Steffen Christ

Operationalizing Dynamic Pricing Models

Bayesian Demand Forecasting and Customer Choice Modeling for Low Cost Carriers

GABLER

RESEARCH

Steffen Christ

Operationalizing Dynamic Pricing Models

GABLER RESEARCH

Steffen Christ

Operationalizing Dynamic Pricing Models

Bayesian Demand Forecasting and Customer Choice Modeling for Low Cost Carriers

With a foreword by Prof. Dr. Robert Klein

GABLER

RESEARCH

Bibliographic information published by the Deutsche Nationalbibliothek
The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at http://dnb.d-nb.de.

Dissertation University of Augsburg, 2009

1st Edition 2011

Editorial Office: Stefanie Brich | Nicole Schweitzer

Gabler Verlag is a brand of Springer Fachmedien.
Springer Fachmedien is part of Springer Science+Business Media.
www.gabler.de

Cover design: KünkelLopka Medienentwicklung, Heidelberg
Printed on acid-free paper
Printed in Germany

ISBN 978-3-8349-2749-1

“It is very difficult to make an accurate prediction,
especially about the future.”

Niels Bohr[1]

[1] Niels (Henrik David) Bohr (October 7, 1885 – November 18, 1962) was a Danish physicist who made fundamental contributions to understanding atomic structure and quantum mechanics. Bohr is widely considered one of the greatest physicists of the twentieth century.

Foreword

Following the success of low cost carriers, dynamic pricing has become one of the most popular fields of research at the interface of Marketing and Operations Management. However, analyzing the available literature reveals that most publications concentrate on the development of optimization models for price variation. The major challenge of forecasting demand is most often ignored. With his dissertation, Steffen Christ aims to close the corresponding gap by examining the applicability of existing optimization models for dynamic pricing on the specifics of highly volatile consumer markets using the example of low cost air carriers. The explicit objective is the operationalization of theoretically sound standing dynamic pricing models using only realistic input assumptions and retracting to factually available data.

The approach chosen by Steffen Christ is rooted in the development of self learning demand models that calibrate their parameters as data becomes available yielding the option of using the returned results as input to conventional dynamic pricing models. It claims that the developed models can provide parts of the necessary input data in a merely plug-and-play fashion. Generally, the work of Steffen Christ targets the understanding of both relevant input values for dynamic pricing models, forecasting of latent price-independent demand and models that estimate the price sensitivity of such latent demand.

For forecasting of latent demand, Steffen Christ develops a method, which is based on a Bayesian interpretation of linear regression modeling. Here, the parameters of a linear function are not considered deterministic, unknown and to be estimated based on a stochastic data set, but are considered random numbers each with a stochastic distribution that is iteratively learned based on collected data itself seen as deterministic. The method is less prone to distortions if few data is available and explicitly allows the inclusion of subjective or expert knowledge into the estimation process.

The concluding evaluation of different self-learning models is done using a proprietary software environment. The forecast error for latent demand ranges between 16.1 – 17.4% based on single values and 11.2 – 12.7% based on the total demand for a single flight leg.

Furthermore, Steffen Christ considers the estimation of purchase behavior for individuals expressing such latent demand. The chosen method Customer Choice Modeling uses disaggregate observations of discrete and individual customer behavior that depend on the different price points for air travel found in the market and their individual attributes (e.g., flight schedule). Through the combination of multiple databases (the demand protocol from the airline's online channel, its computer reservation system and pricing data collected through web crawlers) Steffen Christ is able to construct a comprehensive data field providing revealed preferences as basis for the estimation of individual purchase behavior.

That later model yields a forecast error of 14 – 27% on the completed bookings for the outbound direction and 26 – 39% for the inbound directions, what is considered satisfactory based on the data limitations concerning bookings received by the considered airline's competition.

In summary, Steffen Christ shows with his work that it is indeed possible to develop forecasting models for both, latent demand and purchase behavior, even in highly dynamic and volatile markets. His excellent work is of high relevance for both, researchers and industry experts in the field of dynamic pricing. I hope, that his work will find many readers and will receive the recognition it deserves.

Prof. Dr. Robert Klein

Acknowledgements

It is an honor for me to thank my adviser and primary reviewer Prof. Dr. Robert Klein for the valuable guidance and continuous support throughout the completion of this work. During numerous discussions I received precious advice with regards to both, methodology and subject matter. Furthermore he has always been a truly inspiring and motivational mentor.

I would also like to show my gratitude to Prof. Dr. Michael Krapp for kindly undertaking the second review and for his important methodical advice reflected in all statistic sections of this work.

I am indebted to all colleagues at the chair in "Analytics & Optimization" for their support of many kinds, particularly Dr. Jochen Gönsch and Dr. Claudius Steinhardt for their ample structural remarks.

It is a pleasure to thank Achim Lameyer and Michael Stellwagen, who foremost made this work possible by supplying not only real-life data but also disclosing lots of their practical experience and knowledge.

Special thanks go to colleagues and friends at McKinsey & Company, Inc. for shared times of balefulness and continuous supply with coffee and distraction – with a special mention to Dr. Anja Lindau who is the pick of the bunch.

This work would not have been possible without the unfailing support, encouragement and love of my fiancée Sarah Hein. The completion of this dissertation is eventually due to her constant motivation and belief.

Ultimately I owe my deepest gratitude to my beloved parents Marianne and Manfred Christ, who have continuously supported and encouraged me in all my endeavors – peaking in the completion of this dissertation.

Steffen Christ

Contents

List of Figures

List of Tables

Nomenclature

AA	American Airlines
AIC	Akaike Information Criterion
BIC	Bayesian Information Criteria
BLUE	Best Linear Unbiased Estimator
CAB	Civil Aeronautics Board
CBS	Central Booking System
CNL	Cross-nested Logit
CRS	Computer Reservation System
DP	Dynamic Pricing
GDS	Global Distribution System
GEV	Generalized Extreme Value
IATA	International Air Transport Association
IIA	Independence of Irrelevant Alternatives
IID	Independent and Identically Distributed
LCC	Low-Cost Carrier
MECE	Mutually Exclusive, Collectively Exhaustive
MMNL	Mixed Multinomial Logit
MNL	Multinomial Logit
MRS	Marginal Rate of Substitution

NL	Nested Logit
OLS	Ordinary Least Squares
PNR	Passenger Name Record
RM	Revenue Management
sMAPE	Centered Mean Absolute Percentage Error
TAPE	Total Absolute Percentage Error
US	United States
VFR	Visiting Friends and Relatives
VIF	Variance Inflation Factor

Mathematical Nomenclature

D	The *latent* demand
d	The *realized* demand
ω	The fraction of latent demand that takes a buying decision (so-called book-to-look ratio)
$f(\cdot)$	The true functional form of the latent demand D
$\tilde{f}(\cdot)$	An approximation of $f(\cdot)$ that constitutes the model of D
$g(\cdot)$	The true functional form of the fraction of latent demand ω that takes a buying decision
$\tilde{g}(\cdot)$	An approximation of $g(\cdot)$ that constitutes the model of ω
J	The considered set of flight departure dates, with $j \in \{1, \ldots, J\}$ being the index of a single flight departure day
T	The forecast/planning horizon, dispersed in single observation points, with $t \in \{T, \ldots, 1\}$ (decremental) being the index of a single period
S	The decision time window of customers, dispersed in single days, with $s \in \{1, \ldots, S\}$ (incremental) being the index of a single period, so that $s = 61 - t$
I	The considered set of observation dates, with $i \in \{1, \ldots, I\}$ being the index of a single observation date, so that $t = j - i$
K	The number of externally determined deterministic parameters in a model, with $k \in \{1, \ldots, K\}$ being the index of a single parameter
M	The number of model-specific basis functions and stochastic coefficients, with $m \in \{1, \ldots, M\}$ being the index of a single function
$\mathbf{x}_t$	The vector of externally determined independent/deterministic variables $(x_{1t}, \ldots, x_{Kt})^T$ in a model at time t

$\mathbf{a}_t$	The vector of model-specific stochastic coefficients $(a_{t1}, \ldots, a_{tM})^T$ in a model at time t
$\boldsymbol{\phi}(\cdot)_t$	The vector of basis functions $\big(\phi_1(\cdot)_t, \ldots, \phi_M(\cdot)_T\big)^T$ used in a model at time t
y_t	The dependent/internally determined target variable in a model at time t
C	The choice set available to decision makers in a specific choice situation, with $i \in \{1, \ldots, C\}$ being the index of the alternatives
N	The number of observations in a specific choice situation, with $n \in \{1, \ldots, N\}$ being the index of individual decision makers
U_{in}	True utility that decision maker n attributes to alternative i
V_{in}	Systematic utility that decision maker n attributes to alternative i as observed and possibly modeled by the analyst
$\mathbf{s}_n$	The vector of personal characteristics of decision maker n
$\boldsymbol{\beta}_i$	The vector of alternative-specific coefficients $(\beta_{i1}, \ldots, \beta_{iM})^T$ in a customer choice model
E_{P_i,p_j}	Elasticity of the choice probability P_i for alternative i with respect to a change in fare p_j of alternative j

Mathematical Notation

Vectors are denoted by lower case bold letters such as $\mathbf{x}$ and all vectors are defined as column vectors. A superscript T denotes the transpose of a vector, so that in turn $\mathbf{x}^T$ is a row vector. Henceforth, the notation $(x_1, \ldots, x_K)$ denotes a row vector with K elements, while the corresponding column vector is written as $(x_1, \ldots, x_K)^T$.

Upper case letters such as $\mathbf{X}$ denote matrices. Single elements of a matrix are denoted by X_{kl}, with k being the row and l being the column index of the matrix. Again, a superscript T denotes the transpose of a matrix. The identity matrix is denoted by $\mathbf{I}$.

To distinguish current from past observations (where necessary), vectors or matrices containing new observations are marked with an upside-down bow $\breve{\mathbf{x}}$.

Following literature convention, estimates of model parameters based on observed data are denoted by a hat $\hat{\mathbf{x}}$.

Part I

Dynamic Pricing in the Airline Industry

Chapter 1

Introduction

To lay the foundation for this work's motivation in Chapter 2, this chapter provides a quick overview to the specifics of the passenger airline industry in general and details two recent, significant changes in its structure and conduct in particular.

After a short introduction to the industry and its traditional business model in Section 1.1, the advent of low-cost or low-frills airlines respectively is discussed in Section 1.2. The simplified dynamic pricing schemes introduced by the latter are then explained in Section 1.3.

1.1 The Passenger Airline Industry

The overall airline industry is a typical *service industry*, providing transportation for either goods (cargo airlines) or humans (passenger airlines). It is part of a larger *aviation industry* value chain that also comprises manufacturers, financing companies, distribution companies (including travel agencies and distribution system providers), airport handling services (ground handling, fueling and catering) and the airports themselves.

This work focuses on passenger airlines, as with 401bn USD in 2007, those form the major revenue block of the industry – compared to cargo, with 58bn USD (IATA, 2008a).

The passenger airline industry boosts a variety of different business models (see Table 1.1), from those of the more traditional *global network carriers* that mostly emerged from state-owned flag carriers (oftentimes still with substantial financial backing from the government), via those of the smaller *regional* and dedicated *charter carriers*, to those of the youngest breed of so-called *low-cost carriers* (see also Joppien, 2006 and Pompl, 2006).

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6_1,

Business model	Example	Description
Global network (major)	British Airways	Global route network, typically based on a hub-and-spoke system feeding long-haul flights with regional traffic based on complex integrated service products. Nowadays typically part of a global alliance with full interline capability.
Global network (niche)	Austrian Airways	Own global network confined on selected (niche) markets where differentiated and competitive products can be established. Global network offer is typically based on alliance partners.
Regional (primary)	Eurowings	Connects major regional centers using medium-size turboprop or jet aircraft (up to 120 seats) often providing premium class/full service with comfortable seat pitch. Typically serving hubs of network carriers. Low seat load and smaller market size is offset by higher yields.
Regional (commuter)	Augsburg Airways	Connects regional airports with major cities/hubs using turboprop aircraft (19–80 seats).
Charter	BlueWings	Variable "schedule" depending on demand and seasonality. Distribution of entire seat blocks to tour operators who determine destinations, frequency and flight-times. As nowadays tour operators are reluctant to carry the business risk of fixed inventory, many charter carriers started selling single seats.
Low cost	Ryanair	Point-to-point traffic in selected high-volume markets, typically concentrated on short-haul destinations using secondary airports only. High frequency of flights using standardized equipment. Limited in-flight and ground services as well as absence of transfer traffic, interlining and alliance partnerships.

Table 1.1: European passenger airline industry structure
Source: Following Lawton (2002, pp. 41)

In their study of 25 different industries, Knudsen et al. (2005) report that the airline industry is one of those few where the middle market is "vanishing". Indeed, besides Asia, where low-cost carriers are still in their infancy, mostly incumbent global network players and young low-cost carriers are proving able to survive the current industry consolidation process (see also below Section 1.2 as well as O'Connell and Williams, 2005, pp. 260).

While the emergence of the low-cost business model is mostly due to the deregulation in the 1980s–1990s (see Section 1.2), the ongoing industry consolidation process follows from traditionally low industry profitability in conjunction with rising fuel price pressure and a global financial crisis in 2008. Figure 1.1 plots overall world industry performance, showing two significant characteristics: first, overall profitability is cyclical, with upturn

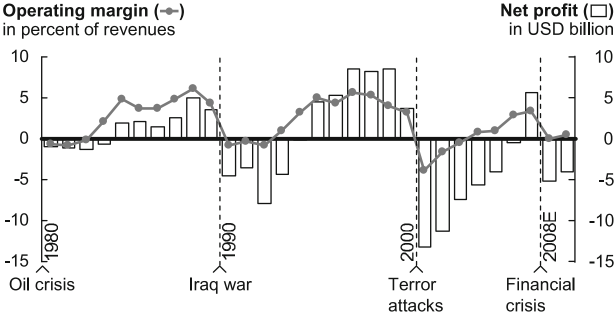


Figure 1.1: Airline industry performance 1980–2009 (estimated)
Source: own design, based on ICAO data for operating profits up to 2007 and IATA estimates for 2006-07 net profits as well as 2008-09 performance forecasts (IATA, 2008b)

cycles typically compensating the effects of preceding downturns; and second, the operating profit margin never rises significantly above 5%.

Although Figure 1.1 only contains estimates for the fiscal years 2008 and 2009, it seems obvious that the last industry upturn cycle was too short for airlines to be able to recover the losses following the September 11, 2001 terror attacks. As a result of heavy oil price increases (and with them, jet fuel price increases) in the summer 2008, the industry is under heavy consolidation pressure, with many incumbent players buying out their rivals or even acquiring low-cost players.

In general, the airline industry faces rather difficult economics: based on expensive assets and high flight-variable cost[1], air transport is a mostly fixed cost-driven business with extremely low marginal costs for an eventually perishable product (i.e., if an empty seat on a plane is not sold, its value perishes at once). Additionally, air travel is considered a mere commodity that therefore is heavily shopped by customers looking for low prices (see Figure 1.2 or Rhoades and Waguespack, 2005 as well as Rothkopf, 2009).

The above-described effects and characteristics foster fierce price competition between carriers through the usage of savvy price management systems employed to optimally adjust prices to the customers' expected willingness to pay.

1 *Flight-variable costs* are variable costs that are incurred in blocks if a particular flight takes place – i.e., they are not variable by single passenger.

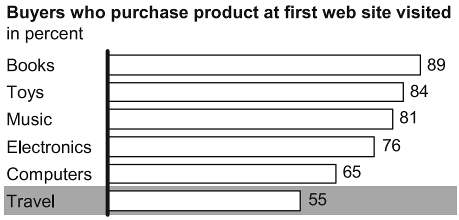


Figure 1.2: Shopping behavior by type of item purchased
Source: following Baker et al. (2001, p. 56)

Before Section 1.3 finally gives a brief introduction to such particular pricing systems, the relatively young business model of low-cost carriers is introduced in the next Section 1.2, as its specifics form the basis for these dynamic pricing schemes.

1.2 The Low Cost Revolution

The concept of low-cost or low-frills airlines, respectively, emerged through the deregulation of air transport in the early 1980s in the United States (US) and the early 1990s in Europe.

Since World War II, the Civil Aeronautics Board (CAB) in the US and its counterparts in Europe did successfully regulate the commercial air transportation market with the objective of *affordable fares* – mostly in the short-haul markets (possibly subsidized by earnings from higher fares on long-haul routes) – and *sufficient flight frequencies* between major source markets. Intended market or route entries by an airline had to be approved by the authorities, with fares regulated and published by an international trade body, the International Air Transport Association (IATA). The resulting industry structure was based on an oligopoly market with inherently low cost-efficiency.

After the advent of full airline liberalization in the US through the Airline Deregulation Act in October 1978, a wide array of new airline start-ups – many solely competing on the basis of price – entered the market. Southwest Airlines is credited by most authors with having been the *first successful* carrier with a consistent focus on keeping down operating costs to efficiently compete with the incumbent players.

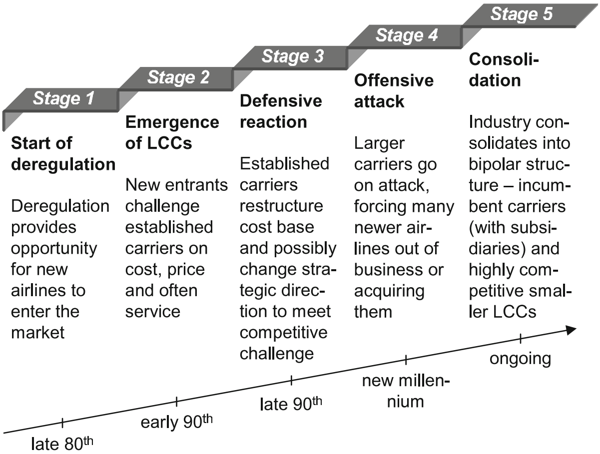


Figure 1.3: European post-deregulation competitive cycle
Source: own design, based on Lawton (2002, pp. 37)

Among the established carriers, the unfamiliar competition led to massive losses for a number of major players: between 1978 and the end of the last century, seven incumbent airlines (America West Airlines, Braniff, Continental, Eastern, Midway, Pan Am, and TWA) plus over 100 smaller airlines went bankrupt or were liquidated, including most of the many new airlines founded in deregulation's aftermath (Lawton, 2002, pp. 35).

In the 1990s, a study conducted by the United Kingdom Civil Aviation Authority described the emergence of a "third way" of air travel in the European aviation industry as well: "Third way (low-fare) airlines bring together costs at the charter level with the convenience (if not the comfort) of scheduled services"(CAA, 1998, p. 125). The undisputed European pioneer in low-cost air service is Ireland-based Ryanair, which heavily draws from the archetypal low-cost model of Southwest Airlines.

Figure 1.3 illustrates the post-deregulation development in European airline industry conduct. The new low-cost carriers (LCC) were able to start from scratch without the burden that the incumbent players carried in terms of route networks, equipment and labor costs. They henceforth often hit those established players simultaneously on three major dimensions: cost, price and sometimes even service. It took the incumbent flag carriers until the new millennium to finally draft responses and re-attack the new players. Since the downturn following the September 11 terror attacks (see Figure 1.1), the industry has found itself in consolidation with flag carriers to some degree even entering low-cost markets with its own subsidiaries.

Beyond other levers (see below), LCCs achieve their competitive advantage by offering fewer "frills" in terms of in-flight and airport services. For that reason, low-cost carriers are sometimes also referred to as *low frills* airlines[2]. In the context of this work, the two terms are taken as synonymous.

The plain Southwest and Ryanair LCC model is defined by the tight management of a number of major cost and complexity levers (see e.g., Lawton, 2002 or Binggeli and Pompeo, 2002):

- **Tariff structure:** Unified tariffs for single-class products without the option to refund (or sometimes even rebook) tickets. Tariffs do not allow connections in general, and there is no interlining with other carriers in particular.

- **Distribution:** Sales are managed exclusively through own channels (primarily online) without system connection to the costly traditional global distribution systems. Tickets are issued as paperless tickets (so called e-tickets).

- **No frills or services:** No additional service measures besides pure air transportation alone:

  - **Before flight:** No advance seat selection and no check-in or boarding priorities that complicate the handling process.
  - **In-flight:** No complimentary meals or beverages and no in-flight entertainment, as those are considered unnecessary frills.
  - **After flight:** No frequent flyer program or other business traveler perks (e.g., airport lounges or dedicated business services).

- **Utilization:** Focus on short-haul point-to-point traffic, resulting in higher utilization of aircraft and crew.

- **Airport charges and handling:** Concentration on smaller secondary airports to reduce taxes and fees as well as ground handling costs and turnaround times.

[2] Europe's Ryanair even calls itself a "*low fares* airline", which indeed typically results from its lower operating cost basis and is naturally a better selling proposition than that implied by the term "low frills".

- **Standardized fleet:** Utilization of single-aircraft typed fleets (typically Boeing 737 or Airbus A320 family) with single class configuration and lower seat pitch (up to 15% more seats per plane than incumbents enjoy).

In 2001, for Ryanair, the tight management of above cost levers resulted in an approximate 63% cost advantage per production unit (i.e., available seat kilometers) compared to the top three incumbent airlines in Europe (see Figure 1.4). This massive advantage in operating costs naturally transfers to lower ticket prices for the customers, potentially turning away business from the incumbent players, what indeed has been the widespread fear in the 90s.

While this effect was partially visible in the early stages of the US industry performance after deregulation, in Europe "the entry of these airlines has generally led to substantial stimulation of *new air traffic* without serious detriment to incumbents' operations" (CAA, 1998, p. ix). Henceforth, European incumbents were mostly pushed to improve the cost-side of their

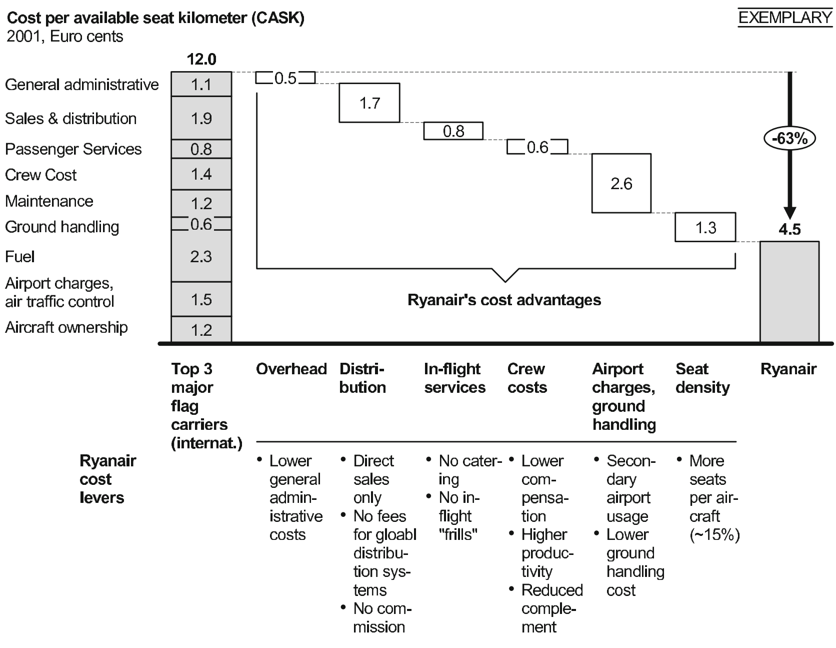


Figure 1.4: Low cost carrier cost advantage
Source: following Binggeli and Pompeo (2002, pp. 90)

businesses to be able to compete on price on selected routes with direct head-to-head LCC competition. However, the feared drain of market share did not materialize (see also Binggeli and Pompeo, 2005).

In recent years, the target customer segment of many LCCs has shifted considerably increasing the dynamics in the market. The traditional segment "visiting friends and relatives" (VFR) has been more and more amended by price-sensitive business customers. Many companies (e.g., self-employed and small corporations) refuse to put employees in business class for less than three hour flights (Lawton, 2002, p. 37). And in fact, short haul business travelers, en mass, are becoming increasingly price-sensitive and move towards LCCs as primary choice for air travel (Mason, 2001, p. 104).

The described customer behavior spurs the alignment of traditional low cost business models and that of global network carriers (see Table 1.1). On the one hand side, incumbents get leaner and more flexible, lowering their cost basis to finally being able to offer extremely low saver fares on a selected basis (e.g., Lufthansa's "betterFly" offer at EUR 99 – round trip). On the other hand, LCCs increasingly target business customers by adding selected services or "bells and whistles" to satisfy particular demand drivers (see Klingenberg, 2005, Sec. 2).

As a result, today many "low cost" carriers have deviated from the pure LCC business model and provide balanced products depending on the stage length of the flight and expected preference structure of their customers: In the US, *JetBlue* boosts leather seats with in-seat entertainment as well as complimentary snacks and beverages, while *Virgin America* even added "first class" seats to the front of the cabin. In Europe, *EasyJet* is flying directly to primary airports and *Air Berlin* also serves complimentary snacks and even provides a frequent flyer program with status amenities (business class check-in, fast lane security and lounge access on long haul flights).

This work refers to the last described carriers as *hybrid LCC*, because their cost structure and management typically mirrors that of traditional LCCs, while their business model in terms of target customer segments is a hybrid between VFR and business travelers.

LCCs are an interesting subject for academic pricing studies because they typically operate in highly dynamic growth markets with scarce historic data or demand information. This in turn prohibits the employment of traditional forecasting and planning techniques from revenue management, which is partially resolved by the development of online pricing support tools in the body of this work.

Besides the tight management of cost drivers as described above, LCCs distinguish themselves by the employed pricing scheme, which is introduced in the next section.

1.3 The Advent of Dynamic Pricing

In the eyes of Dutta et al., pricing in general does not reserve the appropriate attention in many companies, while it should clearly be considered a “strategic weapon” for competitive advantage. In fact, “for many companies, pricing capabilities are increasingly critical to their ability to implement their strategies” (Dutta et al., 2002, p. 62), which is particularly true for airlines – especially low cost carriers.

Based on the low profit margins (see Figure 1.1) in the airline industry – more than in many other industries – price is generally a major profit lever. Across industries, Marn and Rosiello have calculated that a 1% improvement in average price paid can trigger an operating profit improvement of 11.1% (see Figure 1.5). Consequently, “the fastest and most effective way for a company to realize its maximum profit is to get its pricing right” (Marn and Rosiello, 1992, p. 84).

However, an optimal pricing strategy is not necessarily bound to generally high prices or maximum revenues, especially in industries with high fixed costs and perishable goods (also see Section 3.1). “In fact, the opportunity to capture an incremental sale at a lower price may be as compelling as ratcheting up sales to consumers who are less price-sensitive” (Heun, 2001, p. 1).

In 1978, American Airlines’ (AA) vice president of marketing, Robert Crandall, was the first to leverage the above idea to effectively compete with the lower cost and price structures of the newly emerging LCCs, utilizing

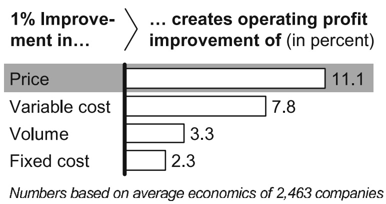


Figure 1.5: Comparison of average profit levers
Source: following Marn and Rosiello (1992, p. 85)

AA's excess inventory, which was already produced at marginal costs near zero to offer what was called "Supersaver fares". The "only" two problems that had to be solved were the following: first, the excess inventory had to be identified correctly on each flight to prevent the displacement of business travelers in favor of selling cheap tickets; and second, a mechanism was needed to prevent the down-selling of business travelers willing to pay the full ticket fares.

A new revenue management system solved precisely these two problems: so-called fencing rules[3] prevented high-value segments from defecting to the cheaper Supersaver fares, and advanced algorithms were introduced to forecast the excess inventory on each flight. In 1985, AA introduced the first full-scale revenue management system called DINAMO that allowed it to basically match or undercut all fares of competing carriers where excess inventory was available. An immediate effect attributed to the new system was the bankruptcy of PeopleExpress, a young and directly competing LCC (see Cross, 1997). At that time, according to Smith et al. (1992), AA estimated the ongoing revenue contribution of its yield management capabilities to be over 500 million USD annually.

American Airlines' DINAMO marked the advent of *revenue management* (RM) systems, which can be considered dynamic pricing schemes in a wider sense. However, while in revenue management, prices vary across differentiated sub-products or customer segments (based on fencing rules), in *dynamic pricing* (DP) in a narrow sense, prices fluctuate for a single product and over all customers – i.e., at any one time, there is only one price posted, and this price is accessible to all customer segments.[4]

Recent years have seen a twofold increase in the adoption of dynamic pricing policies across industries: on the one hand, the spectrum of industries that apply dynamic pricing mechanisms has widened (see, e.g., Coy, 2000), with examples ranging from United Parcel Service Inc. to Amazon.com; and on the other hand, industries whose use of revenue management is an established practice have started to move toward dynamic pricing schemes.

The airline industry is an example of the latter, as the above-discussed

---

[3] Fencing rules segment customers into groups with supposedly different levels of price sensitivity based on characteristics inherent to them. Typical fencing rules are *advance purchase rules* (e.g., the ticket has to be bought a minimum of 30 days in advance), required *Saturday night stay* or *weekend stay*, *minimum stay* (e.g., 7- or 14-day trip length), *round-trip obligations*, *rebooking restrictions* and *non-refundability* of tickets.

[4] Section 3.1.2 gives a more thorough and detailed introduction to revenue management, dynamic pricing and the defining differences.

proliferation of low-cost carriers has sparked the emergence of single-class fare structures that do not support traditional revenue management practices but are rather prone to the dynamic modifications of a single fare (see Elmaghraby and Keskinocak, 2003, p. 1287). "The same rules and restrictions apply to all of the tickets sold; therefore, the tickets are identical and the price is assumed to vary only with the time left until departure" (Anjos et al., 2004, p. 535).

In the airline industry, the emergence of the Internet and increasing computing power have driven the phenomenon of incumbent players moving into DP and academics even discussing it as the basis for entire new business models (see Cortese and Stepanek, 1998 or Burger and Fuchs, 2005). The reasons behind this trend are threefold:

- **Buy-down:** Based on improved price transparency in today's online markets, customers are increasingly buying down in traditional revenue management environments, sometimes even circumventing fencing rules (see, e.g., Boyd and Kallesen, 2004).
- **Menu costs:** The Internet and other direct sales channels allow frequent price changes based on a much lower menu cost (see, e.g., Brynjolfsson and Smith, 2000 or Kannan and Kopalle, 2001).
- **Technology:** The increasing availability and usability of advanced decision-support tools for dynamic pricing and fare monitoring allow well-founded near real-time price adjustments (see, e.g., Elmaghraby and Keskinocak, 2003).

While DP is appealing as a result of its simplicity in terms of basic functionality (i.e., there are no booking classes or fences but only a single posted price), the price-setting mechanism has to be highly sophisticated, as less price-sensitive demand inevitably buys down if that single posted price is set too low.

In academia, the field of dynamic pricing is relatively young, but a considerable amount of literature exists on the general topic (see Chapter 3). However, the recent work of Marcus and Anderson (2008) contains one of the few models that explicitly target *dynamic pricing for low-cost carriers* in consideration of their own particularities.

The next chapter describes the motivation behind this additional work on the topic and also gives an overview of its structure.

Chapter 2

Motivation and Structure

Following the introduction to the relevant underlying subjects of this work that has been given in the previous chapter, the particular motivation for selecting the overall topic and taking the specific approach are discussed here. First, Section 2.1 addresses the question of the topic's overall relevance. This is followed by a justification of the specific focus on the airline industry in Section 2.2. Finally, Section 2.3 details the particular objectives and differentiating elements of this work before its structure is summarized in Section 2.4.

2.1 Relevance of the Topic

The previous chapter has highlighted the major importance of advanced pricing capabilities for any successful business model. Based on the low profitability of airlines in particular (see Section 1.1), smart pricing is an even bigger driver of sustainability for them.

Talluri and van Ryzin (2005, pp. 17) structure the general approach to revenue management and dynamic pricing as comprised of four steps: 1) data collection, 2) estimation and forecasting, 3) optimization and 4) ongoing control. While the overall performance of these steps depends heavily on the quality of each individual measure, "most of the existing revenue management literature focuses on the optimization step and assumes that all parameters are known and demand uncertainty is asystematic and unpredictable" (Xu and Hopp, 2004, p. 2). However, Lin (2006, p. 523) acknowledge that in most models, "there are two major sources of randomness in demand: *customer arrival rate* and *customer reservation price distribution*".

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6_2,

The necessary differentiation between the two gets easily comprehensible by looking at the *Bellmann functional equation* of the general dynamic price optimization problem with the objective of maximizing total revenue[1] (for a general introduction to Bellmann equations, see Bellmann, 1957; for the specific problem at hand, see e.g., Bitran and Mondschein, 1997, p. 67).

Assume a setting where a seller has on sale a finite capacity c of a perishable product. The remaining time until spoilage of this capacity is divided into t equidistant segments, such that at most one customer may arrive at the point of sale in any one period. The actual arrival of customers latently interested in the product is uncertain and defined by some cumulative probability distribution $F(\lambda_t)$ characterized by the average rate of customer arrival λ_t in a single period t (i.e., $\lambda_t \leqslant 1 \; \forall \; t$). Additionally, customers inquiring for the product have a certain reservation price distribution defined by $G_t(p_t)$, returning the probability of an actual purchase at a specific time t under a posted price p_t.

The resulting dynamic program for the maximization of total revenue $R(t, c)$ by variation of the posted price p_t over the remaining time t and available capacity c is then given by

$$\begin{aligned} R(t,c) = max_{p_t}\{ \quad & F(\lambda_t) \cdot G_t(p_t) \cdot [p_t + R(t-1, c-1)] \\ + \quad & F(\lambda_t) \cdot (1 - G_t(p_t)) \cdot R(t-1, c) \\ + \quad & (1 - F(\lambda_t)) \cdot R(t-1, c) \quad \}, \end{aligned} \tag{2.1}$$

with boundary conditions

$$R(0, c) = 0 \; \forall \; c \qquad \text{and} \qquad R(t, 0) = 0 \; \forall \; t.$$

For each period, (2.1) differentiates between three possible outcomes (divided into separate lines above): a) a customer arrives and purchases the product, increasing the total revenue by the posted price and diminishing capacity by one for the next period; b) a customer arrives, but does not purchase the product, yielding no additional revenue but full capacity for the next period and c) no customer arrives, naturally preserving the same capacity for the next cycle. Finally, revenue plumages to zero in case the remaining capacity perishes after t periods or in case remaining capacity has dropped to zero.

---

[1] Under the assumption of a high share of fixed cost, the maximization of revenue can be taken as sufficient approximation of the maximization of profits, what should be the natural focus of any company (see, e.g., Klein and Steinhardt, 2008, Sec. 1.2.1).

For being able to solve (2.1) to optimality (under varying assumptions), "most existing literature concerning dynamic pricing assumes that both customer arrival rate and customer reservation price distribution are well known before the sale begins" (Lin, 2006, p. 523). However, recent work does acknowledge that "the situation faced by revenue managers in practice is different in at least two key regards: Assumptions may be incorrect, and model parameters are not known" (Cooper et al., 2006, p. 968).

At the same time, for many (especially dynamic) industries, it is barely possible to forecast the necessary optimization model inputs before the actual sale begins. This in turn prevents many revenue management and dynamic pricing practitioners from adopting the latest thinking and optimization approaches (see e.g., Lin, 2006 or Lobo and Boyd, 2003).

Spedding and Chan (2000, pp. 331) summarize the inherent limitations of existing forecasting models and techniques, which are even more pronounced in the case of low cost air travel:

> *"Traditional forecasting approaches are based on characterizing the structure of historical time series and then predict future events based on that structure. Obviously, the structure of the time series may change in a volatile business environment."*

In terms of practical usability, the above quotation translates into four significant sources for improvement that are tackled in this work:

- **Dynamics:** Most models are incapable of mastering dynamic or structural changes within time series data.
- **Data requirements:** Typically, model identification or validation is not possible with a limited historical dataset.
- **Competition:** Current models often consider only the monopolistic case with reservation price distributions or purchase probabilities at the maximum implicitly depending on competitive price setting.
- **Cycle length:** Forecast stability usually demands stable patterns over time, and henceforth, most models cannot handle short life cycles of product or service demand.

Henceforth, in many real-life situations caused by the lack of available useful information, forecasts are largely based on subjective considerations. Many price-setting firms would accordingly benefit from forecasting models that rely on less historic information, "especially in a dynamic environment where the quantity demanded may change even at a constant price" (Balvers and Cosimano, 1990, p. 882; see also Spedding and Chan, 2000, pp. 331).

Bayesian forecasting techniques as used in Part II can be employed to solve such problems because Bayesian inference is based on the idea that posterior knowledge (e.g., actual demand to come) can specifically be derived from prior knowledge (e.g., recent demand, demand for substitute products, the management's experience, etc.) and the general likelihood of product demand (see, e.g., Box and Tiao, 1973; Jeffreys, 1961; Lee, 1989 or Press, 1989).

In this context, Bayesian inference is particularly relevant in dynamically changing environments, and "online learning methods are particularly appropriate in situations where historical data are scarce or irrelevant" (see Levina et al., 2006, p. 2; see also Spedding and Chan, 2000, p. 332), which is the case in the airline industry.

Customer choice models as used in Part III allow to model the expected behavior of latent demand explicitly taking into account competitive influence in oligarchic markets. Additionally, meaningfully specified models may also take into account the influence of dynamically changing attributes that drive purchasing behavior.

2.2 Focus on the Airline Industry

Sections 1.1 and 1.2 have already highlighted the specific nature of the airline industry that recommends it as an interesting basis for this study.

Firstly, the inherent low profitability in combination with the advent of the low-cost business model foster price competition as such, with industry characteristics making it prone to revenue management in general (which has already been broadly employed since the 1980s) and dynamic pricing schemes in particular (which have just started to move into the industry).

Additionally, the high share of online or direct bookings for low-cost carriers allow both, the logging of latent demand information and the collection of pricing data from competitors (see, e.g., Nason, 2007, p. 65). While the basic technology is available, only a few low-cost airlines actively and persistently collect such data – not to mention actually using it in a systematic way.

At the same time, other industries with similar characteristics and starting points successfully employ online learning and forecasting methods (see, e.g., Fisher and Raman, 2000). For the retail industry, Fisher and Raman (1996) report a possible increase in profits of 60% solely based on accurate responses to revised forecasts early in the sales season.

This work aims to employ self-learning Bayesian schemes and customer choice analysis to similarly leverage existing real-time data for dynamic pricing models in the airline industry.

The next section introduces the specifics of the approach regarding dynamic pricing for low-cost carriers tackled in this work.

2.3 Objective and Differentiation

This section highlights the differentiating elements of the employed approaches for understanding of latent demand and customer decision making as well as the underlying objectives of the developed Bayesian and discrete choice models.

The work at hand differentiates itself from existing dynamic pricing literature (see Section 3.2.2) by establishing its main focus on the improvement of the *input data* of existing optimization models rather than the optimization scheme itself. As just introduced in Section 2.1, the relevant input is twofold: arrival of latent demand and behavior of such demand depending on its reservation price distribution.

Part II differs from the recent literature on online demand learning (see Section 3.2.1) in that it employs a Bayesian learning scheme for latent demand that is *structurally fixed* but still *adjusts automatically* to micro-environmental changes. The customer choice model in Part III is based on automatically collected revealed preference data and is henceforth adaptable to dynamic markets by its very nature. Both rely on current *real-time information* from each particular sales request and the corresponding possibly completed booking.

The general approach described in this work has emerged based on a series of objectives that may foster its practical applicability while at the same time making it academically challenging and relevant:

- **Compatibility:** The developed forecast models should amend existing optimization schemes discussed in the literature in the sense that they should provide the necessary data input or reduce the uncertainty of it. Additionally, they should be able to act as a replacement for other learning or forecasting models discussed in the literature.

- **Online based:** The solution should rely primarily on recent data collected in real time for the actual sales event under consideration. Functionality should not be hampered by the scarcity of historical information.

- **Intervention capable:** The possibility for subjective intervention or for the amendment of the data should be included in a structured form (i.e., not based on manual manipulation of the results). The model should accommodate the uncertainty of such interventions and accordingly reflect different confidence levels for such.

- **Statistically sound:** The results should be derived from a statistically sound model in an analytically structured way. Also, the stochastic nature of the predictions should be reflected.

- **Industry specific:** The functional composition should exploit the characteristics of the airline industry where appropriate, while being flexible enough to allow for the possible transfer to other industries.

- **Dynamic:** The dynamic nature and changing strengths of market drivers should be reflected, with the forecasting scheme ideally automatically adjusting for changing environments without the need for manual intervention.

- **Data frugal:** Acceptable results and stable models should be achievable with a limited amount of input data or manual compilation.

- **Computationally efficient:** The necessary calculations should be computationally efficient – i.e., the model ought to easily scale to a larger forecasting basis.

Parts II and III address these objectives throughout model development to understand latent demand development (Part II) and customer choice behavior (Part III) respectively. The general structure of the remaining parts is given in the next section below.

2.4 Structure of Work

To build upon the introduction and motivation that has already been given in Chapters 1 and 2, Chapter 3 below furnishes a more detailed introduction to dynamic pricing. The exact definition, together with current limitations and shortcomings, is discussed prior to the presentation of a thorough overview of general dynamic pricing literature and online learning literature in particular. The chapter ends with a global definition of the proposed learning approach to forecasting latent demand in Part II and approximating price sensitivity in Part III.

Part II begins with an introduction to specific self-learning linear models based on Bayesian statistics in Chapter 4. The particular dataset on which the part's analyses are based on is described in Chapter 5, where the characteristics of the observed demand for low-cost air transportation are highlighted. Based thereon, a specific forecasting model following the above objectives is developed and mathematically specified in Chapter 6. After the model has been validated for statistical significance in the same chapter, its actual forecasting performance is evaluated and discussed in Chapter 7. There, potential improvements to the results that might be possible through informed learning and aggregation are also included and tested. Finally, the findings are summarized in Chapter 8, where concrete recommendations for the model's employment are eventually given.

Part III starts with an introduction to discrete customer choice analysis in Chapter 9. The extended dataset on which the part's later analyses are conducted is defined and described in Chapter 10. Based on the identified characteristics and structures, Chapter 11 develops a specific choice model, namely a multinomial logit, for understanding of customer decision making in low cost air travel markets. The resulting directional models are validated and evaluated for practical application in Chapter 12. Finally, Chapter 13 summarizes the findings and gives an outlook on possible further research and developments.

Chapter 3

Dynamic Pricing

The present chapter's objective is threefold. First, it locates advanced pricing and revenue optimization schemes within the general pricing context in order to highlight their application areas and principal characteristics (Section 3.1). Second, it gives an overview of prevalent literature and optimization models within that context (Section 3.2). Third, it discusses the shortcomings and limitations of dynamic pricing models developed throughout the literature (Section 3.3) to finally highlight a suitable scope for the work at hand (Section 3.4).

3.1 Definition and Scope

This section gives a short introduction to pricing in general in Section 3.1.1 and highlights the specifics of dynamic pricing in particular in Section 3.1.2.

3.1.1 Introduction to Pricing

For general service and product pricing, marketing-oriented pricing literature and research mostly distinguish four stages of pricing capability that a firm may traverse (see, e.g., Phillips, 2005, Chap. 2):

- **Cost-plus pricing:** The oldest and still most popular approach to price-setting. Prices are determined based on the cost of production plus a given markup. While this procedure entails a compelling simplicity, it has multiple obvious drawbacks: a) it is entirely inwardly focused, ignoring competition and customers' willingness to pay; b) it relies on the objective and correct calculation of unit costs; and c) it does not allow for price-differentiation.

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6_3,

- **Market-based pricing:** Pricing is solely aligned with the market environment, and hence, it is based on the price of the market leader or on the established market price (e.g., in a commodity market). This ignores individual customer valuation and production cost, assuming that the former is reflected in the prevalent market price while the latter by definition just has to be low enough to facilitate competition in the market.

- **Value pricing:** Based on the proposition that the price of a product or service should relate to its value from the perspective of a particular customer, prices are set to reflect individual willingness to pay. The customer valuation can be determined based on an objective added value (e.g., a process time reduction) or a subjective valuation alone (e.g., through branding). Value pricing is difficult to implement in competitive settings because it naturally invites undercutting. In many situations, it is also virtually impossible to differentiate between individual customers so as to charge different value prices; hence, arbitrage and cannibalization occur.

- **Pricing and revenue optimization:** Here, pricing is also based on different customer valuations, but with an additional value dimension based on time and risk propensity. While customer valuation changes over time, remaining sales time and available capacity are limited so that both customer and seller face a continuous risky and uncertain trade-off situation:

  - *The seller* has to decide whether to close a sure deal early on while possibly yielding a lower price or whether to protect capacity over time for later-arriving customers with higher valuation – both strategies being uncertain and risky.
  - *The customer* has to trade certainty about the availability and price of the product at the time of decision against a possibly lower price in the future where availability might have vanished.

Across industries, based on their specific characteristics, advanced pricing schemes (the last two bullets above) naturally work especially well in Internet markets out of three reasons (see, e.g., Kannan and Kopalle, 2001, pp. 68):

1. Customers get an *immediate response* to their pricing queries and, conversely, sellers get instant feedback on quoted prices.

2. The seller faces extremely *low menu costs*, as compared to those associated with traditional distribution channels, given that the offer can be

changed, extended and adjusted to competitor moves or for individual customers with basically neither time loss nor measurable costs.

3. For most customers, especially for commodity-type products, the Internet has a very high level of *purchasing convenience* because the purchasing process can be completed online and even from home.

The following Section 3.1.2 discusses the characteristics of the fourth of above pricing approaches, highlighting the differences between pricing schemes that explicitly rely on prior customer segmentation (i.e., revenue management) and those that use implicit differentiation based on time of purchase alone (i.e., dynamic pricing).

