is equal to n. Conversely, every

minimal state-space system (5) of state dimension n corresponds to a transfer function in $M(n)$.

Let $S(n)$ denote the set of all state-space systems (A, B, C, D, K) for fixed s, m and n satisfying the stability condition

$$|\lambda_{max}(A)| < 1 \tag{6}$$

and the minimum phase condition

$$|\lambda_{max}(A - KC)| \leq 1 \tag{7}$$

where λ_{max} denotes an eigenvalue of maximal modulus. Let $S_m(n) \subset S(n)$ denote the subset of all minimal $(A, B, C, D, K) \in S(n)$. The mapping

$$\pi : S_m(n) \rightarrow M(n) \tag{8}$$

$$(A, B, C, D, K) \mapsto zC(I - zA)^{-1}(B, K) + (D, I) \tag{9}$$

is surjective, but not injective. This is because the classes of observationally equivalent minimal state-space systems (A, B, C, D, K) are characterized by linear nonsingular state transformations, i.e. all state-space systems in the set $\mathcal{E}(A, B, C, D, K) = \{(TAT^{-1}, TB, CT^{-1}, D, TK), \det(T) \neq 0\}$ are mapped onto the same transfer function by π and, conversely, if $(l, k) = \pi(A, B, C, D, K)$, then the entire inverse image $\pi^{-1}(l, k) \subset S_m(n)$ is given by $\mathcal{E}(A, B, C, D, K)$.

Of course, there are a number of desirable properties of a parameterization; see Deistler (2001). In particular, continuity is important: As is well known, $M(n)$ is a real analytic manifold of dimension $2ns + m(n + s)$ (with boundary points) which cannot be continuously parameterized with one coordinate mapping ψ in the MIMO case. We now discuss the following parameterizations commencing from $M(n)$:

- *Full state-space parameterization*. Here, $S_m(n)$ is used as a parameter space, i.e. all entries in minimal state-space matrices $(A, B, C, D, K) \in S_m(n)$ are considered as parameters; see e.g McKelvey (1995). Note that $S_m(n)$ is open and dense in $S(n)$. Of course in this situation we do not have identifiability; for $(l, k) \in M(n)$ the classes of observational equivalence $\mathcal{E}(A, B, C, D, K)$ in $S_m(n)$ are manifolds of dimension n^2. For criteria functions which are constant along equivalence classes then of course the optimum is not unique. In other words, this approach is particularly simple, but has the drawback that there are n^2 essentially unnecessary coordinates.

- *Canonical forms*. A second possibility is the use of *canonical forms* which reduce the number of free parameters to be estimated. Canonical forms are mappings $c : M(n) \rightarrow S_m(n)$, selecting from every equivalence class $\mathcal{E}(A, B, C, D, K) = \pi^{-1}(l, k)$, $(l, k) \in M(n)$ a unique representative. This approach is currently most often used when maximum likelihood-type estimation procedures are employed. Important examples are the echelon canonical form—see chapter 2 in Hannan and Deistler (1988)—and balanced canonical

forms such as Ober's Lyapunov and stochastically balanced canonical forms—see Ober (1991) or Ober (1996)—and McGinnie's minimum phase balanced canonical form; see McGinnie (1993). Both echelon and balanced canonical forms lead to a partition of $M(n)$ into pieces of different dimension which are parameterized separately; for $s = 1$ and echelon forms, $M(n)$ is parameterized directly.

In case of echelon forms the pieces, $V_\alpha \subset M(n)$ say, are defined by the so called Kronecker indices α which specify a special selection of basis rows from the Hankel matrix of the transfer function. The state-space matrices (A, B, C, D, K) then are defined via these basis rows. A main advantage of echelon forms is that certain entries in (A, B, C, D, K) are fixed to be zero or one and all other entries are free parameters. There also exist echelon ARMAX forms and there is a simple one-to-one relation between the free parameters in both forms. One piece $V_\alpha \subset M(n)$ is open and dense in $M(n)$ and this V_α is called a generic neighborhood. From a numerical point of view, echelon forms seem to be sometimes problematic. In particular, they have been found to be inferior in many simulation experiments when compared to balanced canonical forms; see Ribarits (2000).

Balanced canonical forms are obtained from an SVD of the Hankel matrix of the transfer function by imposing additional restrictions. Again, there is a generic neighborhood, $V_\delta \subset M(n)$ say, (which differs from the generic neighborhood V_α for echelon forms), which is open and dense in $M(n)$. As opposed to echelon forms, here the entries in (A, B, C, D, K) are rather complicated transformations of the free parameters, however as a tradeoff the parameter spaces are simpler; for example, the minimality requirement for the state-space system translates into strict positivity of the singular values of the Hankel matrix, and the differences of these singular values form a part of the vector of free parameters. This simplicity of parameter spaces may also be an advantage for specification search; see e.g. Bauer and Deistler (1999). A disadvantage of balanced compared to echelon forms is that for the first in general more pieces are needed to cover $M(n)$; even for $s = 1$, for balanced forms $M(n)$ cannot be described by a single parameter space.

- *The description of the manifold $M(n)$ by local coordinates.* The most common approach here is a choice of local coordinates obtained from selections of basis rows of the Hankel matrix of the transfer function which correspond to the so called structural indices α. This approach is similar to echelon forms, for instance with respect to properties of parameters and parameter spaces. However, the pieces $U_\alpha \subset M(n)$ are now all of the same dimension $2ns + m(n + s)$ and they are all open and dense in $M(n)$. Hence, they are overlapping. A particular piece U_α and its parameterization coincides with the generic neighborhood V_α for echelon forms and its parameterization. This 'overlapping description' of $M(n)$ is available for state-space and ARMAX systems; see also chapter 2 in Hannan and Deistler (1988).

- *Data driven local parameterizations*. The choice of parameterizations influences in particular the numerical properties of identification procedures. These properties depend in general on the 'true' transfer function. The main idea of data driven local parameterizations is to choose the parameterization in a data driven way out of uncountably many possibilities in order to obtain favorable properties. This was the motivation for the introduction of the parameterization by *data driven local coordinates* or, briefly, `DDLC` by McKelvey et al. (2004). A modification of `DDLC` has been introduced and analyzed in Ribarits (2002) and is called *separable least squares data driven local coordinates*; see also Ribarits et al. (2003). A further extension of `DDLC` called `orthoDDLC` is subject to future analysis; see Ribarits (2002). These parameterizations will be discussed in more detail below.

In the sequel we provide a brief discussion of the data driven local parameterizations `DDLC`, `slsDDLC` and `orthoDDLC`.

3.1 Data Driven Local Coordinates (DDLC)

Here we commence from a given initial estimator and its equivalence class $\mathcal{E}(A,B,C,D,K) \in S_m(n)$. The idea now is to avoid the drawback of n^2 essentially unnecessary coordinates by only considering the ortho-complement to the tangent space to $\mathcal{E}(A,B,C,D,K)$ at the given (A,B,C,D,K) as a parameter space. Clearly, the parameter space will then be of dimension $2ns + m(n+s)$ rather than $n^2 + 2ns + m(n+s)$ and thus has no unnecessary coordinates. (A,B,C,D,K) is called 'the initial system' which may be obtained e.g. by some other preliminary estimation procedure. For a detailed presentation see McKelvey et al. (2004). Note that similar ideas can already be found in Wolodkin et al. (1997) in an LFT-type parameterization setting.

It has been shown in Ribarits et al. (2004, 2002) that `DDLC` can be interpreted as a system of local coordinates for the manifold $M(n)$, containing uncountably many coordinate charts. The use of `DDLC` thus offers the possibility to choose one 'convenient' out of uncountably many coordinate charts, and this choice can be performed in each step of a numerical search procedure: For any given $(A,B,C,D,K) \in S_m(n)$, one can choose an observationally equivalent state-space system and apply the `DDLC` construction at this system, yielding a parameterization for an open neighborhood of transfer functions.

The idea of `DDLC` was to improve numerical properties of estimation algorithms, which was a major motivation for including it into the system identification toolbox in `MATLAB 6.x` in the standard case when no particular parameterization is chosen by the user. However, as has been discussed in Ribarits (2002) and Deistler and Ribarits (2001), there are still numerical difficulties and potential drawbacks of `DDLC`. A main tool for a careful analysis of `DDLC` is the investigation of topological and geometrical properties of the parameterization; see Ribarits and Deistler (2002) and Ribarits et al. (2004).

Recently, a modification of the `DDLC` concept has been introduced in Ribarits et al. (2003):

3.2 Separable Least Squares Data Driven Local Coordinates

The idea here is to combine the DDLC philosophy with the method of 'separable least squares', leading to an alternative analogous parameterization which can be used for a suitable concentrated likelihood-type criterion function. The new parameterization is called slsDDLC (for separable least squares data driven local coordinates). An obvious consequence is the reduction of the dimension of the parameter space, and preliminary simulation studies in Ribarits and Deistler (2003) indicate that slsDDLC has numerical advantages as compared to e.g. the more commonly used echelon canonical form and to conventional DDLC.

Results concerning geometrical and topological properties of the parameterization are derived in Ribarits et al. (2003) and Ribarits (2002).

The new identification method using slsDDLC seems to be very promising, in particular if slsDDLC is used in combination with so called subspace identification techniques (see, e.g., Deistler, 2001) which often yield good starting values for the optimization of the likelihood function. However, the behaviour of both slsDDLC and DDLC when no good initial estimates are available, i.e. when the initial transfer function estimate is 'far' from the 'true' transfer function, is still unclear. There are a number of open questions regarding this 'global' perspective, e.g. the question whether it is possible to obtain state-space systems in the course of the numerical search procedure which lead to more and more ill-conditioned estimation problems if we do not adapt (sls)DDLC.

This question has also been the main motivation for another modification of DDLC:

3.3 Orthogonal Data Driven Local Coordinates (orthoDDLC)

OrthoDDLC has been briefly introduced in Ribarits (2002). The terminology is motivated by the fact that the optimization of the likelihood function is restricted to a subset of the set of balanced stable allpass systems; see e.g. Hanzon and Peeters (2000) or Peeters et al. (1999). These systems have the property that the corresponding state-space matrices (A, B, C, D), if arranged appropriately, form an orthogonal matrix. OrthoDDLC seems to be a promising approach, but still has to be worked out in more detail and simulation studies have to be carried out.

4 Future Research Topics

Currently, the concepts of data driven local parameterizations outlined in Section 3 above have only been considered for the stationary case, i.e. the case of stable transfer functions (l, k). Extensions to the parameterization and estimation problem for unit root systems – with an emphasis on cointegration – are a subject of future research.

The main aim of cointegration analysis is to decompose the observed variables into a stationary part and a part generated by a few common trends, where the economically relevant information lies in the cointegrating relations. DDLC and slsDDLC could be applied e.g. in the following situations:

- The DDLC philosophy could be applied to maximum-likelihood type estima-

tion of cointegration models incorporating restrictions arising e.g. from a priori knowledge on cointegrating relations. In the classical AR-framework of Johansen (1995) this would amount to new `slsDDLC` estimation procedures for reduced rank regression models where additional restrictions are incorporated.