3.1.2 Dynamic Pricing and Revenue Optimization

Desiraju and Shugan (1999) examine the applicability of pricing and revenue optimization schemes from a marketing perspective to determine new *strategic pricing principles* besides the many pure computer-aided techniques. They find that "costly, complex multi-period yield management systems are far more profitable when a service provider faces different market segments arriving at different times to purchase the service" (Desiraju and Shugan, 1999, p. 1). For dynamic pricing in particular, Talluri and van Ryzin (2005, pp. 179) name three specific industries/occasions where dynamically varying prices are a common and most natural mechanism of demand management and pricing:

- **Styles-goods markdown pricing:** Retailers usually use *markdown pricing* toward the end of vending seasons to clear inventory. The literature gives multiple possible explanations as to why the price path in retail is usually monotonic-decreasing: a) retailers are *uncertain* whether items will meet the current fashion trend, so they initially price high and later discount items that have not sold well; b) customers who purchase early in the season usually have a higher *willingness to pay*, as they consider themselves "trend setters" (i.e., first ones to wear a new style); or c) when sales peak, customers naturally spend more time in various stores and hence are more aware and *sensitive to price levels.*

  Heching et al. (2002) report an additional revenue potential of 4% for a surveyed retailer when using an appropriate dynamic pricing model.

- **Air travel ticket pricing:** Most incumbent airlines and especially LCCs dynamically vary their ticket prices to adjust for demand variations. Here the price direction is typically increasing (so-called *markup* *pricing*), although external market dynamics might also lead to lower prices as the departure date approaches. An example for the latter are last-minute tickets sold through dedicated channels, which charter carriers use to fill up pro-rata capacity[1] that is not used in the end by tour operators.

American Airlines, the pioneer of revenue management in air travel, reported an annual benefit of 500 million USD after the introduction of its new revenue management system (Smith et al., 1992).

- **Consumer-packaged goods promotions:** Here, promotions are typically short-term, as customers are well aware of past prices and promotions and hence, their subjective "reference price" or "fair price" for the discounted product is easily affected (in both a beneficial and a negative way). However, promotions that are run well can have a significant impact on demand, although this might come at the cost of demand dilution from other products or later periods.

The scope of the aforementioned *pricing and revenue optimization* examples is still broad, and therefore, a more actionable definition for *dynamic pricing* is required.

Definition According to Klein and Steinhardt (2008, Sec. 5.1.2), dynamic pricing is defined as the process of *tactically adjusting one-sided prices* during the vending period to *react to changes in demand* and *competitor behavior* with the objective of *maximizing total revenue* (see also Talluri and van Ryzin, 2005, Chap. 5).

What remains difficult, even given the above definition, is the differentiation from *revenue management*,[2] a common practice in airline revenue optimization, especially as the two terms are frequently used interchangeably in established literature. While some authors regard dynamic pricing and revenue management as equal-standing concepts of demand management (e.g., Boyd and Bilegan, 2003, pp. 1378), others see dynamic pricing as

[1] European charter carriers typically block capacity for tour operators at special negotiated rates to be sold exclusively by the latter. While *fixed capacity* is ultimately dedicated and has to be settled by the tour operator in all cases, seats from *pro rata capacity* that have not been sold are fully credited by the carrier upon release by the tour operator.

[2] Starting from customer segmentation, first, differentiated products with a predetermined price are created, and afterward, the available capacity is divided up between these. On that basis, the continuative availability of individual products is controlled in an operative and tactical manner. For an introduction to revenue management, see for example Klein (2005).

the dominant concept and interpret classical revenue management (i.e., the steering and management of capacity based on predetermined price levels) as a sub-form of dynamic pricing (e.g., Bitran and Caldentey, 2003, p. 223). Inversely, Talluri and van Ryzin (2005, Sec. 5.1.1) treat dynamic pricing as a special form of revenue management and thus differentiate "quantity-based revenue management" and "price-based revenue management". Similarly, Marcus and Anderson (2008, p. 259) view dynamic pricing as a price-based control problem that is merely a "complement to the allocation-based control historically practiced in the airline industry". Indeed, in the end, the differentiation rests on the "question of the extent to which a firm is able to vary quantity or price in response to changes in market conditions" (Talluri and van Ryzin, 2005, p. 176) – i.e., practical business constraints usually dictate the choice of the appropriate tactical response.[3]

The work at hand is based on the latter views, which have been substantiated in Currie et al. (2008, p. 1) and Klein and Steinhardt (2008, Sec. 5.1).

Prerequisites In their book, Klein and Steinhardt (2008) name three market characteristics required for the possible application of dynamic pricing:

1. There is only *one class of goods* and *no explicit differentiation* between customer segments through fare or booking classes, i.e., no discrimination is in effect besides that entailed by the pure time of purchase.

2. The individual *prices* of services or products *are not fixed in advance* but are allowed to vary over time, both upward and downward (i.e., a certain monotony is not required).

3. Inquiries from customers do not explicitly relate to the availability of specific price-product-combinations but instead refer to given services or products for which a *spot price* is quoted on request.

In addition to these considerations, Talluri and van Ryzin (2005, Sec. 5.1.1) name a few more technical prerequisites for the application of dynamic pricing schemes:

1. The ability to adjust prices in a timely fashion without incurring exorbitant costs or complexity.

[3] Gallego and van Ryzin (1997) point out that price-based revenue management poses the preferred (since more profitable) option: quantity-based revenue management is based on rationing capacity through reduction of sales by limiting supply, which potentially hampers profits. Achieving the same effect by raising prices consequently increases the revenue and benefits (assuming fixed costs) the resulting profit potential.

2. No inherent need for predetermined and long-term fixed price levels (e.g., those caused by customer preference or habit).

3. No shared components with limited capacity between multiple services or products (i.e., no network effects have to be considered).

The given characteristics are not entirely exclusive, as it still seems technically possible to mimic dynamic pricing via a well-configured revenue management system, especially if the price ranges are predetermined and grouped (see McGill and van Ryzin, 1999). However, the techniques employed differ heavily because in dynamic pricing models, demand is a price-sensitive process and therefore the price is typically explicitly incorporated as a control variable. Thus, "the distinction between dynamically adjusting the price of a single product and managing the availability of different products using the same underlying resource is important, because it impacts how the problem is modeled" (Boyd and Bilegan, 2003, p. 1379).

For the example of air transportation, Boyd and Kallesen (2004) postulate that the choice between revenue management and dynamic pricing schemes mostly depends on the answer to the question on whether "airline fare classes [are] *different products* (i.e., yieldable) or *different prices* for the same product (i.e., priceable)" (Boyd and Kallesen, 2004, pp. 172):

- **Yieldable demand** is class-specific and hence can be segmented a priori, i.e., "the Y passenger is specifically interested in the Y-class product and will purchase that product even when a less expensive Q-class product is available". Here, classic revenue management is the tool of choice (see e.g., Klein and Steinhardt, 2008, Sec. 1.2.2).

- **Priceable demand** is highly price-sensitive and hence will not adhere to predefined segments, i.e., "the Y-class passenger is primarily concerned with price and will purchase a Q-class ticket if it is the lowest fare available". Here, dynamic pricing schemes are more appropriate (see Section 3.2.2).

They note that "there is, however, an overall shift toward priceable demand" (Boyd and Kallesen, 2004, p. 173).

At the same time, the underlying objective of dynamic pricing is similar to classic revenue management schemes.

Objective Also, in dynamic pricing, the objective of all pricing decisions is to maximize the resulting aggregated profit. This especially includes the acceptance of foregone profits in the short term, through pricing above the prevalent reservation price[4] when this leads to higher profits in the long term through the sale of retained capacity at higher prices later in the process.

Most dynamic pricing literature assumes inherent fixed costs/capacities, which allows for the alternative usage of a pure revenue maximization approach (see for example Klein, 2005, Sec. 2.1.1 or Talluri and van Ryzin, 2005, Sec. 5.2).

Functionality Dynamic pricing schemes take systematic advantage of two demand properties that cannot be addressed via traditional static pricing:

- **Stochastic demand effects:** In reality, demand is stochastic, and with it, realized revenue for a specified price may deviate from initial expectations. In conjunction with fixed capacity, this may lead to cases in which it is optimal to increase the price, as the remaining (now more scarce than foreseen) inventory is expected to be sold at a higher price.

  Figure 3.1 illustrates the effect: Based on a fixed capacity of 200 units and a time-invariant demand function, the seller sets the price at 200,

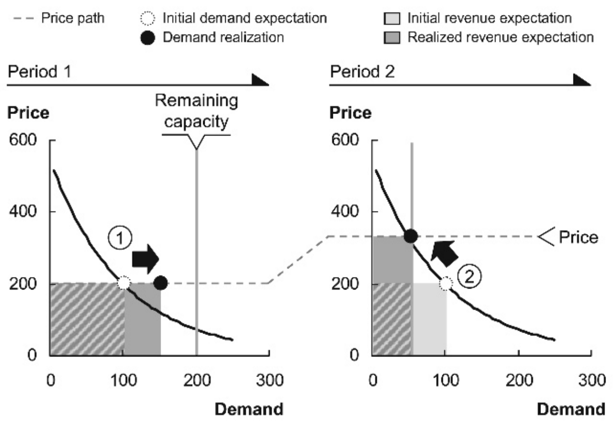


Figure 3.1: Revenue effect by reaction to stochastic variation
Source: own design

[4] Reservation price is the maximum price a prospective customer is willing to pay for a given product or service. It is basically the price-point where the retained utility drops to zero.

expecting to sell 100 units in each of the next two periods. ① Unexpectedly, 150 units are sold in the first period, leaving the seller with only 50 units for the second one. ② Now the much lower capacity can naturally be sold at a higher price.

- **Systematic demand effects:** In cases where price sensitivity is time-variant (i.e., the price-demand function shifts in time), dynamic prices can capture the changing optimums based on the effective price-demand situation per time period.

  Figure 3.2 illustrates the effect: Based on the same assumptions as above, the seller sets the price at 200, expecting to sell 100 units in the next two periods each. While this expectation materializes in the first period, ① after the price sensitivity has decreased in period two, ② the remaining capacity can be sold at a higher price.

The applicability of dynamic pricing schemes is typically not within the sole decision space of the seller itself but is dictated by its specific industry context. For example, in 2008, it was still common for European tour operators to publish fixed rates for their vacation packages in printed catalogs because customers were simply accustomed to it. Hence, to begin a dynamic variation of these prices would simply be infeasible for most single players. In other sectors of the travel industry (e.g., airlines and hotels), which already sell the bulk of their services through computer-based channels that foster

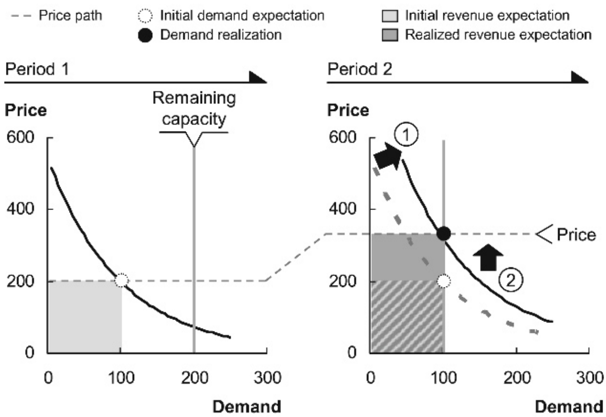


Figure 3.2: Revenue effect by proactive price-variation
Source: own design

immediate price changes (i.e., the Internet or global distribution systems), customers are used to being faced with varying prices.

The next section gives an overview of existing literature on dynamic pricing and the possible learning of input variables or data.

3.2 Literature Overview

The following sections give an overview of the relevant literature in the field of online demand learning in particular (Section 3.2.1) and general non-learning pricing schemes like dynamic pricing (Section 3.2.2).

The intention here is not to give a full-fledged introduction to dynamic pricing and online demand learning but is rather to provide an overview and structure for the most relevant literature. Toward this end, the presented literature forms an informed selection of the vast amount of available work.

The reader is expected to be familiar with the basic concepts (i.e., not all technical terms are explained in detail). Readers unfamiliar with the underlying foundations should refer to relevant introductions on the topic (see, e.g., Klein and Steinhardt, 2008 or Talluri and van Ryzin, 2005).

3.2.1 Demand Learning Models

While most of the cited literature in below Section 3.2.2 acknowledges the fact that both customer arrival rate and reservation price in real-world settings are stochastic, the exact types of their distributions are often predetermined to be of well-behaved mathematical form for the purpose of simplicity (e.g., Gallego and van Ryzin, 1994) or, at a minimum, the value of the distribution parameters is assumed to be known in advance or to be easy to derive from available data (e.g., Feng and Gallego, 1995).

In a separate string of literature on what is termed *online demand learning*, researchers explicitly try to overcome these limitations of revenue management and dynamic pricing models, which in their generic form "assume exact knowledge of the underlying statistical characteristics of the demand" (Aviv and Pazgal, 2002, p. 2). For example, Bertsimas and Perakis "exploit the fact that over time firms are able to acquire knowledge regarding demand behavior that can be utilized to improve profitability. Much of current research does not consider this aspect but rather considers demand to be an exogenous stochastic point process following a certain distribution" (Bertsimas and Perakis, 2006, p. 46).

By way of basic classification, Amman and Kendrick (1994, 1997) distinguish and compare the effectiveness of three types of learning that are also used for classification below:

- **Certainty Equivalence:** Parameter uncertainty is ignored when modeling the decision, e.g., through deterministic substitution models. However, the parameter in question is re-estimated at every periodic decision stage, as the schemes acknowledge the possibility that the parameter might have changed (i.e., that in reality it is uncertain).

- **Passive Learning:** The possibility of future learning is not anticipated and considered when making the current stage decision, e.g., demand information is recorded and evaluated, but price decisions are not influenced as so to increase learning speed.

- **Active Learning:** The decision-making process explicitly takes into account the additional effect of current decisions on future learning (e.g., a lower price might lead to higher demand, yielding more observations that in turn possibly speed up the learning process).

A fundamental prerequisite of any type of demand learning is that for stochastic customer arrival processes, it is *not* assumed that increments are stochastically independent over time. That is, there has to be some sort of *functional dependency* – otherwise, learning is not possible (Lin, 2006). Accordingly, Bitran and Wadhwa (1996) distinguish two types of demand uncertainty:

- **Predictable factors**, like a lack of information about the *attractiveness* of the product or service – i.e., latent demand and its price sensitivity across time.

- **Unpredictable factors**, such as weather and traffic. Naturally, these *cannot* be learned through sales or demand observations.

In the following paragraph, the very early work in the area of online demand learning is surveyed. Based thereon, the later paragraphs introduce extensions, structured along the type of the underlying learning mechanisms as introduced in the list above.

For each group, a table summary is given, highlighting a) whether overall demand is separated into pure arrival of latent demand and price sensitivity; b) the assumptions that are made about the corresponding underlying distributions; c) whether these assumptions are allowed to be time-variant; d) which learning mechanism is employed; e) restrictions to certain market settings; and finally f) other peculiarities.

Early Work There have been several early attempts in the pricing literature, driven mostly by economists, to cope with the problem of information imperfection in regard to true price and demand relations. Clower (1959) was the first to explicitly criticize "the thinking of the majority of professional economists, most of whom, as before, think and work in terms of models which presuppose a world of perfect information, perfect certainty and instantaneous response to changing circumstances" (Clower, 1959, p. 705). At that time, the proposed solution intended to reduce a firm's uncertainty about demand and price sensitivity was to increase expenditures for market research. The marginal benefits of market research were proven to be non-negative and non-increasing in Manning (1979) and were supposed to be "the firm's most common activity of self protection aimed to reduce demand uncertainty" (Manning, 1979, p. 366). On the other hand, Nguyen (1984) notes that a firm also naturally learns about the unknown parameters of its demand through experimentation and observation, and that this learning process is naturally prone to Bayesian treatment. Thus, any firm will incur foregone utility during the learning process but expects to over-compensate for this through additional benefits in later periods. In her early work on dynamic inventory models (i.e., without pricing decisions), Azoury (1985) shows that the resulting multi-dimensional state space of dynamic programs for Bayesian learning models can be reduced to a one-dimensional space that is then also mathematically tractable. With this finding, her work lays the foundation for various future learning work in that direction (see Bertsimas and Perakis, 2006), which is discussed below.

Certainty Equivalence Models A model with deterministic, albeit initially unknown, demand is considered by Petruzzi and Dada (2002). They assume that a deterministic demand function is controlled by a single parameter and reoccurs in later cycles, and hence that the parameters can be learned across periods. Here, pricing is assumed not to directly affect the learning process and arrival rate and price sensitivity are not differentiated – as in most certainty equivalence models.

Aviv and Pazgal (2002) complement the elementary dynamic pricing models (see Section 3.2.2) of Gallego and van Ryzin (1994) and Bitran and Mondschein (1997) with a learning mechanism for the customer arrival rate. The exact (time-invariant) distribution is assumed to be unknown to the seller at the beginning of the sales period. However, the seller has a prior belief about it that is updated as information becomes available throughout the sales period. The seller finds it inessential to consider the impact of the pricing decision on the learning process itself but instead adopts a learning

Article	Demand Differentiation[a]	Arrival Rate		Price sensitivity		Learning Method	Market Setting[b]	Remarks
		Distribution	Time-variant	Distribution	Time-variant			
Petruzzi and Dada (2002)	NO	—	—	Deterministic with unknown noise parameter	NO	Calculus	M	
Aviv and Pazgal (2002)	YES	Poisson	NO	Exponential (not learned)	NO	Bayesian updating	M	
Bertsimas and Perakis (2006)	NO	—	—	Linear with normal distributed noise	YES	Least squares	M/O	Coefficients allowed to vary slowly in time
Zhang and Chen (2006)	NO	—	—	Linear with normal distributed (additive) noise	NO	Bayesian updating	M	Focus on inventory control

[a] Separation of stochastic effects into arrival rate and price sensitivity
[b] M: monopoly, D: duopoly, O: oligopoly, C: competition

Table 3.1: Certainty equivalent learning literature (excerpt, in chronological order)

approach based on a certainty-equivalent heuristic, which disregards the uncertainty of the learned arrival rate during price setting. However, the true parameter of the targeted Poisson distribution is learned using Bayesian updating. The results based on an assumed Gamma distribution (a conjugate prior of the Poisson distribution[5]) complement the results of Gallego and van Ryzin (1994), as dynamic pricing appears to be most valuable "in settings with high but resolvable initial uncertainty about how successful the product is" (Aviv and Pazgal, 2002, p. 28), i.e., the seller may start with a low price to foster learning about the demand state instead of looking for short-term revenue maximization.

Bertsimas and Perakis (2006) explicitly consider competition in a non-cooperative oligopoly, for which purpose they derive a methodology for jointly learning and setting prices without assuming any prior knowledge of demand or price sensitivity. Here price elasticities are even allowed to vary slowly in time. The authors assume a linear price-sensitive demand model with Gaussian noise whose coefficients can then be estimated based on a least squares optimization (i.e., *not* using a Bayesian update scheme). Competitor behavior is modeled in a similar way, assuming rational and also optimal pricing using the same model. The authors develop a certainty equivalent heuristic to solve the model, as the derived dynamic program suffers from

[5] See Section 4.2 for an introduction to Bayesian statistics.

state space explosion in the competitive szenario.

For the case of a linear demand function, Zhang and Chen (2006) consider learning about price-controlled demand using a complex model combining demand uncertainty, dynamic pricing and inventory replenishment. They use a Bayesian updating scheme that can be reduced to a dynamic programming formulation for which they also derive some structural properties.

Table 3.1 gives an overview of literature dealing with certainty equivalent learning models.

Passive Learning Models Based on a case study of a fashion ski-wear manufacturer, Fisher and Raman (1996) describe and test a learning model for suppositionally bivariate normal distributed demand and a rather basic learning mechanism that rests upon the aggregation of expert advice. In their practice oriented paper, they find the proposed scheme to increase profits up to 60%.

Bitran and Wadhwa (1996) develop a passive learning model that is able to cope with a non-stationary demand distribution. Their work extends the literature from the field of demand learning in the context of inventory management. Here, total demand in any sales period is a random variable whose distribution parameter is unknown. However, the planner is expected to have some prior information on it (e.g., from previous periods) and hence is able to specify a prior distribution on the demand parameter. After each sales period, the observed realized demand is used to update the prior belief to a posterior distribution via the application of Bayes' rule. The authors derive learning models under non-stationary arrival and reservation price distributions, the former being Poisson and the latter being exponentially distributed.

A comprehensive approach to learning in retail is presented in Subrahmanyan and Shoemaker (1996), who not only learn the parameters for a single given demand distribution, but also select the most appropriate distribution from a predefined set based on past observations. Besides pure demand learning, the pricing model also accounts for possible reorders of stock.

Also motivated by a retail setting, Raju et al. (2004) propose an agent-based reinforcement learning scheme and compare the performance of multiple reinforcement learning policies under different information settings. Under absence of any market information (i.e., no insight into latent demand or competitor stock and price levels), the tested adaptive learning scheme

Article	Demand Differentiation[a]	Arrival Rate		Price sensitivity		Learning Method	Market Setting[b]	Remarks
		Distribution	Time-variant	Distribution	Time-variant			
Fisher and Raman (1996)	NO	Bivariate Normal	NO	—	—	Aggregation of expert advice	M	
Bitran and Wadhwa (1996)	YES	Poisson	YES	Exponential	NO	Bayesian updating	M	
Subrahmanyan and Shoemaker (1996)	NO	—	—	Multiple negative binomial distributions	NO	Bayesian updating	M	Learning includes choice of distribution
Raju et al. (2004)	YES	Poisson	NO	Arbitrary utility function	NO	Agent-based reinforcement learning	D	
Lin (2006)	YES	Poisson, arrival rate is gamma	YES	—	—	Bayesian updating	M	
Levina et al. (2006, 2007)	YES	Poisson	NO	Variable	YES	Aggregation of expert advice	M	
Kachani et al. (2007)	NO	—	—	Linear with normal distributed noise	YES	Least squares	O	Coefficients allowed to vary slowly in time

[a] Separation of stochastic effects into arrival rate and price sensitivity
[b] M: monopoly, D: duopoly, O: oligopoly, C: competition

Table 3.2: Passive learning literature (excerpt, in chronological order)

(Q-learning[6]) massively outperforms a simple adaptive following scheme. However, in a partial information setting with symmetric sellers, the adaptive policies are found to converge more quickly.

Another passive learning approach is pursued by Lin (2006) who assumes that customers arrive according to a conditional Poisson process, whose rate is not known in advance but can be learned by the seller starting from some prior estimate, where the learning takes place within the sales horizon of a single event. Here the purchasing probability is explicitly split up into arrival rate and reservation price distribution, but learning only takes place for the arrival rate of the Poisson process, which is assumed to follow a gamma distribution. A solution to the dynamic pricing problem under the self-adjusting forecast is derived and numerically shown to be robust when the true arrival rate differs from the initial forecast.

[6] Q-learning is an agent-based technique for learning a so-called action-value function, which returns the expected utility of a given action in a given environmental state when following a fixed action policy thereafter. For an introduction, see e.g., Sutton and Barto (1998).

Levina et al. (2006) propose a methodology based on a variation of the Aggregating Algorithm (see Vovk, 1990) for learning an arbitrary form of the reservation-price distribution as well as the arrival rate of demand (see also Levina et al., 2007, for details on the underlying technique). Specifically, demand is learned through aggregation of individual forecasts that are generated from an online pool of stochastic prediction strategies.

The work of Kachani et al. (2007) extends the above certainty equivalent model of Bertsimas and Perakis (2006) by relaxing the constraint on the linearity of the demand function. The resulting problem is formulated as a so called mathematical program with equilibrium constraints. Competition is modeled as a Cournot model where prices are determined indirectly through capacity allocations that result in a Nash equilibrium. The price-demand parameters are estimated on equilibrium demand levels, and the future price policy is thereby derived for the competition and for the firm itself. The latter benefits from the fact that it knows its past demand and hence can calculate a policy that explicitly takes into account the current level of remaining capacity.

Table 3.2 summarizes the discussed literature that considers passive demand learning.

Active Learning To the author's knowledge, Nguyen (1984) is the first to explicitly consider the effect of active demand learning on a firms "utility" (i.e., revenue) over multiple planning periods and under uncertainty. He shows that a firms intertemporal output may indeed vary depending on the effect of a period's decisions on the expected utilities that are derived from profits to come in future periods.

Easley and Kiefer (1988) as well as Kiefer and Nyarko (1989) describe active learning approaches in a general context (i.e., no explicit assumptions are made about the form of the distributions to be learned) based on agent learning, where the beliefs of the agents are proven to converge – but not necessarily against the true value. Balvers and Cosimano (1990) are the first to extend this model to allow for an evolving demand distribution (i.e., a truly dynamic environment) to illustrate how learning affects pricing decisions over time. For the considered case of a linear demand function with unknown slope and intercept, they find that in the face of fully anticipated parameter changes, it may be optimal for firms to reduce price variations, as these trigger additional demand reactions that are not fully predictable (i.e., additional noise).

Lobo and Boyd (2003) derive a heuristic solution to the stochastic dynamic program that incorporates learning about the underlying demand curve. They find it advantageous to add dithering (i.e., a random perturbation to the price), as this will "excite" the learning process by conditioning the information matrix well. Adversely, too much random price variation destructs profits. They also name suitable extensions to the existing learning literature: a) the consideration of multiple products (although learning between products is not explicitly nominated); b) allowing for time-varying demand functions; and c) the adoption of a non-linear demand model (e.g., a multiplicative log-normal distribution, which is a common assumption for demand structures).

Araman and Caldentey (2005) introduce a more sophisticated heuristic that takes arrival rate uncertainty explicitly into account when pricing. Their setting is based on the retail industry with non-perishable products. Price variation is induced through a particular value function, i.e., opportunity costs that the seller incurs in not switching to other products yielding higher expected revenues instead of keeping high-priced but low-selling items stocked for too long. They employ a Bayesian update scheme – one that is greedy with respect to the described value function – to learn the unknown parameters of the underlying price-sensitive Poisson arrival process.

The same approximate value function is employed by the decay balancing heuristic developed in Farias and van Roy (2007). The authors then compare three heuristics (certainty equivalent by Aviv and Pazgal, 2005, greedy active learning by Araman and Caldentey, 2005 and their own decay balancing heuristic) based on a uniform model: the arrival rate is assumed to be distributed according to a finite mixture of Gamma distributions, reservation prices are assumed to follow an exponential distribution and the objective is to maximize expected discounted revenue over an infinite horizon. They show that decay balancing achieves near-optimal performance and that its generated price policies set higher prices when uncertainty is present in the market.

Aviv and Pazgal (2005) benchmark an active learning formulation against selected passive learning models. Unlike in their early work (Aviv and Pazgal, 2002), now prices may fluctuate in periods without sales as the proposed heuristic for exponentially distributed reservation prices tends to vary prices in order to actively promote learning of demand, especially as in the model demand effects cannot be explicitly attributed to changes in arrival rate or to a shifting reservation price distribution. As expected, learning behavior depends on the length of the sales horizon – i.e., short horizons encourage low initial prices to foster learning early on.

A Bayesian model based on Dirichlet distributed priors for the demand function is defined in Cope (2007). Additionally, several price-testing strategies are developed to learn the shape of the underlying distribution of reserva-

Article	Demand Differentiation[a]	Arrival Rate		Price sensitivity		Learning Method	Market Setting[b]	Remarks
		Distribution	Time-variant	Distribution	Time-variant			
Nguyen (1984)	NO	—	—	Non-linear "von Neumann-Morgenstern" type	NO	Bayesian updating	M	
Easley and Kiefer (1988)	NO	—	—	Variable one-parametric	NO	Agent with Bayesian updating	M	Converges not always to true parameters
Kiefer and Nyarko (1989)	NO	—	—	Linear with normal distributed noise	NO	Agent with Bayesian updating	M	Converges not always to true parameters
Balvers and Cosimano (1990)	(YES)	Linear with normal distributed noise	YES	Linear with normal distributed noise	YES	Bayesian updating	M/C	
Lobo and Boyd (2003)	NO	—	—	Linear with normal distributed parameters	NO	Bayesian updating	M	
Araman and Caldentey (2005)	YES	Poisson (two arrival rates: low vs. high)	NO	Exponential (not learned)	NO	Bayesian updating	M	Infinite horizon, but opportunity costs on low selling items
Aviv and Pazgal (2005)	YES	Poisson	NO	Exponential	YES	Markov decision process - Bayesian updating	M	
Farias and van Roy (2007)	YES	Poisson, arrival rate is gamma	NO	Exponential (not learned)	NO	Bayesian updating and decay balancing	M	
Cope (2007)	NO	—	—	Dirichlet distribution	NO	Bayesian updating	M	

[a] Separation of stochastic effects into arrival rate and price sensitivity
[b] M: monopoly, D: duopoly, O: oligopoly, C: competition

Table 3.3: Active learning literature (excerpt, in chronological order)

tion prices while constantly trading off the competing objectives of exploring additional untested price levels and actually using prices that are known to yield high revenues.

The discussed active demand learning literature is classified in Table 3.3.

Miscellaneous Models Using dynamic simulation with Bayesian updating, Wieland (2000) derives an optimal pricing policy that incorporates a substantial degree of "optimal experimentation" to drastically improve learning.

He determines that policies, which separate control and estimation should "experiment" in order to not induce biases into their system and prevent non-stationary behavior (i.e., continuously changing pricing policies).

Xu and Hopp (2004) consider learning about demand functions belonging to an exponential family. Unlike in the model by Easley and Kiefer (1988), the connection between the stages in Xu and Hopp (2004) is based on the size of the carried inventory. Additionally, the focus is on statistical inference using a rather small sample size, as sales horizons often tend to be small. In their computational experiments, they find the value of learning to be significant if customer arrival is noisy, and that non-learning policies can perform very poorly if the assumed prior is volatile.

The above model is extended to explicitly include the effect of competitor pricing on the probability of a firm's own sales by Currie et al. (2008). They derive the nature of the relationship between the competitors' price structure geometrically and show that under a small set of assumptions, the resulting optimal price function is unique.

Table 3.4 lists the literature that uses miscellaneous learning schemes.

The following paragraphs additionally give an overview of simple retrospective curve-fitting models and of performances evaluations of models.

Curve Fitting Anjos et al. (2004) are among the first to explicitly target the fitting of distributions or demand models to real demand behavior and price sensitivity. They approximate the expected booking behavior of customers by fitting a pre-selected function to real data, minimizing the sum of squared errors. However, the shape of the underlying demand function has to be heuristically specified in advance: "The choice of functions (...) is made by careful examination of the behavior of the bookings as the price (...) and the number of days until departure (...) vary" (Anjos et al., 2004, p. 536). When fitting, the authors do not differentiate between actual customer arrivals and booking probability (based on the actual price), but they note that "if data on hit rates on airlines' websites for particular flights (...) are available, it may be possible to improve the fitting process by fitting the demand function separately from the function describing the probability of purchasing a ticket" (Anjos et al., 2004, p. 540). Accordingly, the fitting is difficult when price structures vary significantly and overlay the pure demand behavior. The proposed model also does not consider systematic seasonal variations, and once estimated, model parameters are only updated between flights but not during the sales process for a particular flight. Finally, Anjos et al. (2004) derive a continuous-time model based on Lagrangian multipliers that is later

Article	Demand Differentiation[a]	Arrival Rate		Price sensitivity		Learning Method	Market Setting[b]	Remarks
		Distribution	Time-variant	Distribution	Time-variant			
Wieland (2000)	NO	—	—	Linear with normal distributed noise	NO	Agent with Bayesian updating	M	
Xu and Hopp (2004)	YES	Piecewise deterministic, exponential family	NO	Exponential	NO	Bayesian updating	M	

[a] Separation of stochastic effects into arrival rate and price sensitivity
[b] M: monopoly, D: duopoly, O: oligopoly, C: competition

Table 3.4: Miscellaneous models learning literature (excerpt, in chronological order)

shown in Anjos et al. (2005) to be optimal for a family of continuous pricing functions that are merely time-dependent rather than stock-dependent alone. Here the underlying assumption is that "the potential sales of a given product is more accurately described by the number of queries on a company's website than by the number of items actually sold (which depends more strongly on the current market pricing)" (Anjos et al., 2005, p. 247).

Kephart et al. (2000) evaluate the collective behavior of dynamic pricing environments driven by self-learning software agents that set prices based on their past experience. They point out that most learning schemes (especially agent-based systems) expect the environment (i.e., the opponent) to behave in a fixed, predetermined way that explicitly does not include learning. They conclude that for learning schemes to be effective, they must incorporate possible learning of the environment, which many agent-based systems do not. In their test, they show that even ordinary Q-learning may not converge to optimality when confronted with self-learning agents.

Based on a basic linear demand model, Spedding and Chan (2000) compare "Bayesian dynamic linear time series forecasting" to auto-regressive integrated moving average (ARIMA)[7] analysis using a software called BATS[8] (developed by Pole et al., 1994). For their data sample and the underlying linear model, they find the forecasting errors associated with their Bayesian approach to be much lower than those for the benchmarked ARIMA model, as by design the latter does not benefit from additional information in the

[7] For an introduction to ARIMA, see for example Christopher Chatfield (1989); Fuller (1996) or Harvey (1993).

[8] BATS offers a variety of functions, including time-series model-building, Bayesian forecasting monitoring, intervention analysis tools, error analysis functionality and graphical visualization tools. For more information on BATS, see Pole et al. (1994).

same way. An in-depth discussion of their insights on the strength and weaknesses of Bayesian models as well as an introduction to Bayesian statistics in general can be found later in Chapter 4.

In an empirical study within a retail setting, Heching et al. (2002) benchmark adaptive and full information policies and find that the latter can increase revenues by up to 13%, while the impact of adaptive policies lies only around 3%.

Related Areas Literature on fashion retailing is currently heavily pushing learning research in the wake of new so-called "quick response" strategies that allow for replenishment even within the sales season and in some cases also permit re-assortment (see, e.g., Caro and Gallien, 2007 or Kök and Fisher, 2007). Here, demand learning can provide additional valuable insights besides information on latent demand and price sensitivity alone.

The next Section 3.2.2 provides a literature overview of conventional non-learning dynamic pricing models (Sections 3.2.2.1 and 3.2.2.2) as well as the newer customer choice models that aim at understanding of customer purchasing behavior (Section 3.2.2.3).

3.2.2 Non-learning Pricing Models

The available literature in the area of dynamic pricing consists of multiple and to some extent fairly separate research strings. Besides the functional areas that allude to the topic (e.g., marketing, economics etc.), a variety of assumptions and restrictions affect the structure and applicability of the models discussed in the following sections. Therefore, a useful structural framework based on underlying model assumptions will first be introduced.

In their comprehensive research overview, Elmaghraby and Keskinocak (2003) provide a categorization framework that is also partly adopted and extended by Talluri and van Ryzin (2005) to classify the various dynamic price-sensitive demand models based on their specific assumptions – which will also form the basis for the following literature review:

- **Myopic vs. strategic customers:** This classification concerns the level of sophistication on the demand side: *Myopic customers* take only the currently quoted price and their own reservation price (i.e., their willingness to pay) as the basis for purchasing decisions – if the latter is lower than the posted price, a purchase is made. *Strategic customers* also consider the expected future price path when making a decision and hence might hold back demand in anticipation of lower prices. Strategic customer models allow demand to adapt to the pricing policy of the vendor; hence, they are more realistic, but they also make the resulting dynamic pricing models more complex (essentially encouraging a strategic game). In contrast, myopic-customer models are usually mathematically more tractable and therefore are widely used throughout the literature.

- **Dependent vs. independent demand over time** or **finite vs. infinite population:** In *dependent-demand* models, the customer population is thought of as *finite*[9], and hence every customer who makes a purchase decision in early periods is removed from the population (sampling without replacement) and affects the composition of the remaining pool of potential customers (i.e., the distribution of their willingness to pay). This is also called the *durable-goods assumption*, as for durable goods, "by definition, the life of the product is longer than the time horizon over which the retailer makes price changes" (Elmaghraby and Keskinocak, 2003, p. 1289) and consequently, sales in a certain period are sales moved forward from subsequent ones. In *independent demand* models, the customer population is treated as *infinite* (sampling with replacement), and hence the assumed customer characteristics do not change depending on whether sales are made in early periods. "The infinite-population model is a reasonable approximation when there is a large population of potential customers and the firm's demand represents a relatively small fraction of this population because in such cases the impact of the firm's past sales on the number of customers and the distribution of their valuations is negligible" (Talluri and van Ryzin, 2005, p. 185).

- **Type of competition:** An important assumption for every dynamic pricing model is the type of competition expected to drive the market. *Monopoly models* assume that the resulting demand is only a function of the seller's own price, and thus these models either assume away competition or see the competitive price reaction empirically contained

[9] "Of course, in reality, every population is finite; the question is really a matter of whether the number and type of customers that have already bought changes one's estimate of the number or type of future customers." (Talluri and van Ryzin, 2005, p. 184)

in the demand sensitivity to the seller's own price.[10] Typically, this assumption is made for the sake of mathematical tractability. *Oligopoly models* explicitly model the expected competitive dynamics but, accordingly, run the risk of using the wrong specification or making erroneous assumptions regarding competitive behavior (e.g., competing firms may not always behave rationally in terms of expected price reactions). Finally, *perfect competition models* see the firm primarily as a price taker in a perfect market setting. Given that the latter model assumption is not applicable to dynamic pricing, it will not be considered in what follows.

As one might expect, the bulk of the dynamic pricing literature assumes monopoly settings with independent demand and myopic customers. Only a few recent papers (see, e.g., Netessine and Shumsky, 2005) consider the oligarchic case, trying to incorporate competitive price reactions and dependent demand to explicitly model customer choice behavior. Similarly, only selected literature deals with the strategic decision-making of customers, who may anticipate last-minute sales and price drops (see Section 3.2.2.2).

Overarching assumptions characteristic of most dynamic pricing literature include the following (see, e.g., Gallego and van Ryzin, 1994; McGill and van Ryzin, 1999 or Weatherford and Bodily, 1992):

- **Imperfect competition:** There exists no global market clearing price determined by market equilibrium; hence, each seller still has individual pricing power that directly affects realized demand.
- **Fix selling horizon:** The sales season is bounded by a certain point in time (e.g., end of season, flight departure, etc.) where all remaining capacity perishes at once and all sales stop.[11]
- **Predetermined capacity:** The seller has a fixed stock of items or capacity that is short-term fixed (i.e., fixed over the selling horizon), either because of physical constraints (e.g., the number of rooms in a

[10] The demand seen by a vendor dependent on its own price policy is also naturally affected by the competitors' price reactions and hence to some extent already includes competitive dynamics. However, it cannot account for changing competitive reactions to the same pricing policy, and therefore, these models are hampered in most competitive settings (see, e.g., Phillips, 2005, p. 55).

[11] In general, for dynamic pricing models to be effective, they need a bounded selling horizon. However, in case of an unlimited horizon, the necessary conditions can also be set through other realities that penalize unsold stock in the long run (see, e.g., Araman and Caldentey, 2005).

hotel) or because of inflexibility (e.g., necessary equipment allocation has a certain lead time).[12]

- **High fixed costs:** All costs related to the potential purchase of products or services are considered fixed or quasi-fixed in cases where the predominant share of costs is fixed up-front.

- **Price sensitive demand:** Time held constant, demand monotonously decreases in price, i.e., there is no paradox behavior as for so called *Veblen goods* (see Veblen, 1899).

- **No backlogging of demand:** Demand can only be satisfied if, at the time of a request, there is capacity or stock left – i.e., each sale in an early period is a potential displaced sale later in the sales season.

- **Low salvage value of unsold goods:** After close of sale, all remaining products or capacity have a fixed salvage value that is considerably lower than the sales price range (i.e., a sale during the sales season is typically superior to retained capacity at the end). Note that this definition includes zero.

Building on this introduction to possible structural characteristics of non-learning models, the following sections highlight relevant literature clustered along that structure.

3.2.2.1 Dynamic Pricing with Myopic Customers

Dynamic pricing research and literature dates back to the 1960s (see Kincaid and Darling, 1963, for the first work) and mostly originated in the field of economics and later marketing (see, e.g., Nagle, 1984; Rao, 1984 and Varian, 1980). Accordingly, marketing scientists were the first to note "the need to develop pricing strategies over a relevant time period and to allow for market *dynamics*" (Monroe and Della Bitta, 1978, p. 426). In his literature overview, Rao (1984, p. S45) states that "recent years have witnessed an intense amount of activity on models for pricing products *over time*" (e.g., Dolan and Jeuland, 1981, derive a general methodology for determining optimal pricing strategies over a product's life cycle) and he concludes that "almost invariably this research relates to new products since dynamic issues are more important in that area." Rajan et al. (1992, p. 241) note years later that "there is, however, little research on prices that change over the

[12] For an overview of relevant work on replenishment models (e.g., for retailers), the reader should refer to Elmaghraby and Keskinocak (2003, pp. 1299).

short term. The traditional argument [being] that companies are unwilling or unable to change prices in the short term", but at the same time recognizing that "short-term price changes not only occur in practice, but are becoming more common" (Rajan et al., 1992, p. 241), especially in fashion retail (see also Pashigian, 1988).

The next paragraph briefly lists very early work on dynamic pricing, after which the genuinely accepted reference model of Gallego and van Ryzin is introduced. The paragraphs thereafter discuss direct extensions and consider models limiting the amount of price changes allowed. Finally, special settings including multiple products or allowing for reordering of stock/capacity are reviewed.

Early Work Kalish (1983) is the first to explicitly acknowledge that uncertainty in demand and price sensitivity causes a firm's costs and revenue to vary over time, although the effects are not explicitly modeled. Chen and Jain (1992) build upon this work and incorporate the effects of uncertainty in their stochastic control problem; however, optimality conditions are only established under severe restrictive assumptions.

Around the same time, Rajan et al. (1992) introduce a dynamic pricing model for a monopolistic retailer facing a known demand function and derive simultaneous pricing and inventory policies for the deterministic case. Dockner and Jørgensen (1988) study a similar setting under oligarchic markets from a marketing perspective and use differential game theory to establish optimal policies.

The Reference Model The well noted work of Gallego and van Ryzin (1994) lays the ground for a whole series of stochastic optimization models with price-dependent demand under various assumptions. They develop a continuous-time optimal intensity control model for dynamic pricing of a single product with stochastic and price-sensitive demand. Here, demand is described as a Poisson process whose intensity is a "known decreasing function of the price" (Gallego and van Ryzin, 1994, p. 999) and hence price is the sole control for demand. They assert that an optimal pricing policy can only be computed for a particular family of exponential demand functions and for the general deterministic case also in closed form, but that these optimal policies then "change prices continuously and thus may be undesirable in practice" (Gallego and van Ryzin, 1994, p. 1002). However, they also show that policies that allow at most one price change are asymptotically optimal when capacity grows to infinity. They develop a fixed-price heuristic with more stable prices that is asymptotically optimal (in increasing expected sales levels). More importantly, assuming time-invariant reservation prices, they derive two structural properties of optimal pricing policies that are also validated in the works of Chatwin (2000); Feng and Xiao (2000b); Lee and Hersh (1993) and Zhao and Zheng (2000):

(P1) **Inventory-monotonity:** With time held constant, the price decreases for the remaining items.[13]

(P2) **Time-monotonity:** With the number of items held constant, the price decreases over time.

A property similar to (P1) was already developed for retail in Lazear (1986), where prices fall constantly when customer valuation is time-invariant during the fixed selling horizon (see also Pashigian, 1988, for an empirical evaluation).

A thorough introduction to the foundational models (including the reference model of Gallego and van Ryzin, 1994) can be found in Gönsch et al. (2009), who also provide an exhaustive literature overview.

Direct Extensions Zhao and Zheng (2000) address optimal dynamic pricing under both, a non-homogeneous arrival process and time-variant reservation price distributions. They show that (P1) still holds under the new assumptions but that (P2) may not hold in the case of an increasing reservation price distribution. Again, known distributions for arrival rate and reservation price are assumed. They conclude that “price changes become even more critical when the reservation price distribution shifts over time”, in which case the impact “using the optimal dynamic pricing policy could be as high as 100% over that achieved by using the optimal single policy” (Zhao and Zheng, 2000, p. 378). This is in line with the results of Gallego and van Ryzin (1994), already suggesting that dynamic pricing is mainly beneficial in cases of shifting reservation price distributions (Gallego and van Ryzin, 1994, p. 1000).

In a similar way, Feng and Gallego (2000) address a model where demand is a general Poison process with Markovian, time-dependent and (non-homogeneous) but predictable intensities. They develop an efficient algorithm for computing the optimal pricing policies, and notably, they already assert the need for “adaptive forecasting methods” (Feng and Gallego, 2000, p. 953) to support the applicability of the model.

[13] The first property (P1) is heavily dependent on the assumption of a fixed selling horizon. For example, Das Varma and Vettas (2001) show the opposite behavior for the retail case with no definite selling horizon.

Walczak and Brumelle (2007) extend the models of Zhao and Zheng (2000) and Feng and Gallego (2000) to semi-Markovian models, allowing multiple fare requests per time period and a more general arrival process. The new formulation then spans both, dynamic pricing and traditional revenue management problems.

Limited Amount of Price Changes For the use of general demand functions and to reduce undesired price fluctuations, Feng and Gallego (1995) develop a continuous-time Markov process for the case where only a single price change is allowed. Under rather mild conditions, they find it to be optimal to decrease (or increase) the price at the point at which the remaining length of the sales horizon falls below (or above) a certain time threshold, which in turn depends on the remaining items to be sold. They postulate that simple seat protection levels that do not take into account remaining sales time may be inefficient in gathering maximum revenue. Like most authors within this body of literature, they too assume that "management knows the expected demand rate at certain prescribed prices" (Feng and Gallego, 1995).

Starting from seasonal price promotions in retailing Bitran and Mondschein (1993, 1997) define another continuous-time model based on the stochastic arrival of customers. Additionally, they benchmark more realistic models featuring periodic pricing reviews, showing that the possible loss compared to that which may occur using continuous policies is small as long as an appropriate number of price reviews is allowed. To the author's knowledge, they are also the first to explicitly distinguish between the pure arrival process of customers (here Poisson), which "is often a response to their regular purchasing patterns (...) rather than a function of individual prices" and actual sales, which in turn depend on the customers' "distribution of reservation prices for a product" (Bitran and Mondschein, 1997, p. 64). They show, however, that under these assumptions, the model is equivalent to the one by Gallego and van Ryzin (1994) with a pure price-dependent Poisson purchasing process. They are also the first to allow the arrival of potential customers to be a non-homogeneous time-variant process. The model is extended in Bitran et al. (1998) to incorporate possible coordination of prices across multiple outlets with different arrival patterns.

Smith and Achabal (1998) extend the early models of Gallego and van Ryzin (1994) and Bitran and Mondschein (1993, 1997) to account for seasonal variations in demand and to explicitly consider the influence of initial inventory levels on demand.[14]

[14] In typical retail settings, the shelf space dedicated to a product can heavily influence demand and sales. This causal relationship is less pronounced in the aviation case.