- Recently, parameterizations for general unit root processes in the state-space framework have been introduced, which highlight the (polynomial) cointegration properties; see Bauer and Wagner (2003). The minimal state-space systems considered are of the form (5), where no observed inputs u_t are present now and the matrix A can have eigenvalues on the unit circle implying that the corresponding transfer function $k(z)$ has poles at these locations. One particular feature of the parameterization is that it splits the parameters into a part corresponding to the stationary subsystem and a part corresponding to the subsystem with the unit roots only. Therefore, the stationary subsystem can be parameterized using any parameterization for stationary systems, and the possibility of using `DDLC` and `slsDDLC` would have to be investigated in this context.

Bibliography

Bauer, D. and Deistler, M. (1999). Balanced canonical forms for system identification. *IEEE Transactions on Automatic Control*, 44:1118–1131.

Bauer, D. and Wagner, M. (2003). A canonical form for unit root processes in the state space framework. *Econometric Theory*. Submitted.

Deistler, M. . (2001). System identification – general aspects and structure. In Goodwin, G. C., editor, *Model identification and adaptive control*, pages 3–26. Springer, London, Berlin, Heidelberg.

Deistler, M. and Ribarits, T. (2001). Parametrizations of linear systems by data driven local coordinates. In *Proceedings of the CDC'01 Conference*, pages 4754–4759, Orlando, Florida.

Hannan, E. J. and Deistler, M. (1988). *The Statistical Theory of Linear Systems*. Wiley, New York.

Hanzon, B. and Peeters, R. L. M. (2000). Balanced parametrizations of balanced siso all-pass systems in discrete time. *Mathematics of Control, Signals and Systems*, 13:240–276.

Johansen, S. (1995). *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models*. Oxford University Press, Oxford.

McGinnie, P. P. (1993). *A Balanced View of System Identification*. PhD thesis, University of Cambridge.

McKelvey, T. (1995). *Identification of State-Space Models from Time and Frequency Data*. PhD thesis, Department of Electrical Engineering, Linkoeping University.

McKelvey, T., Helmersson, A., and Ribarits, T. (2004). Data driven local coordinates for multivariable linear systems and their application to system identification. *Automatica*, 40(9).

Ober, R. (1991). Balanced parametrization of classes of linear systems. *SIAM J. Control and Optimization*, 29(6):1251–1287.

Ober, R. (1996). Balanced canonical forms. In Bittanti, S. and Picci, G., editors, *Identification, Adaptation, Learning*, pages 120–183. Springer, Berlin.

Peeters, R., Hanzon, B., and Olivi, M. (1999). Balanced realizations of discrete-time stable all-pass systems and the tangential schur algorithm. In *Proceedings of the ECC'99 Conference*, Karlsruhe, Germany.

Ribarits, T. (2000). Case studies in system identification. Technical Report CUED/F-INFENG/TR383, University of Cambridge, Cambridge.

Ribarits, T. (2002). *The Role of Parametrizations in Identification of Linear Dynamic Systems*. PhD thesis, TU Wien.

Ribarits, T. and Deistler, M. (2002). Data driven local coordinates: geometrical and topological properties. In *Proceedings of 15th IFAC World Congress*, pages 4754–4759, Barcelona, Spain.

Ribarits, T. and Deistler, M. (2003). A new parametrization method for the estimation of state-space models. In *Workshop 'Econometric Time Series Analysis – Methods and Applications'*, Linz, Austria.

Ribarits, T., Deistler, M., and Hanzon, B. (2002). Data driven local coordinates. In *Proceedings of the 15th Symposium MTNS'02*, South Bend, Indiana.

Ribarits, T., Deistler, M., and Hanzon, B. (2003). Separable least squares data driven local coordinates. In *Preprints to the 13th IFAC Symposium on System Identification*, pages 1922–1927, Rotterdam, The Netherlands.

Ribarits, T., Deistler, M., and McKelvey, T. (2004). An analysis of the parametrization by data driven local coordinates for multivariable linear systems. *Automatica*, 40(5):789–803.

Wolodkin, G., Rangan, S., and Poolla, K. (1997). An lft approach to parameter estimation. In *Preprints to the 11th IFAC Symposium on System Identification*, volume 1, pages 87–92, Kitakyushu, Japan.

Factor Models for Multivariate Time Series

Eva Hamann, Manfred Deistler and Wolfgang Scherrer

1 Introduction

The analysis of multivariate time series is an important issue in many research areas, such as economics, finance, signal processing and medicine. Multivariate time series are modeled jointly when the relation between the single time series or comovements are important. In general, the number of free parameters, which is a measure of the complexity of the model class considered, increases with the number of observed variables. E.g., in a general VAR(p) model, the dimension of the parameter space is proportional to the square of the output dimension. This fact causes problems for actual applications, when the number of parameters to be estimated is large compared to sample size (the so-called *curse of dimensionality*). Factor models mitigate this problem.
In this contribution we give a short introduction to different kinds of factor models, emphasizing problems of identifiability and estimation. The following types of factor models will be discussed:

- Quasi-Static Principal Components Analysis (Quasi-Static PCA)
- Dynamic PCA
- Quasi-Static Frisch model
- Dynamic Frisch model
- Reduced Rank Regression model

There has been an increasing interest in factor models recently. In macroeconomics, based on the seminal paper by Sargent and Sims (1977), factor models have been further developed and used for analyzing a relatively large number of time series. For instance, to study dynamical movements of sectoral employments for the U.S. economy, to analyze business cycles in Europe and the U.S. or to obtain a forecast by combining many predictors, see e.g. Forni and Reichlin (1998) and Stock and Watson (1999). Factor models have a relatively long tradition in finance econometrics, see e.g. Chamberlain and Rothschild (1983). The outline of this contribution is as follows. First, the general model is presented in some detail. In the remaining sections identifiability and estimation problems for the particular model classes considered here are discussed.

2 The Basic Framework

Here we restrict ourselves to the stationary case. For instance, cointegration models, which might be interpreted as factor models generating integrated series are not

considered.

The idea, here is, that the n–dimensional vector of observed variables y_t is driven by a linear combination of a small number, say $r \ll n$, of, in general, unobserved variables, so-called factors, and an n–dimensional noise component u_t. Hence, the model may be written as

$$y_t = \Lambda(z)\xi_t + u_t, \tag{1}$$

where z denotes the backward shift as well as a complex variable, $\Lambda(z) = \sum_{j=-\infty}^{\infty} \Lambda_j z^j$ is an $n \times r$ dimensional linear dynamic filter, $\hat{y}_t = \Lambda(z)\xi_t$ denotes the latent variables and $\mathbb{E}\xi_t u_s' = 0$ for all $s, t \in \mathbb{Z}$. The filter $\Lambda(z)$ can be seen as a dynamic generalization of the loading matrix Λ of a static factor model. If we assume that the spectral densities f_ξ and f_u resp. of (ξ_t) and (u_t) resp. exist, then the spectrum of the process (y_t) is given by

$$f_y(\lambda) = \Lambda(\mathrm{e}^{-i\lambda}) f_\xi(\lambda) \Lambda^*(\mathrm{e}^{-i\lambda}) + f_u(\lambda), \tag{2}$$

where $\Lambda^*(z) = \sum_{j=-\infty}^{\infty} \bar{\Lambda}_j' z^{-j}$. Throughout the contribution, we assume that $f_y(\lambda) > 0$, $\mathrm{rank}(\Lambda(\mathrm{e}^{-i\lambda})) = r$ and $f_\xi(\lambda) > 0$ a.e.
An important special case of (1) is the quasi static factor model, where the factor loading matrix is constant (i.e. not dependent on z),

$$y_t = \Lambda\xi_t + u_t. \tag{3}$$

In this case the variance covariance matrix of (y_t) is of the form

$$\Sigma_y = \Lambda\Sigma_\xi\Lambda' + \Sigma_u, \tag{4}$$

where $\Sigma_y = \mathbb{E}y_t y_t' > 0$, $\mathrm{rank}(\Lambda) = r$, $\Sigma_\xi = \mathbb{E}\xi_t\xi_t' > 0$ and $\Sigma_u = \mathbb{E}u_t u_t'$. All processes considered are assumed to have mean zero. Otherwise, one would have to add a constant on the right hand side of Equations (1) and (3).

For identification of factor models, as described in Equation (1), the number of factors r, $\Lambda(z)$ and in many cases also f_ξ, f_u and the unobserved factor processes (ξ_t) are of interest. Unless stated explicitly otherwise, we consider a semi non-parametric setting. Factor models pose a fundamental identifiability problem. In general, for given f_y or Σ_y the quantities of interest are not unique. Without further assumptions, every $\Lambda(z)$ is compatible with a given f_y, or in other words, f_y (or Σ_y) give no restrictions for $\Lambda(z)$ (or for Λ). In this case, for every $\Lambda(z)$ one can find suitable spectral densities f_ξ and f_u or covariance matrices Σ_ξ and Σ_u which fulfill Equations (2) or (4). Therefore, additional structure has to be imposed in order to make identification meaningful. This leads to the PCA, Frisch and Reduced Rank Regression models discussed below. For these classes the following issues arise, when identifiability is analyzed:

1. Identifiability of r
2. Identifiability of $f_{\hat{y}}(\lambda) = \Lambda(\mathrm{e}^{-i\lambda}) f_\xi(\lambda) \Lambda^*(\mathrm{e}^{-i\lambda})$ and $f_u(\lambda)$
3. Identifiability of $\Lambda(\mathrm{e}^{-i\lambda})$ and $f_\xi(\lambda)$

Once identifiability is guaranteed in estimation and inference one is concerned with:

1. Inference for r
2. Inference for the free parameters in $\Lambda(\mathrm{e}^{-i\lambda})$, $f_\xi(\lambda)$ and $f_u(\lambda)$
3. Estimation of ξ_t

Often factor models are used for analysis, but they may also be used for forecasting. In this case a forecasting model for factors and eventually also for the noise component has to be identified. The models we consider in this context are of ARX type. E.g.

$$\xi_{t+1} = A(z)\xi_t + B(z)x_t + \epsilon_{t+1}, \tag{5}$$

where $A(z)$ and $B(z)$ are polynomial matrices in z, the stability condition

$$\det(I - z(A(z))) \neq 0 \text{ for all } |z| \leq 1 \tag{6}$$

holds, ϵ_t is white noise, x_t are (observed) inputs and $\mathbb{E}x_t\epsilon_s' = 0$ for all t, $s \in \mathbb{Z}$.

3 Quasi-Static Principal Components Analysis (Quasi-Static PCA)

Here we commence from Equations (3) and (4), where the additional structural assumption is imposed that $\hat{y}_t$ is obtained by minimizing $\mathbb{E}u_t'u_t = \mathrm{tr}(\Sigma_u)$ over all rank r matrices Λ. This is done by performing an eigenvalue decomposition of the matrix $\Sigma_y = O\Omega O'$, where $\Omega = \mathrm{diag}(\omega_1, \ldots, \omega_n)$ and $0 < \omega_i \leq \omega_j$ for all $i > j$. It can be shown that the optimal decomposition in this sense is given by

$$\Sigma_y = \underbrace{O_1\Omega_1O_1'}_{\Sigma_{\hat{y}}} + \underbrace{O_2\Omega_2O_2'}_{\Sigma_u}, \tag{7}$$

where Ω_1 is the $r \times r$ north west block submatrix of Ω and Ω_2 is $(n-r)\times(n-r)$ south east block submatrix of Ω and O_1 and O_2 are defined accordingly. This decomposition, i.e., $\Sigma_{\hat{y}}$, Σ_u is unique for $\omega_r > \omega_{r+1}$. However, the factors, ξ_t, and the factor loadings Λ are not unique. This is discussed in detail in Section 5. A special choice is given by

$$y_t = O_1\xi_t + u_t, \tag{8}$$

where $\xi_t = O_1'y_t$ and $u_t = y_t - O_1O_1'y_t = O_2O_2'y_t$. Note that, here, the factors ξ_t are measurable (linear) functions of the observed variables y_t.