Feng and Xiao (1999) extend the two-price model of Feng and Gallego (1995) by incorporating risk. Moreover, they prove that the exact solution to the stochastic problem is attainable in analytical form instead of using a deterministic substitute like Gallego and van Ryzin (1994). Still, demand is supposed to follow a Poisson process with constant intensity.

Feng and Xiao (2000b) extend their work in Feng and Xiao (1999) to allow for multiple price changes while maintaining an implementable solution. Therefore, the model is restrained to monotonic price changes (in either direction – i.e., pure markup or pure markdown policies) based on a predetermined set of prices. They find the resulting value function to be piecewise concave and decreasing in time and inventory, what is consistent with the aforementioned properties.

In a separate work, Feng and Xiao (2000a) allow for multiple (but still limited) numbers of price changes that are also reversible. The underlying model is restricted to allow for time-homogeneous demand only. In the optimal solution, each inventory level has a corresponding set of time thresholds that lead to a price change. These thresholds are again shown to be monotonically decreasing in price and inventory.

A similar modeling approach can later be found in Chatwin (2000) who extends the model to the case where the demand intensities that depend on price vary with the time-to-go, and in which the retailer can restock after the initial inventory has been sold.

For the case of limited price changes, Wen and Chen (2005) derive structural properties of the optimal time threshold for changing the price levels. In Netessine (2006), the number of price changes is also limited in order to derive a piece-wise constant pricing policy where the price is allowed to change directions between periods (i.e., increase and then decrease again). He considers a monopoly player in a dynamic but deterministic environment to derive the structural results of the model.

Multiple Products In Gallego and van Ryzin (1997), the authors' early approach in Gallego and van Ryzin (1994) is extended to cover multiple products in network environments. They formulate a deterministic model that gives a bound on the expected revenue and forms the motivation for two heuristics for the stochastic model (the so-called make-to-stock and make-to-order heuristics), which are shown to be optimal when the expected sales volume tends toward infinity. For a similar setting, Kleywegt (2001) develops a deterministic optimal-control formulation allowing for cancellations where all parameters (i.e., arrival rate, reservation price, number of cancellations) are allowed to be time-dependent. Recently, Liu and Milner (2006) studied multi-item pricing under the presence of a joint pricing constraint and developed an optimal solution for the deterministic formulation as well as a heuristic for the stochastic problem.

A broader spectrum of choice drivers is explored in Chun (2003), who establishes an optimal policy based on the demand rate, customers' other preferences (besides pure price alone) and length of season for a negative binomial demand distribution. Also, the capacity decision is included in the model, as average revenue naturally decreases with available capacity, which therefore has to be considered to find a truly global model optimum (Chun, 2003, p. 74).

A problem related to Gallego and van Ryzin (1994), but based on the seller's ability to approach customers with different demand intensity functions sequentially in an inhomogeneous Poisson type arrival process, is tackled by Lin (2004). He generates a stochastic dynamic program to cope with the random arrival of customers and develops an algorithm for computing an optimal pricing policy. The solution is a threshold-based bid price[15] policy that may even be determined at the beginning of the sales horizon. The model is also applied to a continuous-time problem for which a near-optimal heuristic is developed that does not rely on specific assumptions about the arrival process. Still, one limiting assumption is made, namely that reservation prices do not change over time.

Additional revenue sources are explored by Levin et al. (2007), who incorporate the effect of fee-based price guarantees into their model (i.e., customers pay a fee to receive a price guarantee for the fare on their flight). The resulting discrete-time optimal control model is non-Markovian.

Reordering of Capacity/Stock In the context of dynamic pricing in retail, there also exists a fairly separate string of models combining the problem of optimal price-setting and lot-sizing for reordering. An early deterministic work of combined continuous pricing and inventory decisions is Eliashberg and Steinberg (1987), and comprehensive surveys of early literature on that particular topic can be found in Chan et al. (2004); Eliashberg and Steinberg (1993) and Elmaghraby and Keskinocak (2003).

Uncertainty is later explicitly introduced by Federgruen and Heching (1999). Adida and Perakis (2004) study the simplified problem in its deterministic form using a fluid model. In Adida and Perakis (2006), the authors

[15] A bid price is the minimum amount of revenue to be collected for a single capacity unit at a certain point in time. It is the threshold value against which customer bids can be compared to evaluate whether a transaction is being made. For a more detailed introduction see, e.g., Klein and Steinhardt (2008) or Talluri and van Ryzin (2005).

introduce a robust optimization approach for the stochastic version of their earlier model.

Burnetas and Smith (2000) develop an adaptive pricing and ordering solution to the newsboy problem, including a mechanism for pricing under an unknown arrival rate distribution.

Overviews and Introductions Very thorough and current introductions to dynamic pricing in general can be found in McAfee and te Velde (2006); Phillips (2005) and Talluri and van Ryzin (2005). Recent literature overviews can be found in Chan et al. (2004) and Gönsch et al. (2009), the earlier of which also includes various sources on restocking and inventory decisions; as well as Bitran and Caldentey (2003) and Elmaghraby and Keskinocak (2003), which are more revenue-management oriented.

3.2.2.2 Dynamic Pricing with Strategic Customers

The assumption of myopic customers in above section is widely applied in practical settings and pricing software but is most meaningful when customers make impulse purchases or buy readily consumable goods (food, snacks, etc.). It has the considerable property of leading to mathematically tractable models and formulations; however, it is more realistic to consider the strategic nature of customers explicitly, especially when they are purchasing expensive or durable goods (Liu and van Ryzin, 2008b, pp. 110).

Dynamic pricing models that incorporate strategic customers naturally have to assume a finite population, as “a meaningful model incorporating strategic customers requires consideration of customers individually” (Levin et al., 2006b, p. 2). Unfortunately, finite-population models under strategic customer behavior are more complex than the models from the previous section, and hence, often only the deterministic case is explored. Accordingly, there is only a rather small amount of literature available.

Pure Utility Maximization Besanko and Winston (1990) consider the deterministic case in which a monopolist faces consumers who act strategically and maximize their “intertemporal utility”. They show that prices are generally lower as with myopic customers and, moreover, that failing to use the derived equilibrium policy by the seller results in significantly lower profits.

Exploiting Time sensitivity A model based on strategic customers but infinite time horizon and demand population is presented by Gallien (2006).

Customers are time sensitive (i.e., impatient) and the resulting optimal policy increases the price after each sale and henceforth over time towards departure.

Su (2007) considers a deterministic demand model with heterogeneous customers segmented along two dimensions: valuation of the product and waiting costs (i.e., patience). He finds the degree of heterogeneity to be important, as the two dimensions jointly determine the structure of the optimal pricing policies: when high-valuing customers are less patient, markdown policies are effective, while on the contrary when high-valuing customers are patient, prices should naturally increase over time. Surprisingly, he also finds strategic behavior sometimes to be beneficial to the seller, as waiting customers may later compete for scarce capacity at higher prices.

Creating Rationing Risk Liu and van Ryzin (2008b) consider a model with two periods where control is exercised through quantity decisions rather than price alone. They show that the seller can induce demand in the early period by credibly rationing capacity for the second period if customers are risk-averse and fear that the quantity on sale in the second period might be scarce.

Liu and van Ryzin (2008b) explicitly incorporate risk-averse customers who tend to purchase early to avoid rationing risk. The results show that it is optimal for the seller to either create extensive rationing risk for the customer or none at all in cases where the seller is not able to credibly announce capacity rationing in future periods.

Aviv and Pazgal (2008) study the case of fashion retailing where the valuation of customers deterministically decreases with time. They show that strategic behavior suppresses the benefits of price segmentation, particularly when customers are very heterogeneous and the valuation declines slowly over time. Even when faced with strategic consumers, they find announced pricing policies to be more advantageous to the seller than contingent-based pricing schemes. Finally, in their numerical study, they show that the potential loss of revenue can reach 20% when customers are incorrectly assumed to behave myopically.

A similar model featuring pre-announced markdowns but customers with multi-unit demands who exhibit fixed valuation across the season is derived in Elmaghraby et al. (2008), who present a stochastic model for periodic markdowns with multi-unit demand where prices are updated at fixed times during the sales interval. They find that the optimal policy uses no more than two or three price steps depending on whether the customer knows the clearing price (i.e., the price at which demand exceeds the supply/capacity). In both settings, it is optimal for the buyer to submit so-called all-or-nothing bids (i.e., the buyer always submits a bid for its full – possibly multi-unit – demand).

Strategic Games Levin et al. (2006b) derive two models encompassing strategic behavior by customers and sellers in a unified stochastic game. They consider the monopolistic case, where they prove the existence of a unique subgame-perfect equilibrium pricing policy that is also mathematically tractable for realistic problem sizes. The approach is extended in Levin et al. (2006a) to also cover oligarchic competition.

3.2.2.3 Customer Choice Models

Conventional quantity-based revenue management models treat demand as an independent stochastic process (see, e.g., Talluri and van Ryzin, 2005), as it is assumed that the characteristics of the underlying population do not change when customers arrive and make their purchase decision. Albeit based on simplifying assumptions, these models worked well before the advent of LCCs and the accompanying increased price transparency (Gallego and Hu, 2006, pp. 2), but they now may produce problems in practical settings (see Wilson et al., 2006) as the independent demand assumption becomes untenable (Cooper et al., 2006, p. 969).

Customer Choice Models in Dynamic Pricing Recent papers explicitly introduce customer choice models to dynamic pricing and revenue management to overcome the above-mentioned issue. An early practically oriented approach to the topic can be found in Proussaloglou and Koppelman (1999), who model customer choice behavior between carriers, flights and fare classes. Talluri and van Ryzin (2004) consider the single-leg case in which customers simply choose between open fare classes (in a revenue management type of setting). Gallego et al. (2004) and Liu and van Ryzin (2008a) consider choice-based linear programming models for instances with network effects (i.e., shared scarce resources). A pure airline network problem is resolved in van Ryzin and Vulcano (2006) based on a customer choice model. The authors propose a simulation-based approach to obtaining virtual nesting controls. Zhang and Cooper (2005, 2009) consider multiple, parallel flights for a single carrier between the same origin and destination but without network effects. Gallego and Hu (2006) and Mishra et al. (2005) additionally do explicitly consider competition in the customer choice setting.

Table 9.1 in Part III gives an overview of specific practically oriented work on the actual employment of customer choice models in real-life settings.

Overview and Introductions Good introductions to the topic from the economics side can be found in McFadden's Nobel lecture (McFadden, 2000b) and in Small (2006). For an in-depth review of customer choice behavior, the reader should refer to McFadden (2000a). Introductions to discrete choice models in general, which are probably the most widely used in practice, can be found in Ben-Akiva and Lerman (1985); Koppelman and Bhat (2006) and Train (2003).

The next section discusses the major limitations and shortcomings of the discussed dynamic pricing models with respect to the objectives given throughout earlier Section 2.3.

3.3 Limitations and Shortcomings

This section discusses the limitations and shortcomings of the relevant literature and models discussed in above Section 3.2. It addresses oversimplified modeling approaches and inaccurate assumptions related to dynamic pricing models in general (Section 3.3.1) or demand learning in particular (Section 3.3.2). At the same time it lays the foundation for the presentation of the author's own approaches to demand learning and modeling of price sensitivity, which are finally outlined in Section 3.4.

3.3.1 Dynamic Pricing Models

Numerous work on dynamic pricing does not explicitly differentiate pure customer arrival – i.e., latent demand – from price-dependent buying decisions – i.e., realized demand (see, e.g., Chatwin, 2000; Gallego and Hu, 2006; Heching et al., 2002; Gallego and van Ryzin, 1994 or Liu and Milner, 2006). While this tendency is convenient in terms of model complexity and mathematical tractability (especially in stochastic models), it fails to capture realistic settings appropriately: pure customer arrival is mostly not driven by the seller's pricing policy (see Bitran and Mondschein, 1993, 1997) but mainly rests on external factors (e.g., weather, economic situation, seasonality etc.) and other internal but non-pricing marketing levers (e.g., advertising, distribution strategy, branding etc.).[16]

Many recent publications address this issue (see, e.g., Aviv and Pazgal, 2008 or Lin, 2006) and explicitly differentiate between the pure customer

[16] In the long run, a customer-friendly pricing policy might also increase latent demand (e.g., through goodwill and brand awareness), but in a short and medium term view, pricing affects customers only *after* their arrival at the store.

arrival process (which is typically assumed to follow a Poisson distribution) and the reservation price distribution of customers that already arrived at the point of sale (often assumed to be a member of the exponential family). Besides improved mapping of real-world behavior through the differentiation between internal and external effects or influences, this separation naturally fosters the identification of the underlying distributions and their controlling parameters. This in turn is of tremendous importance in real-world settings where the latter are typically not known in advance but instead have to be derived or learned manually.

This leads to a review of the typically assumed demand behavior and information settings of the presented literature: The reference model of Gallego and van Ryzin (1994) and many extensions to it (see, e.g., Elmaghraby et al., 2008 or Feng and Gallego, 1995) are based on two assumptions that may well be violated in reality:

(A1) **Time-invariance:** The distributions of the customer arrival rate and price sensitivity at any given price or point in time are stationary.

(A2) **Full information:** The seller knows the type of underlying distributions and the defining parameters for the customer arrival rate and price sensitivity at any given price or point in time.

While (A1) is later relaxed (see, e.g., Bitran and Mondschein, 1993, 1997; Lin, 2004 or Zhao and Zheng, 2000), (A2) or variations of it (e.g., the seller knows the pure demand arrival rate at any given point in time) are prevalent even in current models (see, e.g., Aviv and Pazgal, 2008; Liu and van Ryzin, 2008b or Su, 2007). However, naturally in most realistic settings, neither assumption (A1) nor assumption (A2) holds true:

1. Customer arrival depends on various internal and external influences (see above) and typically exhibits some sort of (micro-)seasonality or trend. Additionally, in many industries, customer valuation changes in time (e.g., in retail settings, customers tend to value new products highest at the beginning of the season; conversely, in the air travel industry, high-value customers tend to book at the end of the booking period).

2. Both, customer arrival rate and valuation are barely observable (especially in offline market settings) and often also difficult to estimate because of continuously changing market dynamics (competitor activities, product refinements, etc.) or – in the case of new introductions – scarce historical information. Even the often cited management experience will usually not easily transform into quantifiable knowledge about

the assumed form of stochastic distributions with associated parameter values.

Therefore, dynamic pricing models should account for realistic assumptions in terms of time-variant distributions. Additionally, they should provide systematic approaches for deriving missing information about underlying distributions and time-variant parameters from the available (ideally non-historic) demand and sales information. Ideally this includes the assessment of experience and knowledge from management and pricing analysts.

The described gap is partially bridged by the novel demand learning models discussed in Section 3.2.1 – which, however, still have their limitations, as are discussed in the following Sections.

3.3.2 Demand Learning Models

Similar to most dynamic pricing literature, many demand learning models in dynamic pricing do not differentiate between the stochastic distribution of customer arrival and reservation price (see, e.g., Easley and Kiefer, 1988; Kiefer and Nyarko, 1989; Petruzzi and Dada, 2002 or Subrahmanyan and Shoemaker, 1996), which yields the same downsides as for pure dynamic pricing models (see above Section 3.3.1).

Additionally, most work assumes a specific type of distribution for which the underlying parameters have to be learned in the pricing process. This is typically driven by the selected learning method based on Bayesian updating (see, e.g., Araman and Caldentey, 2005; Aviv and Pazgal, 2002; Cope, 2007; Lin, 2006 or Zhang and Chen, 2006), which – for the sake of mathematical tractability – requires the distribution that is learned to be of a certain conjugate family of distributions (see Section 4.2). Only few authors explicitly study the determination of the appropriate distribution from the data or the fitting of data to a set of pre-selected distributions (see, e.g., Anjos et al., 2004, 2005). At the same time, Bell and Zhang note that "(...) large forecast errors lead to decisions that differ significantly from optimality and carry a high expected contribution penalty" (Bell and Zhang, 2006, p. 385), which is necessarily the case when using erroneous assumptions about underlying demand distributions to derive pricing policies or to learn their parameters.

A fundamental assumption to allow for demand learning is that the underlying distribution is independent and identically distributed (IID) over time (i.e., time-invariant) or that there exists a fixed functional dependency of the driving parameters in the time-variant case (see Lin, 2006). Otherwise, the underlying distribution parameters by definition cannot be learned, which is the reason why most works simply do not allow for time-variant distributions (see, e.g., Araman and Caldentey, 2005; Aviv and Pazgal, 2002; Cope, 2007; Farias and van Roy, 2007; Fisher and Raman, 1996; Raju et al., 2004; Subrahmanyan and Shoemaker, 1996 or Zhang and Chen, 2006) or instead explicitly assume a functional – often linear – dependency between parameters (see, e.g., Balvers and Cosimano, 1990; Bertsimas and Perakis, 2006 or Kachani et al., 2007).

To the author's knowledge, all existing demand learning literature on dynamic pricing only considers learning over time using data and sometimes expert/management advice from a single product or service only. This poses a considerable limitation but offers an additional research opportunity in cases where information about substitute, complementary or parallel services or products could enhance the learning process:

- **Styles-goods retailing:** Demand changes for pivotal products might act as indicators for other related items within a given season. For example, rising demand for swimsuits and bikinis may indicate the start of the summer season's shopping period, followed by increased demand in other related (but possibly time-lagged) categories like towels, beachwear, sandals, etc. Similarly, high demand for orange-colored spring clothing might indicate that orange will be the hot-selling color in the summer as well.

- **Air travel services:** The demand for flights and the price sensitivity of customers might be linked in multiple ways. First, from a micro-seasonal perspective, flights that leave within a certain time-frame might exhibit functional dependency in the sense that demand from the most attractive flights (e.g., Monday morning) might roll over to parallel flights (e.g., Sunday evening). Similar effects could be true for price sensitivity (possible after a time lag). Second, within some macro-season, isochronous flights (e.g., each last departure on Friday nights) might behave in a similar fashion.

- **Consumer-packaged goods:** Demand for complementary products might be tightly linked but exhibit a different timely distribution. For example, demand for breakfast cereals might also trigger demand for milk, but possibly along multiple future purchasing occasions, as the consumption rates might differ and milk is typically not bought in bulk, but rather timely close to consumption. Successive demand behavior could be prevalent for products related to a certain phase in life, like baby food and diapers.

In these scenarios, learning of demand parameters would be possible even in cases of time-variant *and* time-independent stochastic distributions, as learning stems from information about *other products*, whose demand then would have to be dependent on that of the main product – as obviously is assumed in the above examples.

Based on the outlined critique and limitations of the existing literature above, the following section generally describes the proposed enhanced learning approach for latent demand as well as the specific type of choice models for deriving the customers' price sensitivity that are pursued in this work.

3.4 Proposed Approach

This section outlines the approaches to demand learning and modeling of choice behavior in dynamic pricing that are researched in this work. The focus here is on giving a general overview that is mainly independent of a specific industry setting, but highlights its advantages and the differences from the existing literature discussed in above sections.

Note that the objective of this work is *not* to develop a new dynamic price optimization model with a different set of underlying assumptions, but rather to tackle the typically assumed full information setting in the established literature that is critiqued in above Section 3.3.

A new passive demand learning mechanism is presented for learning the unknown functional composition of latent demand while not incurring the drawbacks discussed in Sections 3.3.2. Hence, the presented model is independent from a specific dynamic pricing mechanism and can be employed as substitute for any other passive learning model (see Section 3.2.1). The concept is introduced in its general form in this section and operationalized and tested against a real-world setting in Part II, where the necessary mathematical concepts and methodologies are also explained in detail.

A specific approach for deriving price sensitivities and elasticities from automatically collected data based on an universal formulation of conventional customer choice models is introduced in its general idea below and again operationalized and tested for real-world application later in Part III.

As in other works (see Sections 3.2.2.1 and 3.2.2.2), the presented approaches separate latent demand in the form of pure customer arrival from realized demand that depends on the prevalent pricing policy in conjunction with competitor prices and specific customer behavior. The former is learned in Part II of this work, while Part III illuminates possible learning of price sensitivity parameters. Besides facilitating a better reflection of the real-world context, this procedure allows both models to be based on uncensored data (i.e., recorded through an online sales channel), which is naturally much more exact than any method based on censored realized demand.

Here latent demand D is assumed to be a function $f(\cdot)$ of external or internal but non-price-related factors. In turn, realized demand d is then calculated as the fraction of customers ω of the latent demand that actually take a buying decision (in the airline industry, the so called book-to-look ratio). That in turn is also a function $g(\cdot)$ of internal and external factors (that may well include own and competitor prices) – which is in line with many previous works (see Section 3.2)

$$\begin{aligned} d &= D \cdot \omega && \Leftrightarrow \\ d &= f(\cdot)g(\cdot). && \qquad (3.1) \end{aligned}$$

Latent Demand The function that defines latent demand takes as input a set of external (possibly also functionally transformed) deterministic parameters $\mathbf{x}$ (e.g., weekday, time, etc.) and a set of corresponding stochastic coefficients $\mathbf{a}$

$$D = f(\mathbf{x}, \mathbf{a}). \qquad (3.2)$$

The definition in (3.2) is thereby general enough to allow for a variety of functional demand definitions while explicitly differentiating stochastic and deterministic input.

Each coefficient a_m in $\mathbf{a} = (a_1, \ldots, a_M)^T$, indicating the weight of the corresponding parameter x_m, follows a stochastic distribution $\mathcal{A}_m$ controlled by an associated set of parameter values $\boldsymbol{\psi}_m$ that are unknown in the beginning, although some prior information on the parameter values might exist either from historic observations or from management expertise

$$a_m \sim \mathcal{A}_m(\boldsymbol{\psi}_m) \quad \forall\, m. \qquad (3.3)$$

When the functional form of (3.2) and the distribution types in (3.3) are known, and when the deterministic parameters as well as the actual (de facto realized) latent demand is recorded, then principally the parameter values $\boldsymbol{\psi}_m$ for each distribution $\mathcal{A}_m$ can be learned using Bayesian updating along Bayes' theorem (a detailed introduction to Bayesian inference is given in Chapter 4).

Latent demand is allowed to be time-variant, as the parameter vector $\mathbf{x}_t$ may change over time, while the distribution parameters for the coefficients are assumed to be time-invariant to allow for learning of the latter, so that

$$D_t = f(\mathbf{x}_t, \mathbf{a}). \tag{3.4}$$

Additionally, parameters and coefficients determining demand for a certain product i can be dependent on the product itself $(\mathbf{x}_t^i, \mathbf{a}^i)$ or on related products j in the sense of Section 3.3.2 $(\mathbf{x}_t^j, \mathbf{a}^j)$, so that

$$D_t^i = f(\mathbf{x}_t^i, \mathbf{a}^i, \mathbf{x}_t^j, \mathbf{a}^j). \tag{3.5}$$

Following (3.5), the demand learning process for product i can then benefit from possibly earlier completed learning regarding other products j, i.e., the distribution parameters of $\mathbf{a}^j$ might already be well established through learning in earlier periods, which in turn would reduce the uncertainty surrounding D_t^i.

To the author's knowledge, such consideration of parameter inputs from other, possibly time-lagged products has not been reported in prior work. The advantages of the suggested approach are three-fold (see also Chapter 8):

- **Broad functional definition:** The model factors driving latent demand can be based on a large argument base, allowing for a wider and possibly more accurate functional definition.

- **Exploitation of time lags:** Mutual coefficient values (e.g., the demand effect of weekends) could be derived from other products earlier in time than for the observed product alone.

- **Enhanced learning:** Stochastic parameters that are to be learned might not be IID over time for the same product (which would prevent learning), but within certain time-frames they might be IID to those of other products, so that learning could take place on an inter-product basis.

Price Sensitivity The function to calculate the fraction of customers who actually take a buying decision firstly takes as input the specific characteristics or needs $\mathbf{s}$ of the customers actually having arrived at the point of sale (e.g., weekday of preferred departure, advance purchase time, etc.), which are entirely external to the affected company. Additionally, customer purchase decisions are naturally controlled by the internal input parameters consisting of attributes $\mathbf{x}$ of the sellers' choice selection (e.g., weekday of departure, price, etc.). The actual functional composition is again assumed to be defined by a set of coefficients $\mathbf{a}$ that determine the weight for each individual (possibly functionally transformed) input variable (internal and external)

$$\omega = g(\mathbf{s}, \mathbf{x}, \mathbf{a}). \tag{3.6}$$

The outcome regarding customer decision making is allowed to be time-variant as both, the vector of customer characteristics $\mathbf{s}_t$ and the parameter vector of the available choices $\mathbf{x}_t$ may change over time. The coefficients $\mathbf{a}$ are assumed to be time-invariant to allow for time-independent estimation of the latter, so that

$$\omega_t = g(\mathbf{s}_t, \mathbf{x}_t, \mathbf{a}). \tag{3.7}$$

In reality, parameters and coefficients determining the purchasing behavior of customers depend not only on the available internal choices, but also on the available external ones, reflecting the influence of competition. Therefore, two separate sets of parameters and coefficients are considered, one for the internally provided choices i ($\mathbf{x}_t^i, \mathbf{a}^i$) and one for the external options j ($\mathbf{x}_t^j, \mathbf{a}^j$) – both possibly with multiple entries, if the considered company and its competition offer multiple choice alternatives

$$\omega_t^i = g(\mathbf{s}_t, \mathbf{a}^s, \mathbf{x}_t^i, \mathbf{a}^i, \mathbf{x}_t^j, \mathbf{a}^j). \tag{3.8}$$

Following (3.8), the estimation of the driving coefficients can be conducted without explicit knowledge about the customers' actual choices regarding possible competitive products. Whilst most literature does not explicitly consider competitive attributes in the functional composition of the purchasing probability of a single product, the derived model in Part III necessarily has to consider competition attributes stemming from the source of automatically collected data.

To the author's knowledge, such consideration of competitive attributes using incomplete but automatically collected information about the competition has not been reported in prior work. The advantages of the approach are threefold (see also later Chapter 13):

- **Consideration of competition:** The model explicitly incorporates the effect of competition on customer choice decisions, but achieves this through a formulation that does not depend on a full information setting regarding the actual choice of customers.

- **Broad functional definition:** The functional composition of customer choice behavior may include a large attribute base (internal and external) for each individual choice option.

- **Prevention of bias:** Through the exclusive usage of automatically collected revealed behavior data, the threat of data bias as typically found in stated choice data is avoided.

The remaining chapters detail and operationalize the outlined approaches to demand learning and estimation of purchase probability using a proprietary real-world data sample. First, in Part II, a self-learning Bayesian update scheme is introduced for a definition of latent demand that is linear in its parameters. In Part III, an universal formulation of the multinomial logit model is employed to understand the drivers behind customer price sensitivity and purchase behavior.

Part II

Forecasting Latent Demand

Part II Objective

Following the introduction to dynamic pricing and the airline industry in Part I of this work, this part's objective is to actually develop a forecasting methodology for latent customer demand in low-cost air transportation markets.

Most work on dynamic pricing implicitly assumes the existence of such forecasts, but only a few explicitly develop mechanisms to generate usable models (see Chapter 2). The methodology developed here is general enough to fit most dynamic price optimization models from Section 3.2 and has moreover been developed and tested using real demand data collected by a German hybrid LCC.

In particular, the pursuit approach distinguishes itself based on the following characteristics, which in combination are new to the academic literature:

1. The objective is to build a forecasting model for *latent demand* – not just for the eventually generated bookings, as is done in most of the existing revenue management forecasting systems (see, e.g., Talluri and van Ryzin, 2005, Chap. 9).

2. The forecast is supposed to be based on a fixed model structure, but with self-learning and self-correcting parameters, so that it dynamically adjusts to changing market environments.

3. The employed input data do not contain any historic or stochastic data but solely consist of deterministic inputs and real-time or online data, which is especially important for LCCs where historic data might be scarce.

4. Observations are collected for a particular flight event under forecast and additionally for multiple adjacent flights under the assumption of structural similarities of timely close departures.

5. The learning mechanism is geared towards automatic forecast generation without the need for manual adjustments (e.g., for changing calendar effects).

6. While the resulting forecast can serve directly as input for existing dynamic price optimization systems, the model also generates structural insights regarding the underlying demand behavior that may also be used as a well-founded basis for management decisions and manual pricing interventions.

7. The aim is to provide a 60-day forecast for the last two months before flight departure with daily granularity, as roughly 75% of requests arrive within that period.

However, the developed model inevitably yields forecast errors, as naturally not all factors and external effects driving demand can be exhaustively included.

Chapter 4

Self-Learning Linear Models

This chapter covers the theoretical and technical background of the specific learning method later employed in Chapters 6 and 7 to actually forecast latent demand based on its characteristics as described in Chapter 5. The presented method rests on the *Bayesian* interpretation of probability, which is fundamentally different from the *classical* or *frequentist* interpretation, where probabilities are simply viewed "in terms of the frequencies of random, repeatable events" (see, e.g., Bishop, 2006, p. 21).

Denison et al. (2002, p. 4) note that "there are [a] huge number of ways to approximate the truth and no one single specific approach can be uniformly better than any other in terms of predictive ability." With this in mind, the choice of a Bayesian method in this work is based on three specific reasons:

1. It is understandable, with limited mathematical knowledge, which could foster broad application in real-life settings (see the rest of this chapter).

2. It is technically capable of constituting the functional form exhibited by the collected data (see Chapter 5).

3. It includes a self-learning mechanism that is sufficiently flexible in terms of application breadth, while remaining computationally efficient (see Chapters 6 and 7).

The remainder of this chapter lends a mathematical introduction to the chosen method. Section 4.1 introduces traditional linear basis function models, which later form the basis for Bayesian linear regression in Section 4.3, after a quick introduction to Bayesian statistics in general and Bayesian inference in particular has been given in Section 4.2.

Throughout the remaining sections, the reader is expected to be familiar with basic principles of probability theory and matrix algebra. Introductions

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6_4,

to these topics can be found in Artin (1991), Graybill (2001) or Roussas (1973), amongst others.

4.1 Linear Regression Models

In linear regression modeling, given T observations $\{\mathbf{x}_1, \ldots, \mathbf{x}_T\}$ of sets of external or independent variables $\mathbf{x}_t = (x_{1t}, \ldots, x_{Kt})^T$ together with the corresponding stochastically dependent target values y_t, the goal is to understand the functional dependency between $\mathbf{x}$ and y, so as to later be able to predict the value of y_{T+1} for a new set of input variables $\mathbf{x}_{T+1}$. Ideally, this can be done by constructing a function $f(\cdot)$ whose values for arbitrary inputs of $\mathbf{x}$ yield the prediction for the corresponding value of y.

The underlying rationale in regression analysis is that there exists a fixed functional dependency between the independent input variables and the dependent result variable. Figure 4.1 illustrates an example based on an assumed relationship between flight sector distance and fuel consumption, which obviously exists – although it is visibly non-linear.[1]

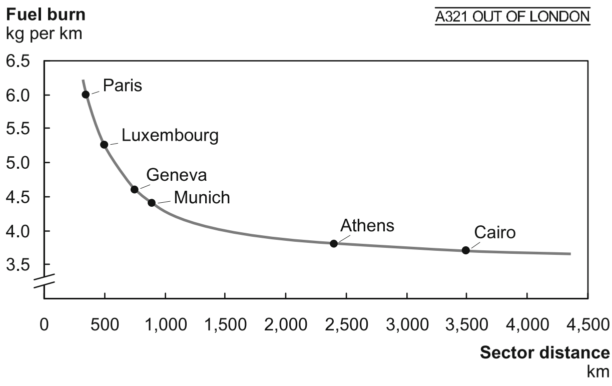


Figure 4.1: Regression example: impact of sector distance on fuel burn – Airbus A321-200 on routes from London
Source: Based on Doganis (2002, p. 129)

[1] The illustrated dependency between sector distance d and fuel burn rate f in Figure 4.1 is polynomial with degree three. The model underlying the observations in the graphic – but not verified using a larger sample – is based on the logarithmized distance: $f = 77.50 - 26.50 \cdot \ln(d) + 3.20 \cdot \ln(d)^2 - 0.13 \cdot \ln(d)^3$.

Basis Function Models The simplest linear model for a regression is constituted by a linear combination of K external input variables

$$y = f(\mathbf{x}, \mathbf{a}) = a_0 + a_1 x_1 + \ldots + a_K x_K = \mathbf{x}^T \mathbf{a}, \tag{4.1}$$

where $\mathbf{x} = (x_0, \ldots, x_K)^T$ is the vector of input variables with $x_0 := 1$ and $\mathbf{a} = (a_0, \ldots, a_K)^T$ is a vector of coefficients defining the linear functional combination of the variables, such that this particular form is simply known as *linear regression*. The individual coefficients in $\mathbf{a}$ indicate the degree to which the dependent variable will change once the corresponding independent variable is altered. While elegant because of its simplicity, the model allows only for a linear combination of the input variables in $\mathbf{x}$, which imposes significant limitations on the constructible models.

However, the capabilities of this class of models can easily be extended by allowing the independent variables to be nonlinear functions $\phi_m(\cdot)$ of the original input variables $\mathbf{x}$, known as *basis functions*. Note that the index $m = \{0, \ldots, M\}$ of the basis functions used in the model is independent of the number of actual input variables $k = \{0, \ldots, K\}$. Defining an additional dummy basis function $\phi_0(\mathbf{x}) := 1$ the extended model can be expressed as

$$y = f(\mathbf{x}, \mathbf{a}) = \sum_{m=0}^{M} a_m \phi_m(\mathbf{x}) = \boldsymbol{\phi}(\mathbf{x})^T \mathbf{a}, \tag{4.2}$$

where $\boldsymbol{\phi}(\mathbf{x}) = \big(\phi_0(\mathbf{x}), \ldots, \phi_M(\mathbf{x})\big)^T$.

Employing nonlinear basis functions in place of $\boldsymbol{\phi}(\mathbf{x})$ allows for the creation of nonlinear models or functions of the input vector $\mathbf{x}$. However, regression models of the form (4.2) are still called linear models because the functional composition is still linear in the coefficients $\mathbf{a}$.

There is a vast choice of possible basis functions, including, but not limited to, the exponential function $\phi_m(\mathbf{x}) = e^{x_k}$, powers $\phi_m(\mathbf{x}) = x_k^j$ or the logarithm $\phi_m(\mathbf{x}) = \log(x_k)$. Bishop (2006, Sec. 3.1) gives a detailed overview of frequently used basis functions. The exemplary regression in Figure 4.1 is based on $\phi_i(\mathbf{x}) = \ln(x_1)^i,\ i = \{0, \ldots, 3\}$ with $\mathbf{x} = (x_1)^T$.

No matter what types of basis functions and particular models are used, the actual observations of the dependent variable will typically not exactly match the function values of the model, as "a regression model is a formal means of expressing the two essential ingredients of a statistical relation" (Neter et al., 1983, p. 26): The dependent variable varies with the independent variables in a systematic or functional fashion, but the actual observations scatter around the expected functional relationship in a stochastic fashion. These characteristics explicitly acknowledge uncertainty in the model by postulating that (see Neter et al., 1983, pp. 26):

1. In the population of observations, there exists a probability distribution $\Psi(\cdot)$ of y for each level of $\mathbf{x}$ with $p(y|\mathbf{x}, \mathbf{a}, \sigma^2) = \Psi(y|f(\mathbf{x}, \mathbf{a}), \sigma^2)$.[2]
2. At the same time, only the mean $\mu = f(\mathbf{x}, \mathbf{a})$ of this probability distribution $\Psi(\cdot)$ varies according to the assumed functional relationship.

Referring back to the example relationship from Figure 4.1, the above implies that the actual fuel burn on the mentioned routes may vary around the expected rate (e.g., depending on weather and traffic situation), as is illustrated in Figure 4.2.

As the true functional relationship is unknown and cannot be exactly derived in analytic form – even if it exists – a reasonable objective is to search for a close approximation $\tilde{f}(\cdot)$, both in functional form and in the estimation of the parameter coefficients (see Denison et al., 2002, p. 14).

The multivariate regression analysis is a confirmatory statistical method; therefore, no straightforward or mathematical method exists for deriving the

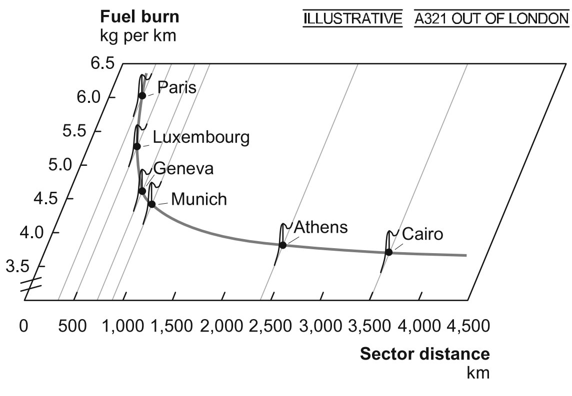


Figure 4.2: Regression example (continued): inclusion of probability distributions in functional dependency model
Source: Extension of Doganis (2002, p. 129)

2 Here σ^2 is the variance of the errors, defined by the differences of true observation y and their functionally explained part $f(\mathbf{x}, \mathbf{a})$.

most exact form of $\tilde{f}(\cdot)$. Instead, the hypothesis of a particular assumed functional relationship can only be tested for validity, as will be discussed at the end of this section. First, the next paragraph presents the most commonly used technique for estimating the vector of coefficients $\mathbf{a}$ for a predetermined functional form of $\tilde{f}(\cdot)$.

Ordinary Least Squares In frequentist interpretation, the independent variables $\mathbf{x}$ are assumed to be observed without error; that is, they are considered deterministic because they are externally determined based on a sufficiently high amount of observations T.

The true functional form of the dependent variable $f(\cdot)$ is then approximated using a deterministic function $\tilde{f}(\mathbf{x}, \mathbf{a})$ – here with additive normal distributed noise or error ϵ (see Bishop, 2006, pp. 140)

$$y = f(\cdot) = \tilde{f}(\mathbf{x}, \mathbf{a}) + \epsilon \qquad \text{with } \epsilon \sim \mathcal{N}(0, \sigma^2), \tag{4.3}$$

implying that

$$y = f(\cdot) \sim \mathcal{N}\big(\tilde{f}(\mathbf{x}, \mathbf{a}), \sigma^2\big). \tag{4.4}$$

Under the assumed normal distributed errors, the regression function $\tilde{f}(\mathbf{x}, \mathbf{a})$ will typically not exactly match the observed dependent values y but will rather be scattered around it (see Figure 4.3 for an illustration). These deviations may have two underlying structural causes (see Backhaus et al., 2006, p. 78):

- **Systematic error:** Structural factors that are either unknown to the analyst or simply cannot be measured and accounted for.
- **Stochastic error:** Truly stochastic variations in the observations based on random effects in reality or coincidental measurement errors.

Under the *ordinary least squares* (OLS) estimation scheme, the objective for a given model and a set of T observations is to find an estimate $\hat{\mathbf{a}}$ so that the squared sum of deviations or residuals e_t within the observation sample is minimized:

$$\min_{\hat{\mathbf{a}}} \sum_{t=1}^{T} e_t^2 = \sum_{t=1}^{T} [y_t - \tilde{f}(\mathbf{x}_t, \hat{\mathbf{a}})]^2. \tag{4.5}$$

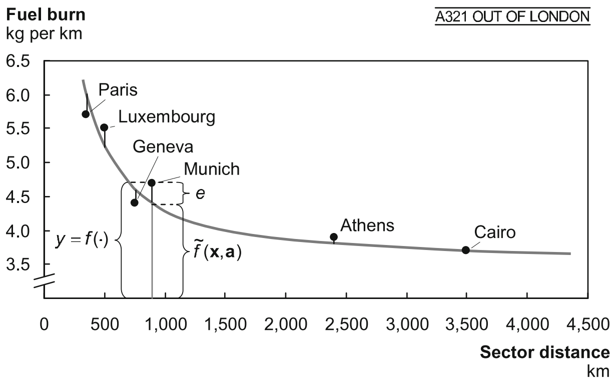


Figure 4.3: Regression example (continued): scattering of observations and geometry of least squares
Source: Extension of Doganis (2002, p. 129)

The OLS scheme, as the most common statistical method for estimating regression parameters, entails two intuitive advantages (see Backhaus et al., 2006, pp. 58):

1. Deviations are recorded in absolute values, which prevents individual but oppositional errors from compensating each other.

2. Extreme discrepancies are heavily penalized by the squared measure.

To minimize (4.5) over a set of T observations, a system of linear equations has to be solved. Assembling the observations of the dependent variable in a vector $\mathbf{y} = (y_1, \ldots, y_T)^T$ and the independent variables in a matrix $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)^T$, that system can be expressed as

$$\mathbf{y} = \mathbf{X}\mathbf{a} + \mathbf{e}, \tag{4.6}$$

where $\mathbf{e}$ is the vector of residuals with dimension $T \times 1$. The minimizing result for the system is then obtained by setting (for a mathematical discussion and proof see, e.g., Lehn and Wegmann, 2006 or Neter et al., 1983)

$$\hat{\mathbf{a}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}. \tag{4.7}$$

To allow for a linear basis function model in the sense of (4.2), the system can be transformed to (see Bishop, 2006, pp. 138)

$$\mathbf{y} = \mathbf{\Phi a} + \mathbf{e} \tag{4.8}$$

with corresponding estimate

$$\hat{\mathbf{a}} = (\mathbf{\Phi}^T \mathbf{\Phi})^{-1} \mathbf{\Phi}^T \mathbf{y}, \tag{4.9}$$

where $\mathbf{\Phi}$ has dimension $T \times M$ and is called the *design matrix*. Its elements are given by $\Phi_{tm} = \phi_m(\mathbf{x}_t)$, so that

$$\mathbf{\Phi} = \begin{pmatrix} \phi_0(\mathbf{x}_1) & \phi_1(\mathbf{x}_1) & \cdots & \phi_M(\mathbf{x}_1) \\ \phi_0(\mathbf{x}_2) & \phi_1(\mathbf{x}_2) & \cdots & \phi_M(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_0(\mathbf{x}_T) & \phi_1(\mathbf{x}_T) & \cdots & \phi_M(\mathbf{x}_T) \end{pmatrix}, \tag{4.10}$$

with M being the index of the basis functions (see (4.2)) and T being the index of the observations. Substituting the coefficient vector in (4.6) by the estimates in (4.7) then results in the forecast or explicable part of the dependent variable

$$\hat{\mathbf{y}} = \mathbf{\Phi} \hat{\mathbf{a}}. \tag{4.11}$$

Under a set of sufficient assumptions for the model and the variables (see below), the estimators in (4.7) and (4.9), respectively, show desirable properties. Namely, they are said to be the *best linear unbiased estimators* (BLUE) (see Backhaus et al., 2006, pp. 79). "Unbiased" means that $E(\hat{\mathbf{a}}) = \mathbf{a}$ and "best" stands for the lowest variance within all unbiased and linear estimators (for proof and details, see e.g., Bleymüller et al., 2004 or Kmenta, 1997). The necessary assumptions and conditions for yielding estimators that are BLUE in above sense are detailed in the following paragraph.

Model prerequisites The linear regression model and the above given estimator $\hat{\mathbf{a}}$ for the coefficient vector $\mathbf{a}$ are based on a set of specific assumptions for the model and its parameters (see e.g., Backhaus et al., 2006, pp. 79):

(A1) The model is correctly specified, meaning that it is linear in $\mathbf{a}$, includes all relevant independent variables $\boldsymbol{\phi}(\mathbf{x})$ and is estimated based on a sample of observations T that is greater than the number of basis functions used $T > M$.

(A2) The residuals have expectancy of zero, $E(e) = 0$, and are also normally distributed, $e \sim \mathcal{N}(0, \sigma^2)$. Here, the latter requirement is not an integral part for the linear regression model itself but is rather a prerequisite for the applicability of the significance tests later needed to verify these assumptions.

(A3) The residuals exhibit a constant variance independent of the particular sample selection (*no heteroscedasticity*): $Var(e_1) = Var(e_2) = \sigma^2$ with $\mathbf{X}_1, \mathbf{X}_2 \subseteq \mathbf{X}$.

(A4) The residuals exhibit no serial correlation according to the order of observations (*no autocorrelation*): $Cov(e_t, e_{t+i}) = 0\ \forall t$ and $i \neq 0$.

(A5) The independent variables $\phi_m(\mathbf{x})$ are not fully linearly dependent on each other (*no multicollinearity*), i.e., each input basis function carries truly additional information: $\forall\ m, \nexists\ \mathbf{k} = (k_1, \ldots, k_{m-1})^T$ so that $\phi_m(\mathbf{x}_t) = \sum_{n \neq m} k_n \cdot \phi_n(\mathbf{x}_t)\ \forall\ t$.

For a particular regression model $\tilde{f}(\mathbf{x}, \mathbf{a})$ with its estimates $\hat{\mathbf{a}}$ to be considered valid, each of the above assumptions (A1) – (A5) has to be verified using the appropriate statistical tests. Otherwise, the model in question might not follow (4.3), and the obtained estimate from (4.9) might not be BLUE, but might instead be heavily biased.

Given a particular regression model has been found to adhere to the necessary preconditions given, the single remaining question is how closely $\tilde{f}(\mathbf{x}, \hat{\mathbf{a}})$ approximates the true model $f(\cdot)$ or, in other words, what part of the overall variance within the observations $\mathbf{y}$ is explained by $\tilde{f}(\mathbf{x}, \hat{\mathbf{a}})$. The appropriate measure for that purpose is explained in the next paragraph, before the appropriate tests to check for above assumptions are introduced below (which partly depend on that measure).

Coefficient of Determination The so-called *coefficient of determination* R^2 states what fraction of total variance within the observations of the dependent variable is explained by a particular regression model. Therefore, $0 \leqslant R^2 \leqslant 1$, that is, $R^2 = 0$ if all coefficients are zero $\mathbf{a} = (0, \ldots, 0)^T$ yielding no explicatory value at all and $R^2 = 1$ if the regression model fully explains the true model $\tilde{f}(\mathbf{x}, \mathbf{a}) = f(\mathbf{x}, \mathbf{a})$ (see e.g., Backhaus et al., 2006). Correspondingly, the coefficient of determination is computed as

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \tag{4.12}$$

with

$$SSR = \sum_{t=1}^{T} (\hat{y}_t - \bar{y})^2 \qquad \text{(explained variance)} \tag{4.13}$$

$$SSE = \sum_{t=1}^{T} (y_t - \hat{y})^2 \qquad \text{(unexplained variance)} \tag{4.14}$$

$$SST = \sum_{t=1}^{T} (y_t - \bar{y})^2 \qquad \text{(total variance)} \tag{4.15}$$

whereby $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$, $\mathbf{y} = (y_1, \ldots, y_T)^T$ and $\hat{\mathbf{y}} = \mathbf{\Phi}\hat{\mathbf{a}}$ as introduced above in (4.11).