The number of factors r cannot be determined from the observed process (y_t), since for every r, $1 \leq r \leq n$, such a decomposition may be constructed. However, there are several rules of thumbs for choosing r from data. A simple but reasonable procedure is to plot the eigenvalues of Σ_y and to look for a natural cutting point. Another procedure is to look at the eigenvalues of the correlation matrix of y_t and to attribute the eigenvalues, which are smaller than one to the noise part.

Estimation in the Quasi Static PCA is straight forward. Λ, Σ_u and ξ_t are simply estimated by replacing Σ_y by the sample variance covariance matrix $\hat{\Sigma}_y = \frac{1}{T}\sum_{t=1}^{T} y_t y_t'$ in Equation (7). Here, T denotes sample size. Also forecasting of factors as described above is straight forward.

4 Dynamic PCA

For the dynamic generalization of the PCA considered above we commence from the spectral density f_y rather than the covariance Σ_y and its eigenvalue decomposition, see Brillinger (1981)

$$f_y(\lambda) = \underbrace{O_1(\lambda)\Lambda_1(\lambda)O_1^*(\lambda)}_{f_{\hat{y}}(\lambda)} + \underbrace{O_2(\lambda)\Lambda_2(\lambda)O_2^*(\lambda)}_{f_u(\lambda)}. \tag{9}$$

The model analogous to Equation (8) then is of the form

$$y_t = O_1(z)\xi_t + u_t, \tag{10}$$

where $\xi_t = O_1^*(z)y_t$ and $u_t = y_t - O_1(z)O_1^*(z)y_t = O_2(z)O_2^*(z)y_t$. It is shown in Brillinger (1981) that this decomposition gives the minimal noise, in the sense that $\mathbb{E}u_t'u_t$ is minimal, among all decompositions where $\mathrm{rk}(\Lambda(z)) = r$ a.e.
Similar to the Quasi Static case the number of factors r is not identifiable from the process (y_t). In practise, r is chosen from data by inspecting the eigenvalues $\omega_i(\lambda)$ of the spectral density $f_y(\lambda)$. The factors $\xi_t = O_1^*(z)y_t$ again are linear functions of the observed process (y_t). However, in general, the filter $O_1^*(z)$ is two-sided and non-rational. Thus, naive forecasting of the factors by a model of the type (5) yield infeasible forecasts for y_t. This problem could be solved, in principle, by restriction to factors obtained by causal filters $O_1^*(z)$. However, this problem has not been solved completely see Forni et al. (2003).
As in the Quasi Static case, estimation of Λ, f_u and ξ_t is done by replacing f_y by some estimate in Equation (9). However, estimation of spectra is more demanding than estimation of variance covariance matrices. In particular, for large n relatively large sample sizes are needed for reliable estimates.

5 Quasi-static Frisch Model

Here we consider Equation (3) together with the following additional assumption:

$$\Sigma_u \text{ is diagonal.} \tag{11}$$

Of course, in the case that u_t is white noise this condition is equivalent to the condition "f_u is diagonal". By this assumption the noise part represents the (static) individual effects for each component of y_t and the factor part the (static) common features between the components of y_t. Note, that for given latent variables $\hat{y}_t$, the components of y_t are conditionally uncorrelated. Identifiability here is more demanding compared to the PCA case. At this point, in more general terms, the question of solvability and of uniqueness of solutions of Equation (4) arises. Consider first the question whether Σ_y can be uniquely decomposed as the sum of $\Sigma_{\hat{y}} = \Lambda\Sigma_\xi\Lambda'$ and Σ_u. For given n and r, the number of equations (i.e. the number of free elements in Σ_y) is $\frac{n(n+1)}{2}$ because of the symmetry of Σ_y. The number of free parameters on the right hand side is $nr - \frac{r(r-1)}{2} + n$. Now, let $B(r) = \frac{n(n+1)}{2} - (nr - \frac{r(r-1)}{2} + n) = \frac{1}{2}((n-r)^2 - n - r)$, then the following cases might occur:

- $B(r) < 0$: In this case we might expect non-uniqueness of the decomposition
- $B(r) \geq 0$: In this case we might expect uniqueness of the decomposition

The argument can be made more precise, in particular, for $B(r) > 0$, generic uniqueness can be shown, see Scherrer and Deistler (1998).
Once $\Sigma_{\hat{y}}$ is uniquely determined from Σ_y suitable conditions on Λ and Σ_ξ may be imposed in order to obtain uniqueness of Λ and Σ_ξ from $\Sigma_{\hat{y}}$. If Σ_ξ is assumed to be the identity matrix I_r then Λ is unique up to postmultiplication with $r \times r$ orthogonal matrices, corresponding factor rotation. For a detailed discussion of different kinds of rotations see Lawley and Maxwell (1971). If no assumption on Σ_ξ is imposed, then Λ is unique up to postmultiplication with arbitrary non singular matrices. Unless stated otherwise we will assume $\Sigma_\xi = I_r$.
Note that in the Frisch model, as opposed to PCA, the factors ξ_t, in general, cannot be obtained as a function of the observations y_t. Thus, for estimation of factors, the factor process has to be approximated by a (linear) function of y_t.

We consider two methods, one has been discussed in detail in Thomson (1951) and the other one has been proposed in Bartlett (1937, 1938).

1. The regression method investigated by Thomson:
   The idea, here, is to estimate ξ_t by a linear function of y_t such that the variance of the estimation error, $\xi_t - \hat{\xi}_t$, is minimal. Therefore, $\hat{\xi}_t$ is given by the regression of ξ_t onto y_t,
   $$\hat{\xi}_t^T = \Lambda' \Sigma_y^{-1} y_t, \tag{12}$$
   since by the above assumptions
   $$\mathbb{E} y_t \xi_t' = \mathbb{E}[(\Lambda \xi_t + u_t)\xi_t'] = \Lambda. \tag{13}$$
   As can easily be seen, this estimator is biased in a certain sense, since $\mathbb{E}(\hat{\xi}_t^T | \xi_t) = \Lambda' \Sigma_y^{-1}(\Lambda \xi_t + \mathbb{E}(u_t|\xi_t)) \neq \xi_t$.

2. Bartlett's method:
   In his method Bartlett suggests to minimize the sum of the standardized residuals with respect to $\hat{\xi}_t$, i.e.,
   $$\min_{\hat{\xi}_t} (y_t - \Lambda \hat{\xi}_t)' \Sigma_u^{-1} (y_t - \Lambda \hat{\xi}_t). \tag{14}$$
   Thus, the estimate for ξ_t is given by
   $$\hat{\xi}_t^B = (\Lambda' \Sigma_u^{-1} \Lambda)^{-1} \Lambda' \Sigma_u^{-1} y_t. \tag{15}$$
   This estimate is unbiased in the same sense as above, if $\mathbb{E}(u_t|\xi_t) = 0$ holds true, since $\mathbb{E}(\hat{\xi}_t^B | \xi_t) = (\Lambda' \Sigma_u^{-1} \Lambda)^{-1} \Lambda' \Sigma_u^{-1}(\Lambda \xi_t + \mathbb{E}(u_t|\xi_t)) = \xi_t$.

There is no general rule which method to apply. The decision can be based upon the properties the factor estimates should possess. Generally one can say, that Bartlett's

estimate is unbiased but at the expense of a higher variability than Thomson's estimate.

In the case where ξ_t and u_t and, thus, y_t are white noise, estimates of Λ and Σ_u may be obtained by Maximum Likelihood (ML) estimation. The negative logarithm of the Gaussian likelihood, up to a constant, has the form

$$\begin{aligned} L_T(\Lambda, \Sigma_u) &= \frac{1}{2} T \log(\det(\Lambda\Lambda' + \Sigma_u)) + \frac{1}{2} \sum_{t=1}^{T} y_t'(\Lambda\Lambda' + \Sigma_u)^{-1} y_t = \\ &= \frac{1}{2} T \log(\det(\Lambda\Lambda' + \Sigma_u)) + \frac{1}{2} T \mathrm{tr}((\Lambda\Lambda' + \Sigma_u)^{-1} \hat{\Sigma}_y). \end{aligned} \tag{16}$$

A likelihood ratio test for the number of factors has been suggested in Anderson and Rubin (1956).

Forecasts for the factors ξ_t may be obtained by ARX models (5). However, note that there is no general rule which factor estimates yield the best forecasts for y_t.

6 Dynamic Frisch Model

Here Equation (1) together with the assumption

$$f_u \text{ is diagonal.} \tag{17}$$

is considered. Again u_t represents the individual influences and ξ_t the comovements. The only difference to the previous section is that Λ is now a dynamic filter and the components of u_t are orthogonal to each other for all leads and lags.

A rather complete structure theory is presented in Scherrer and Deistler (1998). Concerning estimation and specification neither a complete theory nor general methods are available. See, however, Watson and Engle (1983), Geweke (1977), Beghellis et al. (1990), Forni et al. (2000) and Forni and Lippi (2001).

7 Reduced Rank Regression Model

Here we consider a regression model of the form

$$y_{t+1} = F \underbrace{G\tilde{x}_t}_{=\xi_{t+1}} + u_{t+1}, \qquad t \in \mathbb{Z}, \tag{18}$$

where the $\tilde{m}$-dimensional vector process $(\tilde{x}_t)$ of explanatory variables contains possibly lagged inputs x_t and lagged observed variables y_t and (u_t) denotes the n-dimensional noise process. In addition we assume:

(i) (x_t) and (u_t) are uncorrelated, i.e. $\mathbb{E}x_t u_s' = 0 \; \forall s, t$

(ii) (x_t) is (weak sense) stationary with a non-singular spectral density

(iii) (u_t) is white noise with $\mathbb{E}u_t u_t' > 0$

(iv) a stability assumption analogous to (6)

The major assumption however is that $\beta = FG$ is of rank $r < \min(n, \tilde{m})$. Thus, $F \in \mathbb{R}^{n\times r}$ and $G \in \mathbb{R}^{r\times\tilde{m}}$ and $G\tilde{x}_t$ can be interpreted as the r-dimensional factor process (ξ_{t+1}), the matrix F can be interpreted as the corresponding factor loading matrix. Models of this kind have been analyzed in Anderson (1958). In Anderson's paper, where no lagged endogenous variables as regressors are considered, it has been shown that the maximum likelihood estimate is obtained by an OLS estimation of β followed by a weighted singular value decomposition, where only the largest r singular values are kept.
As far as identifiability is concerned, note that F is unique only up to postmultiplication by a nonsingular matrix and an analogous statement holds for G and ξ_{t+1}. We use the singular value decomposition of β

$$\beta = U\Sigma V' \tag{19}$$

where U and V are orthogonal matrices of dimensions n and $\tilde{m}$, resp., and $\Sigma \in \mathbb{R}^{n\times\tilde{m}}$ is the matrix of singular values, σ_i, $i = 1, \ldots, \min(n, \tilde{m})$, arranged in decreasing order. The strictly positive singular values are assumed to be different and the singular vectors, corresponding to these positive singular values, are unique up to sign change and suitably normalized in order to obtain uniqueness.
Let $\hat{\beta}$ denote the OLS estimator of β and let $\hat{\beta} = \hat{U}\hat{\Sigma}\hat{V}'$ denote its singular value decomposition. The reduced rank estimator of β, denoted as *direct* estimator, then is given by

$$\hat{\hat{\beta}}_D = \hat{U}_1\hat{\Sigma}_1\hat{V}_1' \tag{20}$$

where $\hat{\Sigma}_1 \in \mathbb{R}^{r\times r}$ is the matrix formed from the r largest singular values of $\hat{\Sigma}$ and $\hat{U}_1$ and $\hat{V}_1$, resp., are formed from the first r columns of $\hat{U}$ and $\hat{V}$, resp.
An alternative procedure, denoted as the *indirect* procedure, is obtained by performing the SVD for a suitably weighted matrix, see Anderson (1958); Deistler and Hamann (2003). For a canonical correlations analysis one would consider

$$\Sigma_y^{-1/2} y_{t+1} = \Sigma_y^{-1/2}\beta\Sigma_{\tilde{x}}^{1/2}\Sigma_{\tilde{x}}^{-1/2}\tilde{x}_t + \Sigma_y^{-1/2}u_{t+1}. \tag{21}$$

Replacing the population second moments by their sample counterparts, consider the SVD

$$\hat{\Sigma}_y^{-1/2}\hat{\beta}\hat{\Sigma}_{\tilde{x}}^{1/2} = \hat{U}\hat{\Sigma}\hat{V}' \tag{22}$$

where $\hat{\beta}$ is the least squares estimator. Note, $\hat{U}$, $\hat{\Sigma}$ and $\hat{V}$ are different from $\hat{U}$, $\hat{\Sigma}$ and $\hat{V}$ mentioned above. Retaining only the r largest singular values one obtains (using an obvious notation)

$$\hat{\hat{\beta}}_I = \hat{\Sigma}_y^{1/2}\hat{U}_1\hat{\Sigma}_1\hat{V}_1'\hat{\Sigma}_{\tilde{x}}^{-1/2}, \tag{23}$$

where again $\hat{U}_1$, $\hat{\Sigma}_1$ and $\hat{V}_1$ are different from $\hat{U}_1$, $\hat{\Sigma}_1$ and $\hat{V}_1$ in Equation (20). Furthermore, note that (23) is the ML estimate if there are no lagged variables of y_t

contained in $\tilde{x}_t$.

For the reduced rank model considered here, a complete model specification consists of selection of input variables out of a possible large set of candidate inputs, specification of the dynamics of the inputs and outputs and the number of factors. Data-driven model selection may be done by an AIC or BIC-type criterion of the form

$$\begin{aligned} AIC(\tilde{m}, r) &= \log\det\hat{\Sigma}_{u(\tilde{m},r)} + d(\tilde{m}, r)\frac{2}{T} \\ BIC(\tilde{m}, r) &= \log\det\hat{\Sigma}_{u(\tilde{m},r)} + d(\tilde{m}, r)\frac{\log T}{T}, \end{aligned} \tag{24}$$

where $d(\tilde{m}, r) = nr + r\tilde{m} - r^2$ is the number of free parameters in β for a given specification and $\hat{\Sigma}_{u(\tilde{m},r)}$ is the one step ahead (in sample) prediction error variance covariance matrix corresponding to the specification indicated and to one of the estimation procedures described above. For further details on the procedure we refer to Deistler and Hamann (2003).

Closely related to the approach described above is the use of state space models where the state dimension is smaller than the minimum of the input and output dimension resp.

Bibliography

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

Anderson, T. W. and Rubin, H. (1956). Statistical inference in factor analysis. *Proceedings of the third Berkeley Symposium on mathematical statistics and probability*, V:111–150.

Bartlett, M. S. (1937). The statistical conception of mental factors. *British Journal of Psychology*, 28:97–104.

Bartlett, M. S. (1938). Methods of estimating mental factors. *Nature, London*, 141:609–610.

Beghellis, S., Guidorzi, R. P., and Soverini, U. (1990). The frisch scheme in dynamic system identification. *Automatica, Special Issue*, 26:171–176.

Brillinger, D. R. (1981). *Time Series, Data Analysis and Theory*. Holden Day, San Francisco.

Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets. *Econometrica*, 51:1281–1304.

Deistler, M. and Hamann, E. (2003). Identifiaction of factor models for forecasting returns. Mimeo: Institute of Econometrics, Operations Research and System Theory, Vienna University of Technology.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: identification and estimation. *The Review of Economics and Statistics*, 82(4):540–554.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2003). The generalized dynamic factor model, one-sided estimation and forecasting. `http://www.dynfactors.org/papers/papers.htm`.

Forni, M. and Lippi, M. (2001). The generalized dynamic factor model: representation theory. *Econometric Theory*, 17:1113–1141.

Forni, M. and Reichlin, L. (1998). Let's get real: A factor analytical approach to disaggregated business cycle dynamics. *Review of Economic Studies*, 65:453–473.

Geweke, J. (1977). The dynamic factor analysis of economic time-series models. In Aigner, D. J. and Goldberger, A. S., editors, *Latent Variables in Socio-Economic Models*, pages 365–383. North-Holland, Amsterdam.

Lawley, D. N. and Maxwell, A. E. (1971). *Factor Analysis as a Statistical Method*. Butterworth, London, 2nd edition.

Sargent, T. J. and Sims, C. A. (1977). *Business Cycle Modelling Without Pretending to Have Too Much* a priori *Economic Theory*. Federal Reserve Bank of Minneapolis, Minneapolis.

Scherrer, W. and Deistler, M. (1998). A strucure theory for linear dynamic errors-in-variables models. *SIAM Journal of Control and Optimization*, 36:2148–2175.

Stock, J. H. and Watson, M. W. (1999). Diffusion indexes. This is a substantially revised version of NBER Workin Paper 6702, August 1998.

Thomson, G. H. (1951). *The Factorial Analysis of Human Ability*. London University Press, London.

Watson, M. W. and Engle, R. F. (1983). Alternative algorithms for the estimation of dynamic factor, mimic and varying coefficient regression models. *Journal of Econometrics*, 23:385–400.

Detecting Longitudinal Heterogeneity in Generalized Linear Models

Achim Zeileis

1 Introduction

Structural change is of central interest in many fields of research and data analysis: to learn if, when and how the structure of the data generating mechanism underlying a set of observations changes. In many situations, it is known with respect to which quantity the structural change might occur, e.g., over time or with the increase of an explanatory variable like income or firm size. These situations have in common that the structural changes lead to longitudinal heterogeneity within the observations where the timing and pattern of the change is typically unknown. One of the simplest examples for such a structural change is a time series whose mean changes at a single breakpoint. Such a time series is depicted in Figure 1 giving the number of youth homicides per month in Boston between 1992(1) and 1998(5). The plot suggests that the number of homicides varies around a constant mean up to about 1996 but drops to a lower mean at the end of the sample. To assess whether there is evidence for such a structural change or not, a statistical test is needed: given a model (in the example: constant mean number of homicides) it is tested whether the data support the hypothesis that there is a stable structure against the alternative that it changes over time.

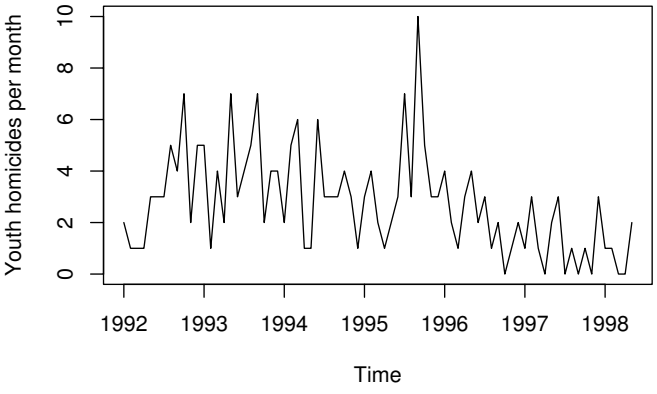

Figure 1: Number of youth homicides per month in Boston

In parametric models, structural change is typically described by parameter instability. If this instability is ignored, parameter estimates are generally not meaningful, inference is severely biased and predictions lose accuracy. For the standard linear regression model, tests for structural change or parameter instability have been receiving much attention, both in the statistics and the econometrics literature. Starting from the recursive CUSUM test of Brown et al. (1975), a large variety of tests has been suggested most of which can be broadly placed into two different classes: generalized fluctuation tests (Kuan and Hornik, 1995) that do not assume a particular pattern of deviation from the hypothesis of parameter constancy, and F tests (Andrews, 1993; Andrews and Ploberger, 1994) that are built for a single shift alternative (of unknown timing).

However, the situation where the relationship between the dependent variable and explanatory regressors is not well described by a simple linear regression and a generalized linear model (GLM) is more appropriate has not yet been studied in detail in a structural change context. In practice, this is important when the longitudinal heterogeneity of choice data, ratings on a discrete scale or counts (as in the Boston homicides example) is to be investigated. Therefore, we describe in Section 2 how the class of generalized fluctuation tests can be extended to the GLM which is subsequently applied to the Boston homicides data in Section 3 and a brief summary is given in Section 4.

2 Generalized Fluctuation Tests in the Generalized Linear Model

In the simple linear regression model, the generalized fluctuation tests fit the model to the data via ordinary least squares (OLS)—or equivalently via maximum likelihood (ML) using a normal approximation—and derive a process which captures the fluctuation of the recursive or OLS residuals (Brown et al., 1975; Ploberger and Krämer, 1992; Chu et al., 1995a), of the recursive or rolling/moving estimates (Ploberger et al., 1989; Chu et al., 1995b), or of M-scores which includes ML scores (Nyblom, 1989; Hansen, 1992; Hjort and Koning, 2002; Zeileis and Hornik, 2003). For these empirical fluctuation processes, the limiting process under the null hypothesis of structural stability is known and therefore boundaries can be chosen that are crossed by the limiting process (or some functional of it) only with known probability α. Hence, if the empirical fluctuation process exceeds these boundaries the fluctuation is improbably large and the null hypothesis of structural/parameter stability has to be rejected. Hjort and Koning (2002) and Zeileis and Hornik (2003) also show how the idea of using ML scores for capturing structural instabilities can be used in more general parametric models and in particular in GLMs. Below, we first construct the empirical fluctuation processes based on ML scores for the GLM and then outline how these can be visualized and aggregated to test statistics.

2.1 Empirical Fluctuation Processes

Consider the GLM like in McCullagh and Nelder (1989). To fix notation, we assume n independent observations of a dependent variable y_i and a vector of regressors or

covariates x_i. The observations of the response variable are distributed independently according to a distribution $F(\theta, \phi)$ where θ is the canonical parameter and ϕ is the dispersion parameter common to all y_i. The following relationship is assumed for the covariates and the mean μ_i of the responses:

$$\mu_i \quad = \quad h(\eta_i) \quad = \quad h(x_i^\top \beta_i) \qquad (i = 1, \ldots, n), \tag{1}$$

where $h(\cdot)$ is the inverse link function, β is the vector of regression coefficients and η_i is the linear predictor. For this model, the hypothesis of structural stability becomes

$$H_0 : \ \beta_i \quad = \quad \beta_0 \qquad (i = 1, \ldots, n)$$

which is tested against the alternative that (at least one component of) β_i varies over "time" i.