Following (4.12), the coefficient of determination may be improved simply by adding input variables to the model, even if they just "explain" the variance in the observation as a matter of chance or of the peculiarities of the specific observation sample in use (so-called *over fitting*). To penalize such behavior, the *adjusted coefficient of determination* should be used in regression models with more than one independent variable:

$$R^2_{adj} = R^2 - (1 - R^2)\frac{M}{T - M - 1} \ . \tag{4.16}$$

The next paragraph now gives the appropriate tests later employed in Chapter 6 to confirm that a model indeed adheres to the assumptions (A1) – (A5) outlined above.

Testing for Assumptions The following list gives an overview of statistical testing methods that are necessary to confirm the validity of a linear regression model. The individual methods are not explained in their very detail, as the reader is expected to be familiar with basic statistical tests (like F-test or t-test). The application of the tests to the data sample considered in this work is discussed in Chapter 6, while for a detailed introduction to the methods being used, the reader should refer to Bamberg et al. (2009), Draper and Smith (1981) or Neter et al. (1983).

(T1) The F-test is used to test the dependent variable's adherence to the assumed functional form of the regression model (i.e., its linearity in **a**), which is the same as testing the significance of R^2:

Hypothesis
$H_0 : a_0 = a_1 = \cdots = a_m = 0.$

Test statistic
$F_T = \frac{\sum_{t=1}^{T} (\hat{y}_t - \bar{y})^2 / M}{\sum_{t=1}^{T} (y_t - \hat{y}_t)^2 / (T-M-1)}.$

Critical values
The null hypothesis H_0 is rejected if the test statistic F_T is greater than the corresponding critical value of the F-distribution, with its degrees of freedom depending on the amount of used basis functions M and the number of available observations $F_T > F_{M-1;T-1;\alpha}$.

After the F-test has rejected the hypothesis that all regression coefficients are zero, and thus that some relation in the postulated form exists, the t-test is used to validate that each particular independent variable $\phi_m(\mathbf{x})$ has a significant influence on the result, i.e., that the corresponding coefficient a_m is significantly different from zero:

Hypothesis
$H_0 : a_m = 0.$

Test statistic
$t_{a_m} = \frac{a_m}{s_{a_m}}$
with
$s_{a_m} = \sqrt{\frac{\frac{1}{T}\sum_{t=1}^{T} e_t^2}{\sum_{t=1}^{T} (x_t - \bar{x}_m)^2}}$ (standard error of a_m).

Critical values
The null hypothesis H_0 is rejected if the absolute value of the test statistic is greater than the corresponding critical value of the students t-distribution with its degrees of freedom depending on the number of observations $|t_{a_m}| > t_{T-2;1-\frac{\alpha}{2}}$.

(T2) A correctly specified regression model (i.e., one that includes all variables with systematic influence) should yield an expectancy of zero for the residuals, as the true values are scattered randomly around the regression line. Conversely, if the model was missing an input factor, the true values would be systematically deviated from the regression line and the expectancy of the residuals would not be zero.

The above-described OLS estimation scheme automatically forces the residuals to have an expectancy of zero, thereby injecting the error into the intercept coefficient. The only indication of such an effect would be that the overall distribution of the residuals would not adhere to the normal distribution assumption from (A2). The appropriate test for this behavior is the Kolmogorov-Smirnov-test, which is described in detail later, in Section 5.2.1.

(T3) If the residuals of two independent samples of the observation show the same variance (as postulated in (A3)), i.e., the residuals exhibit perfect *homoscedasticity*, then the variance of the residuals based on any particular sample $\mathbf{X}_i \subseteq \mathbf{X}$ is equal to the overall regression variance $\sigma_i^2 = \sigma^2$. This hypothesis can be tested using the White-test:

Hypothesis
$H_0 : \sigma_i^2 = \sigma^2 \quad \forall\ \mathbf{X}_i \subseteq \mathbf{X}.$

Test statistic
To calculate the test statistic, a so-called *auxiliary regression* is calculated that uses the squared residuals of the main regression as the dependent variable and the original regressors, their square and their cross product as independent variables:

$$\begin{aligned} e^2 = \alpha_0 &+ \textstyle\sum_{m=1}^{M} \alpha_m \cdot \phi_m(\mathbf{x}) \\ &+ \textstyle\sum_{m=1}^{M} \alpha_{M+m} \cdot \phi_m(\mathbf{x})^2 \\ &+ \textstyle\sum_{m=1}^{M} \sum_{n=m+1}^{M} \alpha_{2M+m+n} \cdot [\phi_m(\mathbf{x})\phi_n(\mathbf{x})]. \end{aligned}$$

The test statistic is computed using the resulting coefficient of determination multiplied by the number of observations considered: $T \cdot R^2$.

Critical values
Under the null hypothesis, the test statistic is asymptotically Chi-squared distributed with g degrees of freedom according to the amount of variables considered in the auxiliary regression above. Thus, H_0 is rejected (i.e., the residuals exhibit heteroscedasticity) with significance level α if $T \cdot R^2 > \chi^2_{g,\alpha}$.

(T4) The assumption of absence of autocorrelation (A4) between close residuals can only be tested through consideration of the order of observations or residuals in the data, i.e., whether the residuals follow a first-order autoregressive process, i.e., $\exists\ \rho \neq 0$, so that $e_t = \rho e_{t-1} + \epsilon$ with $\epsilon \sim \mathcal{N}(0, \sigma^2)$. The Durbin-Watson-test checks for the strength and direction of such a relationship:

Hypothesis
$H_0 : \rho = 0.$

Test statistic
$D = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2}$ for a single series of length T, or
$D = \frac{\sum_{j=1}^{J} \sum_{t=2}^{T} (e_{jt} - e_{j(t-1)})^2}{\sum_{j=1}^{J} \sum_{t=1}^{T} e_{jt}^2}$ for J multiple series of length T.

Critical values
By construction, the level of the resulting test statistic is limited to $0 \leqslant D \leqslant 4$, with $D = 2$ indicating that $\rho = 0$ and hence that the residuals are independent. However, there exists no fixed procedure for obtaining definite results if $D \neq 2$. For smaller sample sizes, Durbin and Watson (1950, 1951) have suggested areas $D_L \leqslant D \leqslant D_U$ where the test is inconclusive and a clear rejection of H_0 only if $D < D_L$. For larger samples, Garson (2008) gives a guiding rule whereby D should be between 1.5 and 2.5 to securely indicate the independence (i.e., absence of autocorrelation) of observations.

(T5) The described regression model, together with its OLS estimator, is only mathematically feasible if the used input variables (i.e., basis functions) $\boldsymbol{\phi}(\mathbf{x})$ are not fully linearly dependent on each other (A5). A suitable test statistic used to detect multicollinearity is the variance inflation factor (VIF):

Hypothesis
H_0: $\phi_m(\mathbf{x}) = f\big(\phi_1(\mathbf{x}), \ldots, \phi_{m-1}(\mathbf{x}), \phi_{m+1}(\mathbf{x}), \ldots, \phi_M(\mathbf{x}), \mathbf{a}\big)$.

Test statistic
$\text{VIF}_m = \frac{1}{1-R_m^2}$,
where R_m^2 is the coefficient of determination for the linear regression specified in H_0.

Critical values
If there exists perfect multicollinearity for a particular input variable $\phi_m(\mathbf{x})$, then $R_m^2 = 1$ for which VIF_m is not defined. However, perfect collinearity is only present when variables have been added multiple times. In other cases, the null hypothesis can confidentially be rejected if $\text{VIF}_m \leqslant 4$ (see Garson, 2008).

After a regression model has been validated against the necessary assumptions (A1) – (A5) and evaluated for explanatory content, it may yield structural insight into the driving factors of the dependent variable that itself can be useful. Additionally, the model may be used to predict new levels of y depending on various input levels of $\mathbf{x}$ or $\mathbf{\Phi}(\mathbf{x})$, respectively.

This section has presented the "classic" linear regression model based on the assumption that fix and deterministic model parameters $\mathbf{a}$ exist for the true model $f(\mathbf{x}, \mathbf{a})$, which can then be estimated using the described OLS method.

Next, Section 4.2 introduces Bayesian statistics and with it a different view on probability theory in Section 4.2.1. Subsequently, Section 4.2.2 discusses Bayesian inference as a basis for Bayesian linear regression in Section 4.3, which loosens the deterministic assumption regarding $\mathbf{a}$.

4.2 Bayesian Statistics

The first section of this chapter has been based on the classical frequentist interpretation of probability (see Section 4.1). The current section now gives an introduction to the Bayesian interpretation, in which probabilities provide a quantification of uncertainty or degree of belief.

Probability in a general sense has been a major subject of study for hundreds of years. However, statistics in particular is a relatively young field. The linear regression models that have been introduced above were first developed in Galton (1889) and later amended with measures for correlation and goodness-of-fit by Pearson (1896). Interestingly, Bayesian methods date back to the work of reverend Thomas Bayes (1763), which was later independently rediscovered by Laplace (1774).

The *frequentist interpretation* of probability expects the driving parameters θ of stochastic events to be deterministic; therefore, it offers a set of tools for one to use to estimate or discover these parameters. Given a sufficiently high number of observations, these estimates $\hat{\theta}$ are thought to converge against the true (i.e., deterministic) value of the parameter θ.

This frequentist interpretation results in two major difficulties (see e.g., Bishop, 2006 or Bradley P. Carlin and Thomas A. Louis, 2000):

1. The parameter estimate may be heavily skewed and extreme given only a small number of observations (see also Example 4.1 below).

2. There exists no useful conclusion about how resilient or certain the estimate is after the data have been recorded (see discussion below).

Example 4.1 *Suppose a fair coin is tossed twelve times and the outcome is evaluated and recorded, with the coin landing head-side up three times. Under a classical maximum likelihood estimate (see e.g., Lehn and Wegmann, 2006, Sec. 3.3), the frequentist probability of such an outcome would be calculated as* $P(X = head) = 3/12 = 0.25$*. A more extreme estimate would follow if the coin landed head-side up three out of three times:* $P(X = head) = 3/3 = 1$*. Obviously, both estimates are heavily skewed compared to the true parameter value of 0.5.*

In the above example, the results are imprecise because of the low number of observations, which apparently do not lead to the assumed convergence of the parameter estimate to the true value $\hat{\theta} \to \theta$. However, the frequentist approach does not give any indication on the degree of belief that the analyst can have in the results after the observations have been made.

The only possible guidance could be derived from a *confidence interval*, which has the property that on average over repeated applications, the interval will fail to capture the true parameter value θ only with a certain probability α

$$P\big(l(\mathbf{y}) \leqslant \theta \leqslant u(\mathbf{y})\big) = 1 - \alpha, \tag{4.17}$$

with $l(\mathbf{y})$ being the lower and $u(\mathbf{y})$ the upper bound of the confidence interval, depending on the observations $\mathbf{y} = (y_1, \ldots, y_t)^T$.

Now, as the true parameter θ is supposed to be deterministic, the notion of a confidence interval in any *single* data-analytic setting is somewhat difficult to understand: *After* collecting the data and computing $l(\mathbf{y})$ and $u(\mathbf{y})$, the interval defined in (4.17) either covers the true θ or it does not, but in Example 4.1 above, no statement can be made concerning how certain it is that 0.25 is the true probability of the observed coin's landing head-side up (for a more detailed discussion see, e.g., Bradley P. Carlin and Thomas A. Louis, 2000).

The next section illustrates how Bayesian probabilities can help here.

4.2.1 Bayesian Probabilities

The above discussion illustrates why the "Bayesian methodology has been increasing in popularity for the past half-century (...). Firstly, Bayesian methods take an axiomatic view of uncertainty allowing the user to make coherent inference. Secondly, Bayesian modeling is particularly well suited to incorporating prior information, which is often available" (Denison et al., 2002, preface).

While the frequentist view takes the underlying parameters of a stochastic system as deterministic and hence considers the available data or observations as a stochastic realization, the Bayesian interpretation of probability views the available information (i.e., the dataset or observations) as deterministic and regards the driving parameters themselves as stochastic. As described below, this approach allows for a quantification of the remaining degree of uncertainty about a model's parameters *after* observations have been made (see Denison et al., 2002, pp. 12). By construction, following Bayes' theorem, Bayesian interpretation of probability also allows for the inclusion of prior knowledge about the model's parameters.

Bayesian statistics is based on *Bayes' theorem*, which relates the conditional and marginal probabilities of two random events based on the conventional definition of conditional probability[3]

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}, \tag{4.18}$$

where A and B are stochastic events whose outcomes are interdependent in some fashion.

Now, if the underlying parameter θ of a stochastic system is *not* deterministic, it is feasible to capture the assumptions about its stochastic behavior in a probability distribution $p(\theta)$. Henceforth, a probability corresponding to the actual realization of observed data y based on the assumption for $p(\theta)$ can be calculated as a function of θ with $p(y|\theta)$. Bayes' theorem then takes the form

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}, \tag{4.19}$$

allowing for the evaluation of the uncertainty in θ *after* the observation of y in the form of the probability $p(\theta|y)$.

In (4.19), $p(\theta)$ captures the so-called *prior* belief or distribution of θ *before* any observations are made, while $p(\theta|y)$ is supposed to capture the *posterior* belief *after* the observation of y (i.e., conditional on the observation). The quantity $p(y|\theta)$ is evaluated for a particular observation y and can be viewed

[3] Given two interrelated events A and B, the conditional probability of A given that B already occurred is defined by $P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B \cap A)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$.

as a function of θ, in which case it is called the *likelihood function*.[4] It expresses how probable the observation y is for various levels of the parameter vector θ. Using these definitions, Bayes' theorem can be stated as

$$\textit{posterior} \propto \textit{likelihood} \times \textit{prior} \tag{4.20}$$

or be paraphrased as follows: 'The posterior distribution of the parameter is proportional to the likelihood of the observed data times the prior distribution', whereby all elements are considered to be a function of θ (see, e.g., Bishop, 2006; Bradley P. Carlin and Thomas A. Louis, 2000 or Lee, 1989, each in Chap. 1).

The following example following Bishop (2006, pp. 12) illustrates the usage of Bayesian probabilities along (4.18) – (4.20) in a simple setting.

Example 4.2 *Suppose that in an experiment, a single fruit is drawn from either of two bowls (black and white), which each contain an unequal number of apples and pears (see Figure 4.4). The bowls are selected at random, so that the probability of selecting either one is supposedly* $P(black) = P(white) = 0.5$. *Figure 4.4 illustrates that the probability of selecting an apple differs depending on which bowl has been selected* $P(apple|black) = 3/4 = 0.75$ *and* $P(apple|white) = 2/8 = 0.25$.[5]

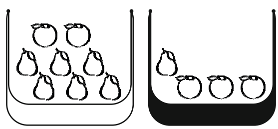


Figure 4.4: Fruit example

If we had been asked about the probability that the white bowl had been chosen, before knowing which fruit had been selected, then the most complete information available would have been $P(white) = 0.5$ *(the prior probability). However, if we know that the selected fruit is an apple, the posterior probability* $P(white|\ apple)$ *can be updated using Bayes' theorem to*

$$\begin{aligned} P(white|apple) &= \frac{P(apple|white)P(white)}{P(apple)} \\ &= \frac{P(apple|white)P(white)}{P(apple|black)P(black) + P(apple|white)P(white)} \\ &= \frac{0.25 \cdot 0.5}{0.75 \cdot 0.5 + 0.25 \cdot 0.5} = \frac{0.125}{0.375 + 0.125} = \frac{0.125}{0.5} = \mathbf{0.25}. \end{aligned}$$

The result is in line with the intuition that it is less likely that the white bowl had been selected, as the proportion of apples is much lower in it.

[4] The likelihood function is *not* a probability distribution, and its integral therefore does not necessarily equal one.

[5] Note that the two conditional probabilities for selecting an apple add up to one only incidentally.

Example 4.2 above illustrates the particular strength of Bayesian models: parameters are not thought to be fixed and deterministic but instead are believed to posses a distribution that is updated and refined based on available deterministic information. Spedding and Chan (2000, p. 332) summarize the advantages as follows:

> *"Traditional forecasting approaches are based on characterizing the structure of historical time series and then predict future events based on that structure. Obviously, the structure of the time series may change in a volatile business environment. The parameters of the time series model would then need to be re-estimated based on the new structure of the time series. Bayesian forecasting, however, is based on the principle that routine forecasts can be updated by subjective intervention as external information becomes available. In this context* Bayesian learning *is particularly relevant to applications in a dynamically changing environment."*

Section 4.2.2 now gives a short introduction to how actual inferences can be drawn along (4.20) using probability distributions as prior and posterior.

4.2.2 Bayesian Inference

While the basic idea of Bayesian inference is rather intuitive and simple (see Example 4.2 above), it can be drawn from (4.19) that the conclusion can become mathematically challenging when based on arbitrary distributions.

Firstly, the form of the likelihood function in Bayes' theorem is typically predetermined by the data or observations themselves. For example, in an experiment involving recurring rolls of a die, the resulting probability distribution for the likelihood function would be a binomial distribution. Now, depending on the particular chosen prior distribution for the underlying parameters, the derivation of the resulting posterior distribution might be mathematically difficult, if not numerically impossible, depending on the form of $p(y) = \int p(y|\theta)p(\theta)\ d\theta$ in the denominator of (4.19).

For simplicity's sake in Bayesian learning, the employed prior distributions are often selected to be well-behaved in the sense that the resulting posterior distribution can be derived numerically and itself will also be well-behaved. For continuous or recurring updating of a particular prior, the necessary prerequisite is raised further, as here the posterior distribution ideally should be of the same probabilistic type as the prior distribution. In such particular cases, the prior is said to be a *conjugate prior* (see e.g., Pilz, 1991, Chap. 4.1).

In the following, a normally distributed likelihood function is used as an exemplary case to illustrate Bayesian updating or learning in the sense of (4.19). This is a valid assumption under various observation settings, especially under the central limit theorem (see e.g., Lee, 1989, Sec. 1.3).

Assuming a normally distributed random variable y whose variance σ^2 is known but whose mean μ will be estimated based on a set of T observations $\mathbf{y} = \{y_1, \ldots, y_T\}$, the likelihood function – that is, the probability of the observations given μ – is then given by

$$p(\mathbf{y}|\mu) = \prod_{t=1}^{T} p(y_t|\mu) = \frac{1}{(2\pi\sigma^2)^{T/2}} \exp\left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^{T} (y_t - \mu)^2 \right\}. \tag{4.21}$$

As noted before, the likelihood function $p(\mathbf{y}|\mu)$ is *not* a probability function by itself. However, in the above example, it can be derived easily from the normal distribution.

Assuming that the unknown mean μ is itself normally distributed, with the parameters μ_0 and σ_0^2 containing the available prior knowledge, then

$$p(\mu) \sim \mathcal{N}(\mu|\mu_0, \sigma_0^2). \tag{4.22}$$

Following (4.20) above, the posterior distribution is defined by (see Bishop, 2006, p. 97)

$$p(\mu|\mathbf{y}) \propto p(\mathbf{y}|\mu)p(\mu). \tag{4.23}$$

The resulting posterior apparently is a product of two exponentials of quadratic functions of μ and hence will also be normal

$$p(\mu|\mathbf{y}) \sim \mathcal{N}(\mu|\mu_T, \sigma_T^2), \tag{4.24}$$

wherefore Bishop (2006, pp. 97) shows that the moments are defined by

$$\mu_T = \frac{\sigma^2}{T\sigma_0^2 + \sigma^2}\mu_0 + \frac{T\sigma_0^2}{T\sigma_0^2 + \sigma^2}\bar{\mu} \qquad \text{and} \tag{4.25}$$

$$\sigma_T^2 = \left(\frac{1}{\sigma_0^2} + \frac{T}{\sigma^2} \right)^{-1}, \tag{4.26}$$

with $\bar{\mu} = \frac{1}{T}\sum_{t=1}^{T} y_t$ being the maximum likelihood estimate for μ, given by the sample mean. For a complete proof of (4.24) – (4.26) see, e.g., Bishop (2006, pp. 98) or Lee (1989, Sec. 2.2).

Taking a closer look at (4.25) and (4.26), the moments of the posterior distribution exhibit some interesting properties: each is some form of compromise between the information contained in the prior and the new information contained in the observations. For $T = 0$ (no additional observations), both parameters equal their original value in the prior $\mu_T = \mu_0$ and $\sigma_T^2 = \sigma_0^2$. In contrast, for the number of observations approaching infinity $T \to \infty$, the mean equals the maximum likelihood solution, while the variance tends to zero $\mu_T = \bar{\mu}$ and $\sigma_T^2 = 0$ (see also Bishop, 2006, p. 152).

Next, Section 4.3 introduces a particular Bayesian learning scheme specifically exhibiting the advantages named by Spedding and Chan (2000) as cited in Section 4.2.1, because the introduced Bayesian regression models are capable of updating their parameters continuously as new information becomes available.

4.3 Bayesian Linear Regression

This section extends the Bayesian inference regarding single parameters for a selected distribution (see Section 4.2.2) to the case where a full parameter set of a linear basis function model has to be inferred from observation data.

Bayesian linear regression is a form of so-called supervised learning from the field of machine learning, as the model itself has to be constructed by the "supervisor", while the parameters can then be "learned" automatically from the data themselves (see Bishop, 2006, p. 137).

4.3.1 Parameter Distribution

Following (4.3) in Section 4.1, a linear basis function model $\tilde{f}(\boldsymbol{\phi}(\mathbf{x}), \mathbf{a})$ with independent basis functions $\boldsymbol{\phi}(\mathbf{x})$ and corresponding (but unknown) parameter values $\mathbf{a}$ is assumed. Based on T observations, the matrix $\boldsymbol{\Phi} = (\boldsymbol{\phi}(\mathbf{x})_1, \ldots, \boldsymbol{\phi}(\mathbf{x})_T)^T$ contains the input basis functions and the vector $\mathbf{y} = (y_1, \ldots, y_T)^T$ the corresponding dependent values.

Unlike in the frequentist regression model, the parameters $\mathbf{a}$ here are *not* expected to be deterministic but are instead assumed to follow a multivariate normal distribution with a corresponding set of moments $\mathbf{m}, \mathbf{S}$, where $\mathbf{m} = (m_1, \ldots, m_M)^T$ is a $M \times 1$-dimensional vector of means corresponding to the parameters in $\mathbf{a}$ and $\mathbf{M}$ is the $M \times M$-dimensional covariance matrix.

The choice of normally distributed parameter values $\mathbf{a}$ is typical for linear regression models, especially under the central limit theorem for larger observation sets (see Denison et al., 2002, p. 13).

Under the above assumptions, the likelihood function $p(\mathbf{y}|\mathbf{a})$ along (4.21) is the exponential of a quadratic function of $\mathbf{a}$ and hence, the corresponding conjugate prior distribution of $\mathbf{a}$ is multivariate normally distributed and given by

$$p(\mathbf{a}) = \mathcal{N}(\mathbf{m}_0, \mathbf{S}_0), \tag{4.27}$$

with initial mean $\mathbf{m}_0$ and covariance $\mathbf{S}_0$. On this basis, the posterior distribution of $\mathbf{a}$ is also normally distributed, with the form

$$p(\mathbf{a}|\mathbf{X}, \mathbf{y}, \beta) = \mathcal{N}(\mathbf{m}_T, \mathbf{S}_T), \tag{4.28}$$

where

$$\mathbf{m}_T = \mathbf{S}_T(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\mathbf{\Phi}^T\mathbf{y}) \qquad \text{and} \tag{4.29}$$

$$\mathbf{S}_T^{-1} = \mathbf{S}_0^{-1} + \beta\mathbf{\Phi}^T\mathbf{\Phi}, \tag{4.30}$$

with β being the precision (inverse variance) of the noise or uncertainty in the observed data $\mathbf{X}, \mathbf{y}$.

The mathematical proof and derivation are omitted here but can be found in Lee (1989, Sec. 6.3) amongst others. Instead the properties and behavior of the obtained posterior distribution will be examined somewhat closer. Taking an infinitely broad prior $\mathbf{S}_0 = \alpha^{-1}\mathbf{I}$ with $\alpha \to 0$ as a starting point, the mean of the resulting posterior distribution (4.29) then matches the OLS estimate from (4.9). On the contrary, if no new observations arrive ($T = 0$), the posterior equals the prior distribution. In a typical update, the posterior distribution can be interpreted as a weighted average of the OLS and the prior moments.

It is intuitive that the estimation process can be repeated multiple times, whenever new data arrive, using $\mathbf{m}_T, \mathbf{S}_T$ as prior distribution in place of $\mathbf{m}_0, \mathbf{S}_0$ (see, e.g., Bishop, 2006, Sec. 3.3). In a learning setting, this process is referred to as "training" the regression model because observations are used in an incremental fashion to improve the accuracy (i.e., reduce the variance) of $p(\mathbf{a})$.

Finally, an example based on Bishop (2006, pp. 154) illustrates the effects of recurring updates or learning on a parameter distribution in Bayesian linear regression.

Example 4.3 *Consider a linear model with the form* $y = f\left(\binom{1}{x}, \binom{a_0}{a_1}\right) = a_0 + a_1 x = \mathbf{x}^T\mathbf{a}$*, where the relationship between* y *and* x *is apparently assumed to be linear in the* 2×1*-dimensional parameter vector* $\mathbf{a}$*.*

To train the corresponding model, test data are generated using parameter values $\binom{a_0}{a_1} = \binom{-0.3}{0.5}$ *on values* x_t *drawn from the universal distribution* $\mathcal{U}(-1, 1)$ *with additive normally distributed noise* $\mathcal{N}(0, 0.02)$*.*

Now, using Bayesian regression estimators along (4.28) – (4.30) *the original parameters will be learned from that data. The bivariate prior distribution for the parameter vector* $\mathbf{a}$ *is initially set to be*

$$\mathbf{a}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{S}_0) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.5 & 0 \\ 0 & 0.5 \end{pmatrix}\right),$$

meaning that, based on prior knowledge, both parameters are expected to be zero, but with a low degree of certainty, as indicated by the high variance on the diagonal of $\mathbf{S}_0$*. Additionally, the uncertainty of* a_0 *taking on a particular value is independent of the expected value of* a_1*, as the covariances are set to zero in* $\mathbf{S}_0$*. The above assumptions regarding the prior basically allow a variety of relationships between* y *and* x *and signify that there is no sound prior information available. Both are illustrated in Figure 4.5.*

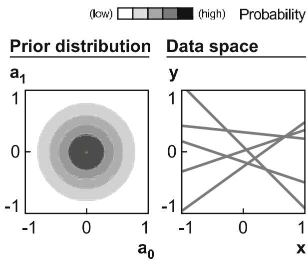


Figure 4.5: Illustration of Bayesian regression learning – step 1
Source: Own illustration of Bishop (2006, p. 155)

Now, the first observation $x = 0.9; y = 0.0$ *is made as indicated in Figure 4.6 (the dotted line shows the real relationship between* x *and* y*; because of normal noise, the observation is slightly off the line). The associated likelihood function illustrates that the probability of yielding the particular data point is highest given certain parameter combinations along a diagonal line passing close to the real values. Using the prior distribution from Figure 4.5 and this likelihood function, the posterior in Figure 4.6 is generated along*

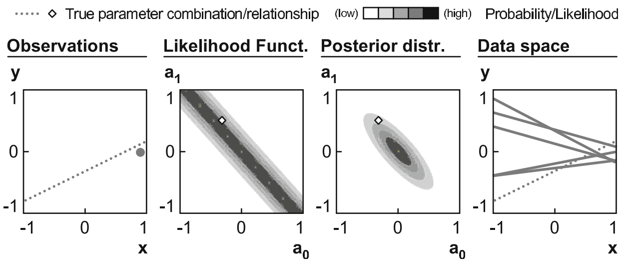


Figure 4.6: Illustration of Bayesian regression learning – step 2
Source: Own illustration of Bishop (2006, p. 155)

(4.28) – (4.30), *which is essentially the prior from Figure 4.5 squeezed along the likelihood function*

$$\mathbf{a}_1 \sim \mathcal{N}(\mathbf{m}_1, \mathbf{S}_1) = \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.10 & -0.09 \\ -0.09 & 0.12 \end{pmatrix}\right).$$

The improved assumptions regarding the parameter distribution restrict the possible relationships between x and y within the data space, as a_1 and a_2 are still both expected to be zero, but the partial distribution of each is now dependent on that of the other.

After a second observation $x = -0.7; y = -0.6$ is made in Figure 4.7, the update procedure described above (using the posterior from Figure 4.6 as prior) considerably restricts the parameter distribution through the application of the new likelihood function. Under the resulting posterior distribution, the choice of relationships is already fairly restricted and close to the original values. In addition, the posterior distribution of **a** *has contracted to close to the true values (see Figure 4.7)*

$$\mathbf{a}_2 \sim \mathcal{N}(\mathbf{m}_2, \mathbf{S}_2) = \mathcal{N}\left(\begin{pmatrix} -0.30 \\ 0.32 \end{pmatrix}, \begin{pmatrix} 0.02 & 0.00 \\ 0.00 & 0.03 \end{pmatrix}\right).$$

Intuitively, it seems that the quality of the parameter distribution will improve further (the means will improve and the variances will decline) as one adds more observational data.

The above example illustrates the Bayesian learning mechanism for parameter distributions of linear regression models along (4.28). However, in practical applications of regression modeling, the interest is eventually on the

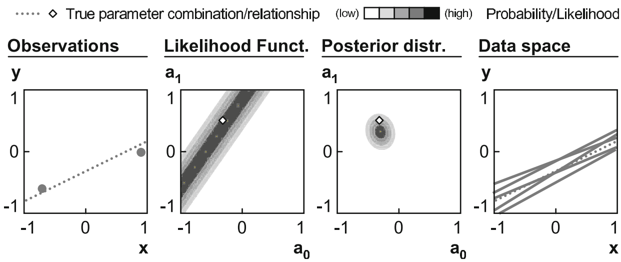


Figure 4.7: Illustration of Bayesian regression learning – step 3
Source: Own illustration of Bishop (2006, p. 155)

functional relationship between the independent and dependent variables or the prediction of y for new input variables $\mathbf{x}$. This requires an evaluation of the so-called predictive distribution, which will be introduced in the following section.

4.3.2 Predictive Distribution

Bayesian regression modeling, as shown in Section 4.3.1, assumes probability distributions for two elements that constitute the model (see discussion above): first, future observations themselves are expected to be stochastic, and second, the driving parameters behind the model also follow a probability distribution that depends on recorded past observations. However, of ultimate interest in practice is the prediction of new dependent variable values based on a pre-selected set of independent input variables rather than solely knowing the models' underlying stochastic drivers.

In Bayesian linear regression, the prediction of a new value $\breve{y}$ depends on the convolution of two probability distributions: the first one gives the conditional probability of possible outcomes $\breve{y}$ under expected parameters $\mathbf{a}$, while the second one gives the posterior distribution of these parameters given past observations $\mathbf{X}$ and $\mathbf{y}$. Correspondingly, the *predictive distribution* for $\breve{y}$ is defined by

$$p(\breve{y}|\mathbf{X}, \mathbf{y}, \mathbf{m}_0, \mathbf{S}_0, \beta) = \int p(\breve{y}|\mathbf{X}, \mathbf{a}, \beta) p(\mathbf{a}|\mathbf{X}, \mathbf{y}, \mathbf{m}_0, \mathbf{S}_0, \beta) \mathrm{d}\mathbf{a}, \tag{4.31}$$

where $\mathbf{X}$ and $\mathbf{y}$ are (dependent and independent) past observations, $\mathbf{m_0}, \mathbf{S_0}$ are the moments of the prior distribution of $\mathbf{a}$ and β is the precision or noise respectively of the data (see, e.g., Bishop, 2006).

The conditional distribution of the target variable $p(\breve{y}|\mathbf{X}, \mathbf{a}, \beta)$ is given by (4.4) with $\sigma^2 = \beta^{-1}$, and the posterior distribution of the parameters $p(\mathbf{a}|\mathbf{X}, \mathbf{y}, \mathbf{m}_0, \mathbf{S}_0, \beta)$ is given by (4.28) – (4.30) in the preceding Sections 4.1 and 4.3.1. Hence, (4.31) involves the convolution of two normal distributions that – for a particular set of new independent variables $\breve{\mathbf{x}}$ – can be proven to take the form (see e.g., Bishop, 2006, Sec. 8.1.4)

$$p(\breve{y}|\breve{\mathbf{x}}, \mathbf{X}, \mathbf{y}, \mathbf{m}_0, \mathbf{S}_0, \beta) = \mathcal{N}(\mathbf{m}_T^T \phi(\breve{\mathbf{x}}), \sigma_T^2(\breve{\mathbf{x}})), \tag{4.32}$$

where

$$\sigma_T^2(\breve{\mathbf{x}}) = \frac{1}{\beta} + \phi(\breve{\mathbf{x}})^T \mathbf{S}_T \phi(\breve{\mathbf{x}}) \tag{4.33}$$

with $\mathbf{m}_T$ and $\mathbf{S}_T$ according to (4.29) and (4.30), respectively, and prior $p(\mathbf{a}) = \mathcal{N}(\mathbf{m}_0, \mathbf{S}_0)$ as defined in (4.27).

Apparently, the predictive distribution (4.32) for a *new* dependent variable $\breve{y}$ according to a specific linear regression model depends on five inputs:

- $\breve{\mathbf{x}}$ The set of *new* independent variables, which are input to the basis functions $\phi_m(\cdot)$, which in turn are entered into the regression model to calculate the new dependent value or distribution of $\breve{y}$.

- $\mathbf{X}$ The set of T *historic* observations $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T)^T$ of the independent variables, which are used to train the regression model, i.e., to derive the moments $\mathbf{m}_T, \mathbf{S}_T$ of the parameter distribution for $\mathbf{a}$ along (4.28) and (4.29).

- $\mathbf{y}$ The set of T *historic* observations of the dependent variable that, together with $\mathbf{X}$, are used to train the model.

- $\mathbf{m}_0, \mathbf{S}_0$ The moments of the prior distribution of $\mathbf{a}$, assumed before the incorporation of historic observations $\mathbf{X}$ and $\mathbf{y}$.

- β The noise precision parameter of the observations reflecting the uncertainty and noise in the data – i.e., its variance.

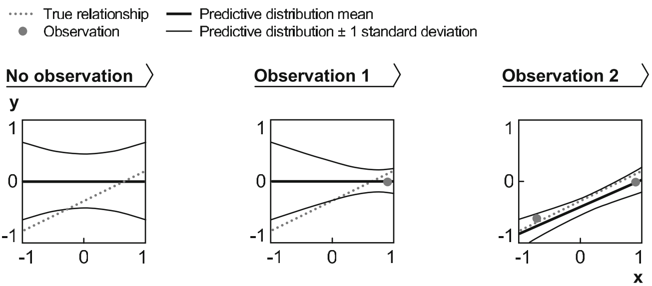
It can be shown (see Qazaz et al., 1997) that (4.32) becomes narrower as more observations are added $\sigma_{T+1}^2(\mathbf{X}) \leqslant \sigma_T^2(\mathbf{X})$, which limits the uncertainty in the predictive distribution of the dependent variable, so that eventually $\lim_{T\to\infty} \sigma_T^2(\mathbf{X}) = 0$.

Following the above, the predictive distribution in a Bayesian regression setting naturally does not return just a point estimate for a new dependent variable $\breve{y}$, but rather its distribution (here, a normal distribution) as a function of the independent input parameters $\breve{\mathbf{x}}$ (see (4.32)). Correspondingly the certainty of the prediction (i.e., the variance of $\breve{\mathbf{y}}$) may vary across the domain of $\breve{\mathbf{x}}$ and may also be dependent on the number of observations used to train the model. The following continuation of Example 4.3 from Section 4.3.1 illustrates the implications.

Example 4.3 (continued) *In Example 4.3 above, Bayesian learning regarding the parameter distribution in a simple example has been illustrated. While possible implications for the relationship between the independent and dependent variables have been depicted in "data space", actual predictions regarding the dependent variable for various levels of the independent variable have not yet been derived; this is now done in the following.*

Before any observations are made, the assumed multivariate prior distribution for the parameters in Example 4.3 peaks around $\binom{0}{0}$ *with a wide variance, basically allowing for a variety of relationships, with* $y = 0 + 0 \cdot x$ *being somewhat more certain (see Figure 4.5). Following* (4.32)*, the actual predictive distribution can be derived, whose mean is zero for all values of* x *with a wide variance – albeit slightly lower around zero (see Figure 4.8(a)).*

After the first observation is made, the parameter distribution is still cen-



(a) Starting situation before data recording

(b) After recording of 1st observation

(c) After recording of 2nd observation

Figure 4.8: Illustration of Bayesian regression learning – changes in predictive distribution

Source: Own illustration

tered around $\binom{0}{0}$, but the variance has changed, rendering certain combinations more probable. In particular, relationships close to actual observations seem more likely, as is indicated in the data space in Figure 4.6. The predictive distribution's mean has not changed (following the mean of the parameter distribution), but the variance has retracted considerably close to the last observation (see Figure 4.8(b)).

Through the second observation, the mean of the parameter distribution is finally moved considerably towards the true values, and its variance is further retracted, leaving less probable choice for the relationship in the data space (see Figure 4.7). This is reflected in the predictive distribution (see Figure 4.8(c)), whose mean already matches closely with the true relationship. Moreover, variance has been reduced considerably along the line, with predictions around zero still exhibiting slightly lower variance, driven by the remaining influence of the initial prior parameter distribution.

The introduced Bayesian regression and learning mechanism is employed in Chapter 7 to train a predefined model from Chapter 6 of iteratively arriving observations of latent demand for low-cost airline tickets. However, beforehand, critique of Bayesian statistics in general and the method's limitations in particular are discussed in the following section.

4.4 Critique and Limitations

There has been ongoing discussion in academia associated with the relative merits of the frequentist and Bayesian paradigms (see, e.g., Efron, 1986); this dispute is addressed in this section, together with apparent limitations of Bayesian statistics.

Subjective Selection of Prior Distributions A common criticism of Bayesian inference is that the prior distribution is often selected based on mathematical convenience alone (i.e., that the distribution is conjugate), rather than as a reflection of true prior beliefs – while the effect on the resulting posterior distribution is known to be significant. "Indeed, Bayesian methods based on poor choices of prior can give poor results with high confidence" (Bishop, 2006, p. 23).

The choice of prior distribution in any Bayesian setting is undoubtedly an extremely difficult element that tends to invite controversy based on its subjective nature, even if there is substantial reason (e.g., past experience) for the particular choice (see, e.g., Gelman, 2008 or Lindley, 1983). However, inappropriate choices regarding priors can lead to incorrect inferences: “It is common to use priors that assign zero probability outside of some range (e.g., a uniform prior). However, this should be done with care because if some value is assigned zero probability a priori, Bayes theorem (...) ensures that the posterior distribution also assigns it zero probability” (Punt and Hilborn, 2001, Sec. 4.1).

To prevent the above issue from arising, prior distributions should be objective in the sense that they assign a non-zero (although possibly very small) probability to all plausible values if there is not a fundamental reason to restricting their domain (see, e.g., Efron, 1986). To differentiate cases where prior information is available and it is thus reasonable to include this information in the prior distribution from those where such is not the case, Box and Tiao (1973) define uninformative and informative priors:

- **Uninformative priors** provide little information about the parameter distributions relative to the experiment. This can be achieved through the assignment of uniform priors (see, e.g., Punt and Hilborn, 1997) or at a minimum through the usage of a sufficiently broad prior that is not truncated at the extreme ends of its domain (e.g., a normal distribution).

  However, the use of uninformative priors is still controversial because the posterior results may be sensitive to the prior input in cases where the observed data do not affect the parameter distribution in the prior (i.e., the model is poorly specified or the data simply provide very little information about the model parameters).

- **Informative priors** intentionally carry prior beliefs about the parameter distribution, both in the priors’ driving moments (e.g., the mean and variance) and in possible restrictions of the parameter domain made by explicitly assuming a probability of zero at its extreme ends. As discussed above, poorly specified or restricted priors can heavily bias the posterior results, as either vast amounts of data are needed to sufficiently influence the prior distribution or the true parameter distribution can never be reached because the prior already assigns it a fixed zero probability a priori.

Punt and Hilborn (2001) advocate a rather pragmatic approach to the choice between uninformative and informative priors, preferring to select an uninformative prior and then test its sensitivity to the data rather than to “invent” an informative prior that perhaps markedly biases the results. However, they also note that well developed informative priors can reduce uncertainty considerably.

The application of a Bayesian regression in Chapter 6 will employ both approaches. First an uninformative prior will be used to train the model, while later, an informed prior (based on supposed historic experience) will be used to illustrate possible enhancements to forecast accuracy. However, the underlying definition of the informative prior still allows considerable flexibility in the posterior distribution in cases where the observed data do not follow the prior assumptions.

Data Limitations As already noted in Example 4.1, Bayesian inference may produce more credible or accurate results with limited data as the inherent uncertainty is captured in the breadth (i.e., the variance) of the posterior distribution. Bayesian posteriors or corresponding predictive distributions (see Section 4.3.2) directly represent the uncertainty of the estimate and do not provide supposed point estimates based on insufficient information.

Additionally, in the event of limited observations, the specification of the prior distribution allows for the inclusion of historical data or management expertise to substitute for missing data. The apparent downside being that this subjective approach may seriously bias the resulting posterior distributions, especially when there are few observations (see also above discussion).

However, depending on the actual objective of the analysis, challenging and altering current beliefs through Bayesian inference might be superior to drawing point conclusions with low confidence based on limited data in a frequentist interpretation (see, e.g., Efron, 1986, p. 3).

Missing single values in observation data are automatically handled by Bayesian analysis, as posterior beliefs by definition are made equal to prior beliefs in the absence of new information. "Mathematically, if y_t is missing or unreliable and construed to represent no useful information, then the current mean and variance are equal to previous mean and variance under different conditions: $p(\theta_t|\, y_t) = p(\theta_t|\, y_{t-1})$" (Spedding and Chan, 2000, p. 333).

Model Selection Bayesian statistics is widely criticized – especially by practitioners – for its supposed promises of automatic inference mechanisms up to self-acting model selection (see, e.g., Gelman, 2008), which seem theoretically feasible if a flexible definition of model parameters is employed.

In a reasonable application setting, the automated inference of model parameters is only applicable where the model structure is fixed and can be determined beforehand; otherwise, learning is more complex, and advanced methods like neural networks have to be employed (see Bishop, 2006, pp. 172). In reference to Section 4.3 this specifically means that the structure of the regression model – i.e., the precise selection of independent input variables together with their corresponding basis functions – has to be specified manually, while the parameters for the composition can be inferred following Bayesian mechanisms. Here, the major advantage of Bayesian methods is the ability to consider model uncertainty *within a single framework* (see also Punt and Hilborn, 2001, Sec. 5.4).

Particularly in a regression setting, the Bayesian method is insensitive to the threat of over-specification, as it exhibits no inherent trend toward over-fitting when employed for models whose number of parameters greatly exceeds the number of data points. "Indeed, in a Bayesian model, the effective number of parameters adapts automatically to the size of the data set" (Bishop, 2006, p. 9).

Summary Applying the above discussion on the controversy around Bayesian statistics to the topic of this work, a Bayesian approach is used in Chapter 6 because it facilitates the representation of uncertainty related to the model's parameter values, where "in contrast, most decision analyses based on maximum likelihood (or least squares) estimation involve fixing the values of parameters that may, in actuality, have an important bearing on the final outcome of the analysis and for which there is considerable uncertainty" (Punt and Hilborn, 2001, Sec. 5.1). Here, Bayesian statistics provides insight even based on limited data, while indicating that the inference is uncertain – i.e., that the variance in the driving parameters is large (see also Spedding and Chan, 2000, pp. 333).

An additional major benefit of the Bayesian approach is the ability to incorporate prior information (see Punt and Hilborn, 2001, Sec. 5.1), and even Efron notes that "subjectivism is undoubtedly useful in situations involving personal decision making, for example business and legal decisions" (Efron, 1986, p. 3). Particularly for the demand forecast in the dynamic area of LCCs, the inferences are conditional on the *actual data* collected shortly before the considered flight event (see Bradley P. Carlin and Thomas A. Louis, 2000, pp. 9), unlike the typically employed models based on historical information (see Chapter 2).

The next chapter introduces the employed dataset and gives a descriptive overview of possible drivers for the Bayesian regression model that is later developed in Chapter 6.

Chapter 5

Demand in Low Cost Markets

This chapter describes the structure and characteristics of the collected demand data, which form the basis for the development of the corresponding forecasting model in Chapter 6.

Section 5.1 first describes the data source (Section 5.1.1) and the necessary data cleansing (Section 5.1.2). Based on the clean data, long- and short-term characteristics are identified in Sections 5.2 and 5.3 respectively, before the implications of the findings for the forecasting model are finally presented in Section 5.4.

The objective of this chapter is *not* to exhaustively test and evaluate *all possible* demand drivers but rather to discover all *relevant* factors that appear to influence long- or short-term demand behavior. Where reasonable, the assumed relationship is verified using the appropriate statistical tests. The applicable tests for correct and coherent model specification are conducted thoroughly in Chapter 6.

5.1 Experimental Data Set

Typically, revenue management and dynamic pricing systems use recorded bookings as base data to reconstruct demand based on unconstraining of such "demand data" simply because true latent demand cannot be observed as it appears in intermediary channels like travel agencies or online distribution systems (see, e.g., Zeni and Lawrence, 2004). Today, in the case of airlines, real demand can be recorded as clicks in the online sales channels or similarly as fare requests through reservation systems. Figure 5.1 illustrates the conventional and novel measuring points used in this work along the purchase funnel.