The true vector of regression coefficients β_0 under the null hypothesis is usually unknown and estimated by ML whereas ϕ is treated as a nuisance parameter (or is known anyway). The resulting score function for β is

$$\psi(y_i, x_i, \beta) \quad = \quad x_i \, h'(x_i^\top \beta) \, V(\mu_i)^{-1} (y_i - \mu_i), \tag{2}$$

where $h'(\cdot)$ is the derivative of the inverse link function and $V(\mu)$ is the variance function of the model. These scores yield the usual ML estimate $\hat{\beta}_n$ via the first order condition

$$\sum_{i=1}^{n} \psi(y_i, x_i, \hat{\beta}_n) \quad = \quad 0. \tag{3}$$

The corresponding covariance matrix J_n is given by

$$J_n \quad = \quad \frac{1}{n} \sum_{i=1}^{n} h'(x_i^\top \beta)^2 w(\phi) \, V(\mu_i)^{-1} x_i x_i^\top . \tag{4}$$

where the function $w(\cdot)$ is determined by the distribution F (for more details see McCullagh and Nelder, 1989).

It can be easily seen that the scores ψ have zero mean under the null hypothesis and one would expect systematic deviations from zero in the case of longitudinal structural changes over i. To capture such changes it is natural to consider the empirical fluctuation process of partial sums of the scores.

$$W_n(t, \beta) \quad = \quad n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} \psi(y_i, x_i, \beta). \tag{5}$$

Under H_0, the behavior of the empirical fluctuation process $W_n(\cdot, \hat{\beta}_n)$ is governed by a functional central limit theorem:

$$\hat{J}_n^{-1/2} W_n(\cdot, \hat{\beta}_n) \quad \xrightarrow{\text{d}} \quad W^0(\cdot), \tag{6}$$

where $\hat{J}_n$ is some consistent covariance matrix estimate and $W^0(\cdot)$ denotes the standard Brownian bridge. For a proof see Zeileis and Hornik (2003).

2.2 Test Statistics

Given an empirical fluctuation process $efp(t) = \hat{J}_n^{-1/2} W_n(t, \hat{\beta}_n)$, we have to judge whether the fluctuation in $efp(\cdot)$ is extreme compared to the fluctuation of a Brownian bridge $W^0(\cdot)$. To accomplish this, the empirical fluctuation process is typically aggregated to some test statistic by some scalar functional $\lambda(\cdot)$ which is able to measure the fluctuation. This is then compared to the distribution implied by applying the same functional to a Brownian bridge. Formally, it can be shown that

$$\lambda(efp) \quad \xrightarrow{\text{d}} \quad \lambda(W^0). \tag{7}$$

A typical choice for λ would be the maximum of the absolute values of the empirical fluctuation process which is also used in the application to the Boston homicides data in Section 3. Given a critical value c_α for this test statistic, every component of the process which exceeds c_α can be seen to violate H_0 at level α which can be used for carrying out the significance test visually as described in the following.

2.3 Visualization

In their seminal paper Brown et al. (1975) point out that the generalized fluctuation test framework

> [...] includes formal significance tests but its philosophy is basically that of data analysis as expounded by Tukey. Essentially, the techniques are designed to *bring out departures from constancy in a graphic way* instead of parametrizing particular types of departure in advance and then developing formal significance tests intended to have high power against these particular alternatives. From this point of view the significance tests suggested should be regarded as yardsticks for the interpretation of data rather than leading to hard and fast decisions. (Brown et al., 1975, pp. 149–150)

This emphasizes two points we will use in the application in the following section: first, visualization of fluctuation tests is important, and second, the tests are not only significance tests but also explorative tools. In a structural change framework, it is usually not only of interest to find some model that fits the data but also to be able to identify and interpret the structural changes which are of high interest for practitioners and researchers. One example for such a situation is the Boston homicide data described above where it is not only of high interest *if* but also *when* the intervention became effective.

As indicated above, rejection of the null hypothesis by the generalized fluctuation test is closely related to the situation that the empirical fluctuation process exceeds a boundary that was derived from the limiting process W^0. In the simplest case, this boundary is constant and equal to c_α. If the empirical fluctuation process crosses this boundary there is evidence (at significance level α) for a structural change. Furthermore, the shape of the process conveys information about the type of change: in the case of a single abrupt shift in the coefficients β_i the mean function of the process will

have a single peak at about the time of the shift, thus allowing for visual identification of the breakpoint.

3 The Boston Homicides Data

To address the problem of continuing high homicide rates in Boston, in particular among young people, a policing initiative called the "Boston Gun Project" was launched in early 1995. This project implemented what became known as the "Operation Ceasefire" intervention in the late spring of 1996 which aimed at lowering homicide rates by a deterrence strategy (Kennedy et al., 1996). To test whether this project was effective in lowering the number of youth homicides we apply the results established for the generalized fluctuation test above to the data from Figure 1. As is natural for count data, a Poisson GLM is used to describe the mean monthly number of youth homicides and its stability is assessed using the empirical fluctuation process $efp(t)$ which can be seen in Figure 2. As the process crosses its boundary $c_{0.05}$ indicated by the horizontal lines, there is evidence for a decrease of the number of homicides. Furthermore, the intervention seems to have become effective gradually as there is no sharp peak in the process but rather a longer time of increased fluctuation. However, the change seems to have occurred around early 1996 when the Operation Ceasefire was implemented (dotted vertical line). The corresponding p value is < 0.0001.

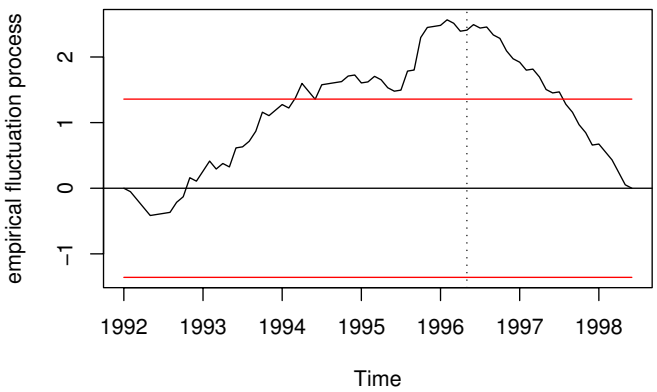

Figure 2: Poisson CUSUM test for Boston homicides data

All computations have been carried out using the R system for statistical computing (R Development Core Team, 2003) and in particular the package `strucchange` (Zeileis et al., 2002).

4 Summary

In this article, we address the problem of detecting longitudinal heterogeneity in generalized linear models which is accomplished by a formal significance test for parameter instability accompanied by visualization techniques which allow for carrying out the test graphically and exploring the structure of the data.

Bibliography

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61:821–856.

Andrews, D. W. K. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62:1383–1414.

Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society B*, 37:149–163.

Chu, C.-S. J., Hornik, K., and Kuan, C.-M. (1995a). MOSUM tests for parameter constancy. *Biometrika*, 82:603–617.

Chu, C.-S. J., Hornik, K., and Kuan, C.-M. (1995b). The moving-estimates test for parameter stability. *Econometric Theory*, 11:669–720.

Hansen, B. E. (1992). Testing for parameter instability in linear models. *Journal of Policy Modeling*, 14:517–533.

Hjort, N. L. and Koning, A. (2002). Tests for constancy of model parameters over time. *Nonparametric Statistics*, 14:113–132.

Kennedy, D. M., Piehl, A. M., and Braga, A. A. (1996). Youth violence in Boston: Gun markets, serious youth offenders, and a use-reduction strategy. *Law and Contemporary Problems*, 59:147–183.

Kuan, C.-M. and Hornik, K. (1995). The generalized fluctuation test: A unifying view. *Econometric Reviews*, 14:135–161.

McCullagh, P. and Nelder, J. A. (1989). *Generalized Linear Models*. Chapman & Hall, London, 2nd edition.

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, 84:223–230.

Ploberger, W. and Krämer, W. (1992). The CUSUM test with OLS residuals. *Econometrica*, 60(2):271–285.

Ploberger, W., Krämer, W., and Kontrus, K. (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics*, 40:307–318.

R Development Core Team (2003). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-00-3.

Zeileis, A. and Hornik, K. (2003). Generalized M-fluctuation tests for parameter instability. Report 80, SFB "Adaptive Information Systems and Modelling in Economics and Management Science", Vienna University of Economics and Business Administration.

Zeileis, A., Leisch, F., Hornik, K., and Kleiber, C. (2002). `strucchange`: An R package for testing for structural change in linear regression models. *Journal of Statistical Software*, 7(2):1–38.

Ensemble Methods for Cluster Analysis

Kurt Hornik and Friedrich Leisch

1 Introduction

Ensemble methods create solutions to learning problems by constructing a set of individual (different) solutions ("base learners"), and subsequently suitably aggregating these, e.g., by weighted averaging of the predictions in regression, or by taking a weighted vote on the predictions in classification. Such methods, which include Bayesian model averaging (Hoeting et al., 1999), bagging (Breiman, 1996) and boosting (Friedman et al., 2000) have already become very popular for supervised learning problems (Dietterich, 2002).

Employing ensemble methods for cluster analysis can be attractive or even necessary for several reasons, the main three being as follows (see e.g. Strehl and Ghosh, 2002):

- To improve quality and robustness of the results. In general, aggregation yields algorithms with "low variance" in the statistical learning sense so that the results obtained by aggregation are more "structurally stable". For example, many clustering algorithms are sensitive to random initializations, choice of hyperparameters, or the order of data presentation in on-line learning scenarios. An obvious idea for possibly eliminating such *algorithmic* variability is to construct an ensemble with (randomly) varied characteristics of the base algorithm. This idea of "sampling from the algorithm" is used e.g. in the voting and voting/merging approaches of Dimitriadou et al. (2002, 2001), see also Section 3. Another idea is to try to improve quality via varying the data by resampling or reweighting. "Bagged Clustering" (Leisch, 1999), see also Section 2, constructs bootstrap samples; a similar approach is used in Dudoit and Fridlyand (2002). Other possible strategies include varying the "features" used for clustering (e.g., using various preprocessing schemes), and constructing "meta-clusterers" which combine the results of the application of different base algorithms as an attempt to reduce dependency of results on specific methods.

- To reuse existing knowledge. In applications, it may be desired to reuse legacy clusterings in order to improve or combine these. Typically, in such situations only the cluster labels are available, but not the original features or algorithms.

- To accommodate the needs of distributed computing. In many applications, it is not possible to use all data simultaneously. Data may not necessarily be available in a single location, or computational resources may be insufficient to use a base clusterer on the whole data set. More generally, clusterers can have access to either a subset of the objects ("object-distributed clustering") or the features ("feature-distributed clustering").

To fix notations and terminology, suppose we are given a set $\mathcal{X}$ of n objects, each holding the measurements on the same variables or features. A K-*clustering* of $\mathcal{X}$ assigns to each x_i in $\mathcal{X}$ a (sub-)probability K-vector $C(x_i) = (\mu_{i1}, \ldots, \mu_{iK})$ (the "membership vector" of the object) with $\mu_{i1}, \ldots, \mu_{iK} \geq 0$, $\sum_k \mu_{ik} \leq 1$. Formally,

$$C : \mathcal{X} \to M \in \mathbb{R}^{n \times K}; \qquad M \geq 0, \quad M1 \leq 1.$$

This framework includes both "crisp" (where each $C(x_i)$ is a unit vector) and fuzzy clustering, as well as incomplete (e.g., completely missing) results where $\sum_k \mu_{ik} < 1$. Changing the labels (which correspond to the columns of the membership matrix M) amounts to replacing M by $M\Pi$, where Π is a suitable permutation matrix. Finally, a *clusterer* is an algorithm producing a clustering.