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6_5,

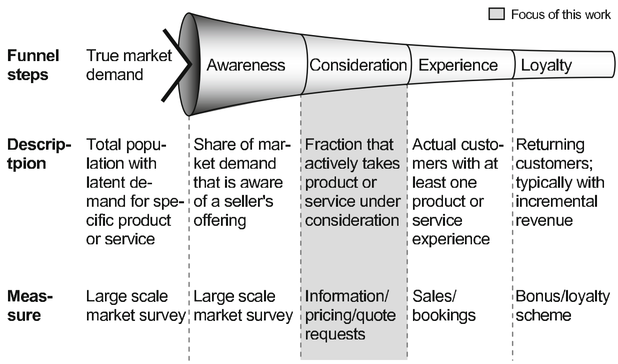


Figure 5.1: Customer purchasing funnel
Source: own design follow. Riesenbeck and Perrey (2008, Sec. 2.4)

5.1.1 Data Collection

The data used in this work were collected in cooperation with a German hybrid LCC (see Section 2). Using their proprietary Computer Reservation System (CRS), it was possible not only to log clicks on their web portal but actually to log all fare requests arriving at the centralized CRS. The data therefore include all relevant sales channels (tour operator/charter, online, call center and intermediary web portals) as illustrated in Figure 5.2. Only requests via the Global Distribution Systems (GDS) could not be recorded because the interface between the in-house CRS and the GDS is based on periodic replication, with the GDS holding a copy of all available fares and seat inventory so that fare requests are answered by the GDS directly.

The share of lost demand data can be estimated based on the recorded share of bookings or revenue made through such traditional GDS. In the considered case, this share amounts to about 13%, with no visible skew in the distribution over routes or destinations.[1]

Therefore, the data is assumed to be representative of true latent market demand as a) the missing data from GDS channels can be rescaled based on the known booking share and b) the brand awareness[2] of the carrier under

[1] This information was relayed by the controlling department and could not be verified using independent sources; nevertheless, it seems reasonable based on the recorded requests attributable to other channels.

[2] Brand awareness refers to the share of a population that exhibits an unaided awareness of a specific brand under observation when plainly asked to give names of providers in the specific service field – here, air transportation.

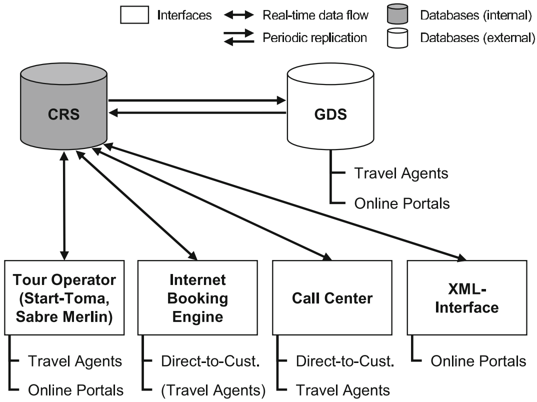


Figure 5.2: System environment for data collection
Source: own design

observation is much greater than 90% – i.e., if demand exists, it will trigger a measurable request in >90% of cases.

While the data allow an unprecedented view of latent flight demand, they still entail a few limitations, mostly due to technical constraints:

- The recorded observations are logged with a time stamp of daily granularity only, for technical reasons (i.e., performance and space constraints). Hence, no timely distribution of demand throughout the day can be derived, though here, this is not considered a considerable limitation.

- The logging mechanism is not able to distinguish repeat requests from the same source or customer. This could potentially lead to double counts if a single source inquires multiple times about the exact same event (same route, same departure date, same request date). This possibility cannot be eliminated entirely, but it is not considered of major incidence and therefore is not deemed to be of relevance here.

- The customer-facing availability display for a particular flight event automatically displays relevant information regarding neighboring events, i.e., if a fare request is made for a particular route on a specific departure date, inventory and available fares for the same route but for neighboring dates are displayed automatically. As a result, corresponding demand within that neighboring range will not be logged, as the information is already supplied along the way without explicit request.

The data analyzed in the following were collected during the period from September 10, 2007 to May 28, 2008 (request date range) for all bookable (i.e., within 365 days) future flight events (departure date range) for a popular short-haul route. The LCC under observation hosts a monthly sales event with special fares for selected routes lasting three days (Thursday to Saturday). To prevent server overload on these occasions, the logging mechanism was disabled by the IT department on Thursday mornings and reactivated on the following Monday, resulting in a total data gap of typically five days, as the collected data for Thursday and Monday had to be disregarded entirely because of the daily granularity of the time stamp (see above). Additionally the logging mechanism had to be disabled between December 20, 2007 and January 4, 2008 because of scheduled server maintenance. In total, 109,612 individual data tuples (date of request, date of departure, number of requests) equaling 2,450,637 single requests were collected.

Figure 5.3 displays the distribution of requests over time to actual departure (advance request) for the entire request date range. It is apparent that the majority of requests arrive within 120 days to actual departure, with >75% arriving within the last 60 days. This finding is comparable to the booking request development observed at easyJet by Barlow (2004, p. 15).

Based on the request distribution observed in Figure 5.3 the analysis

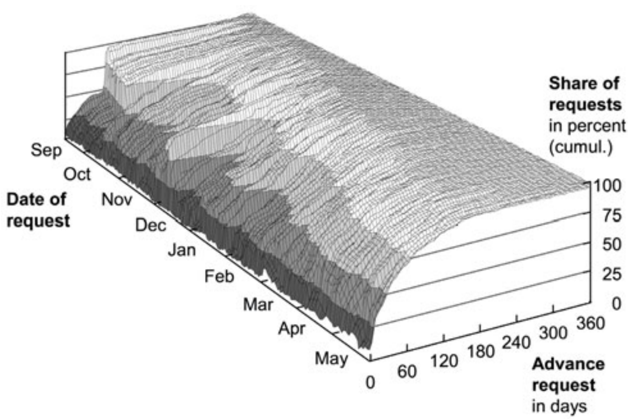


Figure 5.3: Distribution of availability requests over time to departure
Source: own design based on collected data

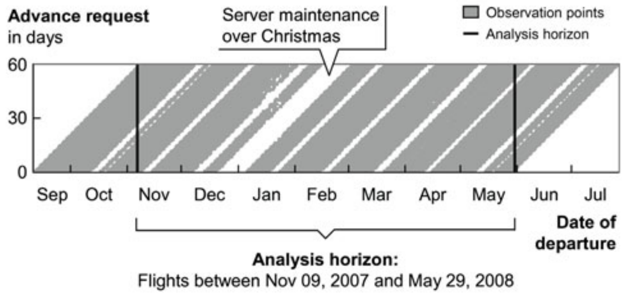


Figure 5.4: Considered analysis horizon
Source: own design based on collected data

horizon has been limited to November 09, 2007 to May 29, 2008, achieving a maximum horizon length over the departure date range, while including the relevant development time-frame over advance request days (1–60 days). The final dataset used from the collected base data is highlighted in Figure 5.4, with the distribution of missing observations also visible.

Before the actual analysis of structural characteristics in Sections 5.2 and 5.3, the collected data has to be sanitized of possible irregularities, which is described in the following section.

5.1.2 Data Cleansing

The analyzed data is collected in a fully automated manner using a logging mechanism built into the CRS of the examined air carrier. Thus, the possibility of manual data collection errors can be neglected, but nevertheless, the process is susceptible to unmeant accounting for artificial demand induced by electronic bots or web crawlers. These systems check fare availability for selected routes and flight dates systematically to feed anything from simple fare databases (e.g., www.airfarewatchdog.com) to sophisticated airfare prediction systems (e.g., www.farecast.com). However, requests from such systems do not represent true demand in the sense discussed in the above section, as data is stockpiled for future use.

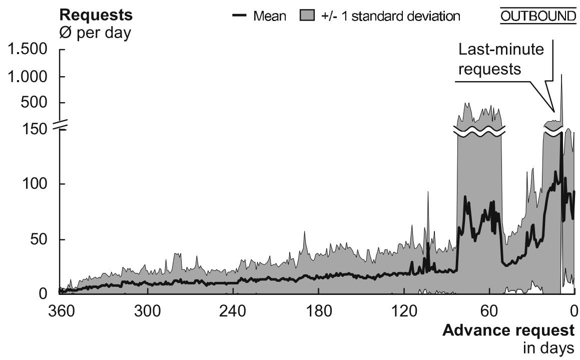


Figure 5.5: Development of requests towards departure – uncorrected data
Source: own design based on collected data

The development of average requests together with its standard deviation toward departure (see Figure 5.5) exhibits structural disruptions that give rise to the assumption of crawler effects within the data. This is especially true because the increased demand seems concentrated around 60 days before and close to departure, with peaks reaching >500 requests per day, while the average is typically well below 100 daily requests.

To reveal systematic disturbances with a heavy effect on the collected data, unusually peaked demands with >100 requests per day are plotted exclusively over the departure date range and request date range in Figure 5.6. The graphic reveals two distinct and systematic particularities:

1. Exorbitant demand for flights around the first weekend in December (outbound: November 29 – December 03, 2007; inbound: December 05 – December 09, 2007) that are not reflected in recorded bookings (i.e., realized demand) and that also do not correspond with special events in the destination city/region. The peaks occur only based on requests in September and October 2007, start with a few thousand requests per day and drop abruptly to exactly 1,080 daily requests on October 10, 2007 (outbound and inbound).

2. The same amount of total daily requests (1,080) is observed for a moving 14-day trip length throughout the entire data set. The particular queried flight dates typically remain fixed for two weeks, then are increased by a day for the following two weeks, and remain stable again.

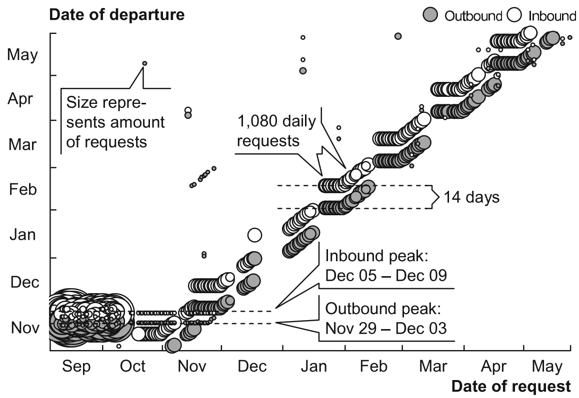


Figure 5.6: Distribution of unusually peaked requests (>100 per day)
Source: own design based on collected data

Both effects exhibit the same set of source channel flags: all requests are made for flexible fares in conjunction with a system-specific best price flag through offline channels. This combination is rare in the remaining data, as LCC tickets are typically requested as base fares through online channels that do not carry a best-price flag, as base fares are always quoted as the best price.

Therefore, excessive observations (> 300) with the above-mentioned channel characteristics are considered crawler activity and are excluded from the dataset to prevent such effects from overlaying the true demand patterns to be detected and forecasted in Chapter 6.

Next, Section 5.2 examines overarching long-term characteristics of the crawler-corrected data that could form the basis for demand forecasts. On that basis, Section 5.3 takes a closer look at short-term characteristics with possible further influence on demand.

5.2 Overarching Long-term Characteristics

As a first step in the analysis of the demand structure within the collected data, this section examines long-term characteristics that are either persistent over the entire data horizon or can only be uncovered through long-term changes in its structure.

The dataset (see Section 5.1) consists of 382 basically individual and independent time series, 191 for outbound and 191 for inbound flight dates. These are indexed with j in the following, whereby J denotes the set of all flight dates. Each time series is composed of 60 single request observations for the period from 60 days to 1 day before departure (i.e., decremental), indexed by t (also decremental).

The key questions that are eventually answered across these series within this chapter are then:

- What is the collective development structure over all series, i.e., in an aggregated view? (Section 5.2.1)
- Does the structure of the series exhibit long-term seasonalities or long-term trends? (Section 5.2.2)
- To what degree is the timely development of separate but adjacent flights possibly related? (Section 5.2.3)

5.2.1 Log-linear Demand Structure

In a first step, the typical development of flight demand time series is examined. The development of mean requests towards departure over all flight dates under observation has been plotted in Figure 5.7.

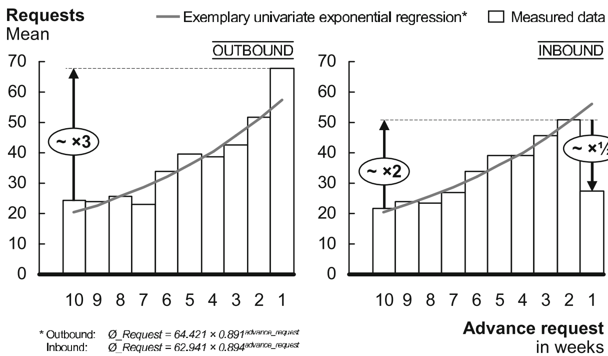


Figure 5.7: Development of average requests towards departure – corrected data

Source: own design based on collected data

Based on the depicted weekly aggregate, it is apparent that demand intensity grows exponentially towards the actual date of departure, with outbound demand roughly tripling during the last 10 weeks. The development on the inbound side is slower, with demand roughly doubling until the second-to-last week and demand dropping thereafter by $\approx 50\%$ in the last week. The slower development on the inbound side could be attributed to customers' first only checking outbound flights and only considering returns after the selection of an appealing outbound leg. Correspondingly, the sharp drop during the last week reflects the typical vacation span of one week between outbound and inbound travel.

Figure 5.7 also illustrates an exemplary exponential regression line where mean demand is simply modeled based on advance request weeks – i.e., mean demand growth is explained by time to departure alone, not yet accounting for possible micro-variations.

Figure 5.8 gives an impression of the corresponding standard deviation on the same aggregation level, exhibiting the same structure (exponential growth) and an apparent high level of variance.

Taken together, Figures 5.7 and 5.8 clearly indicate a multiplicative demand structure with the typical effect of increasing the standard deviation alongside the mean values (see Harvey, 1993, pp. 106).

Similar conclusions can be drawn from the overall histogram of total requests in the dataset (see Figure 5.9).

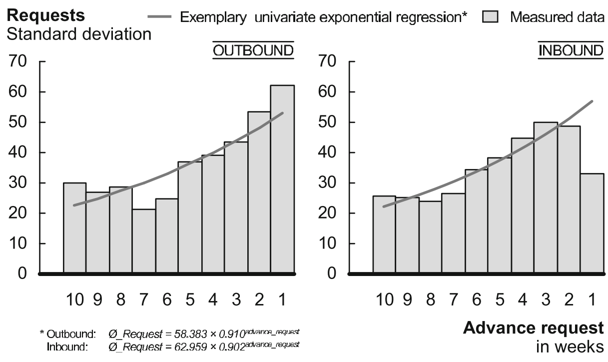


Figure 5.8: Development of standard deviation towards departure – corrected data

Source: own design based on collected data

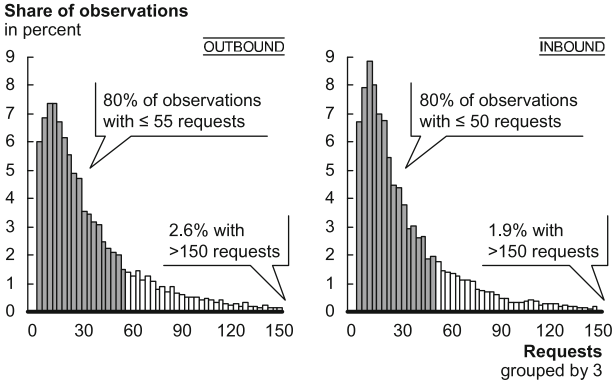


Figure 5.9: Histogram of request observations – corrected and untransformed data
Source: own design based on collected data

The distribution is heavily right-skewed, with ≈ 80% of individual observations below 55 requests on outbound and 50 requests on inbound flight events, respectively. It is obvious that the data distribution is far from a normal distribution; hence, the corresponding test is omitted here.

The above findings yield a set of downsides and restrictions for further analysis and, moreover, for the development of a proper forecasting model, as most methods are based on additive structures (see e.g., Harvey, 1993, pp. 107). Additionally, the instability in variance is undesirable for most statistical method tests.

A possible solution to the above problem is a transformation of the data in the sense explored by Box and Cox (1964), with

$$\phi(y_t) = \begin{cases} \frac{y_t^\lambda - 1}{\lambda} & \text{if } 0 < \lambda \leqslant 1 \\ \ln(y_t) & \text{if } \lambda = 0, \end{cases} \tag{5.1}$$

where $\lambda = 1$ results in a simple linear offset and $\lambda = 0$ yields a logarithmic transformation of the data. Harvey (1993) recommends setting $\lambda = 0$ for heavily right-skewed data to possibly achieve multiple effects on the underlying data (see also Christopher Chatfield, 1989, Sec. 2.4):

- **Stabilize variance:** Particularly in time series exhibiting a trend where "the standard deviation is directly proportional to the mean, a logarithmic transformation is indicated" (Christopher Chatfield, 1989, p. 11) to stabilize the variance.

- **Make multiplicative effects additive:** For example, seasonal effects on data containing a trend are usually multiplicative. Here, a logarithmic transformation results in an additive relationship.[3]
- **Achieve normal distribution:** The logarithmic transformation of right-skewed data may yield normal distributed inputs, as lower source values are more heavily scaled than higher base values.

Following the above discussion, the original observation data is transformed using $\lambda = 0$, i.e., $\phi(y) = \ln(y)$. Figure 5.10 shows the resulting changes on request development in the transformed data set, with the mean demand now growing linearly over advance request time and, moreover, the standard deviation being flat over time.

Thus, the effect of variance stabilization and the multiplicative catenation of demand drivers have been achieved through the transformation. The transformed distribution in Figure 5.11 also gives rise to the assumption of a normal distribution.

As the assumption of normal distributed observations is important not only for the later-introduced outlier detection technique but also, more im-

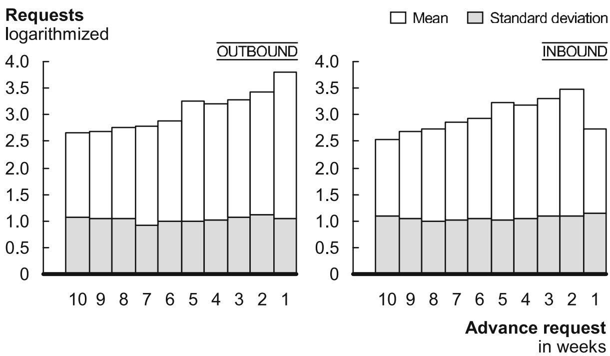


Figure 5.10: Development of requests towards departure – corrected and logarithmized data
Source: own design based on collected data

3 If $c = a \cdot b$, then following the logarithmic rules $\ln(c) = \ln(a) + \ln(b)$.

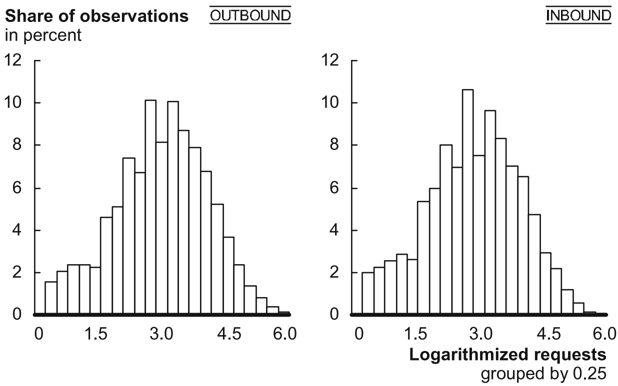


Figure 5.11: Histogram of request observations – corrected and logarithmized data
Source: own design based on collected data

portantly, for the modeling approach described in Chapter 6, the necessary test procedure and its results are described in somewhat more detail in the following paragraph.

Test for normal distribution When one compares Figures 5.9 and 5.11, it becomes apparent that the test has to be based on the logarithmized observations, whose histogram seems closest to a normal distribution just by visual inspection alone. For later use in the forecasting model, each individual time series (i.e., for each flight departure date) is tested for its adherence to the normal distribution assumption.

Here, the Kolmogorov-Smirnov-test is used to test the null hypothesis that the distribution $F_j(\ln(y_{jt}))$ of the logarithmized demand observations $\{\ln(y_{jT}), \ldots, \ln(y_{j1})\}$ of each flight departure event j is sufficiently close to a normal distribution F_{N_j} with parameters μ_j and σ_j^2, whereby these moments are estimated based on $\hat{\mu_j} = \bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_{jt}$ and $\hat{\sigma^2} = s^2 = \frac{1}{T-1}\sum_{t=1}^{T}(y_{jt} - \bar{y})^2$.

Hypothesis
$H_0 : F_j(\ln(y_{jt})) = F_{N_j}$ with $F_{N_j} \sim N(\mu_j, \sigma_j^2)$ against
$H_1 : F_j(\ln(y_{jt})) \neq F_{N_j}$.

Test statistic
$D_j^1 = max_t \; |F_j(\ln(y_{j(t-1)})) \; - F_{N_j}(\ln(y_{jt}))|$
$D_j^2 = max_t \; |F_j(\ln(y_{jt})) \quad - F_{N_j}(\ln(y_{jt}))|$
$D_j = max \; (D_j^1, D_j^2)$
with values $\ln(y_{jt})$ ordered in ascending fashion.

Critical values
The null hypothesis H_0 is rejected if the test statistic is greater than the corresponding critical value $D_j > D_{n;1-\alpha}$.

Figure 5.12 demonstrates the functionality of the Kolmogorov-Smirnov-test: the distances between the cumulative distribution function F_j of the logarithmized observations for a particular flight date j and the assumed normal distribution F_{N_j} are calculated, and the maximum of these deviations D_j is not allowed to exceed a certain threshold $D_{n;1-\alpha}$ that depends on the number of observations n and the desired significance level α.[4]

The test has been applied separately to all sets of observations relating to specific flight departures. For all but one series (outbound departure on December 08, 2007) the null hypothesis could not be rejected based on a

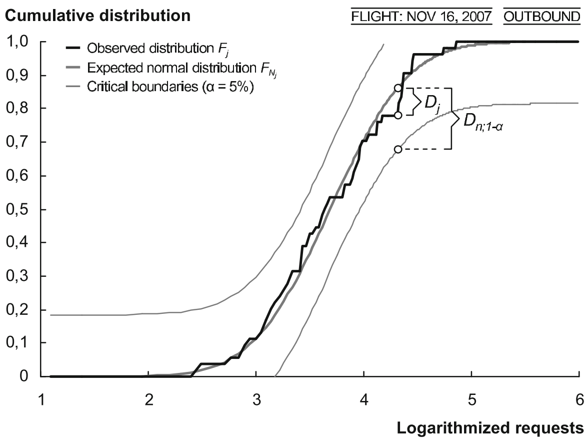


Figure 5.12: Illustration of Kolmogorov-Smirnov-test
Source: own design based on collected data

[4] For a detailed discussion of the Kolmogorow-Smirnow-test see, e.g., Sachs (1997) or the original sources Kolmogorov (1933) and Smirnov (1939). The critical values for $D_{n;1-\alpha}$ are taken from Sachs (1997), p. 427.

significance level of $\alpha = 5\%$, and hence, the assumption of log-normal distributed requests is kept for these time series for the remainder of the work, while the set of December 08, 2007 has been excluded from further analysis.

The next section examines long-term characteristics throughout the dataset to detect possibly changing seasonal trends throughout the entire departure dates horizon. Short-term seasonal patterns between adjacent departure dates and within single flight events are examined in Sections 5.3.2 and 5.3.3, respectively.

5.2.2 Macro-Seasonalities and Trends

The objective of this section is the analysis of long-term seasonalities or trends across the entire dataset (i.e., over the series of all flight events) to define the appropriate base point for forecasting flight-individual time series.

Section 5.2.1 highlights that the overall development of requests within the period starting 60 days to departure seems to follow a log-linear structure, where as open question remains the definition of an appropriate anchor point (i.e., the appropriate intercept) for such a log-linear growth model.

Figure 5.13 plots the average requests within the targeted forecasting horizon over the departure dates contained in the observation data (grouped by calendar weeks). It is fairly obvious that this average demand level follows

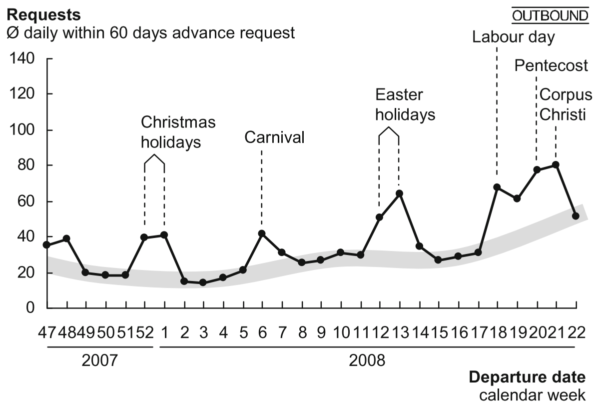


Figure 5.13: Long-term seasonality of requests – outbound
Source: own design based on collected data

an inert curve of around 20 to 30 daily requests that is slowly rising towards the summer main season (gray shade). The trajectory exhibits minor seasonality with a local low end in December/January, slowly rising levels towards March and Easter and finally, a growing level at the end of the dataset.

The major deviations from the described long-term trend are invariably driven by holidays and vacations (as highlighted in the graphic). This finding gives rise to the typical traditional air traffic demand forecasting schemes, based on transferring historic seasonal structures with subsequent adjustments for calendar effects (see Talluri and van Ryzin, 2005, Sec. 9.3).

A further investigation of the described effects, splitting the 60-day aggregate from Figure 5.13 into three equally broad groups (early, mid-term and last-minute demand), reveals the following findings (see Figure 5.14):

- **Early demand** is least susceptible to fluctuations induced by vacations or holidays. The overall variation is smallest between the three groups with the early demand level typically being the lowest – i.e., demand usually increases towards departure. Hence, early demand may be a good and stable indicator of the starting level of log-linear demand growth (see Section 5.2.1) and long-term demand trends, but it carries limited information about mid-term seasonality, as this element is mainly driven by short-term demand.

- **Last-minute demand** is the most prone to massive increases induced

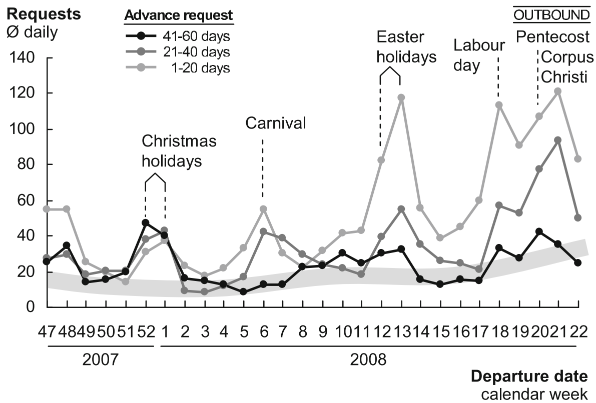


Figure 5.14: Long-term seasonality of requests – outbound, grouped by advance request (early, mid-term and last-minute demand)
Source: own design based on collected data

by spontaneous customers. Thus, based on the overall long-term seasonality typified by the early demand level, micro-seasonal indicators like holidays or extended weekends are the driver of demand growth close to departure.

Here, the effect is most visible on holidays surrounding weekends in May. It is less visible at Christmas, a holiday around which people tend to either make definite plans to visit relatives and thus book early or simply stay at home, with no late changes in demand.

The exact same effects are visible for inbound flight events (see Figure 5.15), with the unsurprising difference that fluctuations are lagged by about one week, reflecting a typical vacation stay. The amplitude is also smaller, possibly driven by select customers' only querying outbound flights in a first step and – if the prices are outside their reservation price range – skipping the search for return flights.

Figure 5.16 finally compares the 60-day aggregate of outbound and inbound flights directly to visualize the described effects.

The above analyses give rise to the assumption that demand development for flight events within a limited time-frame might exhibit similarities, e.g., the outbound flights directly before Labor Day might follow a similar growth pattern. This matter is illuminated in the following section.

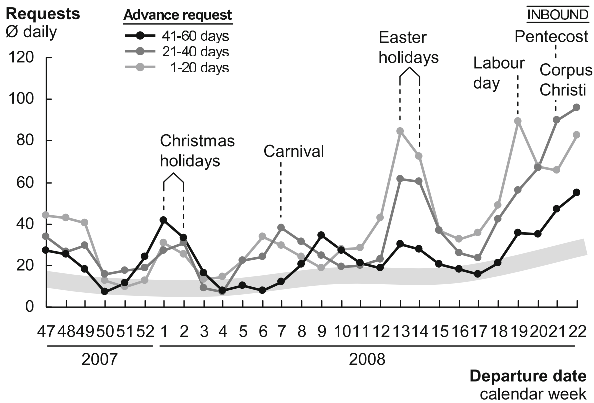


Figure 5.15: Long-term seasonality of requests – inbound, grouped by advance request (early, mid-term and last-minute demand)
Source: own design based on collected data

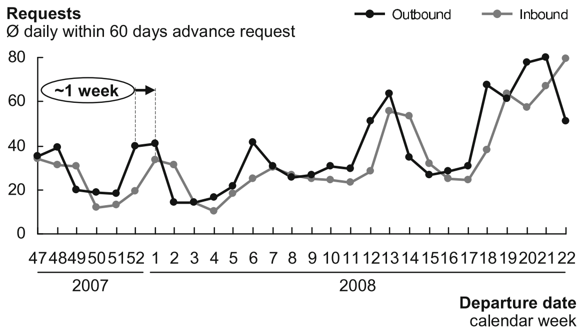


Figure 5.16: Long-term seasonality of requests – outbound vs. inbound
Source: own design based on collected data

5.2.3 Similarities of Adjacent Flights

The previous section has highlighted that demand development is driven by seasonality in general and air travel-inducing events in particular.

Above, Figure 5.14 shows similar demand effects for holidays and vacations: while the variational effect on early demand (60 to 41 days before actual departure) is rather low, the amount of requests significantly increases towards departure. This section now examines whether this behavior also translates into parallel behavior regarding adjacent flights.

As seen, overall aggregated demand around special occasions tends to exhibit an exponential and massive growth pattern. Looking at the disaggregated demand for individual flight events, one sees that this finding could potentially stem from two sources:

1. Dependent on the particular vacation or holiday, there exists a certain preferred travel date on which supply will eventually not be able to fulfill demand. At that point, original demand for that particular date spills over to adjacent days.

2. Initially, the original demand for vacation travel is dispersed across multiple departure dates around the particular holiday as different customer groups favor diverse travel plans (e.g., short weekend trips vs. full-week club vacations).

In both cases, the time series of adjacent flights viewed by request dates should exhibit analog behavior, where the absolute level of demand could

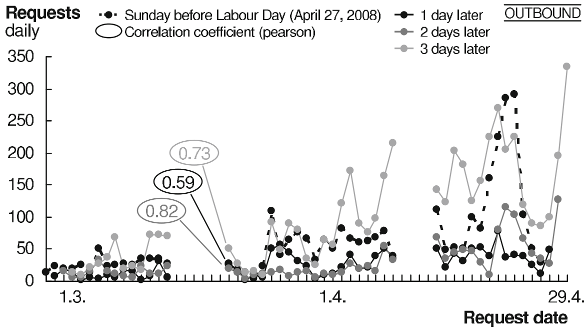


Figure 5.17: Demand development of adjacent flight dates – Labour Day 2008, outbound
Source: own design based on collected data

vary but the structural development should be similar.

Figure 5.17 exemplary shows the demand development for the last four outbound flights *on and before* Labor Day 2008. The particular demand on Sunday, April 27 is taken as a reference curve, against which the demand for the following days is correlated. As indicated in the graphic, all three departure dates exhibit a considerable correlation between 0.59 and 0.82, with Monday explicitly yielding the lowest correlation.

All lines show a fundamental parallel behavior potentially following the second of the above rationales. Interestingly, the Monday departure exhibits massive demand growth and variation starting at the beginning of April, which could be an effect of the first reasoning above: as the Sunday departure fills up and prices are increasing, demand spills to the next best departure date.

Figure 5.18 exemplary shows the same analysis for the corresponding inbound flights for the days *following* Labor Day 2008. Here, demand on the first Sunday after Labor Day (May 04) is taken as a baseline against which the following days' demand is correlated. Again, a significant correlation is found, ranging between 0.69 and 0.82.

The above results exemplary illustrate that there exist potentially strong similarities between adjacent flight events. The implications of this are twofold; first, demand forecasting could benefit from existing information or data from flight departures close to the considered base case. Second, the

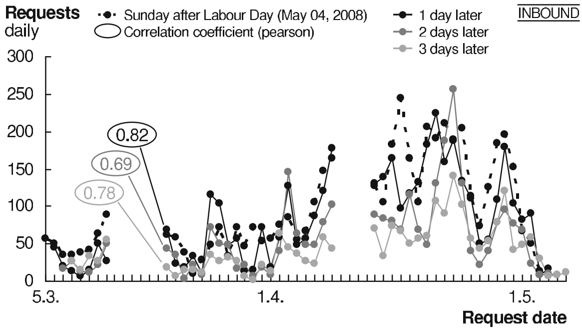


Figure 5.18: Demand development of adjacent flight dates – Labour Day 2008, inbound
Source: own design based on collected data

generation of aggregated forecasts could potentially benefit from the discovered structure as well. Both effects are later considered and tested in the developed forecasting model.

The next section takes the analysis of individual flight demand one step further, examining short-term characteristics within single time series.

5.3 Short-term Characteristics

Taking a closer look at the individual time series for the demand of single flight dates (Figure 5.19 gives an example), one realizes that the data may be exhibiting four different kinds of systematic and stochastic variations (gray shade) that time series in general might contain (see Christopher Chatfield, 1989, pp. 9):

1. **Seasonal effects:** Structural and recurring variations with a time-dependent pattern, which often occur in real time series data due to macro-seasonalities (e.g., increased demand for energy/heating during the winter months) or micro-seasonalities (e.g., electricity demand peaks in the morning, around noon and in the evening).
2. **Time trends:** Non-stationarity over time constituted by a typically monotonic chronological development of measurements, as common in many econometric datasets (e.g., the gross domestic product typically exhibits a positive trend in most developed countries).

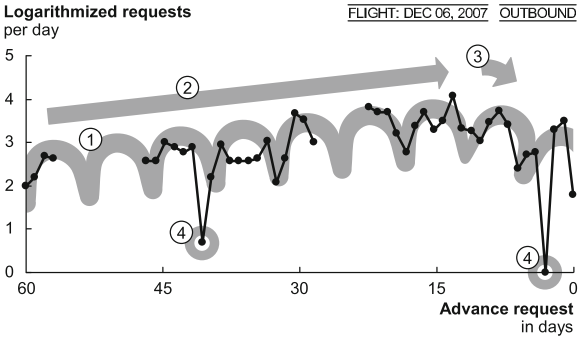


Figure 5.19: Example of flight request time series exhibiting ① seasonal effects, ② a time trend, ③ a structural change and ④ stochastic fluctuations
Source: own design based on collected data

3. **Structural interruptions:** Abruptly occurring level shifts or inversions of trends, which can occur stochastically after external shocks (e.g., the September 11th, 2001 terror attacks) or predictably upon certain predetermined events (e.g., start of the summer season sales in retailing).

4. **Singularities:** Random and extreme intermittent observations that deviate from a series' remaining replicate determinations, which can be due to measurement errors or induced by unexplainable or unrecognized external shocks.

The first three types of series variation are structural and should be reflected in any forecasting model. Their treatment will be discussed in Chapter 6. Conversely, extreme and singular stochastic fluctuations in the sense of the above definition should be treated before the fitting of models – which is discussed in the next section – as their inclusion might seriously disturb the process.

5.3.1 Time Series Disruption Through Outliers

Although the analyzed data is free from measurement errors in this work, outlying observations should still be excluded from the following analysis of short-term effects and the underlying trigger be excluded from the forecasting model:

> *"If (...) we could be sure that an outlier was caused, not by any large error, but some peculiarity (nonnormality) of the inherent variability of the population under study, then it might still make good sense to discard the observation from a statistical analysis based on the method of least squares, but the observation should not be forgotten." (Anscombe and Guttman, 1960, p. 124)*

Anscombe and Guttman propose an impartial rejection rule[5] that must be applied iteratively to any individual time series because it rejects single outliers based on their deviation from the series' mean, which in turn is affected by the outlier itself. Underlying assumptions are that the observations $\{y_{j1}, \ldots, y_{jt}\}$ of the particular time series under consideration for a specific flight j follow a normal distribution $N(\mu_j, \sigma_j^2)$ and that the series is stationary over time.

Under the above assumptions, let $\bar{y}_j = \frac{1}{T}\sum_{t=1}^{T} y_{jt}$ be the estimate of the mean μ_j of the time series j and $z_{jt} = y_{jt} - \bar{y}_j$ denote the deviation of observation number t from that mean. The series' variance σ_j^2 is estimated by $s_j^2 = \frac{1}{T-1}\sum_{t=1}^{T}(y_{jt} - \bar{y}_j)^2$. Looking at the single observation with number m having the greatest deviation z_{jm} from the mean, so that $|z_{jm}| \geqslant |z_{jt}|\ \forall t \neq m$, it should be rejected if it is "excessively large":

> **Step 1:** For a given factor C, reject the observation m with the highest deviation from the observation mean if $|z_{jm}| > C \cdot \sigma_j$.
>
> **Step 2:** If an observation has been rejected in Step 1, consider the remaining observations as a time series of size $T-1$ and start over. The mean and standard deviation of the new series are reestimated using the retained observations only.

Before applying the described procedure to the dataset, its adherence to the necessary prerequisites has to be checked, that is, it has to be tested whether the data seem normally distributed (already done in Section 5.2.1) and exhibit stationarity.

Achieving Stationarity An intuitive view on stationarity is given in Christopher Chatfield (1989, p. 10): "Broadly speaking a time series is said to be stationary if there is no systematic change in mean (no trend), if there is no systematic change in variance, and if strictly periodic variations have been

---

[5] "No prior knowledge concerning the means or regression coefficients that are to be estimated from the data is incorporated in the rejection rule." (Anscombe and Guttman, 1960, p. 128)

removed." In mathematical terms, for a stochastic process to be stationary, the following conditions have to be met (see e.g., Christopher Chatfield, 1989; Fuller, 1996 or Harvey, 1993):

$$E(y_t) = \mu \tag{5.2}$$

$$E[(y_t - \mu)^2] = \sigma_y^2 = \gamma(0) \tag{5.3}$$

$$E[(y_t - \mu)(y_{t-\tau} - \mu)] = \gamma(\tau) \qquad \tau = 1, 2, \ldots \tag{5.4}$$

Expressions (5.2) and (5.3) state that the mean and variance of the series have to be independent of time, while (5.4) requires covariance only to be dependent on the size of the time-lag – not on the specific anchor-point in time.

The described outlier rejection rule put forth by Anscombe and Guttman depends on the first two requirements of stationarity: namely, a time-invariant mean and variance, which will therefore be of main concern here.[6] While the originally time-dependent variance has been stabilized in the course of the series' transformation (see Section 5.2.1), the series clearly still exhibits a trend (for an example, see Figure 5.19).

A simple and (for the purpose of outlier detection) adequate trend removal technique is to model the time series using an univariate linear regression based on the time variable alone and then take the residuals as stationary

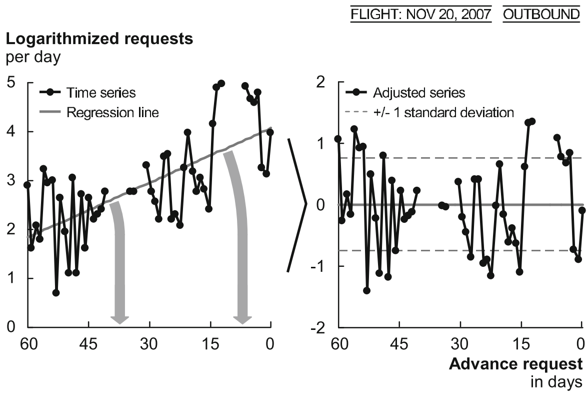


Figure 5.20: Example of trend removal for outlier detection
Source: own design based on collected data

6 A closer examination of the autocorrelation and covariance of the time series is conducted later in Chapter 6.

time series for outlier detection (NIST/SEMATECH, 2008).[7] Figure 5.20 gives a sense of the procedure: the fitted regression line is folded down onto the horizontal axis, thereby removing the trend from the series, while the residuals – and with them, the series' structure – remain intact.

The actual outlier identification can now be conducted based on the adjusted series of logarithmized requests, while the original series will be used for later analysis, as their inherent trend is of particular interest.

Outlier Detection In the following, the trend-adjusted series are used to detect outliers in the sense put forth by Anscombe and Guttman. While the estimation of means and variances is straightforward, the determination of C (the exact cutoff point in terms of multiples of the series' standard deviation) is tedious. As a rule of thumb, many statistical textbooks recommend setting $C = 3$ (see, e.g., Draper and Smith, 1981 or Fuller, 1996), which, under the assumption of normal distribution, would leave out approximately 0.135% of the observations at the lower and upper end each (see Figure 5.21). While this is adequate when the original source data is normally distributed, in the case of log-normal distributed data (as is the case here) it only rejects outliers that are a daunting multiple above the series mean, while readily restricting observations that are still considerably above zero on the lower end (also see Figure 5.21).

After careful examination, in this work, the outlier rejection rule there-

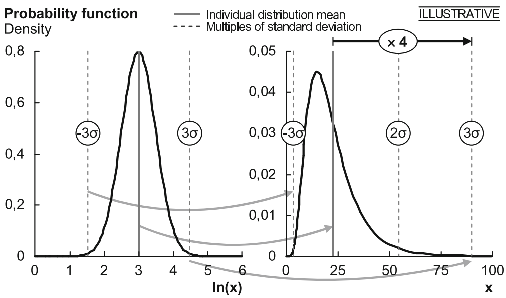


Figure 5.21: Comparison of outlier rejection limits for normal (left) and log-normal (right) distributions
Source: own design

[7] A thorough introduction to linear regression modeling has already been given in Chapter 4 and is therefore omitted here. For a quick introduction to the topic, the reader should refer to Draper and Smith (1981) or Neter et al. (1983).

fore is based on two separate values for the lower ($C_L = -3$) and upper bounds ($C_U = 2$), so that only observations that come very close to zero are rejected, while not allowing excessively massive positive deviations from the stationary mean.

Anscombe and Guttman's rejection rule is applied iteratively over all individual time series. Figure 5.22 shows an example in which the iterative process successfully detects two outliers, where the second one is only visible after the adjusted test measures have been re-computed, excluding the rejected first outlier.

After the processing of the dataset and a visual examination of accumulations of rejections, a total of 417 single outliers were marked for rejection.

Having removed potential disturbances from within the individual time series, we can proceed to Sections 5.3.2 – 5.3.4, which take a closer look at the demand effect of different departure and observation weekdays.

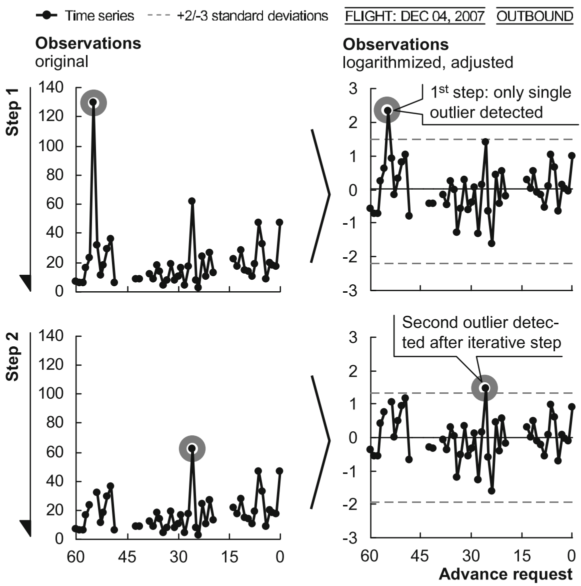


Figure 5.22: Example of successful iterative outlier detection
Source: own design based on collected data

5.3.2 Patterns Based on Departure Weekdays

Section 5.2.2 highlights long-term variations in demand driven by overall seasonality (e.g., winter vs. summer) and vacations or public holidays. This section now examines micro-patterns based on demand variations between departure weekdays.

Figure 5.23 shows a detailed plot of (logarithmized) average requests within the last 60 days to departure by departure date. While the overall demand level follows the described macro-seasonality from Figure 5.13, the graphic also clearly exhibits micro-seasonal demand fluctuations following a weekly pattern (easy to spot based on the highlighted Sundays).

Especially in the lower part of Figure 5.23 (showing the inbound demand), a recurring pattern is clearly visible. While the amplitude varies with macro-seasonality and specific departure date, the underlying weekly structure repeats, with Sunday departures typically exhibiting peak demand

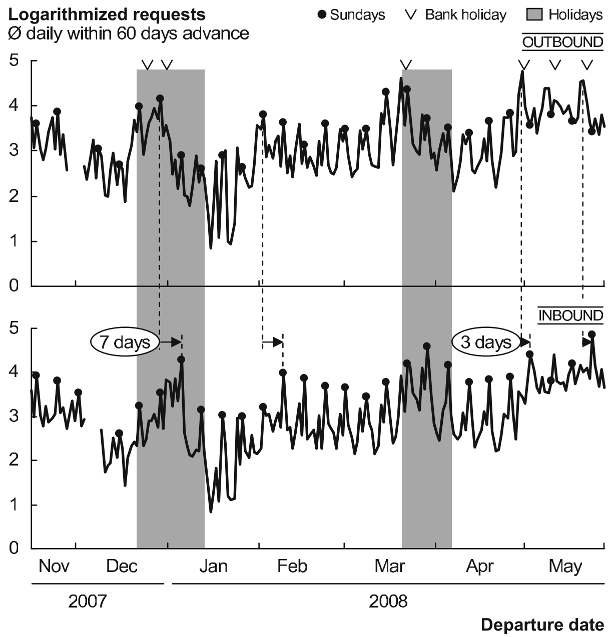


Figure 5.23: Seasonality of requests along departure date
Source: own design based on collected data

(indicated by dots). The understandable exception to this pattern forms outbound demand around the May bank holidays where such peaks occur on the last workdays before the actual holiday.

A comparison of the upper and lower parts of Figure 5.23 also unveils a timely lag between outbound and inbound demand peaks of typically one week. Again, the exception being the May departures where extended weekends invite demand for shorter trip lengths around three to four days.

Although the addressed pattern based on departure weekdays seems visually obvious, it is not necessarily statistically significant. Thus, this assumption will be tested using a two-sided t-test, comparing the mean demand levels of different departure weekdays. For this purpose, the overarching effects of macro-seasonalities and trends (see Section 5.2.2) have to be removed beforehand. The average demand $\bar{y}_j$ of a particular flight event j (within the last 60 days to departure, as shown in Figure 5.23) is indexed on the average demand of all flight departures within ±3 days

$$\bar{\bar{y}}_j = \frac{\bar{y}_j}{\frac{1}{7}\sum_{k=-3}^{3} \bar{y}_{j+k}} \qquad \text{with}$$

$$\bar{y}_j = \frac{1}{60}\sum_{t=1}^{60} y_{jt}.$$

The indexed demand $\bar{\bar{y}}_j$ for a flight j equals one when the demand is exactly the average demand of the entire week centered on the specific departure day j under observation. If demand exceeds average weekly demand, $\bar{\bar{y}}_j$ is greater than one, and vice versa for lower demand.

Figure 5.24 shows boxplots of the distribution of indexed demand $\bar{\bar{y}}_j$ over the entire dataset, differentiated by flight direction and grouped by weekday of departure.

It seems obvious that the average demand level varies by weekday. However, there is still a considerable level of variance (indicated by the 10% and 90% quantiles), so that a test for actual significance is in order.

Two-sided t-test for independent samples The null hypothesis that the mean demand for flights on Mondays is equal to the mean demand on other weekdays is tested using the specifications of the t-test.[8] A necessary prerequisite is that the standard deviations of both samples be equal; this

[8] Monday is chosen as the foundational case because, looking at the mean, it exhibits one of the lowest deviations from the index.

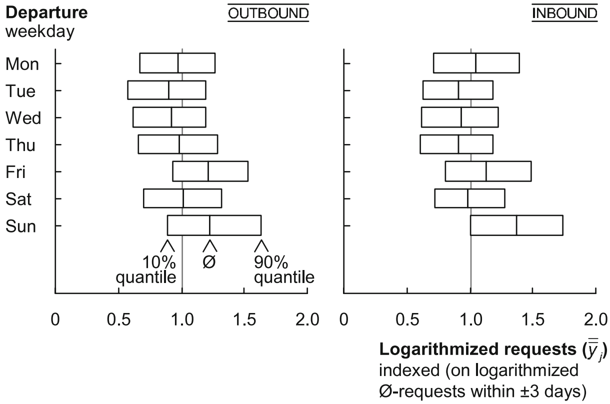


Figure 5.24: Boxplots of request distribution along departure weekdays
Source: own design based on collected data

is tested beforehand using Levene's test.[9] If the standard deviation is not found to be equal at a significance level of $\alpha = 5\%$, a modified test statistic is used for the t-test. The t-test itself is employed as follows:

Hypothesis

$H_0 : \mu_{Mon} = \mu_k \quad k \in \{\text{Tue, Wed, Thu, Fri, Sat, Sun}\}$ against
$H_1 : \mu_{Mon} \neq \mu_k$.

Test statistic

$$t = \sqrt{\frac{MN}{M+N}}\,\frac{\bar{\bar{\bar{y}}}_{Mon} - \bar{\bar{\bar{y}}}_k}{s} \qquad k \in \{\text{Tue, Wed, Thu, Fri, Sat, Sun}\}$$

with

$$\bar{\bar{\bar{y}}}_{Mon} = \frac{1}{M}\sum_{m=1}^{M} \bar{\bar{y}}_m \qquad \text{where } \ weekday(m) = Mon$$

$$\bar{\bar{\bar{y}}}_k = \frac{1}{N}\sum_{n=1}^{N} \bar{\bar{y}}_n \qquad \text{where } \ weekday(n) \neq Mon$$

$$s^2 = \frac{(M-1)s_M^2 + (N-1)s_N^2}{M+N-2},$$

9 For a thorough introduction to Levene's test, see NIST/SEMATECH (2008, Sec. 1.3.5.10), amongst others.

where s^2 is the weighted variance of the individual sample variances s_M^2, s_N^2.

Critical values
Under the null hypothesis, the test statistic is t-distributed with $M + N - 2$ degrees of freedom; hence, H_0 is rejected if $|t| > t(1 - \frac{\alpha}{2}, M + N - 2)$.

Table 5.1 shows the test results for Levene's test on variance equality and the corresponding values for the t-test on equality of sample means. In the four cases where Levene's test rejects the null hypothesis of variance equality at the significance level $\alpha = 5\%$, an appropriate modified test statistic[10] is used for the computation of the t-test on the right side of the table.

The results in Table 5.1 clearly show that the demand level indeed differs based on departure weekdays in all but one case. Only the mean of Thursday outbound departures exhibits a level equal to that of the mean demand on Mondays (base case) at a significance level of $\alpha = 5\%$. Nevertheless, for the forecasting model developed in Chapter 6, the differentiation between departure weekdays should be considered an important input factor.

The next section examines similar effects for weekdays where demand is actually articulated; afterwards, possible cross-effects between the two drivers are examined in Section 5.3.4.

			Levene's test for variance equality			t-test for mean equality				
Direction	Weekday	Mean	Test statistics	Significance of H_0	Reject H_0 at α=5%	Test statistics	Significance of H_0	Reject H_0 at α=5%	Confidence interval	
outbound	Tuesday	0.897	0.005	0.946	–	4.919	0.000	✓	0.043	0.108
	Wednesday	0.927	0.140	0.286	–	3.089	0.002	✓	0.017	0.078
	Thursday	0.980	0.058	0.809	–	-0.406	0.685	–	-0.037	0.024
	Friday	1.215	0.691	0.406	–	-14.760	0.000	✓	-0.273	-0.209
	Saturday	1.019	0.161	0.688	–	-2.807	0.005	✓	-0.076	-0.014
	Sunday	1.227	13.704	0.000	✓	-14.561	0.000	✓	-0.287	-0.219
inbound	Tuesday	0.913	15.602	0.000	✓	8.320	0.000	✓	0.106	0.171
	Wednesday	0.931	6.669	0.010	–	6.667	0.000	✓	0.085	0.156
	Thursday	0.911	12.397	0.000	✓	8.308	0.000	✓	0.107	0.174
	Friday	1.127	0.086	0.769	–	-3.985	0.000	✓	-0,113	-0.038
	Saturday	0.987	15.798	0.000	✓	3.899	0.000	✓	0.032	0.096
	Sunday	1.374	4.574	0.033	–	-14.887	0.000	✓	-0.366	-0.280

Table 5.1: t-Test results of mean equality between departure weekdays

[10] For details on the modidified test statistic to be used, see NIST/SEMATECH (2008, Sec. 1.3.5.3).

5.3.3 Micro-Seasonalities along Observation Weekdays

The above section highlights possible demand differences stemming from unequal departure weekdays, with Sunday departures typically exhibiting the highest demand within a given week.

This section now takes a look at possibly similar micro-seasonalities based on the specific weekday for which demand is articulated, i.e., where a request against the CRS is actually made independent of the particular flight date. For this purpose, the average daily requests for flights within the next 60 days are plotted along the dates when the requests were articulated in Figure 5.25.

First, the graphic unveils an overall but weak trend in demand showing slowly increasing levels towards the summer. More interestingly, the demonstrated development of demand articulation also exhibits micro-seasonality similar to that encountered in Section 5.3.2, but with less extreme amplitudes. However, a recurring weekly structure is visible, here indicating a

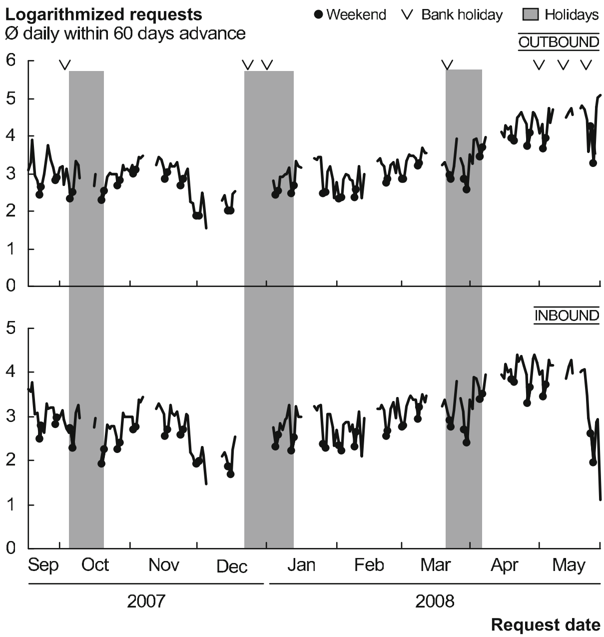


Figure 5.25: Seasonality of requests along request date
Source: own design based on collected data

generally lower demand expression on weekends (highlighted by dots), with a substantial upward shift on Mondays.

It would seem that customers are using time during working hours to search for fares and seat inventory, with a particular spike on the first workday. Thus, the short-term demand structure should measurably vary according to the weekday on which demand is articulated.

To check this assumption, the average demand $\ddot{y}_i$ for flights within 60 days, articulated on a specific observation date i, is calculated and indexed on the average of such demands for observations within a ±3-day range

$$\overset{::}{y}_i = \frac{\ddot{y}_i}{\frac{1}{7}\sum_{k=-3}^{3} \ddot{y}_{(i+k)}} \qquad \text{with}$$
$$\ddot{y}_i = \frac{1}{60}\sum_{s=1}^{60} y_{(i+s)\,s} \ .$$

Figure 5.26 shows the boxplot of the resulting distribution of indexed demand $\overset{::}{y}_i$ grouped by observation weekday and flight direction. It seems obvious that demand varies heavily over weekdays where demand is expressed, starting above average on Mondays, slowly declining throughout the week and ending with a sharp drop towards the weekends.

Again, the two-sided t-test from Section 5.3.2 is used to check the statistical significance of the above findings.

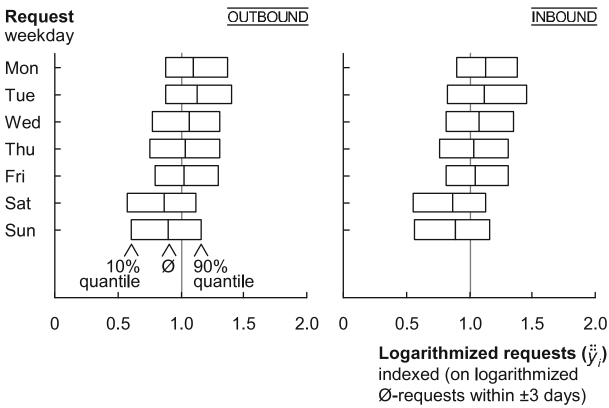


Figure 5.26: Boxplots of request distribution along request weekdays
Source: own design based on collected data

Hypothesis
$H_0 : \mu_{Mon} = \mu_k \quad k \in \{\text{Tue, Wed, Thu, Fri, Sat, Sun}\}$ against
$H_1 : \mu_{Mon} \neq \mu_k.$

Test statistic

$$t = \sqrt{\frac{MN}{M+N}} \frac{\ddddot{y}_{Mon} - \ddddot{y}_k}{s} \quad k \in \{\text{Tue, Wed, Thu, Fri, Sat, Sun}\}$$

with

$$\ddddot{y}_{Mon} = \frac{1}{M} \sum_{m=1}^{M} \ddddot{y}_m \qquad \text{where } weekday(m) = Mon$$

$$\ddddot{y}_k = \frac{1}{N} \sum_{n=1}^{N} \ddddot{y}_k \qquad \text{where } weekday(n) \neq Mon$$

$$s^2 = \frac{(M-1)s_M^2 + (N-1)s_N^2}{M+N-2},$$

where s^2 is the weighted variance of the individual sample variances s_M^2, s_N^2.

Critical values
The test statistic is t-distributed with $M + N - 2$ degrees of freedom; hence, H_0 is rejected if: $|t| > t(1 - \frac{\alpha}{2}, M + N - 2)$.

Table 5.2 shows the results for Levene's test of variance equality and the corresponding t-test. The results clearly confirm that the particular weekday on which demand is articulated also has a significant impact on demand.

Although mean demand on Tuesday observations in both flight directions is not significantly ($\alpha = 5\%$) different from that on Mondays, the results still clearly indicate that the particular weekday on which demand is articulated could be an import driver of overall demand in Chapter 6.

As both, the weekdays associated with flight departure and the weekdays associated with demand observation were found to be significant drivers of the demand structure, the following section briefly examines possible cross-effects between the two – i.e., whether the demand effects are expedited for selected weekday combinations (for flight departure and day of booking/request).

Direction	Weekday	Mean	Levene's test for variance equality			t-test for mean equality				
			Test statistics	Significance of H_0	Reject H_0 at α=5%	Test statistics	Significance of H_0	Reject H_0 at α=5%	Confidence interval	
outbound	Tuesday	1.128	4.395	0.036	–	-1.738	0.082	–	-0.057	0.004
	Wednesday	1.066	3.738	0.053	–	2.120	0.034	✓	0.003	0.067
	Thursday	1.038	1.148	0.284	–	4.055	0.000	✓	0.032	0.093
	Friday	1.030	0.465	0.496	–	5.217	0.000	✓	0.044	0.097
	Saturday	0.866	2.861	0.091	–	17.152	0.000	✓	0.209	0.262
	Sunday	0.904	2.265	0.133	–	14.550	0.000	✓	0.171	0.224
inbound	Tuesday	1.122	4.054	0.044	–	0.257	0.797	–	-0.029	0.037
	Wednesday	1.081	0.009	0.925	–	2.929	0.003	✓	0.015	0.076
	Thursday	1.041	0.153	0.696	–	5.814	0.000	✓	0.057	0.114
	Friday	1.048	0.082	0.774	–	5.638	0.000	✓	0.051	0.106
	Saturday	0.869	0.661	0.416	–	18.361	0.000	✓	0.230	0.285
	Sunday	0.885	2.149	0.143	–	16.790	0.000	✓	0.213	0.269

Table 5.2: t-Test results of mean equality between observation weekdays

5.3.4 Cross-Effects of Departure and Observation Weekdays

Previously, Sections 5.3.2 and 5.3.3 have highlighted that demand variations are statistically dependent (beyond other factors) on the weekday on which demand is exercised (i.e., departure weekday/date) and the weekday on which demand is articulated (i.e., demand observation weekday/date). The pressing question of whether the two effects are possibly interdependent is examined in the following.

Therefore, individual demand y_{jt} for flight j observed at day i is indexed on the average demand for flights within ± 3 days of j that was articulated within ± 3 days of the observation day i

$$\mathring{\mathring{y}}_{ji} = \frac{y_{ji}}{\mathring{y}_{ji}} \qquad \text{with}$$

$$\mathring{y}_{ji} = \frac{1}{49} \sum_{k=-3}^{3} \sum_{l=-3}^{3} y_{(j+k)\,(i+l)} \ .$$

Figure 5.27 plots the combined effects of the actual departure weekday and the weekday on which the demand was articulated, showing the average level of $\mathring{\mathring{y}}_{ji}$ grouped by its specific weekday combination. In all cases, the bubble color indicates whether the indexed demand for a specific weekday combination is above (grey) or below (white) the average (indexed to 1.0), with the bubble size indicating the degree of deviation.

Major deviations occur at the expected departure days Friday to Sunday (also see Figure 5.24) and observation days Saturday and Sunday (also see

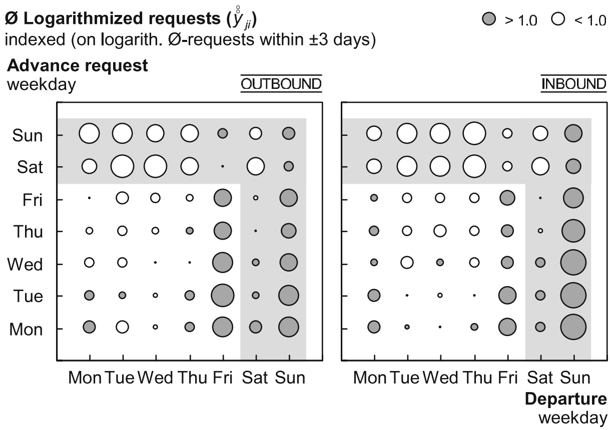


Figure 5.27: Cross-effects of departure and request weekdays
Source: own design based on collected data

Figure 5.26). Here, the effects are obviously additive, without any cross-effect-driven deviations. The same is true for the other workday combinations where the minor variations are also driven by the combined effects of departure and observation weekday, without any additional drivers.

Based on the above findings, cross-effects of departure and observation weekday will not be considered explicitly in Chapter 6 but rather will be modeled independently so that additive effects, as seen in Figure 5.27, can still be contained.

The following section summarizes the findings of this chapter and the implications for the construction of the forecasting model in the next chapter.

5.4 Implications for Forecasting Model

This section summarizes the findings regarding the data's underlying demand structure as have been developed in Sections 5.2 and 5.3 with the aim of outlining implications for the construction of the appropriate forecasting model in Chapter 6.

The following guidelines can be derived for the creation of the demand forecasting model, based on the above-discussed analyses:

- **Data transformation:** The model should be based on logarithmized demand data for three major reasons (see Section 5.2.1):
  - The original demand data exhibit an exponential growth structure over time and therefore, the underlying demand drivers seem to be multiplicatively linked; according to the logarithm rules, these drivers will then be additively linked for the transformed data.
  - Logarithmized demand will follow a normal distribution, which is of particular interest for the application of regression models.
  - The transformed data will exhibit a time-independent variance structure whose overall level will have been stabilized.
- **Separate modeling:** Demand for individual flights should be forecasted separately, as the time-dependent short-term growth structure may vary by season or even by particular flight date (see Section 5.2.2).
- **Season indicator:** Average demand levels early in the booking process (i.e., around 60 days advance request) exhibit fairly smooth seasonal behavior that could act as a trend or macro-season indicator (also see Section 5.2.2).
- **Training sample:** Adjacent flights often exhibit similarities in demand development or structure, respectively. Thus, it could be beneficial to include data from adjacent flights in the data used to train a particular demand forecasting model (see Section 5.2.3).
- **Model parameters:** The model should consider multiple parameters with possible effects on short-term or micro-seasonal demand structure:
  - The particular weekday when the requested flight event takes place seems to be a driver of micro-seasonal demand fluctuations (see Section 5.3.2).
  - The weekday on which such demand is articulated should also be an element of short-term demand structure (see Section 5.3.3).
  - Additionally, the singular demand drivers for holidays or structural effects on full weeks before departure (i.e., 7, 14 or 21 days) should be considered (see Section 5.3).

Chapter 6 now develops a forecasting model using Bayesian regression methods from Chapter 4 based on the structural findings within the demand data taken from the current Chapter 5.

Chapter 6

The Demand Forecasting Model

This chapter's objective is to develop an overarching linear basis function model to forecast demand, which is done in Section 6.1. To ensure that it includes all relevant demand drivers, the model is validated following a typical frequentist interpretation and using the appropriate tests in Section 6.2 before the approach is finally extended to allow for Bayesian learning in Section 6.3.

The performance of the developed model is analyzed in detail later in Chapter 7, where extensions and improvements to the initial Bayesian mechanism are also introduced.

The calculations and tests of the fundamental linear regression model (i.e., based on a frequentist view) have been executed in SPSS 14.0. The actual experiments based on Bayesian regression modeling have been run in a proprietary environment that has been deliberately coded by the author using Microsoft's Visual Basic for Applications on top of an Microsoft Access database for storage of observations and results (see also snapshots of the implementation in Figures 6.3 – 6.5 below). To circumvent possible numerical instabilities in matrix calculations (i.e., inversions), Bluebit's Matrix Active Component has been used in the Advanced Version.

6.1 Linear Basis Function Model

In this section, an overarching linear basis function model (according to Section 4.1) is developed based on the findings in the previous Chapter 5.

The basic linear model structure is justified by the results from Section 5.2.1, which, together with the discovered macro-seasonality from Sec-

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6_6,

tion 5.2.2, lead to the transformations in Section 6.1.1. The discovered short-term characteristics driven by departure (see Section 5.3.2) and articulation weekdays (see Section 5.3.3) are then considered in Section 6.1.2.

6.1.1 Indexing and Data Transformation

The objective of the overall chapter is to actually define an overarching *model structure* for demand forecasting – i.e., the selection of input variables and their functional composition (here additive) is fixed, while the weights of the drivers are later estimated or learned using a Bayesian scheme, separately for individual flight events.

Parameter Indexing Section 5.2 illustrates that total demand levels vary according to a smooth macro-seasonal variation (see Figures 5.13 and 5.14). In the last 60 days before departure, particular growth patterns start from this basic level whose characteristics are discussed in Section 5.3. Thus, for the definition of an overarching model structure, this macro-season level should act as an indexing basis to level long-term influences and to provide a consistent starting point for learning regarding individual growth patterns.

To achieve a robust forecast level, the possible micro-seasonal effects from Section 5.2.3 have to be mitigated. Therefore, a macro demand level $\bar{\bar{D}}_j$ around a particular flight j is defined as the average of the seven micro demand levels $\bar{D}_k$ of flights centered around j (i.e., $j-3 \leqslant k \leqslant j+3$). In turn, $\bar{D}_k$ reports the average demand for an individual flight k in the week before the 60-day forecasting horizon begins

$$\bar{\bar{D}}_j = \frac{1}{7}\sum_{k=j-3}^{j+3} \bar{D}_k \qquad \text{with} \tag{6.1}$$

$$\bar{D}_j = \frac{1}{7}\sum_{t=61}^{67} D_{jt}. \tag{6.2}$$

Consequently, the resulting macro-season indicator is robust against micro-seasonal effects (i.e., departure and articulation weekdays) and can serve as deterministic input variable at the beginning of the actual forecasting horizon (60-day advance request).

The season indicator (6.1) can now be used to index the individual demand observations D_{jt} for later use as dependent variables y_{jt} in the regression model. In addition, the micro-seasonal demand (6.2) is indexed for use as an independent input variable $\tilde{D}_j$ (see below Section 6.1.2)

$$y_{jt} = \frac{D_{jt}}{\bar{\bar{D}}_j} \quad \text{and} \tag{6.3}$$

$$\widetilde{D}_j = \frac{\bar{D}_j}{\bar{\bar{D}}_j}. \tag{6.4}$$

Figure 6.1 illustrates both, the calculation of the micro- and macro-season indicators (6.1) and (6.2), as well as the inference of the indexed point demand (6.3) and indexed micro-season level (6.4).

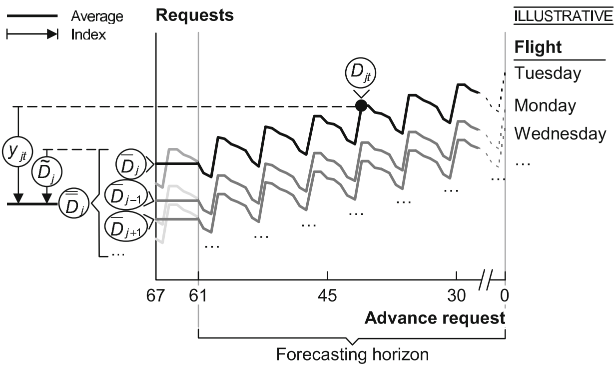


Figure 6.1: Illustration of indexed demand variables
Source: own design

Data Transformation Based on the original observations of the collected data, Section 5.2.1 reveals that the demand levels for individual flights seem to be *log-normally distributed*.[1]

The indexing of demand as discussed above does not affect the normal distribution assumption of the transformed indexed data, as according to the algebraic product rules for logarithms

$$\begin{aligned} \ln(y_{jt}) &= \ln(D_{jt}/\bar{\bar{D}}_j) \qquad \Leftrightarrow \\ &= \ln(D_{jt}) - \ln(\bar{\bar{D}}_j), \end{aligned} \tag{6.5}$$

1 A random number x is said to follow a log-normal distribution if $\ln(x)$ follows a normal distribution.

with

$$\begin{aligned} \ln(D_{jt}) &\sim \mathcal{N}(\mu, \sigma^2) && \text{(see Section 5.2.1)} \\ \ln(\bar{\bar{D}}_j) &= M && \text{(M deterministic)} \end{aligned}$$

so that

$$\begin{aligned} \ln(y_{jt}) &\sim \mathcal{N}(\mu - M, \sigma^2) && \Leftrightarrow \\ &\sim \mathcal{N}(\mu - \ln(\bar{\bar{D}}_j), \sigma^2). && (6.6) \end{aligned}$$

The calculated distribution (6.6) for the logarithmized indexed demand (i.e., the dependent variable in the regression model) firstly shows that the pure indexed demand y_{jt} is also log-normally distributed (similar to the non-transformed raw data) and secondly unveils that the demand structure of the transformed demand again seems additive, as the mean of the resulting normal distribution is linearly shifted by the macro-seasonal level $\bar{\bar{D}}_j$, such that the log-normal distribution of the indexed demand y_{jt} is also called a *shifted log-normal*.

The next section explains the independent model parameters for the regression model to forecast the above-defined logarithmized indexed demand. The micro-seasonal indexed demand $\widetilde{D}_j$ from (6.2) is used as an additional location parameter for individual flight demand.

6.1.2 Driving Model Parameters

This section defines the independent model parameters, and with them, the overall regression equation used in the following to forecast latent demand.

As explained in Section 6.1.1 above, the dependent variable to be forecasted by the model is the logarithmized indexed demand $\ln(y_{it})$. Correspondingly, all independent variables or their respective basis functions also have to fit with the logarithmic transformation.

Advance request time The most obvious independent input parameter (see the plots in Section 5.2.1) is the remaining time to departure or advance request time t, as the particular demand growth for a flight is dependent on time: linearly in the transformed (i.e., logarithmized) demand or exponentially in the original data.

Linear regression models that are based on time as an independent variable may suffer from autocorrelation (see Section 4.1 or Backhaus et al., 2006, Sec. 1.2.5.5) because strong trends in the data typically result in timely close residuals to be dependent on each other.

The "threat" of autocorrelation is not problematic in the model being considered for three major data-inherent reasons:

1. The regression is based not on a single time series with an overarching trend but on multiple micro-series – one for each individual flight event.
2. Although having the same length (60 days or observations respectively), the individual series are shifted in time because of incrementing flight departure dates (the series end one day before the actual departure).
3. The series exhibit strong seasonality on top of the basic time trend, which is also shifted between series, i.e., the specific weekday of the day where the last observation was made within the micro-series, varies depending on the departure date.

Nevertheless, the model is later tested for possible autocorrelation in Section 6.2.2 to assure its adherence to the regression's underlying assumptions.

Micro seasonality Section 5.2.2 reveals that the overall demand level for flights exhibits a macro-seasonal behavior, which is visible early in the booking process, i.e., before the actual forecasting horizon begins. In a single time series setting, the intercept variable would adjust the regression line to the appropriate overall level. Here, the micro-season level $\bar{D}_j$ (see 6.2) is indexed against the macro-season level, and the result is used as additional independent variable $\widetilde{D}_j$ (see 6.4) to define a location parameter for individual flight demand. As the micro-seasonal flight demand is calculated before the actual forecasting horizon begins, it is considered a deterministic input variable for the regression model and has to be logarithmized to the same metric to which the dependent variable y_{jt} is transformed.

Departure weekday In addition to the macro characteristics, Section 5.3.2 unveils a seasonal pattern along the specific weekdays of flight departure. As weekdays follow a so-called nominal scale[2], the characteristics of departure

[2] A nominal scale is not continuous, but values can be distinguished into a finite amount of groups that are not ordered in rank (see, e.g., Stevens, 1946).

weekdays cannot be considered within a single variable, but have to be codified into so-called *dummy variables*, each having the binary values

$$FT_j^W = \begin{cases} 1 & \text{if } weekday(j) = W, \\ & W \in \{Mon, Tue, Thu, Fri, Sat, Sun\}, \\ 0 & \text{else.} \end{cases} \tag{6.7}$$

Notably, the set of dummy variables FT_j^W contains only six variables, as one value is not coded to prevent perfect multicollinearity (see Section 4.1). For this purpose, Wednesday is chosen simply because the deviation from the average demand level is lowest (see Figure 5.24).

Weekday of demand articulation Similarly, the effects of weekdays where demand is articulated are modeled using a separate set of dummy variables RQ_{jt}^W, which are dependent on the particular flight date and observation time

$$RQ_{jt}^W = \begin{cases} 1 & \text{if } weekday(j-t) = W, \\ & W \in \{Mon, Tue, Thu, Fri, Sat, Sun\}, \\ 0 & \text{else.} \end{cases} \tag{6.8}$$

Again, Wednesday is not coded to prevent perfect multicollinearity, as it exhibits the lowest deviation from the average request articulation (see Figure 5.26).

Interaction effects between the departure's weekday and the weekday of request articulation are not considered separately, as they were not found to be significant when tested explicitly (see Sections 5.3.4 and 6.2.1).

Finally, two additional binary variables are included to codify isolated demand effects based on specific events.

Request on bank holiday Section 5.2.2 (see Figure 5.14) illustrates the expedited demand growth effects during public holidays, when demand exhibits strong peaks close to departure as potential customers seek last-minute getaways. A corresponding binary input HY_{jt} is defined as

$$HY_{jt} = \begin{cases} 1 & \text{if } weekday(j-t) \text{ is a public holiday,} \\ 0 & \text{else,} \end{cases} \tag{6.9}$$

which – based on the multiplicative composition (see below Section 6.1.3) – allows for enhanced growth solely for flights requested around public holidays.

Request on full weeks multiple The individual time series also exhibit considerable demand peaks at observation times exactly one, two or three weeks before departure and also close to one month to departure (i.e., 30 and 31 days). As these specific request days show a significant influence on the demand model (see also Section 6.2), an additional binary identifier is added to the model with

$$FW_{jt} = \begin{cases} 1 & \text{if } t \in \{7, 14, 21, 30, 31\}, \\ 0 & \text{else.} \end{cases} \tag{6.10}$$

Table 6.1 finally summarizes all described variables and their coding used in the model following below.

The detected similarities of adjacent flights (see Section 5.2.3) are not contained *within* the model, but are considered through the particular usage of the *learning mechanism*, which is described in Section 6.3.2 below.

The next section now composes the full (transformed and indexed) model and its equivalent true (log-normal) demand equation.

Driver	Type	Coding	Variable definition
Advance request time	continuous	real	t
Micro-seasonality	continuous	real	$\ln(\widetilde{D}_j)$
Weekday of flight departure	nominal	binary	$FT_j^W = \begin{cases} 1 & \text{if } weekday(j) = W, \\ & W \in \{Mon, Tue, Thu, Fri, Sat, Sun\}, \\ 0 & \text{else.} \end{cases}$
Weekday of demand articulation	nominal	binary	$RQ_{jt}^W = \begin{cases} 1 & \text{if } weekday(j-t) = W, \\ & W \in \{Mon, Tue, Thu, Fri, Sat, Sun\}, \\ 0 & \text{else.} \end{cases}$
Request on public holiday	nominal	binary	$HY_{jt} = \begin{cases} 1 & \text{if } weekday(j-t) \text{ is a public holiday,} \\ 0 & \text{else.} \end{cases}$
Request articulated on full weeks multiple	nominal	binary	$FW_{jt} = \begin{cases} 1 & \text{if } t \in \{7, 14, 21, 30, 31\}, \\ 0 & \text{else.} \end{cases}$

Table 6.1: Modeled demand drivers with variable coding

6.1.3 Model Specification and Re-transformation

With the model parameters now having been introduced, this section, in summary, defines the final model as well as the corresponding re-transformation to the actual underlying log-normal demand structure.

Taking the defined independent variables from Table 6.1, the linear basis function model for logarithmized and indexed demand (see Section 6.1.1) is as such

$$\begin{aligned}\tilde{f}(\mathbf{x}, \mathbf{a}) = \ln(y_{jt}) = a_0 + a_1 \cdot t + a_2 \cdot \ln(\widetilde{D}_j) \\ + \sum_{k=1}^{6} a_{2+k} \cdot FT_j^{w_k} + \sum_{l=1}^{6} a_{8+l} \cdot RQ_{jt}^{w_l} \\ + a_{15} \cdot HY_{jt} + a_{16} \cdot FW_{jt} \ , \end{aligned} \tag{6.11}$$

with

$$w_{\{k|l\}} \in W = \{Mon, Tue, Thu, Fri, Sat, Sun\}.$$

This – in matrix notation following (4.2) – corresponds to

$$\tilde{f}(\mathbf{x}, \mathbf{a}) = \boldsymbol{\phi}(\mathbf{x})^T \cdot \mathbf{a}, \tag{6.12}$$

with

$$\mathbf{a} = (a_0, \ldots, a_{16})^T, \tag{6.13}$$

$$\begin{aligned}\mathbf{x} = (j, t, y_{(j-3)61}, \ldots, y_{(j-3)67}, \\ \ldots \\ y_{(j+3)61}, \ldots, y_{(j+3)67})^T,\end{aligned} \tag{6.14}$$

$$\begin{aligned}\boldsymbol{\phi} = \big(\phi_0(\mathbf{x}) &= 1, \\ \phi_1(\mathbf{x}) &= t, \\ \phi_2(\mathbf{x}) &= \ln(\widetilde{D}_j), \\ \phi_k(\mathbf{x}) &= FT_j^{w(k-2)} \quad k = \{3, \ldots, 8\}, \\ \phi_l(\mathbf{x}) &= RQ_{jt}^{w(l-8)} \quad l = \{9, \ldots, 14\}, \\ \phi_{15}(\mathbf{x}) &= HY_{jt}, \\ \phi_{16}(\mathbf{x}) &= FW_{jt}\big)^T.\end{aligned} \tag{6.15}$$

Taking a closer look at (6.13) – (6.15), it becomes apparent that the number of input variables in $\mathbf{x}$ (51 in total) heavily exceeds the amount of actually used basis functions in $\boldsymbol{\phi}$ (16 in total). Notably, based on a 60-day forecast horizon, the actual variable inputs (6.14) and with them the basis functions (6.15) do not reflect any changes through the Bayesian learning mechanism, i.e., their values are considered deterministic at all times throughout the forecast horizon. The learning effect is fully contained in the changing coefficients (6.13) or their distribution, respectively, which are in turn thought of as being stochastic.

The particular usage of the linear basis function model (6.11) in conjunction with the Bayesian learning mechanism is explained in the later Section 6.3.

For the actual demand forecast (see Chapter 7), the model or its predictive results each have to be re-transformed to the original data metric through exponentiation

$$\begin{aligned} y_{jt} = {} & e^{a_0} \cdot e^{a_1 \cdot t} \cdot \widetilde{D}_j^{a_2} \\ & \cdot \prod_{k=1}^{6} e^{a_{2+k} \cdot FT_j^{w_k}} \cdot \prod_{l=1}^{6} e^{a_{8+l} \cdot RQ_{jt}^{w_l}} \\ & \cdot e^{a_{15} \cdot HY_j} \cdot e^{a_{16} \cdot FW_{jt}}, \end{aligned} \tag{6.16}$$

with

$$w_{\{k|l\}} \in W = \{Mon, Tue, Thu, Fri, Sat, Sun\}.$$

Obviously, (6.16) is indeed a multiplicatively linked and exponential growth model over t.[3] In addition, the effect of the binary dummy variables is intuitive: if a particular dummy variable evaluates to one, a factor depending on the corresponding coefficient is multiplied to the model; if it evaluates to zero, the multiplier has no effect on the model whatsoever ($e^{a_k \cdot 0} = e^0 = 1$).

Finally, y_{jt} still represents the indexed demand so that the forecast for the actual latent demand for flight j at time to departure t evaluates to

$$D_{jt} = y_{jt} \cdot \bar{\bar{D}}_j. \tag{6.17}$$

[3] Note that the exponential effect is mitigated for deterministic figures like $\widetilde{D}_j$ or the intercept, as these evaluate to constants, independent of the specific remaining time to departure.

Correspondingly, the above D_{jt} is used in Chapter 7 to evaluate the performance of the forecasting model against the true observed latent demand.

With the full model specification now given, the next Section 6.1.4 describes sample weights for the model coefficients based on conventional OLS estimation and a limited subset of the available data under the assumption of deterministic coefficients. Based thereon, the model is validated against the general regression requirements in Section 6.2 before it is taken to a Bayesian interpretation over the full data set in Section 6.3.

6.1.4 Frequentist Coefficient Weights

To test the plausibility of the model in terms of coefficient significance in the following and to validate the model along the assumptions of Section 4.1 in the next Section 6.2, in this section a frequentist view is taken. That is, the basic log-linear model structure is estimated based on an observation data subset (November and December 2007), for which stable coefficients can reasonably be assumed following conventional OLS estimation and frequentist interpretation (see Sections 4.1 and 4.2, respectively).

However, the derived coefficient weights only serve as proxy to test the model significance and credibility in a conventional frequentist and *retrospective* setting in Section 6.2. Later in Chapter 7, a *prospective* view is taken, and the coefficients are assumed to be stochastic and time-variant. There, forecasts derived under a Bayesian setting are used for evaluation based on the full dataset to allow for time-varying coefficients and macro-seasonal effects.

Based on the limited dataset and using the OLS method described in (4.7) of Section 4.1, frequentist estimates $\hat{\mathbf{a}}$ for the coefficient vector $\mathbf{a}$ of (6.12) can be derived independently for outbound and inbound flight data (see Table 6.2).

As a lookahead on the validation (i.e., the test of actual coefficient significance) in the next section, the plausibility of the coefficients' directional effect is evaluated and found to be consistent with the findings described in Chapter 5 (see remarks and explanations in Table 6.2).

It is important to note that Table 6.2 shows deterministic coefficients derived under a retrospective view after consideration of the entire data subset for November and December 2007 and under a frequentist interpretation of probability. The results are presented in anticipation of Section 6.2, where the model structure is examined for explanatory value. However, the actual

Input variables according to (6.15)	Outbound: Model coefficient value	Outbound: Standardized value	Inbound: Model coefficient value	Inbound: Standardized value	Remarks/explanation
Intercept	0.357	—	0.422	—	
Advance request time	-0.008	-0.090	-0.008	-0.095	Negative in declining time
Micro-seasonality	1.070	0.859	1.039	0.859	Expected strong effect
Monday flight departure	0.042	0.009	0.039	0.009	Demand varies by flight departure weekday, depending on flight direction
Tuesday flight departure	-0.030	-0.007	-0.152	-0.036	
Thursday flight departure	0.243	0.051	-0.601	-0.127	
Friday flight departure	0.270	0.065	0.008	0.002	
Saturday flight departure	0.109	0.026	-0.184	-0.044	
Sunday flight departure	-0.349	-0.073	-0.059	-0.013	
Monday request articulation	0.073	0.016	0.146	0.033	Less demand articulated on weekends, with heavy peaks on Mondays
Tuesday request articulation	0.030	0.008	0.109	0.027	
Thursday request articulation	-0.042	-0.009	-0.118	-0.025	
Friday request articulation	-0.133	-0.027	-0.185	-0.039	
Saturday request articulation	-0.542	-0.119	-0.565	-0.125	
Sunday request articulation	-0.447	-0.099	-0.538	-0.117	
Request on public holiday	-0.451	-0.031	-0.128	-0.010	Less requests on holidays
Request articulated on full weeks multiple	0.516	0.092	0.458	0.081	Strong peaks on selected days

Table 6.2: Frequentist coefficient values (November and December 2007)

intention of this work is to derive a stochastic model that provides a sufficiently good forecast based only on a limited information setting, where the actual demand drivers are not fully known and assumed to be stochastic, which is approached in Chapter 7.

The next Section 6.2 now validates the linear basis function model (6.12) – (6.15) in terms of its basic assumptions from Section 4.1, before the particular Bayesian learning mechanism employed is explained in Section 6.3.

6.2 Model Validation

In this section, the linear basis function model described in (6.12) – (6.15) is validated against the assumptions underlying any regression model (see Section 4.1), in terms of the linearity of the model and the significance of the used input variables (Section 6.2.1) as well as its adherence to the assumptions for the employed solution method (Section 6.2.2).

This validation section is based on the data subset already introduced and used in the previous Section 6.1.4. Henceforth, the coefficient values from Table 6.2 – estimated based on conventional OLS – are used to determine the test statistics and goodness-of-fit values below.

The overall adjusted coefficient of determination for the described model and coefficients is evaluated to $R^2_{adj\ OUT} = 0.778$ for outbound flights and $R^2_{adj\ IN} = 0.747$ for inbound flights, implying that the model seems capable of explaining 77.8% or 74.7%, respectively, of observed daily latent demand variance. However, as a simple evaluation of the coefficient of determination may be misleading in the case of violated model assumptions (see, e.g. Garson, 2008), the latter have to be tested before a final assessment of the model performance can be made.

6.2.1 Model and Coefficient Significance

Two statistical tests are generally used (see, e.g., Backhaus et al., 2006, Sec. 1.2.3 and 1.2.4) to verify that the regression model seems indeed linear in the used parameters and that each single parameter contributes significantly to the overall model performance. The tests have already been introduced and described in Section 4.1 under (T1).

F-test on significance of the linear model The first test investigates the data's adherence to the assumed functional form, i.e., whether the log-linear model assumption seems valid and whether the model as a whole contributes significantly to explaining the overall observation structure.

The null hypothesis to be tested is that all parameter coefficients are zero, that is, the model is not capable of explaining any variation in the underlying data. The corresponding test statistics (see (T1) in Section 4.1) are calculated to 249.717 (outbound) and 213.851 (inbound), both of which are significant at $\alpha = 5\%$. Therefore, the null hypothesis is rejected and the model is principally assumed to be of explicatory value.

t-test on significance of the individual parameters The second test examines whether each individual input parameter or basis function contributes to the overall explanatory quality of the model. Therefore each coefficient a_m is tested for significance (i.e., statistically verifiable difference from zero). The corresponding test statistic and critical values are described in (T1) of Section 4.1.

However, the t-test cannot be used to evaluate the contribution of dummy variables, as these (by definition) are interdependent and henceforth cannot be evaluated separately but only as a group. The recommended method here (see, e.g., Garson, 2008) is to use the so-called *incremental F-test* on the change in R^2 induced by the inclusion of all dummy variables belonging to a certain nominal variable in the model (see Section 6.1.2). The coefficient of determination is calculated for the model excluding the dummy variables $R^2_{without}$ and for the same model, but including the variables R^2_{with}. The corresponding test statistic is then computed to

$$F_{incr} = \frac{\frac{1}{k}(R^2_{with} - R^2_{without})}{\frac{1}{N-m-1}(1 - R^2)} \tag{6.18}$$

with

k : number of dummy variables in the input set
N : number of observations in the data set, here $T \cdot j$
m : number of parameters in the model.

The following Table 6.3 lists the appropriate test statistics for each input variable and the results on whether those are significant to the model at the chosen level of $\alpha = 5\%$.

The tested variables all exhibit significant explanatory effects on the model, except 'request on public holiday' for demand on inbound flights, which could therefore be omitted from the inbound forecast model, but is kept here for possible increasing effects in the remaining data.[4]

In addition to the listed input variables, a set of dummy variables rep-

Input variables according to (6.15)	Type	Outbound		Inbound	
		Test statistic	Significant $\alpha = 5\%$	Test statistic	Significant $\alpha = 5\%$
Advance request time	t-test	-7.398	✓	-6.849	✓
Micro-seasonality	t-test	56.045	✓	51.496	✓
Weekday of flight departure[a]	incr. F-test	11.524	✓	9.910	✓
Weekday of request articulation[a]	incr. F-test	16.502	✓	12.169	✓
Request on public holiday	t-test	-2.140	✓	-0.640	–
Request articulated on full weeks multiple	t-test	6.490	✓	5.373	✓

[a] Multiple dummy variables, tested en block.

Table 6.3: Significance of individual model parameters

[4] While the consideration of non-significant basis functions could harm the regression model in a conventional frequentist setting, Bayesian models are less susceptible to such over-fitting, i.e., the affected variable simply does not get assigned a significant coefficient value (see, e.g., Bishop, 2006).

resenting cross-effects of departure and request articulation weekdays[5] has been tested using the incremental F-test but were not found to be of significant influence to the model, as already postulated in Section 5.3.4.

After the model itself has been tested for validity, that is, it contains only relevant input variables in correct functional composition, its adherence to the assumptions of underlying regression modeling in general has to be tested, which is done in the following section.

6.2.2 Prerequisites and Assumptions

This section reports the test results on whether the described model adheres to the assumptions underlying OLS coefficient estimation and with it – in a weaker form – Bayesian linear regression.

Normal distribution of residuals The Kolmogorov-Smirnov-test described in Section 5.2.1 is used to test the distribution of the resulting residuals for normality. Both test statistics, $D_{e_{out}} = 1.946$ for the outbound model and $D_{e_{in}} = 1.329$ for the inbound model, do not lead to the rejection of the null hypothesis of normal distribution. Hence, the assumptions $e_{out} \sim \mathcal{N}(0, \sigma^2)$ and $e_{in} \sim \mathcal{N}(0, \sigma^2)$ are deemed to be valid.

Homoscedasticity of residuals Following (A3) the distribution of residuals should exhibit homoscedasticity, i.e., the deviation of the regression function from the true observed values should not be dependent on any independent variable of the model (e.g., time or observation order) or the magnitude of the dependent variable. A first indication of whether the model might suffer from heteroscedasticity can be drawn from the plot of the dependent variable over residuals, as shown in Figure 6.2.

The residuals seem randomly scattered around zero and do not exhibit any strong systematic pattern that indicates homoscedasticity (see Backhaus et al., 2006, pp. 86).[6]

Additionally, the White-test for homoscedasticity (see (T3) in Section 4.1) is used to test the null hypothesis $\sigma^2 = \sigma_i^2$ for corresponding samples $\mathbf{X}_i \subseteq \mathbf{X}$. For the outbound flight observations, the test statistic is calculated to $T \cdot R^2 =$

---

5 $$CE_{jt}^{W_1,W_2} = \begin{cases} 1 & \text{if } weekday(j) = W_1 \text{ and } weekday(j-t) = W_2, \\ & W_i \in \{Mon, Tue, Thu, Fri, Sat, Sun\}, \\ 0 & \text{else.} \end{cases}$$

6 The scatter decomposes in two groups (higher and lower levels of demand), which is fine, as both groups' residuals seem evenly distributed around zero.

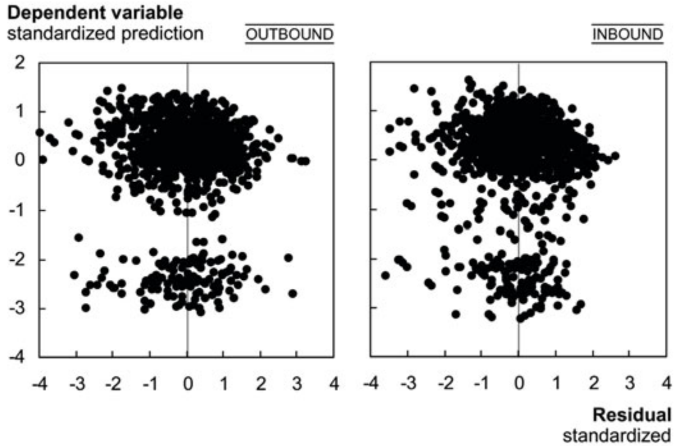


Figure 6.2: Plot of dependent variable over residuals
Source: own design based on collected data

$1,138 \cdot 0.130 = 147.940$, and for the inbound model it computes to $1,153 \cdot 0.115 = 132.595$. The corresponding Chi-squared distribution is $\chi^2_{152;5\%} = 181.770$ so that, for both models, the null hypothesis of homoscedasticity cannot be rejected.