Given an ensemble of clusterings, the following key distinction can be made for possible aggregation strategies to determine a "consensus" clustering. If each clustering is of the vector quantization type, aggregation can be based on the underlying prototypes (provided that these are available). Otherwise, if only the memberships are available, aggregation can proceed by finding a suitable clustering which "optimally represents" the base clusterings. "Bagged Clustering" and "Voting" are two very promising examples from these two aggregation categories, and will be described in more detail in Sections 2 and 3, respectively.

2 Aggregation Based on Prototypes

A large number of partitioning algorithms represent clusters by one prototype c_k for each cluster and (for crisp partitions) assign each observation x_i to the cluster of the closest prototype. Hence, the cluster memberships μ_{ik} can be written as

$$\mu_{ij} = \begin{cases} 1, & \Delta(x_i, c_j) = \min_k \Delta(x_i, c_k) \\ 0, & \text{otherwise} \end{cases}$$

for a suitable distance measure Δ. Fuzzy partitions assign memberships inversely proportional to distance (or ranks of distance). The well known K-means algorithm uses Euclidean distance as Δ and cluster means as prototypes.

If we are given B clusterings of the same set of objects $\mathcal{X}$ with K prototypes each, we may view the set of $B \times K$ prototypes c_{bk} as a new data set that can be used to assess the structural stability of the clusterer. Prototypes that show up often indicate "typical" clusters, while rare prototypes may indicate random fluctuations.

The bagged clustering algorithm (Leisch, 1999) uses this approach to find structurally more stable partitions: By repeatedly training on new data sets one gets different solutions which should on average be independent from training set influence and random initializations. A collection of B training sets can be obtained by sampling from the empirical distribution of the original data $\mathcal{X}$, i.e., by bootstrapping (Efron and Tibshirani, 1993).

The complete algorithm works as follows:

1. Construct B bootstrap training samples $\mathcal{X}_1, \ldots, \mathcal{X}_B$ by drawing with replacement from the original sample $\mathcal{X}$.

2. Run the base clusterer (K-means, competitive learning, ...) on each set, resulting in $B \times K$ prototypes $c_{11}, c_{12}, \ldots, c_{1K}, c_{21}, \ldots, c_{BK}$ where K is the number of prototypes used in the base method and c_{ij} is the j-th prototype found using $\mathcal{X}_i$.

3. Combine all prototypes into a new data set $\mathcal{C} = \{c_{11}, \ldots, c_{BK}\}$.

4. Run a hierarchical cluster algorithm on $\mathcal{C}$, resulting in the usual dendrogram.

5. Let $c(x) \in \mathcal{C}$ denote the prototype closest to x (minimum distance Δ). A partition of the original data can be obtained by cutting the dendrogram at a certain level, resulting in a partition $\mathcal{C}_B^1, \ldots, \mathcal{C}_B^m$, $1 \leq m \leq BK$, of set $\mathcal{C}$. Each point $x_i \in \mathcal{X}$ is now assigned to the cluster containing $c(x_i)$.

Bagged clustering combines the prototypes using hierarchical clustering because different data structures (convex, not convex, ...) can be accounted for using different linkage methods and the resulting dendrograms can be easily interpreted by practitioners. But many other techniques could be used instead. Unfortunately it is not possible to compare prototypes of partitions directly (e.g., c_{11} with c_{21} etc.) due to the relabelling problem, see also Section 3 below for a more detailed discussion of this most important problem for all cluster ensemble methods.

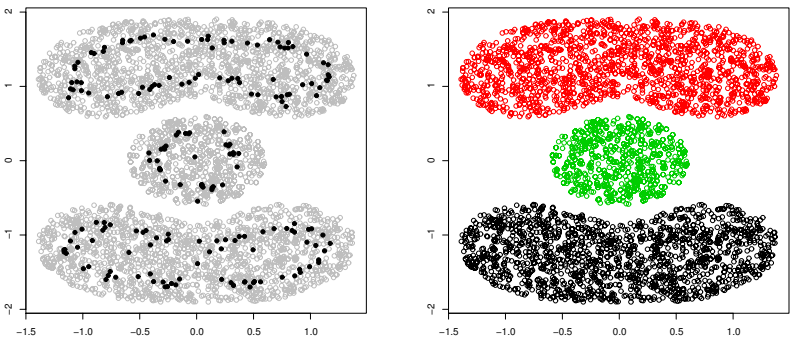

Figure 1: Cassini problem: 200 centers placed by bagged clustering (left) and final solution (right) by combining the 200 centers using hierarchical clustering.

Throughout this paper we use a 2-dimensional artificial data set called "Cassini" to demonstrate some aspects of cluster ensembles. The data set has 3900 objects in 3 groups (see Figure 1): 900 in the interior, 1500 each in the outer groups, all drawn uniformly from the respective shapes. The problem is "hard" for VQ-type base clusterers due to non-convexity of the outer groups, e.g., for the K-means algorithm with $K = 3$ the correct solution is a local minimum only, the global minimum of the K-means objective function splits one of the outer groups into two halves and ignores the inner group.

We apply the bagged cluster algorithm to this data using $B = 10$ bootstrap training samples and K-means as base method with $K = 20$ centers in each run. The

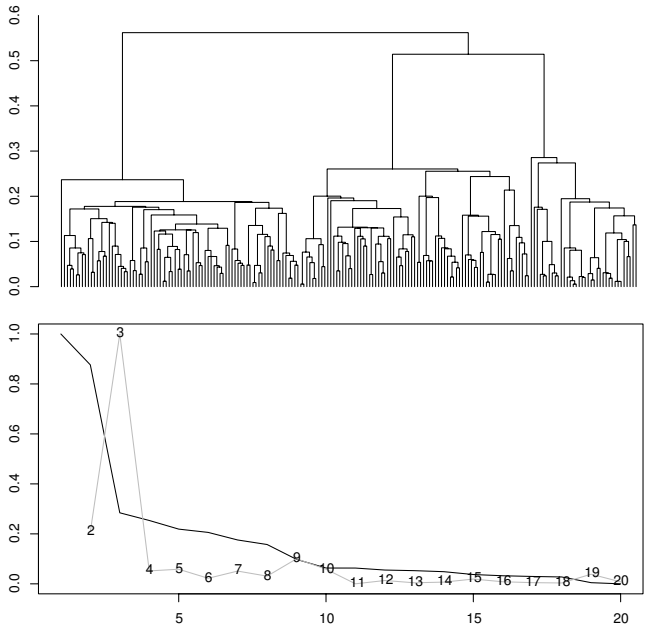

Figure 2: Cassini problem: Hierarchical clustering of 200 bagged cluster centers using single linkage. The upper plot shows the usual dendrogram. The lower plot shows the height of the splits together with their first differences, which can be used to determine the number of clusters (here 3).

left plot in Figure 1 shows the resulting 200 centers. We then perform hierarchical clustering (Euclidean distance, single linkage) on these 200 points, see Figure 2. The three-cluster partition which results from cutting the dendrogram into its three main branches can be seen in the right plot in Figure 1. It recovers the three clusters without error. Note that direct hierarchical clustering of this data set is infeasible due to its size.

3 Aggregation Based on Memberships

If aggregation is to be based on cluster memberships $M_1, \ldots, M_B$ only, a natural way to proceed is by looking for clusterings which "optimally represent" the ensemble. (Note that due to possible relabeling we cannot simply compute average memberships.) Suppose that $d(M, \tilde{M})$ measures dissimilarity (or distance) between two clusterings C and $\tilde{C}$ with corresponding membership matrices M and $\tilde{M}$, respectively. Given d, we can for example look for Ms which minimize average dissimilarity, i.e., which solve

$$M^* = \operatorname*{argmin}_{M \in \mathcal{M}} \sum\nolimits_{b=1}^{B} d(M, M_b)$$

over a suitable set $\mathcal{M}$ of membership matrices M. If $\mathcal{M}$ contains all crisp clusterings, Gordon and Vichi (1998) call M^* the *median partition*, if $\mathcal{M} = \{M_1, \ldots, M_B\}$ it is the *medoid partition*.

We refer to this minimization problem as the (simple) *cluster ensemble problem*. Many extensions are possible, such as minimizing $\sum_b \omega_b d(M, M_b) + \lambda\Phi(M)$, where the ω_b are weights quantifying "importance", and Φ can e.g. measure fuzziness, thus converting the above hard-constrained simple problem into a soft-constrained extended one. Also, one could consider criterion functions resulting in yet more robust solutions, such as the median or trimmed mean of the distances $d(M, M_b)$.

Unfortunately, the simple cluster ensemble problem is computationally very hard. Even if "only" crisp solutions are sought, it would in general be necessary to search all possible crisp clusterings (the number of which is of the order $(K+1)^n$) for the optimum. Such exhaustive search is clearly impossible for most applications. Local strategies, e.g. by repeating random reassigning until no further improvement is obtained, or Boltzmann-machine type extensions (Strehl and Ghosh, 2002) are still expensive and not guaranteed to find the global optimum.

Gordon and Vichi (1998) use the Rand index (Rand, 1971) as distance measure d, while Krieger and Green (1999) use the Rand index corrected for agreement by chance (Hubert and Arabie, 1985). Solving for M^* is NP-hard in both cases, hence the corresponding mathematical programming problems scale bad in the number of observations. Krieger and Green (1999) propose a greedy search algorithm together with "smart" initialization.

The situation can considerably be improved if more information on the structure of the optimal clustering is available. Dimitriadou et al. (2002) use the distance measure

$$d_{\mathrm{DWH}}(M, \tilde{M}) = \min_{\Pi} \|M - \tilde{M}\Pi\|^2$$

where the minimum is taken over all permutation matrices Π. In the crisp case, d_{DWH} counts (a multiple of) the number of differently labeled objects after optimal relabeling. For this distance measure, one can show that the optimal (fuzzy) solution M^* to the cluster ensemble problem is of the form

$$M = \frac{1}{B} \sum_{b=1}^{B} M_b \Pi_b$$

for suitable permutation matrices $\Pi_1, \ldots, \Pi_B$. In the all-crisp case, the aggregated memberships are obtained by simple majority voting after relabeling, which motivates the name "voting" for the proposed framework. Simultaneous determination of the permutation matrices still being computationally hard, the above representation motivates a greedy forward aggregation algorithm where in each step b, a locally optimal Π_b^* for relabeling is determined, and the optimal aggregation M_b^* of $M_1\Pi_1^*, \ldots, M_b\Pi_b^*$ is obtained by on-line averaging. The locally optimal permutation matrix can be determined via linear programming using the so-called Hungarian method for solving the weighted bi-partite graph matching problem (e.g., Papadimitiou and Steiglitz (1982)).

Formally,

$$\Pi_b^* = \underset{\Pi}{\operatorname{argmin}}\, d_{\mathrm{DWH}}(M_{b-1}^*, M_b\Pi)$$
$$M_b^* = (1 - 1/b)M_{b-1}^* + (1/b)M_b\Pi_b^*$$

The final M_B^* is the consensus clustering obtained by "voting".

Figure 3 shows the improvements on the Cassini problem obtained by successive voting using K-means with $K = 3$ clusters as the base learner. Figure 4 demonstrates

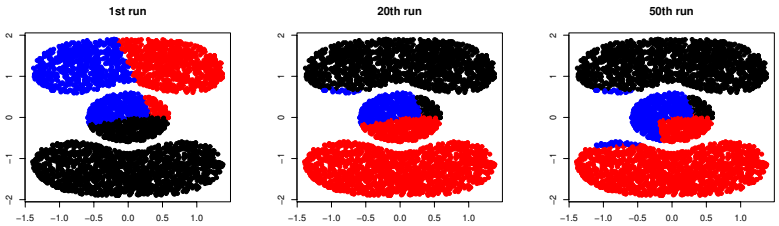

Figure 3: Aggregation by voting on k-means

how voting on the aggregated results of $B = 50$ runs of K-means, Hard Competitive Learning, and an on-line version of fuzzy C-means further improves performance, resulting in (almost) perfect learning of the underlying structure.

4 Summary and Outlook

The main focus of this paper is on aggregation strategies for cluster ensembles. It may be desirable to subject the thus obtained consensus clustering to further computations, such as for collapsing labels representing similar groups, using for example the "Merging" procedure in Dimitriadou et al. (2001). Cluster ensembles can also be used for tuning hyper-parameters of clustering algorithms, such as determining the number of clusters to be employed (Dudoit and Fridlyand, 2002).

Cluster ensembles have already been successfully employed in a wide range of application domains, including market segmentation (Dolničar and Leisch, 2003) and the analysis of fMRI data (Barth et al., 2003). Nevertheless, there is still room for substantial improvements of the underlying theory. For example, it is currently not known under which conditions to solutions to the (unconstrained) cluster ensemble problem can be represented as convex combinations ("weighted voting") of the possibly relabeled membership matrices (including the result of Dimitriadou et al. (2002) as a special case), or can be computed in polynomial time. Such knowledge could result in the construction of substantially more efficient aggregation algorithms, making large-scale application problems computationally tractable.

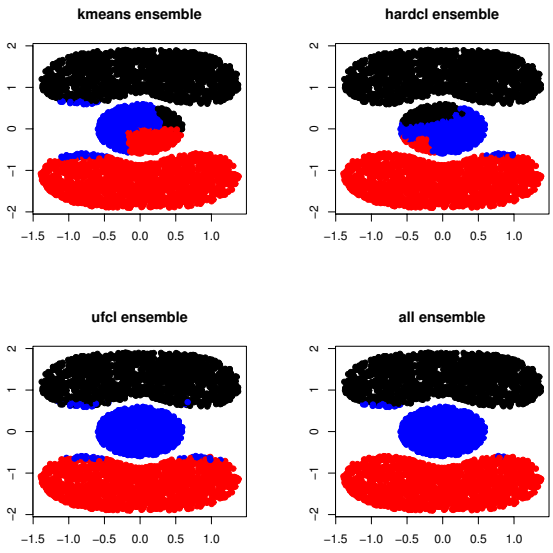

Figure 4: Aggregation by voting on voting on k-means, Hard Competitive Learning, and fuzzy c-means

Bibliography

Barth, M., Dimitriadou, E., Hornik, K., and Moser, E. (2003). Ensemble clustering of fMRI data. In *20th Annual Meeting of the European Society for Magnetic Resonance in Medicine and Biology (ESMRMB 2003)*.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24:123–140.

Dietterich, T. G. (2002). Ensemble learning. In Arbib, M. A., editor, *The Handbook of Brain Theory and Neural Networks*, pages 405–408. MIT Press, Cambridge, Mass.

Dimitriadou, E., Weingessel, A., and Hornik, K. (2001). Voting-merging: An ensemble method for clustering. In Dorffner, G., Bischof, H., and Hornik, K., editors, *Artificial Neural Networks – ICANN 2001*, volume 2130 of *LNCS*, pages 217–224. Springer, Berlin.

Dimitriadou, E., Weingessel, A., and Hornik, K. (2002). A combination scheme for fuzzy clustering. *International Journal of Pattern Recognition and Artificial Intelligence*, 16:901–912.

Dolničar, S. and Leisch, F. (2003). Winter tourist segments in Austria: Identifying stable vacation styles using bagged clustering techniques. *Journal of Travel Research*, 41:281–292.

Dudoit, S. and Fridlyand, J. (2002). A prediction-based resampling method to estimate the number of clusters in a dataset. *Genome Biology*, 3:0036.1–0036.21.

Efron, B. and Tibshirani, R. J. (1993). *An introduction to the bootstrap*. Monographs on Statistics and Applied Probability. Chapman & Hall, New York.

Friedman, J., Hastie, T., and Tibshirani, R. (2000). Additive logistic regression: A statistical view of boosting. *The Annals of Statistics*, 28:337–407.

Gordon, A. D. and Vichi, M. (1998). Partitions of partitions. *Journal of Classification*, 15:265–285.

Hoeting, J., Madigan, D., Raftery, A., and Volinsky, C. (1999). Bayesian model averaging: A tutorial. *Statistical Science*, 14:382–401.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2:193–218.

Krieger, A. M. and Green, P. E. (1999). A generalized Rand-index method for consensus clustering of separate partitions of the same data base. *Journal of Classification*, 16:63–89.

Leisch, F. (1999). Bagged clustering. Working paper, SFB "Adaptive Information Systems and Modeling in Economics and Management Science", Vienna University of Economics and Business Administration.

Papadimitiou, C. and Steiglitz, K. (1982). *Combinatorial Optimization: Algorithms and Complexity*. Prentice Hall, Englewood Cliffs.

Rand, W. M. (1971). Objective criteria for the evaluation of clustering methods. *Journal of the American Statistical Association*, 66:846–850.

Strehl, A. and Ghosh, J. (2002). Cluster ensembles – a knowledge reuse framework for combining multiple partitions. *Journal on Machine Learning Research*, 3:583–617.

Open and Extensible Software for Data Analysis in Management Science

Kurt Hornik and Friedrich Leisch

1 Introduction

Management science deals by definition with the analysis and possible solution of managerial problems using "scientific" methods. In the information age, knowledge is increasingly stored and provided by computers. Terms like *"knowledge warehouse"* for the database of an enterprise are a good symptom for the central place data play in management today. In many cases making managerial decisions is not constrained by the *availability* of data, but by the ability to *efficiently use* the information flood modern computers create.

This turns (significant branches of) management science more and more into a computational science relying on efficient software for data analysis and decision support. Of course management science is not the only discipline affected by this change of paradigm. E.g., statistics also has shifted from a mainly mathematical discipline to a science with strong roots in both mathematics and computer science. The computer is not a mere tool for efficient data handling, it has enabled us to think about problems in new ways and many techniques would be impossible without it. The popular Bayesian models in marketing science rely on Markov Chain Monte Carlo sampling. Simulation-based techniques can help to fill gaps in understanding theoretical and mathematical procedures as well as provide numerical approximations to computationally infeasible exact solutions.

Complicated numerical algorithms must often be used even when we have sound theoretical results. Implementation of these procedures can be just as difficult as the construction of proofs. However, while publication of research papers is based on the verification or proper referencing of proofs for every theorem, there is a tendency to accept seemingly realistic computational results, as presented by figures and tables, without any proof of correctness. Yet, these results are critical for justifying the proposed methods and represent a substantial percentage of the content in many journal articles. In this paper we will discuss how an open source environment for data analysis can help to efficiently reuse knowledge concerning the computational aspects of data analysis in management science, and how to make computational research reproducible (Buckheit and Donoho, 1995; Schwab et al., 2000; Leisch and Rossini, 2003).

2 R: An Environment for Statistical Computing

2.1 The language S

Statistical computing is concerned with turning statistical ideas into software. In doing so, state-of-the-art methodology from theoretical statistics, mathematics and computer science is to be utilized. To be able to satisfy the typical needs in statistical computing, software environments employed must provide a variety of features, including a scripting language for extending the base system and interactive usage, interfaces to programming languages such as C or Fortran, support for adequate high-level representation of complex statistical data types and the "generic" statistical operations performed with these (i.e., support for functional object-oriented programming), and a powerful graphics system.

The most prominent software environment for data analysis and graphics which meets the above requirements is the S system which has been developed by John Chambers and colleagues at Bell Laboratories over the past 30 years. The results of this development effort have been recognized by the Association of Computer Machinery Software System Award 1998 to John Chambers ("for the S language ... which has forever altered the way how people analyze, visualize and manipulate data ..."). The books by Becker et al. (1988), Chambers and Hastie (1992) and Chambers (1998)—also commonly referred to as the "Blue", "White" and "Green" books—describe key features of the S language. A commercial implementation of the S language called 'S-PLUS' is available from Insightful Corporation (`http://www.insightful.com`).

R (R Development Core Team, 2003) is an open source implementation of the S language released under the GPL ("GNU S") and freely available from `http://www.R-project.org`. It is being developed for the Unix, Windows and Macintosh families of operating systems by an international development team ("R Core"), currently consisting of 16 individuals, including the authors of this paper. New versions of R are made available twice per year (spring/fall). Since 2003 the Vienna-based "R Foundation for Statistical Computing" acts as copyright holder, see also Hornik and Leisch (2002). It is a not for profit foundation whose general goals are to provide support for the continued development of R, the exploration of new methodology, and teaching and training of statistical computing.

2.2 Features of R

An R distribution provides a run-time environment with graphics, a debugger, access to certain system functions, the ability to run programs stored in script files, and functionality for a large number of statistical procedures, see Section 3 below. The data analytic techniques described in such popular books as Venables and Ripley (2002), Pinheiro and Bates (2000), or Therneau and Grambsch (2000) have corresponding R packages (**MASS**, **nlme**, and **survival**). In addition, there are packages for bootstrapping, various state-of-the-art machine learning techniques, and spatial statistics including interactions with GIS. Other packages facilitate interaction with most commonly used relational databases, importing data from other statistical software, and

dealing with XML.

R provides a rich resource of both low-level graphics functions and high-level visualization tools such as contour, mosaic, and condition plots. Low-level functionality includes drawing of lines, points and polygons with fine control of color and line types. Mathematical annotation in plot text is available via input expressions reminiscent of TEX constructs (Murrell and Ihaka, 2000). Users can create new visualization tools using the existing graphics components as building blocks. Production quality output can be rendered on-screen or in a variety of common formats, e.g., Postscript, PDF, PNG and JPEG.

Recent developments of R's graphics capabilities feature the introduction of two new packages called **grid** and **lattice**, which represent a move away from the traditional pen-and-paper rendering model. They involve a redesign of the graphics layout capabilities plus some support for interaction, and an implementation of Trellis graphics.

The core of R is an interpreted computer language which allows branching and looping as well as modular programming using functions. Most of the user-visible functions in R are written in R. It is possible for the user to interface to procedures written in the C, C++, or FORTRAN languages for efficiency. Data structures include arrays, lists, and data frames (a construct designed to hold measurements in a cases by variables layout). Many statistical models can be expressed compactly using formula notation. R supports an object-oriented paradigm which allows users to define their own classes and methods, however, R is a functional language with an evaluation model which is based on Scheme (Gentleman and Ihaka, 2000). In essence, everything in R is an object and hence users can easily perform computations on the language.