Autocorrelation Following (T4) in Section 4.1, the Durbin-Watson-test for multiple time series of length T is used to test whether the residuals of the regression model exhibit autocorrelation, i.e., whether the model error at a specific observation time seems to be dependent on the error of preceding observations.

The corresponding test statistic evaluates to $D_{out} = 1.788$ for the outbound model and $D_{in} = 1.919$ for the inbound model. Both values are close to $D = 2$, i.e., they lie within the range of $1.5 \leqslant D \leqslant 2.5$, which has been given in (T4) as range indicating a sufficiently low level of autocorrelation.

Multicollinearity The OLS estimator (see Section 4.1) is only mathematically defined if the independent variables (i.e., the regressors) are not fully linearly dependent on each other. Although complete multicollinearity typical only exists in erroneous specified models, multicollinearity up to a certain extent is common in regression models but – if too strong – might harm the BLUE properties of the parameter estimates. Following (T5) in Section 4.1

	Variance inflation factor	
Input variables according to (6.15)	Outbound	Inbound
Advance request time	1.053	1.057
Micro-seasonality	1.989	1.873
Monday flight departure	1.948	2.197
Tuesday flight departure	1.701	2.242
Thursday flight departure	1.873	1.900
Friday flight departure	2.467	2.066
Saturday flight departure	2.105	2.069
Sunday flight departure	2.502	1.968
Monday request articulation	1.566	1.588
Tuesday request articulation	1.691	1.718
Thursday request articulation	1.482	1.524
Friday request articulation	1.479	1.512
Saturday request articulation	1.548	1.578
Sunday request articulation	1.560	1.561
Request on public holiday	1.072	1.104
Request articulated on full weeks multiple	1.035	1.030

Table 6.4: Multicollinearity test results (variance inflation factors)

the variance inflation factors VIF_m of the M regressors are used as indicators for multicollinearity, i.e., the null hypothesis that multicollinearity is present in the model can be rejected if $VIF_m \leqslant 4 \, \forall \, m$. Henceforth, for the considered models, the hypothesis of multicollinearity is rejected for all basis functions according to the results in Table 6.4.

As the basic linear regression model has now been validated, the next section describes how the Bayesian regression scheme from Section 4.3 is finally employed to learn the model coefficients in real time while incorporating the treatment of similar demand structures within the forecasting model.

6.3 Bayesian Learning Mechanism

Up to now, the demand forecasting model from Section 6.1 has been examined under a typical frequentist interpretation of probability: in retrospective view, the observations of latent demand together with the selected input variables are used to determine presumably deterministic parameter values for the underlying model.

Following the motivation from Chapter 2, the derived model structure (i.e., the linear combination of the defined basis functions) is now examined under a Bayesian interpretation of probability, i.e., the assumption that the driving demand parameters are not deterministic but stochastic in the sense that they may vary over time and may be stochastically influenced by external shocks (e.g., weather or economic crisis).

The next Section 6.3.1 describes the overall online demand learning mechanism based on the Bayesian regression, after which the last Section 6.3.2 extends the general approach to consider cross-flight similarities and prior demand knowledge in particular.

6.3.1 Online Demand Learning

This section illustrates how the demand model from Section 6.1 can be employed to learn its driving parameters in real time (online) so as to finally forecast latent demand within a 60-day period before actual flight departure.

The overall demand structure (i.e., the linear composition of the selected basis functions) derived in the above sections is reused as the underlying model for latent demand. However, in the absence of historic data, only the basis functions $\boldsymbol{\phi}(\mathbf{x})$ and the observed data points $\mathbf{y}$ are considered deterministic, while the defining coefficients $\mathbf{a}$ are assumed to follow a multivariate normal distribution according to the definitions in Section 4.3. This in turn yields three important implications for the coefficients and the model itself:

1. The weights of individual demand drivers are *stochastic* and may vary with external shocks. For example, the demand on Mondays may typically be 25% higher than on weekends, but the ratio may vary depending on the weather conditions and is thus uncertain.[7]

2. Following the joint distribution of the coefficients, some of them may be *more certain* than others. That is, the exhibited variance of a coefficient's partial distribution may be lower than that of others.

3. The driving moments of the coefficient distribution (i.e., mean and variance) may *change over time* as more observations or data become available and increase the confidence for certain coefficient ranges.

[7] One may argue that if the weather conditions impact demand, a corresponding input variable should be added to the model. Albeit the weather is an external parameter, it is stochastic and therefore cannot serve as meaningful input parameter for any forecasting model.

In summary, the model is structurally fixed but is variable in the coefficients, which allows for the online learning of demand by continuous adjustment of the model coefficient's distribution moments based on the Bayesian regression mechanism described in Section 4.3. Thus, the model can adapt to changing environmental conditions (i.e., calendar effects), which is of utmost importance as "obviously, the structure of (...) time series may change in a volatile business environment" (Spedding and Chan, 2000, p. 33).

Model initialization A separate set of coefficients $\mathbf{a}_j$ is used for each individual flight departure j to be forecasted, as in principle the demand structures may vary by flight. Based on the findings in Section 5.2.3 that adjacent flights may exhibit similar demand behavior, Section 6.3.2 explains the possible coupling of models or usage of observations from adjacent flights to calibrate a particular model.

In a first step, noninformative priors (see Section 4.4) are used to initialize the joined distribution of the coefficients $\mathbf{a}_j$ around zero means with a considerable variance ($\sigma^2 = 0.2$), so that the possible domain is not restricted and with the individual coefficients' distributions not being interrelated

$$\mathbf{a}_j = \begin{pmatrix} a_0 \\ \vdots \\ a_{16} \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} 0.2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0.2 \end{pmatrix} \right) \forall j. \tag{6.19}$$

Following 6.16 this initializes the dependent variable y_{jt} to one ($y_{jt} = e^0 \cdot \ldots \cdot e^0 = 1$) for all flights and observation times. However, as y_{jt} represents indexed point demand (see Figure 6.1), the initial demand forecast is set equal to the macro demand level $\bar{\bar{D}}_j$ around the considered flight j.

Figure 6.3 shows a screen shot of the application that has been implemented to run the Bayesian regressions on the observation data. The left-hand side shows the controls where the specifications of the described demand model are entered, while the chart on the right-hand side illustrates the true demand development over the last 60 days to departure for the chosen flight (shown as dashed line, because those observations have not yet been considered, i.e., are unknown to the system) and the initial forecast, which based on (6.19) and the predictive distribution (4.32) is initially calculated to be $\bar{\bar{D}}_j$ (shown as continuous line).

Under (6.19) the model expects the macro demand level (before the 60-day forecast period) to sustain throughout the entire forecast period. However, as the coefficients are considered stochastic, this is not a point estimate

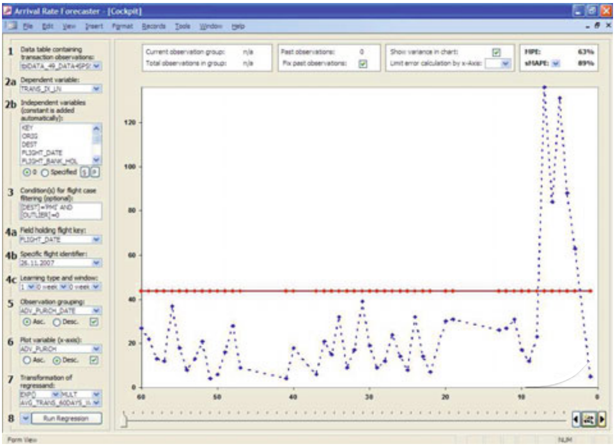


Figure 6.3: Screen shot of implementation – initial forecast with noninformative prior
Source: own programming

(see Section 4.3.2) but rather the average expectation with a significant variance throughout the forecast period (set to be indicated by a fine line, but beyond the chart's x-axis range of Figure 6.3).

Henceforth, the initial forecast exhibits a minimum level of rationality, while at the same time (employing a noninformative prior) being sufficiently broad and general to allow for changing expectations based on future data.

Online learning steps The initial prior distribution of the coefficients (6.19) is updated using the Bayesian regression mechanism (4.28) from Section 4.3.1. As the normal distribution is conjugate to itself, the updated posterior distribution is also normal and can thus directly serve as a prior distribution input for a successive learning step.

An incremental update is carried out online every time a new observation is recorded, i.e., every single day throughout the 60-day forecast period. Depending on the additional wealth of information an incremental data point carries (e.g., information about a particular request articulation weekday is only contained every seven days), the coefficient distribution is updated and with it, the predictive distribution (mean and variance) changes. The effect for the first incremental update on the initial forecast from Figure 6.3 is shown in Figure 6.4.

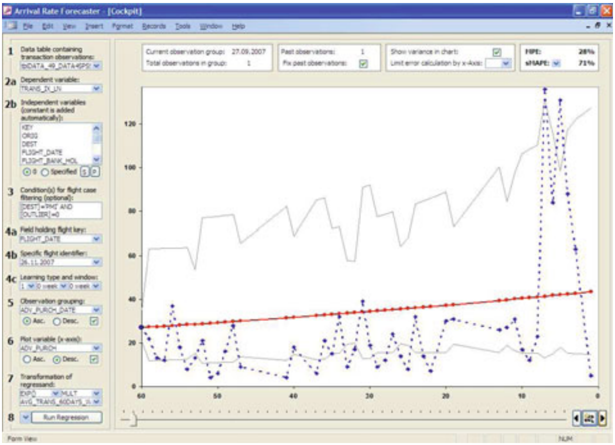


Figure 6.4: Screen shot of implementation – first incremental update
Source: own programming

As can be seen, even a single data point carries enough information to already massively impact the forecast; the regression mechanism immediately adjusts the level of the forecast line and bends it into the expected slightly exponential growth curve. While the mean forecast still forms a smooth line, the inherent uncertainty of the calculated values differs along the forecast horizon, indicated by the varying variance (shown as fine line). For example, the considered observation also contained information about the request structure on a particular articulation weekday (here, 60 days before the considered departure at November 26, 2008, which is September 09, 2008 – a Sunday). Therefore, the uncertainty of the generated forecast (i.e., the variance) is lower for the following Sundays, which is indicated by a lower variance on these days. Naturally, the forecast variance is higher approaching departure.

The described effect increases with additional incremental updates (see Figure 6.5). After a total of seven observations have been loaded, the demand effect of all articulation weekdays has been initially considered, giving the curve a full micro-seasonal pattern.

Apparently, the demand forecast has changed considerably after seven data points, and with it the variance (i.e., the certainty) has improved significantly throughout the entire forecast period, with predictions close to

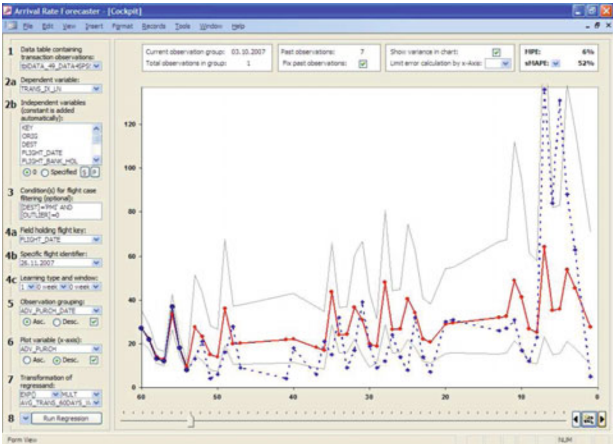


Figure 6.5: Screen shot of implementation – seventh incremental update
Source: own programming

already-observed data points being the most certain. The dark line indicates past observations of demand, which have already been fed into the system (i.e., are known to the learning mechanism).

Naturally, through the addition of further observations, the coefficient distribution and with it the predictive distribution illustrated in the chart will improve further. The coefficient improvement and learning speed is evaluated in detail in Section 7.1.1.

Advantages of the Bayesian learning scheme Independent of the particular forecast performance, which is later evaluated in Chapter 7, the described Bayesian learning mechanism stands out by virtue of inherently advantageous characteristics that stem from its mathematical foundation:

- **Efficient:** New observations can be considered iteratively without the threat of state-space explosion (as for example in Bertsimas and Perakis, 2006) or the burden of extensive data history (as is necessary in conventional regression).

- **Simple:** The update mechanism for the coefficients' and predictive distribution is mathematically simple and tractable, limiting the computational effort – even for larger sets of time series – to forecast them individually.

- **Online:** Information is incorporated online or possibly in real time as it becomes available, without the need for preceding data collection. Through the usage of conjugate priors, the update mechanism can be automated and operates with limited supervision.
- **Robust:** Bayesian regression models are robust against over-fitting or bias, i.e., if data does not contain relevant information for a particular coefficient, its partial posterior distribution will be set equal to the prior – the same is true in the case of missing data points.
- **Flexible:** The Bayesian scheme explicitly allows for the consideration of prior information that is relevant to the model; that is, if substantial information is available, coefficient estimation may not start from scratch (or based on a noninformative prior) but can incorporate such data. Additionally, if only a limited amount of observations are available, the estimates are not as massively biased as in conventional regression.
- **Intuitive:** The predictive distribution, which basically substitutes the point forecasts from conventional regression, is an intuitive vehicle to assess forecast accuracy or certainty: its breadth (i.e., its variance) provides direct guidance on the level of confidence reasonable for a particular value prediction.

Key questions for performance evaluation While the described Bayesian online learning mechanism is appealing based on its characteristics (as described above), three questions remain for the final evaluation of its factual performance in terms of forecast accuracy:

1. **Convergence:** Is the model effective, i.e., does the coefficient distribution converge against the supposed true (in retrospective) values? (see Section 7.1.1)
2. **Quality:** Which quality level can be reached with the computed forecast after sufficient time for convergence? (see Section 7.1.2)
3. **Prior information:** What is the impact when considering possibly available prior demand information on the forecast quality? (see Section 7.2.1)

Before Chapter 7 finally gives answers to the above questions by looking at the computational results, the next section briefly describes how overarching demand structures and prior information can actually be included in the described Bayesian mechanism.

6.3.2 Overarching Demand Structures and Prior Demand Knowledge

Two additional input factors can be included in the described Bayesian model, simply by adjusting the employed data input – on the one hand in the amount of included deterministic observations, and on the other hand in terms of the information level contained in the employed stochastic prior distribution.

The amount of available data or observations can be controlled by considering adjacent flight demand referring back to the detected overarching demand structures from Section 5.2.3, while the initial information level is controlled by the specific form of the coefficients' prior distribution used.

Learning windows As discussed in the previous section, Bayesian regression principally allows for a separate modeling of individual flight events, as it is barely biased in cases where the amount of available observations undershoots the count of employed basis functions $T < M$ (see Bishop, 2006).

However, Section 5.2.3 showed that there exist demand similarities between adjacent flights that might potentially enhance forecast accuracy, as the consideration of such observations enhances the data basis (increases stability of the regression) and moreover exhibits a timely offset (advanced learning effect).

The technical implementation within the described learning mechanism from Section 6.3.1 is rather intuitive. Still, for each individual flight, a separate regression model is employed whose coefficients are learned over time based on recorded observations. However, the considered input data may not only be composed of observations for the particular flight that is learned, but may also be taken from within a learning window centered around that flight to include input data from adjacent departure dates. Figure 6.6 illustrates the resulting effects:

- **Advanced learning:** Depending on the size of the learning window k (i.e., number of adjacent departure dates included) and conditioned on a particular observation day, adjacent flights that depart earlier or later have henceforth reported k more or less observations, respectively, than the actual flight under forecast. Thus, the corresponding $t \pm k$ observations carry timely lagged information with a considerable projection. For example, if an "adjacent" flight departs one week earlier, its demand observations carry information about the effect of advance purchase time of a whole week (the same applies for weekday effects).

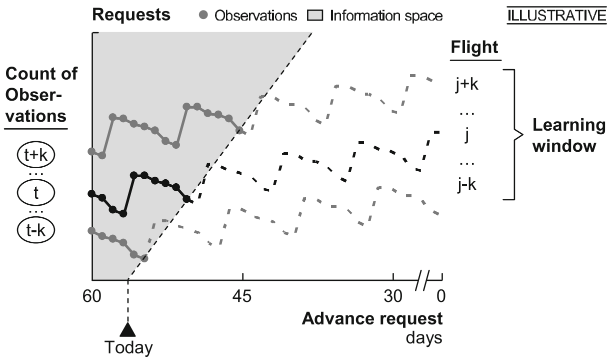


Figure 6.6: Learning window effect on information capture
Source: own design

- **Increased stability:** Depending on the size of the actual learning window, multiple data points are collected and included into the model at once on a single observation day. Through the window size, the contained information increases considerably. For example, the data could contain information on the effect of different departure weekdays, despite being collected on a single day.

The final impact on forecast performance of the discussed implications still has to be determined in Chapter 7. While stability is an overall desirable property for the coefficient distribution, excessive stability may hamper necessary adjustments in the case that individual parameter impact is changing over time, thus worsening the forecast quality. Additional problems may arise when the learning window is too wide. Naturally, the model will consider and thus adjust its parameters to observations that are too far from the flight event under forecast in the sense that the obtained information is misleading and may thus also worsen forecast quality.

Optimal impact on forecast performance should result from rather small learning windows with a maximum of two week distance from actual flight under forecast, while the positive effect is probably most visible early in the forecast horizon, when data from the actual flight under observation are still scarce.

The computational evaluation in Chapter 7 includes performance results for various sizes of the learning window together with a concrete recommendation for its particular choice.

Prior knowledge The earlier introduced forecasting model based on a noninformative prior relies on current real-time information only and thus represents an extreme counterpart to the conventional demand or booking forecasts common in the airline industry, which typically fully rely on historic information. However, the Bayesian framework is prone to considering available prior information from either the analyst's experience or such historic data. In this sense, the model allows for a combination of both worlds, i.e., online learning of model parameters while starting from an informed prior based on established knowledge.

To include existing information on the coefficients, their initial prior distribution from (6.19) has to be modified in two ways: first, the partial mean corresponding to the particular coefficient where additional information is available has to be adjusted. Second, depending on how certain that adjusted mean seems, the coefficient variance has to be modified to decelerate adjustments through the online learning process.

The described effect is exemplary shown in Figure 6.7 based on two forecasts for the same flight date as in Figure 6.5, which were both derived after the observation of 10 data points but with different specifications for the

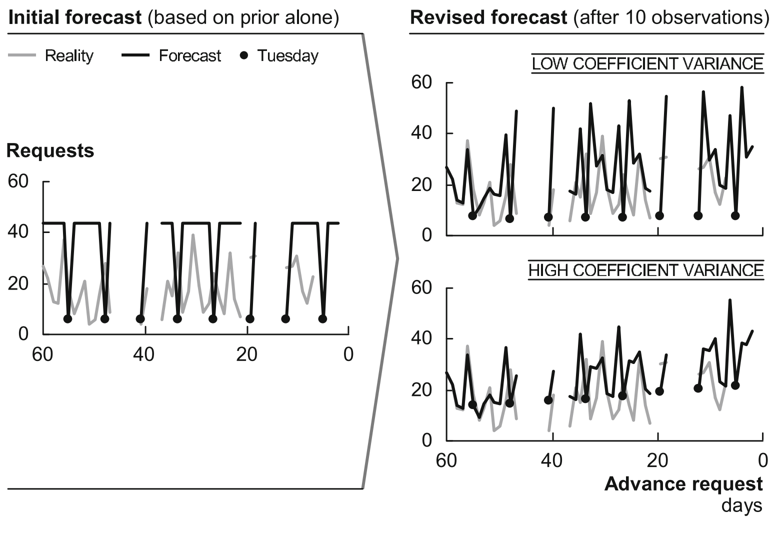


Figure 6.7: Prior knowledge under varying coefficient certainty
Source: own design based on collected data

prior distribution. For illustrative purposes, the effect of Tuesday request articulations has been predefined in the initial prior with an extremely negative impact ($a_{10} = -2.0$). To slow down the online learning effect in the figure's upper forecast, the coefficient's variance has been set close to zero ($\sigma^2_{10,10} = 0.0002$). Thus, an erroneous amplitude is visible on Tuesdays even after 10 observations have been fully considered. In the lower forecast, the variance is kept at the higher level ($\sigma^2_{10,10} = 0.2$), i.e., the prior information is considered uncertain and henceforth quickly corrected by the learning mechanism based on the online-collected data.

Figure 6.8 depicts the online learning effect on the mean of the partial prior a_{10} over the available observations. If the prior information is classified as certain (indicated by a low variance) the coefficient is adjusted only marginally following each observation (barely visible in the chart), although it is wrongly specified to be negative. However, if the prior distribution's mean is considered uncertain (indicated by a high variance), the true coefficient moment is quickly learned.

The above example illustrates the power of knowledge inclusion in the prior distribution. A high confidence in wrong driver parameters may trigger inferior forecast performance and vice versa. The model evaluation in the following chapter revisits the effect and its possible impact on forecast accuracy based on prior information about overarching demand trends.

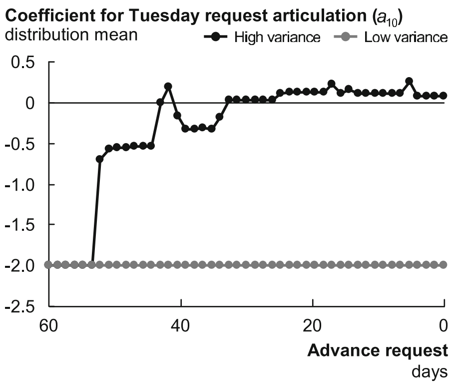


Figure 6.8: Coefficient learning under varying prior confidence
Source: own design based on collected data

After the functionality of the online learning mechanism has now been described together with its advantages and possible enhancements, the next chapter provides the computational results of the model in terms of its forecast performance, depending on different levels of prior information and varying sizes of the employed learning window.

Chapter 7

Computational Results and Evaluation

Following the introduction of the Bayesian self-learning forecasting scheme underlying this work (see previous Chapters 4 – 6), this chapter now provides the computational results and takes a look at the overall predictive performance of the model in Section 7.1 as well as its sensitivity to using informative priors, changing learning window sizes and different forecast granularities in Section 7.2.

Again, the results in this chapter are entirely based on a Bayesian perspective: observed data is considered deterministic, while the functional model composition (i.e., the coefficients' distributions) itself is assumed to be stochastic.

7.1 Performance of the Naïve Bayesian Scheme

This section explores the model's basic forecast performance under naïve or noninformed learning: first, Section 7.1.1 evaluates the convergence speed of the coefficients' distributions, i.e., how quickly the learning mechanism picks up a particular influence on the considered demand drivers. Second, after allowing for sufficient convergence time, the resulting forecast accuracy is evaluated in Section 7.1.2.

7.1.1 Distribution Convergence Speed

As a starting point, this section evaluates the convergence or learning speed of the Bayesian forecasting scheme described in Chapter 6, that is, how quickly the estimates of the model's coefficient means converge to sufficiently precise

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6_7,

values. Naturally, the derived forecast then also converges, so as soon as the learning effect on the individual coefficient distribution parameters has reached a satisfactory level, the absolute forecast accuracy can be evaluated based thereon.

Here, the objective is to determine the necessary lead time in the number of observations that have to be considered until forecasts can meaningfully be evaluated against the true latent demand in the next Section 7.1.2.

To evaluate the convergence speed of the plain Bayesian estimation scheme, a naïve or noninformed learning approach (see Section 4.4) is employed, which does not leverage any prior information on the coefficient means. That is, the initial individual coefficient prior distributions are centered around a zero mean with large variances. The key question to be answered then is:

> *How many observations from within the forecast horizon need to be considered until the posterior distribution means have sufficiently converged?*

The answer to above question requires the constitution of two benchmark figures: first, individual *convergence targets* for the coefficients' means. Second, an overarching *threshold* for the convergence being identified as "sufficient".

In a real-world setting, the factual target (i.e., the true latent demand) is only known after having observed the full forecast horizon, i.e., in retrospect. However, even in such a full-information setting, the model's results will typically not exactly match in all data points for two reasons, which have already been discussed in Section 4.1: stochastic variations (i.e., $e \sim \mathcal{N}(0, \sigma^2)$) and principal model deviation (i.e., $f(\mathbf{x}, \mathbf{a}) \approx \tilde{f}(\mathbf{x}, \mathbf{a})$), as a model is never an exact image of reality. Thus, an upper bound for the goodness that a predefined model can reach is given by its coefficient distributions after *all* finally available information from the full data set has been processed.

Here, the convergence targets for the coefficient distributions are defined solely on the distributions' mean, as the objective at this stage is not to assess the certainty of the forecast but merely to determine the specific point in time where an evaluation is adequate based on the amount of considered information. As soon as the distributions' means have moved sufficiently close to that full-information benchmark, a forecast evaluation is possible, independent of the variance present. However, the variance then gives an indication on the certainty or confidence of the derived forecast. If the observed data do not follow the assumed model structure, the forecast result will be uncertain, which is indicated by a high variance.

Throughout the total 60-day learning period, the coefficient distributions' means continuously change and converge towards their final values, which are eventually reached after all available information has been considered. Here, this convergence shall be considered *sufficient* when the resulting posterior mean is within a 95% confidence interval around the full-information mean.

As both, the final mean and its surrounding confidence interval, can only be calculated in retrospective, the sought-after threshold cannot be calculated online during individual forecasts. Thus, the data subset from Section 6.1.4 is used to derive typical global thresholds based on retrospective knowledge. Using these uniform thresholds, the overall model performance is evaluated for individual prospective forecasts over the entire dataset in Section 7.1.2.

Figures 7.1 and 7.2 on pages 162 and 163 depict the convergence speed of the coefficient distributions' means, calculated collectively based on data of all departure dates within the data subset from Section 6.1.4 (i.e., a single model is estimated for all departure dates within the dataset). Here, the Bayesian learning mechanism does not consider single observations incrementally (as in the described online forecast model), but is simultaneously fed with all observations for the same observation date.

In the charts, the continuous gray line shows the particular coefficient's retrospective distribution mean (i.e., under full information) and the corresponding 95% confidence interval (gray dashed lines). The dark continuous line depicts the online learning progress of the distribution means, which finally converge against their retrospective targets in all cases. The vertical lines mark the stage where the estimated mean falls within the 95% confidence interval for the first time (the circled number indicates the precise amount of necessary observations).

Apparently, most coefficient means converge fairly quickly, with the dummy variables "request articulated on Monday/Tuesday" or "public holiday" showing the longest lags for the outbound model (see Figure 7.1) and "flight departure on Friday", "request articulation on Monday/Sunday" or "public holiday" exhibiting the need for considerable training time for the inbound model (see Figure 7.2).

Notably, there appears to be no development in learning for the early observations on "request articulated on full weeks multiple" and "request articulated on public holiday", whereby this is because the Bayesian scheme does not update its expectation on a particular distribution mean until the first observation containing relevant information has actually been recorded. In the specific case, there are no observations made for a public holiday until the 44^{th} observation is reached.

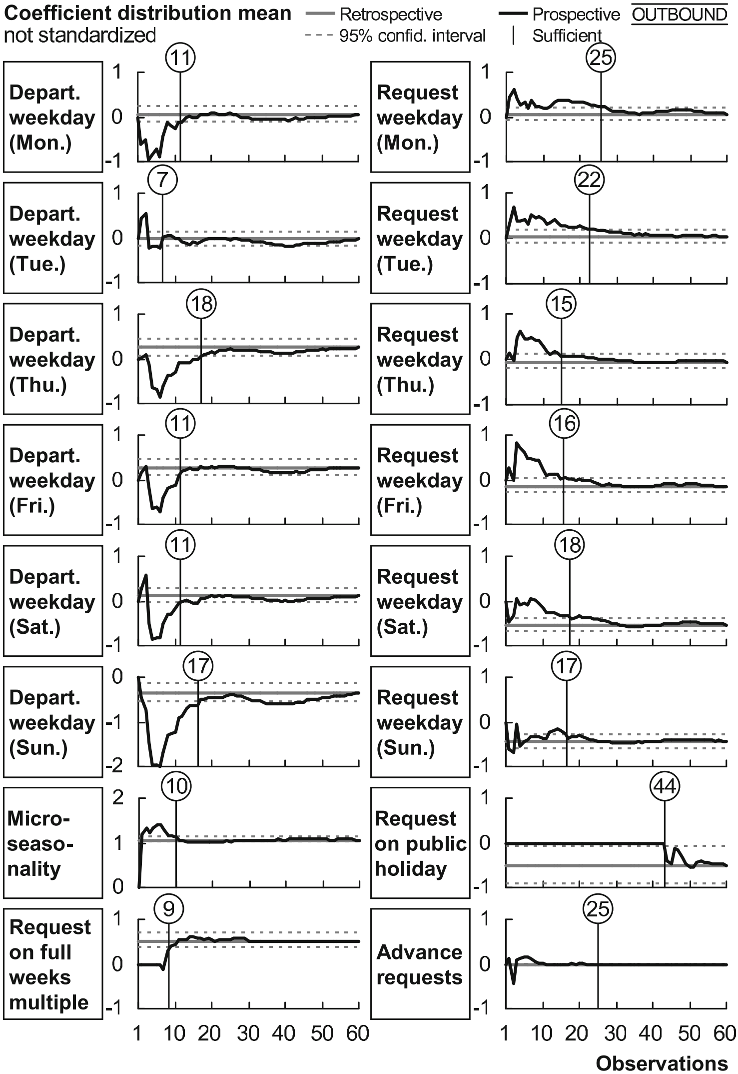


Figure 7.1: Convergence of regression coefficients – outbound
Source: own design based on collected data

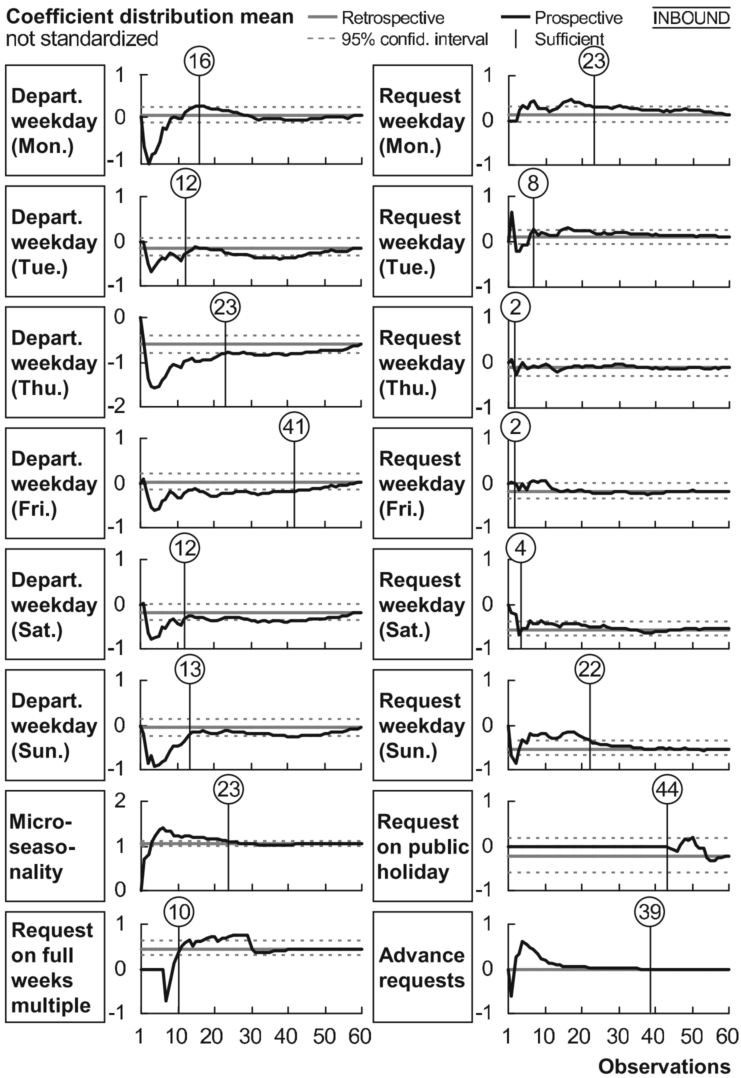


Figure 7.2: Convergence of regression coefficients – inbound
Source: own design based on collected data

Based on the convergence results from Figures 7.1 and 7.2, a uniform threshold of 15 observations is chosen for the evaluation of the models' forecast performance in the next section. That is, the coefficient distributions of the flight-individual models are updated 15 times using the Bayesian learning mechanism described in Chapter 6. The resulting distributions are then used to calculate the predictive distributions for the specific models (according to Section 4.3.2), which are finally compared against the true latent demand values along the entire forecast horizon, i.e., the full 60-day time series. However, in an operational model, learning would never stop as the accuracy of the model's coefficients would still improve with every additional observation until finally the day of departure.

7.1.2 Forecast Quality and Accuracy

The objective of this section is to assess the Bayesian model's forecast performance under a naïve or noninformed prior, or more specifically, to determine whether the learning process of the scheme produces credible results after the consideration of 15 data points only.

To finally calibrate the performance of the model, it is important to note the particular properties of the generated forecasts and the underlying mechanism: based on the collected data (see Section 5.1) 382 *flight-individual forecasts* are generated (one per departure date and direction) with *daily granularity* over a *60-day forecast horizon* each using a training set of only *15 data points*. The model is thereby based on an *uninformed or naïve prior* (i.e., there is no reliable prior or historic information assumed about demand behavior) and includes *no learning from adjacent flights*. Notably, the training of the model itself takes place *without manual intervention* and is hence *fully automatic* at this point.

The overall model validation in Section 6.2 has shown that $\approx 80\%$ of total variation in the collected data can be directly explained by the model in a full-information, retrospective setting. The reminder is due to stochastic variations or factors not included in the model.

While the initial validation of the model structure (see Section 6.2) has included only a subset of the data, the full evaluation of the Bayesian forecast after all 60 observations have been considered (i.e., in retrospective view) naturally exhibits analogous results (see Figures A.1 and A.2 in the Appendix).

These full-information results now form the basis against which the further forecast results are evaluated, i.e., against which the model is supposed to have converged after the consideration of the first 15 data points.

For the evaluation of the forecast performance, two different relative error measures are used throughout this chapter:

- **Centered Mean Absolute Percentage Error (sMAPE):** First of all, a measure for the accuracy of granular point predictions is necessary. It should account for the possible deviation of individual day-specific forecasts from the real observations independent of the direction of deviation (overestimation or underestimation). For true latent demand D_{jt} at a particular departure date j articulated t days in advance and with corresponding predictions y_{jt}, the sMAPE is defined as

$$sMAPE_j = \frac{1}{T} \sum_{t=1}^{T} \frac{|D_{jt} - y_{jt}|}{\frac{1}{2}(D_{jt} + y_{jt})}. \tag{7.1}$$

  Here, sMAPE is chosen over the uncentered MAPE, as it is less sensitive to small prediction errors for low base values (see, e.g., Hyndman, 2006 or Makridakis and Hibon, 2000). Notably, in the observation dataset, $\approx 50\%$ of all data points (daily articulated demand for a single departure date) exhibit a latent demand $\leqslant 20$, for which the smallest uncentered error necessarily would be 5% (as demand is discrete).

- **Total Absolute Percentage Error (TAPE):** In addition to the measure of point estimate precision, a more general rating for the total error per departure date over the full 60-day forecast horizon is nec-

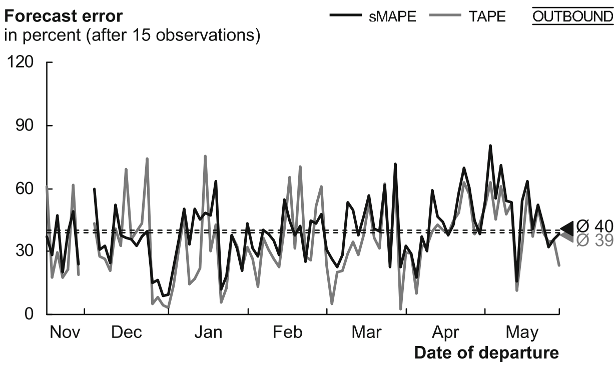


Figure 7.3: Forecast error along flight departures (sMAPE and TAPE) under noninformative learning – outbound
Source: own design based on collected data

essary to assess the level of long-term deviation in addition to pure short-term stochastic variations.

Therefore, the following measure has been newly defined to calculate the forecast deviation in terms of *total latent demand-to-come* for an individual departure date's time series j as

$$TAPE_j = \frac{\left|\sum_{t=1}^{T} D_{jt} - \sum_{t=1}^{T} y_{jt}\right|}{\sum_{t=1}^{T} D_{jt}}. \tag{7.2}$$

Figure 7.3 plots both error measures for the naïve model across the full dataset for the outbound direction.[1] The goodness of neither error measure seems satisfactory, as both average $\approx 40\%$ and also fluctuate heavily. The forecast for the inbound direction in Figure 7.4 shows similar behavior.

The unsatisfactory performance of the forecasts or the model can be explained by looking at one of the major demand drivers, namely "advance request" (i.e., the remaining time to departure).

Section 5.2.1 has identified an overall log-linear demand structure – that is, demand grows exponentially (but linearly in the case of logarithmized demand) towards departure. While this effect has been included in the forecast model (see Section 6.1), the previous section (Figures 7.1 and 7.2) has already

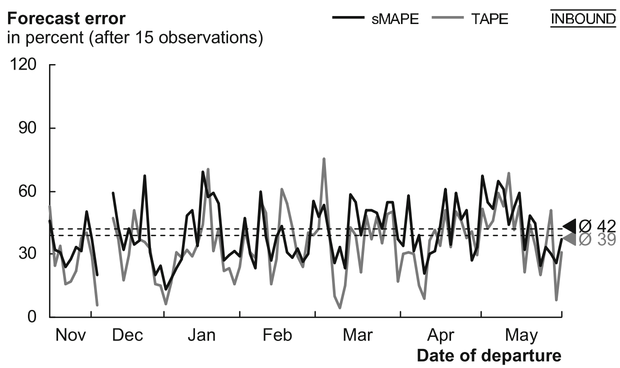


Figure 7.4: Forecast error along flight departures (sMAPE and TAPE) under noninformative learning – inbound
Source: own design based on collected data

[1] The two error measures are plotted as continuous lines over all departure dates for simplicity and readability, instead of as individual points or bars.

indicated that the convergence speed of this particular parameter is rather slow. However, what has not been possible to answer concisely is whether after consideration of only 15 observations, the still limited convergence will already produce credible forecast results – which seems not to be the case.

For the outbound direction, Figure 7.5 shows a possible segmentation of departure dates according to their demand growth structure.[2] Evidently, most groups indeed exhibit strong exponential growth, but which may set in rather late. Moreover, at an earlier point in the 60-day period (60 to 45 days to departure), flight dates exhibiting strong growth may behave similarly to flights with moderate growth at that time.

A few groups also do not exhibit continuous growth during the low season or off-season but instead show a slight demand dip around 30-16 days before departure, which might be due to flexible travelers holding off demand in anticipation of last-minute price reductions.

The model validation (see Section 6.2) has shown that the Bayesian learning scheme is principally capable of capturing the growth trend in the data.

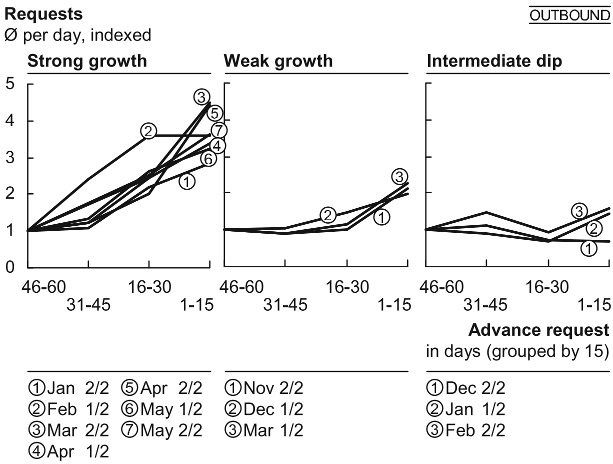


Figure 7.5: Distinct types of demand growth trends
Source: own design based on collected data

2 For readability and simplicity, the data have been grouped by half-months so that weekday effects do not overlay the overall growth effects.

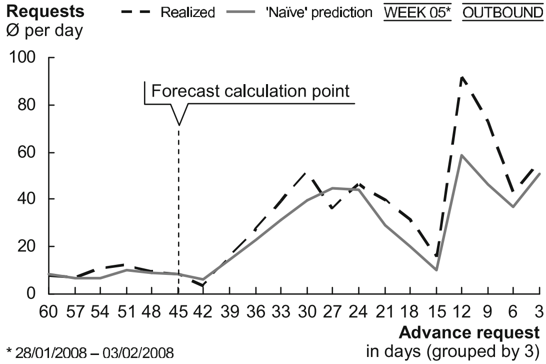


Figure 7.6: Capture of trends through noninformed learning methods – outbound (example: calendar week 05, 2008)
Source: own design based on collected data

Indeed, Figure 7.6 shows an exemplary case of successful learning about a strong demand trend even early in the learning process.

However, a deeper analysis of the demand forecasts resulting from the above naïve or uninformed learning reveals that in nearly all cases with high forecast error, strong and undiscovered demand trends are responsible for the massive defects (especially in TAPE). For example, Figure 7.7 in Section 7.2.1 below shows an exemplary case where such a wrongly predicted trend yields roughly 75% deviation at the end of the forecast horizon.

The forecasts yielding the results in Figures 7.3 and 7.4 are based on pure full automatic learning, starting from a naïve prior distribution for all coefficients, as this defines the most challenging case and does not require manual intervention (e.g., for the adjustment of calendar effects or trends). While the presented model seems limited in terms of its forecast accuracy, it allows for a variety of amendments to improve its performance. The possible effects of such model extensions on forecast accuracy are discussed in the next section.

7.2 Sensitivity of Forecast Accuracy

The objective of this section is the assessment of the forecast accuracy's sensitivity to changing the conditional framework of the model. In the preceding section, specifically the late recognition of exponential growth trends inherent to the data has been identified as the major source of forecast error.

This section now explores model extensions that may improve accuracy. First, Section 7.2.1 evaluates the model's sensitivity to the injection of additional knowledge in the form of an informed prior. Second, Section 7.2.2 illustrates and evaluates the effect of larger training samples (based on the findings from Section 5.2.3). Third, Section 7.2.3 shows the effect of lower granularity for the forecast target, and finally, Section 7.2.4 takes a look at possible combined effects.

7.2.1 Improvement Through Informed Priors

The above results from Section 7.1 on forecast quality are based on a naïve learning process that, for each coefficient, starts from a sufficiently broad prior distribution (i.e., an uninformative one) that does not itself contain any previous information on its likely final form (see Sections 4.2 – 4.4). This approach is especially appealing because it does not influence or bias the coefficient learning, and the results are therefore purely dependent on the observed data (i.e., objective).

However, Bayesian updating does provide a systematic way to include *subjective information* or data into the learning process (see Section 4.4). The inclusion of such prior knowledge may help the model to appropriately assess the likelihood of skewed observations (see Example 4.2 in Section 4.2.1) and, moreover, allows for the anticipation of influences that are known to be present but may not yet be reflected in the data (see Section 6.3.2).

While Figure 7.6 has shown an example where the naïve model is capable of picking up a rather strong demand trend early in the process, Figure 7.7 below depicts the more common case in which the overarching exponential demand trend takes off rather late in the forecasting horizon (here ≈ 35 days before departure) and can therefore not be appropriately predicted by the model after only 15 observations. At the same time, if the goal is ultimately to provide a 60-day forecast, it does not seem feasible to train the model for full 25 days until the convergence of the trend parameter is sufficient.

The dashed line in Figure 7.7 represents the average latent demand in observed calendar week 12, 2008, and the gray line depicts the forecast derived from naïve Bayesian learning using an uninformed prior at 45 days to departure. While local oscillations (e.g., at around 36 days before departure) are predicted correctly, the overall trend is apparently derived from the preceding flat demand development and then extrapolated towards the end – yielding $\approx 75\%$ point error.

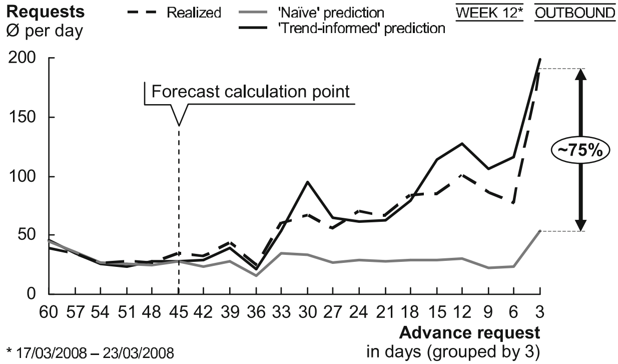


Figure 7.7: Capture of strong trend comparison – trend-informed vs. non-informed learning methods (example: calendar week 12, 2008)
Source: own design based on collected data

To showcase the effect of *subjective learning* and to improve forecast accuracy by anticipating the overall trend, the full model has been re-run using an informed prior solely for the trend parameter: all other coefficients are treated as before (uninformed), while the mean value μ for the trend coefficient's prior distribution ($a_1 \sim \mathcal{N}(\mu, \sigma^2)$) is set to values following Table 7.1.

The values in Table 7.1 serve as prior knowledge that has been fictively derived from historical information, management knowledge or similar and therefore should entail considerable certainty or confidence, as is reflected in a lowered distribution standard deviation ($\sigma = 0.001$). Nevertheless, over time, real observations can still affect coefficient development through the

Departure dates		Prior μ-values	
Start date	End date	outbound	inbound
Nov 16, 2007	Jan 31, 2008	0.15	0.15
Feb 01, 2008	Feb 15, 2008	0.35	0.35
Feb 16, 2008	Mar 15, 2008	0.15	0.15
Mar 16, 2008	Mar 31, 2008	0.35	0.35
Apr 01, 2008	Apr 30, 2008	0.25	0.25
May 01, 2008	May 31, 2008	0.35	0.35

Table 7.1: Prior-values for trend-informed learning

Bayesian updating scheme if reality deflects from the assumed prior.