2.3 R Package Management

In addition to providing an implementation of the S language, two of the key strengths of R are its system for bundling extensions into so-called "packages", and its facilities for quality control which can be applied to both the base system and the extension packages. Packages gather meta-information, R code and documentation, and possibly much more, including data sets, demos, additional (non-R) code that needs to be compiled or interpreted, additional documentation, and self-tests for correct functionality, in a structured layout. There are tools for "checking" them and "building" them into a package source file which subsequently allows for distribution and typically also for plug-and-play installation of the package on all supported platforms, also from within R. It is even possible to automatically update installed packages via the web.

The quality control suite include various tools for analyzing R code and documentation which detect common coding errors or problems and verify completeness and correctness, facilities for computing on packages, and run-time mechanisms for the documented examples as well as additional regression or verification tests. They form the core of a rigorous quality assurance system for both the R base system and existing repositories of R extension packages, most notably the CRAN repository: changes to the R source code are acceptable only if they pass a standard suite of quality control tests ("make check"). Packages distributed via CRAN must pass a similar suite ("R

CMD check", the mechanism for checking packages mentioned above) upon entry, and the tests are re-run on a daily basis for both the current release and development versions of R on all packages in the repository, with various procedures to remedy detected problems in place. By the time of a release (currently, new R releases are made every 6 months), all distributed packages must pass the test suite, ideally without any warnings.

3 R and Management Science

R offers researchers in management science access to an amazingly complete collection of state-of-the-art computational tools for data analysis, including

- linear and generalized linear models, nonlinear regression, random and mixed effects, multinomial logit models
- classical parametric and nonparametric tests
- time series modelling (ARIMA, GARCH, ...)
- hierarchical, partitioning and fuzzy clustering, feature maps
- finite mixtures models and latent class regression
- factor analysis, correspondence analysis, structural equation models
- regression and classification trees, bagging, random forests
- neural networks, support vector machines
- smoothing, generalized additive models, MARS
- bootstrap

and much more.

As a comprehensive overview of all R packages that may be useful in management science is far beyond the scope of this paper, we will concentrate on two particular examples which are very important, e.g., in marketing science: market segmentation and structural equation models.

3.1 Market Segmentation, GLIMMIX and FlexMix

Market segmentation tries to group heterogeneous consumer populations into (more) homogeneous subpopulations. There are two fundamentally different ways of determining segments in a marketplace: If management knows exactly which consumer characteristics are relevant for grouping individuals, the market can simply be split up on the basis of these criteria., this also known as a priori segmentation (e.g., Mazanec, 2000). If, however, management does not have enough prior knowledge to suggest relevant grouping characteristics, segments have to be found or constructed on the basis of information provided by those customers. Typically such information would be the result of an empirical survey or, for instance, a database resulting from a customer loyalty program. This information is used to compute groups of similar customers based on available data and known as a posteriori segmentation.

Most clustering methods can be (and have been) used for market segmentation, recent improvements include our own work on cluster ensembles, see chapter *"Ensemble Methods for Cluster Analysis"* in this volume. Wedel and Kamakura (2001) recommend finite mixture models and latent class regression as state-of-the-art, see

also Williams (2000).

Numerous applications of mixture models have been proposed in the marketing literature, including many extensions and special cases. E.g., mixtures of multinomial logit models can be used for estimating choice models that account for unobserved consumer heterogeneity and mixture models with concomitant variables were used by Gupta and Chintagunta (1994) for profiling market segments based on purchase behavior with respect to demographic variables.

There exists an abundance of software packages for estimating finite mixture models. Most packages, including the commercial GLIMMIX package[1] accompanying Wedel and Kamakura (2001), are stand-alone applications which can handle a few special cases only. Usage requires familiarization with the software and the tedious task of data import/export into the mixture modelling software. Especially a comparison of different model specifications or model estimation methods is difficult. Furthermore, for post-processing and visualization of results these have to be exported to the user's favorite data analysis environment. This is at least inconvenient, because users have to learn many different tools and spend a lot of time transferring data from one program to another. However, the mere existence of software implementing the methods proposed by Wedel and Kamakura (2001) is probably one of the key factors that makes the book a central references in market segmentation, as interested readers can easily try the models on their own data.

FlexMix (Leisch, 2003) is an R package for finite mixtures of regression models containing the GLIMMIX models as a special case. The main design principle was to create an open framework for fitting finite mixtures with the EM algorithm that can easily be extended. If an R function for maximum likelihood estimation of the parameters of a mixture component is implemented in R, plugging it into FlexMix is straightforward. This makes it an ideal tool for both practitioners and method developers. FlexMix uses the usual formula notation for model specification in S (Chambers and Hastie, 1992), hence integrates smoothly into the comprehensive computational environment R and is easy to use for practitioners already using R (or S-PLUS) for data analysis. In addition, method developers can rapidly try out new mixture models, as they only have to provide the M-step of the EM algorithm, with all data handling and user interfaces already taken care of by FlexMix.

Two of the main advantages of using an R extension package rather than a stand-alone application are: R is a general purpose statistical computing environment, hence it can be used for all steps of data analysis, not only the fitting of mixture models. Both R itself and most extension packages (including FlexMix) are open source, hence can be re-used and easily be extended by other users.

3.2 Graphical Models

Structural equation models (SEMs, e.g., Bollen, 1989) are a good example how strong the interaction between theoretical research and applications on the one hand, and software implementations on the other hand, can be. This class of models is more commonly known as "LISREL models" in management science, named after the most

[1] Science Plus Group (formerly ProGAMMA), `http://www.scienceplus.nl/`

popular software package[2] implementing the model. SEMs analyze the relationships among observed and unobserved variables, e.g., the effects of consumer attitudes and intentions on product choice, see Leeflang et al. (2000) and references therein for examples.

Closely related to SEMs are the so-called graphical models (e.g., Edwards, 2000), which can be used to deal with concepts like causality and conditional independence (e.g., between brands and choices given the partitions in a market segmentation). However, although graphical models have now been around for a long time and have shown to have a wide range of potential applications, software for graphical models is currently only available in a large number of specialized packages, such as BUGS, CoCo, DIGRAM, MIM, TETRAD and others[3]. Therefore, we have started an initiative called "gR" ("gRaphical Modeling in R", (`http://www.R-project.org/gR/`) by an SFB-sponsored workshop (gR2002) in Vienna with the purpose of providing next generation graphical models in R.

Bringing graphical modeling into R provides much more than "just" algorithmic challenges. The current modeling facilities of R are based on a variant of the formula calculus by Wilkinson and Rogers (1973) also used in GLIM. This is at least inconvenient, but in fact insufficient, for specifying graphical models. In addition, modeling currently always occurs as fitting models to given data: there is no notion of computing on abstract model objects per se. In a Bayesian context, fitting can be seen as determining a sampler, which does not integrate with the current views of fitted models. Via the gR project, a new system for statistical modeling with R will have to be developed.

4 Conclusions

Software development has become an important part of management science, although often only "behind the scenes". More and more papers present results that are based on complicated simulations or numerical techniques and the percentage of research related to computations is continuously increasing. Although many researchers would never call themselves "software developers" they spend considerable amounts of time programming simulation studies or writing a prototype implementation of a new algorithm to be able to benchmark it against the state-of-the-art.

Reproducibility of results and *incremental development* by reuse of available knowledge have been two of the main principles of science throughout the centuries. If we want to make the computational aspects of management science adhere to these principles, open source software can help a lot. To reproduce tables and figures which are the result of complicated computations, a superficial description of how the results were obtained is in many cases not sufficient to reproduce them. In fact, often only the code used to obtain the results will allow others to verify them.

Proposing a new data analytic technique in a paper without providing a reference implementation of the algorithm means that everybody who wants to try the method

[2] Scientific Software International, `http://www.ssicentral.com/`

[3] See `http://www.R-project.org/gR/` for a link collection

has to reinvent the wheel by reimplementation. Providing code as proprietary software is much better than nothing, but hinders incremental improvements because it is hard to try modifications of existing techniques. The success of open source software projects like the Linux operating system or the R environment for statistical computing are based on the fact that anybody can and may improve the existing code (Torvalds, 1999; Dafermos, 2001). Obviously closed source code cannot be part of the scientific output if computations shall be reviewed by the scientific community, hence convenient tools for distributing scientific open source software are needed. R provides both a comprehensive environment for data analysis as well as the mechanism to easily share methodology via its packaging system.

Bibliography

Becker, R. A., Chambers, J. M., and Wilks, A. R. (1988). *The New S Language*. Chapman & Hall, London.

Bollen, K. A. (1989). *Structural Equations with Latent Variables*. Wiley, New York.

Buckheit, J. and Donoho, D. (1995). WaveLab and reproducible research. Statistics Department, Stanford University, Calif., USA.

Chambers, J. M. (1998). *Programming with Data*. Springer, New York.

Chambers, J. M. and Hastie, T. J. (1992). *Statistical Models in S*. Chapman & Hall, London.

Dafermos, G. N. (2001). Management and virtual decentralised networks: The Linux project. *First Monday*, 6(11).

Edwards, D. (2000). *Introduction to Graphical Modelling*. Springer, Berlin.

Gentleman, R. and Ihaka, R. (2000). Lexical scope and statistical computing. *Journal of Computational and Graphical Statistics*, 9:491–508.

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

Hornik, K. and Leisch, F. (2002). Vienna and R: Love, marriage and the future. In Dutter, R., editor, *Festschrift 50 Jahre Österreichische Statistische Gesellschaft*, pages 61–70. Österreichische Statistische Gesellschaft, Wien.

Leeflang, P. S., Wittnik, D. R., Wedel, M., and Naert, P. A. (2000). *Building Models for Marketing Decisions*. Kluwer Academic, Boston.

Leisch, F. (2003). FlexMix: A general framework for finite mixture models and latent class regression in R. Report 86, SFB "Adaptive Information Systems and Modeling in Economics and Management Science", Vienna University of Economics and Business Administration.

Leisch, F. and Rossini, A. J. (2003). Reproducible statistical research. *Chance*, 16:46–50.

Mazanec, J. (2000). Market segmentation. In Jafari, J., editor, *Encyclopedia of Tourism*. Routledge, London.

Murrell, P. and Ihaka, R. (2000). An approach to providing mathematical annotation in plots. *Journal of Computational and Graphical Statistics*, 9:582–599.

Pinheiro, J. C. and Bates, D. M. (2000). *Mixed-Effects Models in S and S-Plus*. Springer, Berlin.

R Development Core Team (2003). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.

Schwab, M., Karrenbach, M., and Claerbout, J. (2000). Making scientific computations reproducible. *Computing in Science & Engineering*, 2:61–76.

Therneau, T. M. and Grambsch, P. (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York.

Torvalds, L. (1999). The Linux edge. *Communications of the ACM*, 42:38–39.

Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Fourth Edition. Springer, Berlin.

Wedel, M. and Kamakura, W. A. (2001). *Market Segmentation – Conceptual and Methodological Foundations*. Second Edition. Kluwer Academic, Boston.

Wilkinson, G. N. and Rogers, C. E. (1973). Symbolic description of factorial models for analysis of variance. *Applied Statistics*, 22:392–399.

Williams, J. (2000). Fitting regression models to finite mixtures. In *ANZMAC 2000—Visionary marketing for the 21st century: Facing the challange*, pages 1409–1414.