To prevent over-fitting through the usage of retrospective information, the values in Table 7.1 have been derived by taking a very rough cut of aggregated demand development, looking at larger groups of departure dates by half-months. This approach mimics the knowledge that management fictively could posses about expected overall demand development: *'Demand growth is rather slow in the off-season before Christmas, but heavy during Carnival at the beginning of February, the Easter holiday season in March and the May bank holidays.'*

Figure 7.7 above shows the effect of the described purely *trend-informed* learning on forecasting performance for calendar week 12, 2008. Apparently, the informed scheme is able to capture the overall trend while still correctly anticipating the micro variations.

Taking an overall look at the total forecasting performance under a trend-informed prior along Table 7.1 (see Figures 7.8 and 7.9), one sees that the total error has significantly improved, especially for peak season departure dates. In addition, the point error (sMAPE) has stabilized. Notably, both errors still exhibit high overall levels and significant variation during the end of December and February, which is induced by demand dips during the forecast horizon (see right-hand chart of Figure 7.5), which cannot be predicted by the model, even in combination with an informed prior.

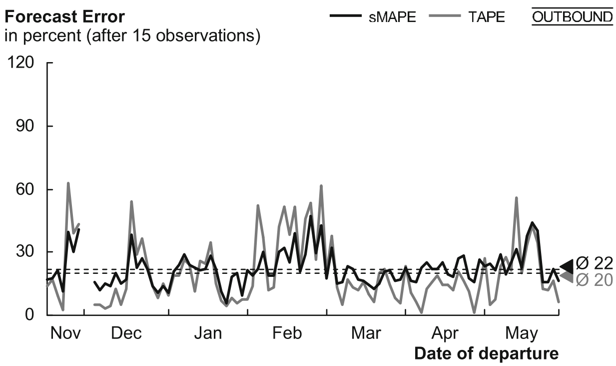


Figure 7.8: Forecast error along flight departures (sMAPE and TAPE) under trend-informed learning – outbound
Source: own design based on collected data

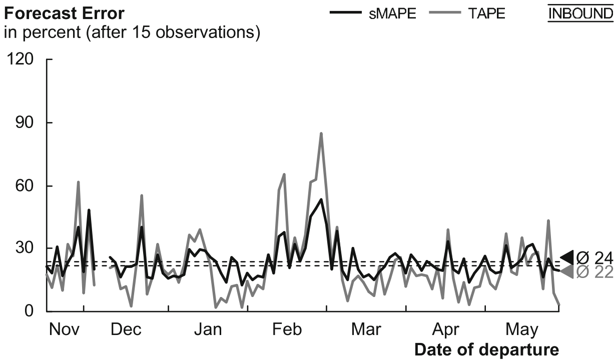


Figure 7.9: Forecast error along flight departures (sMAPE and TAPE) under trend-informed learning – inbound
Source: own design based on collected data

Although both errors have already been significantly reduced through the use of an informed prior for the trend parameter, the following section now takes a look at an alternative measure to capture trends early in the learning process by introducing a wider training data set.

7.2.2 Sizing of Learning window

Up until this point, the Bayesian updating or training has used only demand data explicitly recorded for the particular departure date under consideration (see Section 7.1), which naturally hinders the learning process because the factual model training cannot start before the 60-day forecast horizon is reached within the collected data.

At the same time, Section 5.2.3 has already highlighted the apparent similarities of adjacent flights in terms of latent demand behavior, which are not yet leveraged by the model – especially with respect to the overall growth trend towards departure. This section now evaluates a methodology for using information from adjacent or neighboring flights as proposed in Sections 3.4 and 6.3.2 – that is, additional data are used to train a naïve model, while the generated forecasts still remain flight-specific.

Figure 6.6 in Section 6.3.2 illustrates the possibly advantageous effect of taking data from neighboring flight departures as input for a particular departure date's forecasting model. The preceding flights enter the relevant forecast horizon with a considerable time lag that is dependent on the size of the learning window – e.g., a time series of a flight that departs 14 days in advance of the particular flight under consideration is already 14 days into the 60-day forecast horizon when the very first observation is collected for the latter. In particular, the forecast of varying and delayed exponential growth trends, which has been difficult to deal with in the preceding sections, could benefit from exploiting such time lags.

On the downside, data from adjacent departure dates might contain misleading information that could also negatively affect the learning process. In addition, as the definite demand nature of the particular days under observation can barely be known in advance, there exists no satisfactory basis for an informed selection of sufficiently "similar" dates.

Because of this lack of information, here no single flights are picked to constitute the learning window; instead, complete ranges centered on the particular departure date under forecast are tested. This approach mirrors the findings from Section 5.2.3 that flights departing soon before or after one another exhibit related demand development. To compensate for and incorporate the possibly diverse effects of differing departure weekdays, the model already contains a corresponding set of dummy variables (see Section 6.1.3).

Figure 7.10 exemplary shows the total forecast error for departures in calendar weeks 48/49, 2007 for models trained using varying learning window sizes of ±3 days, ±1 and ±2 weeks in comparison to a naïve model without additional learning. Both weeks are part of the low demand season before Christmas, with slow demand growth towards departure. Here, in both cases, the expected effects from Section 6.3.2 are clearly visible:

1. Large window sizes (±1 and ±2 weeks) already yield relatively low error levels after very few observations. The subsequent forecast improvement thereafter is rather slow.

2. Smaller windows (±3 days and ±1 week) tend to catch up with a lag of a few observations, depending on the window size (the smaller the window, the longer the training time needed to improve the error). Eventually, smaller windows seem to outperform larger ones because less (possibly misleading) information is used to train the model.

3. The usage of no adjacent flight information leads to relatively late catch-up in terms of the error level. However, after 15 observations, the non-learning forecast nearly outperforms all others, independent of their specific learning window size.

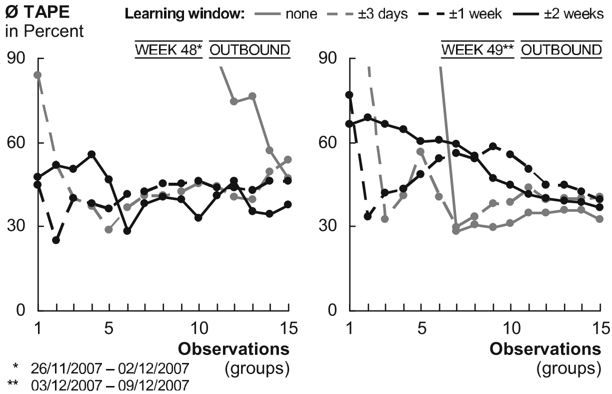


Figure 7.10: Expected effect of different learning window sizes on forecast error (TAPE) using noninformed priors
Source: own design based on collected data

Section 7.1 above has reported massive errors for naïve models in the case of strong demand growth where said growth is not clearly visible in early observations. Figure 7.11 shows two departure weeks in April 2008 that exhibit this type of strong demand growth towards departure. Here, the use of broad learning windows can apparently accelerate the training process substantially, which leads to an enduring improvement in forecast error.

For the considered departure weeks, the broader learning windows (±1 and ±2 weeks) draw additional and time-lagged information about the strong overarching demand trend from neighboring observations apparently exhibiting similar behavior. This additional training basis leads to strong improvements with regard to forecast error only a few observations into the forecast horizon (especially for the ±2-week learning window).

Because of the still-changing (i.e., further-increasing) demand at the error calculation point (after consideration of 15 observations), the naïve forecast is still heavily outperformed by the models leveraging broad learning windows with a corresponding look ahead.

However, in the context of changing trends, the described learning effect can be adverse. Figure 7.12 exemplary shows the forecast performance of different learning window sizes for calendar weeks 10/11, 2008, which contain the last off-season departure dates before the strong demand growth of the Easter season sets in. Here, the larger windows (±1 and ±2 weeks) pick

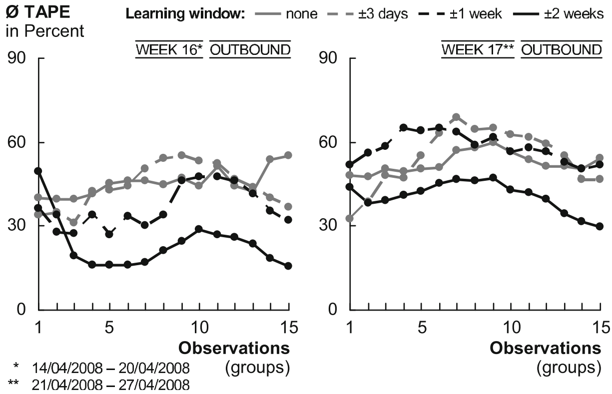


Figure 7.11: Peak-season effects of learning windows on forecast error (TAPE) using noninformed priors
Source: own design based on collected data

up the changing trend from later departure dates too early and thus, the calculated forecasts deviate massively from the true observations.[3]

Figure 7.12 illustrates that in so-called shoulder-seasons, smaller learning windows or even naïve forecasting can be beneficial. In both given weeks, the naïve scheme exhibits the best performance from the start onward, only partially undercut by the smallest learning window of ±3 days.

To conclude the above discussion of exemplary cases, a holistic forecast evaluation of naïve learning with broader training windows that include data from adjacent departure dates is necessary.

Figure 7.13 provides a full glance at the performance of differently sized learning windows for the outbound demand forecast. For the first 15 observations, the heat map on the left side color-codes the best performing learning window sizes (grouped by departure week). The bar graph on the right side illustrates the resulting average TAPE after consideration of 15 observations for training, together with the error improvement gained over the performance of the naïve scheme through the usage of additional learning windows.

[3] In Figure 7.12, the error line for the learning window with ±2 weeks exceeds the chart's scale (0–90%) and is therefore not shown. However, the scale has not been adjusted for chart readability and its comparability with others.

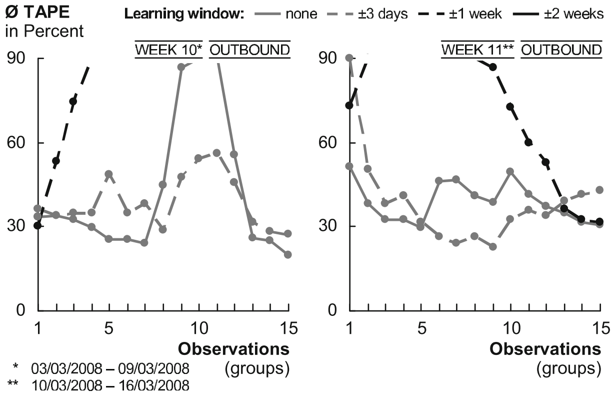


Figure 7.12: Border-season effects of learning windows on forecast error (TAPE) using noninformed priors
Source: own design based on collected data

The following three key findings are easily derived from studying the results in Figure 7.13:

- **Differentiation:** After a full training period (i.e., the consideration of all 15 observations), in case adjacent learning is beneficial, the extreme learning window size mostly performs best: either the large ±2-week window or the usage of no additional learning window at all.
- **Seasonality:** The performance of the largest learning window is heavily dependent on demand seasonality. It can only benefit from clear and persistent demand trends (e.g., the low season before Christmas, the high season around Easter and the May bank holidays).
- **Effectiveness:** The ±2-week window yields notable improvements in total error whenever its usage is beneficial, resulting in a reduced average TAPE of 26% when selecting the best performing learning window breadth after 15 observations.

Similar findings can be derived from Figure 7.14 for the inbound directed demand. Here, the differentiation is even stronger, and the overall improvements are highest in the peak season starting in mid-April.

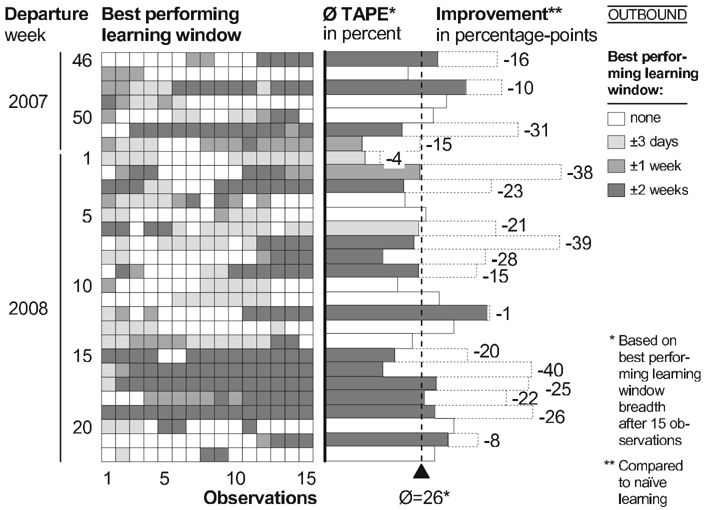


Figure 7.13: Performance of learning window sizes over training period with total error improvement – outbound
Source: own design based on collected data

The presented forecast performance improvements are already substantial. However, the resulting average total errors, at around 26%, still represent insufficient adoption and seem high compared to the results achieved using an informed learning approach (see Section 7.2.1). However, the learning windows all use naïve or noninformative priors without the need for subjective management intervention; henceforth, a possible combination of learning windows and informed learning is later evaluated in Section 7.2.4.

The discussed learning window approach, as well as the usage of an informed prior distribution in Section 7.2.1 above, have both aimed to derive early insights into the growth trend of individual time series to yield substantial improvements in total error. The next section now examines an additional measure that can merely improve performance with regard to individual forecast deviation.

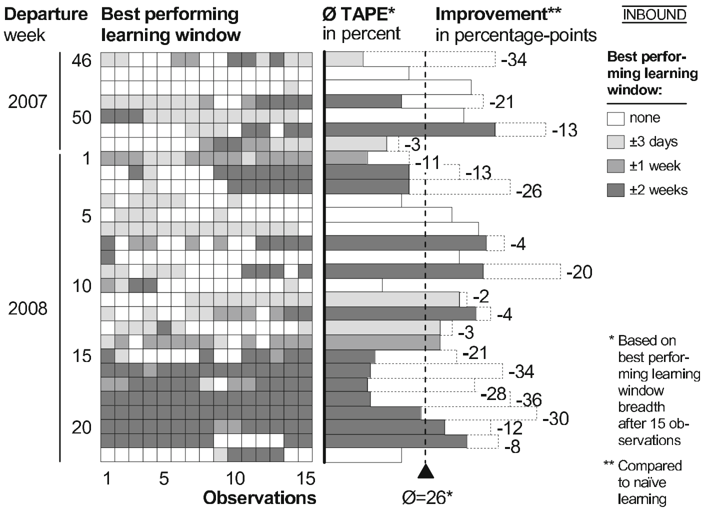


Figure 7.14: Performance of learning window sizes over training period with total error improvement – inbound
Source: own design based on collected data

7.2.3 Granularity of Forecasting Basis

Section 7.1.2 has shown that the pure naïve forecasting approach yields considerable point deviations (sMAPE) that may not be solely attributable to the late recognition of the overall growth trend but could also be due to external influences and stochastic variations in the individual time series. While the preceding techniques have already partially provided improvements to the sMAPE, this section now explicitly takes a leap in dealing with such stochastic variations.

A relevant and common measure used to filter out stochastic variations in the forecast target is to reduce its level of granularity. Naturally, higher data aggregates are easier to forecast because stochastic fluctuations might counterbalance each other.

Above sections have aimed to provide 60-day forecasts of daily demand at a specific departure date. To raise the aggregation level, either the dimension when demand is articulated (i.e., advance request time) or the dimension when demand is exercised (i.e., departure date) could be aggregated.

Based on the apparent structure of demand in low-cost markets (see Chapter 5), multiple reasons in favor of aggregating along the departure date dimension (i.e., when demand is exercised) emerge:

1. Neighboring departure dates might represent valid alternatives or could even be considered a single alternative to the customer (e.g., Thursday or Friday departure for a weekend trip); here, an aggregation could well compensate for corresponding stochastic fluctuations.

2. For dynamic pricing to be effective, prices need to be adjusted to changing demand and reservation prices in a timely fashion. Therefore, maximal granularity of the time dimension where demand is articulated is essential.

3. The structure of latent demand (see Section 5.3.3) shows the considerable effect of the particular weekday where demand is articulated, which is therefore also reflected in the model. A possible aggregation along that dimension would blur this distinctive variable.

To evaluate the effect of reduced data granularity on forecast performance, a new aggregated forecast target is generated following the above rationale: Instead of forecasting point demand for a selected departure date, one attempts to determine average demand for a range of departure dates centered around a particular day. That is, a 60-day forecast is still derived, yielding the expected average latent demand for each of the 60 days until departure for a window with size k around the departure date j under forecast (see Figure 7.15 for an illustration).

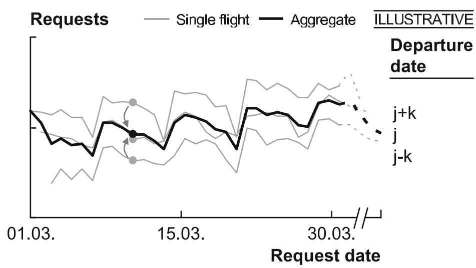


Figure 7.15: Illustration of forecast target aggregation scheme
Source: own design

Figure 7.16 depicts the resulting improvements in forecast error when using the naïve prior from Section 7.1.2 in combination with aggregated targets. Here, the pure aggregation effects seem rather small and limited to improvements in sMAPE, as mainly stochastic fluctuations in point demand are mitigated.

Unsurprisingly, the positive effect of balancing point deviations is heavily overlaid by the not-predicted exponential demand growth trend over the forecast horizon. As the aggregation scheme purposely maintains the 60-day forecast granularity, this trend and its late recognition by the naïve forecasting scheme are not affected by the aggregation. This also explains the constriction of the effect to being mainly a reduction of sMAPE only.

The above sections have shown that the model's forecast performance reacts sensitively to varying the prior information base (naïve vs. informed distributions) of the model (see Section 7.2.1), extending the observation basis using learning windows that include adjacent observations (see Section 7.2.2) or aggregating the forecast target to reduce stochastic fluctuations (see this Section 7.2.3).

All measures have individually proven to be beneficial, mostly either in capturing demand trends early in the forecast horizon or in improving point accuracy. Therefore, the next section examines a possible meaningful combination of these methods that will ultimately yield the best possible forecast performance in terms of both error dimensions, sMAPE and TAPE.

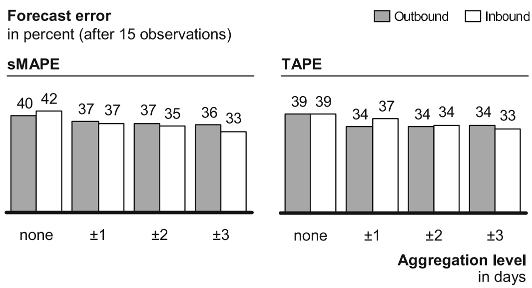


Figure 7.16: Improvement of forecast errors on reduced target granularity
Source: own design based on collected data

7.2.4 Combined Effects

After Section 7.1 has shown that a pure self-adjusting and online learning Bayesian scheme does not provide satisfactory forecasts after the consideration of only 15 observations, Sections 7.2.1 – 7.2.3 evaluated different model extensions to improve the model's accuracy, which mostly affected either the total error or the individual point deviation.

This section now combines the above findings into a single model to finally recommend a specific forecast methodology, as detailed in the next Section 7.3, that provides the best results in terms of both error dimensions. Thus, the objective here is to select the appropriate measures from the above sections that, in combination, will result in *maximum error reduction* (compared to Section 7.1) and *balancing positive effects* on total error (TAPE) and point deviation (sMAPE).

Figures 7.17 and 7.18 report the average forecast errors for all possible method combinations from the above sections over the full dataset, assuming that one particular combination suits all departure dates.

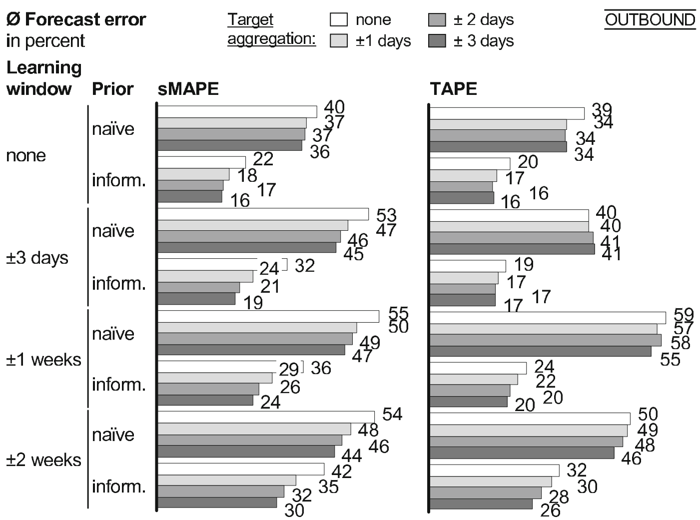


Figure 7.17: Forecast error for combinations of improvement techniques – outbound

Source: own design based on collected data

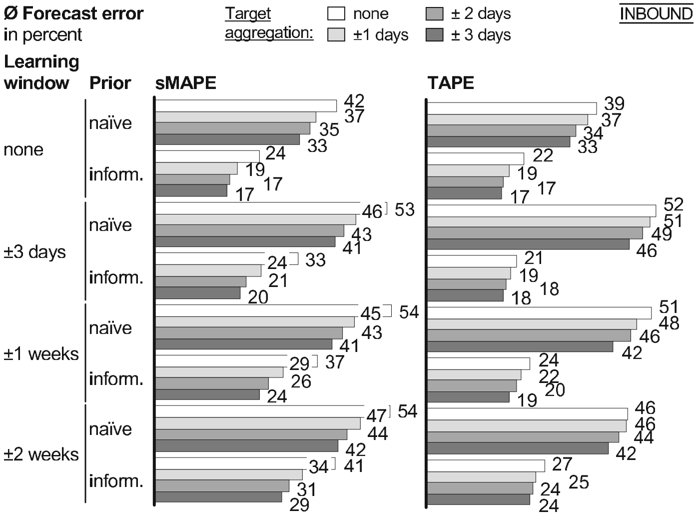


Figure 7.18: Forecast error for combinations of improvement techniques – inbound
Source: own design based on collected data

Obviously, the forecast schemes employing an informed prior as in Section 7.2.1 heavily outperform the naïve schemes. The sMAPE also exhibits expected improvement, with increasing aggregation levels based on the usage of learning windows or an informed prior. Only the different learning window sizes do not favor an obvious best selection, especially on the side of TAPE, which is most affected by varying learning window sizes (see Section 7.2.2).

Tables A.1–A.8 in the Appendix report the detailed error results for all method combinations and departure dates, which have led to the aggregated results in Figures 7.17 and 7.18.[4] Both tables visualize in detail the findings from above while also giving a clear indication that the performance of the different learning window sizes depends heavily on the prevailing seasonality – i.e., whether demand growth trends are persistent over adjacent dates (similar to the results from Section 7.2.2).

Therefore, a sound forecasting scheme should vary the employed learning

[4] The provided tables are amended with color shades highlighting the best, top-3 and top-5 method combinations per departure date to foster better and intuitive readability.

window size while keeping the aggregation level constant (not least to provide steady data granularity for the actual usage of the forecast) and should invariably employ an informed prior.

The detailed results in the Appendix and the findings from Section 7.2.2 give clear indications of seasons, where diverging learning windows can yield significant forecast improvements. Table 7.2 reports the particular selection that has been made based on a weighted combination of errors from Tables A.1–A.8.[5]

While the particular learning window selection does rest on retrospective error results, it has been chosen based on rough period cuts and with an eye toward external demand drivers (like vacations, religious holidays, and so forth). Henceforth, it should not be overly fitted to the particular results and should, moreover, be reproducible for different datasets.

The selection in Table 7.2 extends the results from Section 7.2.2: periods where an informed prior does not provide enough information regarding the time series' trend development (as over the Christmas period or in the February low season) benefit from broader learning windows. Notably, the transition between periods is again steep, mostly switching directly between none and the large ±2-week learning window.

To evaluate the overall goodness of the combined forecast, a benchmark is generated from the results in Tables A.1–A.8: For each departure date,

Direction	Start date		End date		Learning window	Remarks/events
Outbound	Fri	11/16/2007	Sun	12/16/2007	none	
	Mon	12/17/2007	Tue	01/15/2008	±2 weeks	Christmas vacation
	Wed	01/16/2008	Fri	02/08/2008	none	Carnival
	Sat	02/09/2008	Sun	03/02/2008	±2 weeks	Low season
	Mon	03/03/2008	Thu	03/20/2008	±3 days	Start of Easter season
	Fri	03/21/2008	Sat	05/31/2008	none	May bank holidays
Inbound	Fri	11/16/2007	Wed	12/19/2007	none	
	Thu	12/20/2007	Tue	01/15/2008	±2 weeks	Christmas vacation
	Wed	01/16/2008	Tue	02/12/2008	none	Carnival
	Wed	02/13/2008	Tue	03/04/2008	±2 weeks	Low season
	Wed	03/05/2008	Sat	03/22/2008	±3 days	Start of Easter season
	Sun	03/23/2008	Sat	05/31/2008	none	May bank holidays

Table 7.2: Selected learning window sizes by departure period

[5] The differences in forecast error between learning window sizes have been evaluated with weights for sMAPE/TAPE ranging from 30%/70% – 50%/50% to reflect the elevated importance of reasonable total time series error.

the particular method combination that produces the most accurate forecast (i.e., the black-shaded combination in the appendix tables) is selected.

Naturally, the recorded lower bounds are an artificial measure depicting the case in which it would have been possible to select the best-performing method combination at the beginning of the forecast horizon, even if the particular choice shows no clear structure and alternates heavily between methods (as is the case here).

Figure 7.19 reports the resulting average forecast accuracy of the described lower bound benchmark in comparison to the performance of the particular learning window selection defined in the above Table 7.2.

Obviously, an individual combination of the Bayesian model from Chapter 6 and the introduced extensions from Sections 7.2.1 – 7.2.3 can yield a high level of forecast accuracy when the best-performing methods are continuously combined – which would, however, not be feasible in real-life settings.

Still, even a mostly fixed model selection with periodically changing learning window sizes, following Table 7.2, can yield a high point accuracy and a satisfactory total error level. Based on the analyzed dataset, a sMAPE of $16-17\%$ and a TAPE of $11-13\%$ are reported, with deviations mostly resulting from individual cases where single departure dates still exhibit a different demand growth behavior than expected by the prior that can also not be learned from adjacent flights.

Notably, these results are obtained from automated learning amended with very rough prior information and similar selection of appropriate learning window sizes; they do not result from artificial over-fitting of the model.

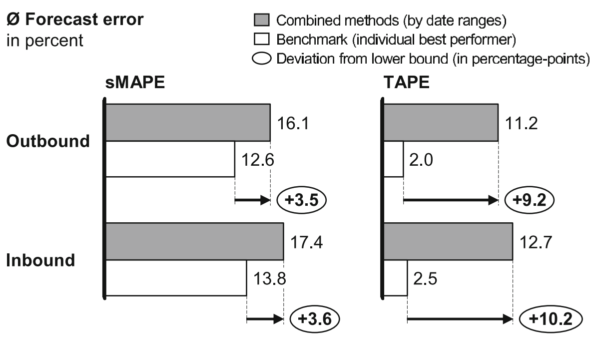


Figure 7.19: Forecast error comparison – final model vs. lower bound
Source: own design based on collected data

This section has surveyed the performance of the Bayesian forecast model when amended with a combination of the improvement methods introduced in Sections 7.2.1 – 7.2.3 and has reported respectable results. The next section now attempts to derive more general recommendations for the generation and employment of such Bayesian schemes to forecasting demand in dynamic markets.

7.3 Recommended Approach

Building on the previous section's reports and discussion of the specific results of the analyzed dataset, this section aims to draw more general conclusions for Bayesian forecasting in similar dynamic markets as the examined air travel market.

While Section 7.1 has reported a quick and accurate convergence of most model coefficients, even a Bayesian model does not include foresight into demand development in the future. It would seem that its performance heavily depends on relevant information either a) to be contained in the particular training data (naïve Bayesian learning process), b) to exist in substitute data that can serve as an extended training set (the novel learning window scheme) or c) to be externally introduced into the model by a pre-specified management prior (informed learning).

In the considered case of demand in dynamic LCC markets, the time-dependent trend parameter has suffered largely from low information value early in the training process, leading to slow convergence of the corresponding model parameter, which is finally reflected in unsatisfactorily high total error levels.

Taking an influential time parameter that exhibits exponential demand growth as given in a dynamic low-cost market, the following guidelines for the creation of a well performing Bayesian learning scheme can be derived from the results:

- **Mandatory usage of informed learning:** The usage of an informed prior distribution for the time parameter heavily improves both sMAPE and TAPE (see Section 7.2.1), which is a general finding independent of any additional techniques used (see Figures 7.17 and 7.18). Naturally, the performance improvement depends on the accuracy of the employed management prior but already reaches notable levels even if used on rough assessments of likely values. Additionally, by construction, the

Bayesian mechanism will update and correct heavy misjudgments, as the latter will continuously exhibit low likelihood given the data.

Moreover, the described positive effects of informed learning cannot be substituted by the employment of learning windows – independent of their size.

- **Selective employment of learning windows:** The punctual extension of an informed learning scheme with learning windows can exhibit an additional positive effect on the total time series deviation (TAPE) and even on the detection of micro-variations. Resulting performance depends on the prevalent macro-season exhibiting a stable growth trend. In such cases, the use of large learning windows can support the identification of inherent demand growth trends, and the enhanced data basis can serve to further calibrate the prior knowledge.

  However, the selection of appropriate learning window sizes should be monitored and adjusted continuously. Changing macro-trends can heavily worsen results under large learning windows, as deceptive information may be falsely included in the training sample (see Section 7.2.2).

- **Careful choice of aggregation level:** As expected, the aggregation of the forecast target as noted in Section 7.2.3 has a generally positive and beneficial effect on point deviation (sMAPE) because stochastic variations and demand deflections may balance each other out. Naturally, the aggregation level does not significantly affect the total absolute error (TAPE).

  The relative error improvement decreases with enlarging aggregation levels, and possible benefits should therefore be weighed against the resulting loss in forecast granularity. Ideally, the chosen level accounts for the successive usage of the forecast. If the results are to be used in a merely qualitative way – i.e., to educate agents with regard to manual price intervention, lower granularity may be sufficient. However, automatized pricing systems often need a higher degree of data granularity as input basis.

Unfortunately, the composition of a Bayesian demand forecast model still needs considerable manual adjustments and cannot reach peak performance levels without close control and according intervention. Nevertheless, the approach entails major advantages over traditional forecasting, which is often purely based on historic data and time series extrapolation.

The following chapter now gives an outlook on this train of thought and comprises a concluding summary of Part II on forecasting latent demand in low-cost markets.

Chapter 8

Summary and Outlook

Preceeding Chapters 4 – 7 developed, validated and finally evaluated a Bayesian forecast model for latent demand in low-cost air travel markets; that is, uncensored demand – unbiased by pricing decisions – is forecasted directly online from deterministic input data.

The current chapter concludes Part II of this work, giving a final summary of the findings and results of its chapters as well as an outlook for possible further research.

Model Description The derived model is capable of providing forecasts for 60-day time series with daily granularity for individual departure dates within a horizon of the next two months or (in the case of aggregation) for individual groups of three, five or seven departure days within the next two months' time (but still with daily granularity in the time series).

The overall time series structure is found to be log-linear in deterministic input variables (linear in logarithmized basis functions) and is deemed to be fixed in its particular selection of input functions. However, based on the employed Bayesian regression methodology, the structural composition may vary through its dynamic coefficients, which are assumed to be stochastic and to follow a multivariate normal distribution.

As a first step, no prior knowledge about the particular parameter values of these coefficient distributions is assumed, which is indicated by the employment of naïve or uninformed prior distributions – that is, the initial parameter distributions are assumed to be heavily tailed and centered around zero.

The employed Bayesian regression mechanism continuously updates these distributions of all model coefficients through the incremental consideration of daily observations regarding the relevant input variables. In this way, the model learns from its changing and dynamic environment, as its train-

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6_8,

ing relies exclusively on current or real-time data. Moreover, no extensive history of preceding demand development needs to be stored. Only through the optional usage of so-called informed management prior distributions does additional anterior information get included in the model; otherwise, it relies exclusively on online information and real-time data.

The coefficient distribution means are found to adjust and converge quickly within a 95% range of their final (full-information) value after the consideration of 15 observations from a naïve or uninformative initial prior. The convergence – which happens synchronous with the retraction of the distribution's variance – requires the recorded observations to carry sufficient and relevant information about the particular input variable or basis function whose coefficient is to be learned. If such information is entirely missing, the distribution does not change; in cases where the information is scarce, the convergence is slowed considerably.

In the presented log-linear model, for pivotal input values, the effect may seriously skew the results, especially in terms of total time series error. However, in such cases, the model can be amended in two ways:

- **Informed prior:** Instead of the particular distribution parameters of the affected coefficient being learned from scratch, existing management or other anterior knowledge can be systematically included in the model by employing an informed distribution in place of the initial naïve prior. That is, existing knowledge is transformed into a likely location parameter for the prior distribution with considerably reduced variance to indicate that it represents an informed choice and therefore is more certain than the uninformed one before.

  However, the Bayesian updating mechanism will still update and adjust the informed distribution when new information arrives. If relevant data renders a different distribution parameter more certain, the prior distribution is updated accordingly – i.e, the model corrects heavily deflecting historic information automatically.

- **Learning window:** The standard Bayesian model only leverages training information that has been explicitly collected for the time series of the departure date under forecast. While this naturally results in forecasts that are tailored to a particular departure day, it neglects possibly relevant information from neighboring dates. Latent flight demand on adjacent dates is often heavily interrelated as these represent valid alternatives to the customer. Henceforth, their time series may carry valuable information about demand development, while being time-lagged (time series of earlier departure dates are more advanced in terms of remaining time to departure).

The newly employed learning windows for the Bayesian forecast model take advantage of this effect by enlarging the data basis that is employed for training the model, while keeping the actual forecast target (a full time series for a particular departure date).

Both measures described can be employed separately or in combination, as the individual effects may complement one another to produce improved forecast results.

An additional method used mainly to reduce point error is the aggregation of forecast targets, as individual stochastic deviations may balance each other out within higher aggregates. In the described model, data is aggregated according to the date of demand exertion (i.e., flight departure) – that is, the daily granularity of the 60-day horizon for which demand is articulated is kept, while demand is forecasted for a range of three, five or seven days centered around a particular distant departure date within a two-month future.

Ultimate Usage The structure of the derived online learning model and its forecast results are general enough to provide input for three different cases that are simultaneously relevant to the literature, as well as practically useful:

- **Optimization input:** Depending on the chosen granularity of the forecast, the model results can serve as direct input for most optimization models in the academic literature.[1] In this spirit, the model closes prevalent gaps in the literature where authors typically assume the existence of highly granular (up to micro-periods) forecasts over extended time horizons (see Section 3.2.2). In these cases, the results may plug directly into existing optimization models, completing the existing overall dynamic pricing approaches to actual usability.
- **Alternative demand learning:** While recent literature has acknowledged the need for integrated approaches that incorporate forecasting of latent demand or online learning regarding demand, most of them assume particular stochastic types for demand or employ specific learning

[1] Particular models might not allow the actual usage of the generated forecasts, as the underlying assumptions and prerequisites only allow for selected stochastic types of latent demand.

schemes – neither of which might fit reality (see Section 3.2.1). In such cases, the developed model may act as a replacement for the employed learning or forecasting schemes, as it is conceptually flexible, allowing for a variety of functional demand structures (through the usage of varying basis functions, e.g., the described log-linear structure).

- **Management insight:** In addition to fully computerized pricing engines, today many companies still heavily rely on manual price intervention and configuration, as well as direct adjustments of these systems or of the offer price. This results, at least in part, from the deficiencies of automated systems when faced with realistic environmental conditions. Here, in addition to indicating the pure forecast itself, the model may provide a structural understanding of latent demand.

  In the considered example, the structural information about how the particular weekdays where demand is articulated affect its development can provide valuable insight (e.g., when rating weekend bookings).

The above list highlights three distinct areas where the developed model could be employed favorably. Naturally, the items are not segregative but may rather be used in conjunction with one another – potentially even amplifying each other.

Model Advantages The general Bayesian demand forecasting model developed in this portion of the work offers significant advantages over existing online learning schemes and specifically over conventional time series models. These advantages mostly come into effect in highly dynamic markets with considerable limitations on the availability of historic data (like the low cost travel market):

- **Functionally broad:** The underlying structure of the derived Bayesian regression model is linear in its basis functions and hence is highly flexible and easy to adapt to differing demand behaviors. While, in the presented case, demand is identified to be log-normal, the variation of the employed basis functions makes the model functionally flexible and adaptable to other forms of demand.

- **Computationally efficient:** The presented Bayesian regression scheme is efficient in its application in two ways. First, very few data are needed to sufficiently train the model (especially in comparison to conventional regression models); in the specific case, only 15 observations were needed to initially train the model. Second, the model accepts new observations incrementally and only stores the current state of the model coefficients in order to run the next regression cycle – not the entire data history (as in conventional regression).

- **Dynamic/self learning:** Through incremental training as previously described, the model continuously adapts to the current – possibly changing – environment. In each cycle, the previously learned coefficient or model state is evaluated for plausibility and certainty, depending on recent observations, and then possibly adjusted if the collected information indicates higher likelihood with regard to a differing state.

- **Controlled subjective:** Naturally, Bayesian learning or updating provides a systematic way to include subjective or management information before or throughout the training process. The presented model adopts this scheme by starting from objective and broad prior distributions for all model coefficients and then selectively incorporating subjective information for specific coefficient distributions (i.e., for the trend parameter).

- **Enlarged training basis:** Based on the structural findings regarding demand in dynamic low cost markets, the training of the proposed model is not necessarily restricted to direct observations collected specifically for the forecast target but may also be comprised of additional information from adjacent events. Using this enlarged field of data, the model can significantly benefit from possible time lags of such observations. Collected data from the latter may contain information that is time-shifted depending on the particular time lag of the event and henceforth may expedite the training process.

In addition to the listed major advantages of the particular derived model, Bayesian learning schemes in general also provide an intuitive and mathematically comprehensible approach to automated machine learning (in comparison to neuronal networks, for example).

Performance and Accuracy The model's overall performance is assessed in two parts. First, based on the model's functional composition and the particular selection of input variables (i.e., basis functions), its performance and global fit are evaluated based on the fraction of total variance in the data that the model is capable of capturing. Second, the convergence speed of the individual forecasts towards this global lower bound benchmark is rated to assess the quality of early demand forecasts that are necessarily based on a limited amount of online observations.

The evaluation of the full model against the entire collected dataset of roughly 15,000 data points shows that the derived functional composition is capable of explaining $\approx 80\%$ of total point variance in the data. In terms of average total error for the individual 382 time series (one per departure date and travel direction) within the dataset, the model's accuracy is higher, yielding only $8-14\%$ error as individual stochastic fluctuations within the series may balance each other out.

The evaluation of the Bayesian learning mechanism employed with the model revealed that a training period of 15 observations should suffice for the majority of the coefficient parameters to converge within a 95% confidence interval against their full-information value. Naturally, this convergence rate depends on the information density of the considered data – i.e., for satisfactory convergence, the observations need to convey the most relevant information on the particular coefficients being trained.

However, fully automated learning over 15 data points in particular is not found to acquire enough information to sufficiently adjust the time-dependent trend parameter (advance request time). This insufficiency results from the data-inherent exponential demand trend, whose magnitude is insufficiently visible before an average of 25 observations have been considered to train the particular time coefficient. For a total time frame of 60 days, the necessary training time would be roughly half of the entire forecast horizon; thus, being not a feasible solution.

To substitute for the missing information, the model is amended with an informed prior distribution for the affected coefficient, which as single standing measure already yields significant improvements to the forecast accuracy. The evaluation of the particular enhancement, i.e., the selection of crude prior distribution means with reduced variance (to signify the increased confidence in the prior information), already reports an improved average total error per departure time series of $20-22\%$ and an average point deviation of $22-24\%$.

The research unveils that this sort of informed forecast model can benefit further from the use of additional supplemental input data throughout the training period. The consideration of observations from adjacent departure dates yields significant error improvements in cases where persistent demand growth trends draw from consistent macro-seasons, i.e., where adjacent dates exhibit sufficiently similar demand development.

While an ideal selection of such learning windows can result in extremely low total errors of $2-3\%$ with considerable point forecast accuracy as low as 12–14%, the prospective selection of the appropriate sizes is challenging. For the final evaluation of the model, learning window sizes have been selected based on predictable macro-season indicators like holidays and special events (e.g., Carnival) to yield realistic results and prevent retrospective over-fitting.

In general, the employment of large learning windows seems primarily beneficial in periods when strong demand trends persist over a considerable span of time, specifically allowing time-lagged learning of the inherent trend parameter.

Naturally, higher aggregation of the forecast target improves point deviation, as truly stochastic deflections and errors may balance each other out within the aggregate. Indeed, the analyzed model shows such effects when switching from point forecasts of latent demand for a specific departure time series to forecasts for a range of three, five or seven departures (centered around the specific departure date under observation).[2] However, the magnitude of the reported effect heavily depends on the specific combination of employed model amendments, i.e., the particular usage of the informed prior and learning window size.

The final model leverages a combination of crudely selected prior distributions for the trend parameter, selective employment of additional learning windows and a target aggregation up to a ±3-day range; it then exhibits an average total error for the departure time series of 11.2−12.7%, with an average point deviation of 16.1−17.4% across series (see Figure 8.1).

Both errors are considerably close to the reported lower error bounds for the full-information model. Figure 8.1 also summarizes the model's performance under individual method amendments in comparison to the model's theoretical lower bound.

Outlook and Further Research Although revenue management in general and dynamic pricing in particular have been the subject of intensive academic study and research, there still exist rather few price optimization models that are based on realistic assumptions with factual applicability to real-world settings. Moreover, mathematically optimal price-setting does not necessarily result in customer-accepted pricing schemes.

Moving forward, there still is ample need for further research moving those models towards practical usability and acceptance. Self-learning models for the required inputs of such price optimization models are a fundamental step in that direction, but they can only form a necessary basis for further model improvements. Optimization models that are proven to result in increased

[2] The specific aggregation scheme has been chosen because neighboring days may be considered valid travel alternatives, such that demand deflections are balanced.

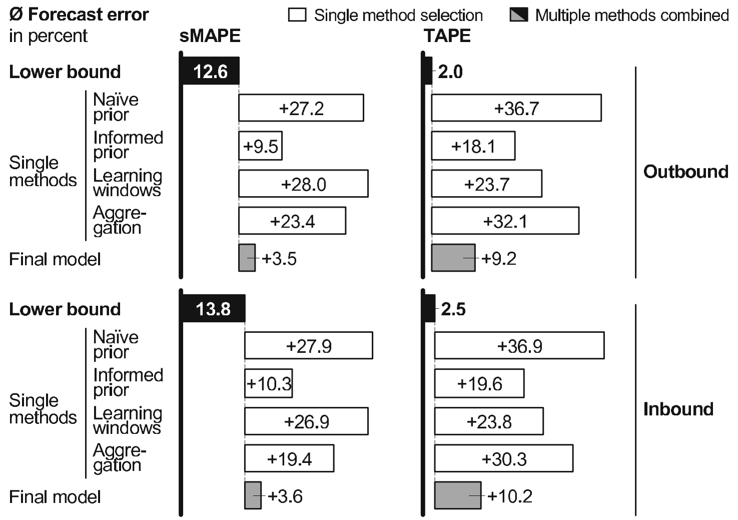


Figure 8.1: Model forecast performance under tested method amendments compared to lower bound
Source: own design based on collected data

revenue and profit under realistic conditions are still scarce, especially for highly dynamic and price-sensitive markets like low-cost travel.

The described model results still leave room for extensions and improvements of the forecast model for latent customer demand – e.g., through the inclusion of additional, possibly external factors. In addition, the forecasting of demand provides only a portion of the necessary input for most price optimization models. Equally important is the forecast of time-variant price sensitivities or reservation price distributions. Part III, which follows this chapter, provides an introductory view of the topic and a possible solution based on customer choice analysis.

Regarding the specific model derived in this part, the above chapters have traced some specific areas for improvement:

- **Learning window selection:** Learning window selection in the above scenario has been based on anterior knowledge and the manual interpretation of holiday seasons and specific events. A further improvement to the model could be made through the invention of a (possibly also self-learning) algorithm for selection and continuous adjustment of the appropriate learning window size without the need for manual intervention.

- **Prior construction:** Similarly, the specific construction of informed priors based on management knowledge could potentially benefit from systematic support – that is, in terms of mapping qualitative and subjective knowledge (*'peak season typically exhibits strong short-term and last-minute demand'*) into concrete parameters to yield an informed prior distribution for the affected parameters.

- **Inclusion of associated industries:** The model has been tested and evaluated based on data from a low-cost air travel market. Demand should exhibit dependencies on associated industries like the hotel, hospitality or rental car industry. A model extension to include data from these business areas to create a further broadened learning window could potentially increase forecast accuracy, especially because related demand in such industries again might be time-lagged.

Ultimately, research in the area of price discrimination will continue to benefit from new opportunities that the ongoing development of the Internet and new online business models provides; it cannot be considered complete until corresponding optimization models are seen as a commodity in every marketing department in suitable industries.

Part III

Estimating Price Sensitivity

Part III Objective

Following the introduction to *dynamic pricing* and the *airline industry* in Part I, and the proposition of a forecasting model for *latent customer demand* in Part II, this part's objective is to develop a specific model to reveal the drivers behind customer choice behavior in low-cost air transportation markets. The final objective here being to understand the price-dependent functional composition of *realized customer demand* based on the underlying latent demand and the fraction of purchasing customers.

Most research on dynamic pricing implicitly assumes the existence of such models, but only a few studies explicitly specify mechanisms to obtain such functional relationships (see Chapter 2). The model developed here is intended to be general enough to fit into most dynamic pricing models from Section 3.2, with the additional advantage of having been developed and tested using *real demand data* from a *highly dynamic market*.

In particular, this approach distinguishes itself from existing ones by the following characteristics, which are in conjunction new to the literature:

1. The approach derives a functional relationship between the expressed latent demand D and the actually realized demand d in parametric form, whereas that relationship's dependence on competitive dynamics is explicitly acknowledged. That is, realized demand is defined as the share of latent demand that actually realizes $d = \omega \cdot D$, where the latter ($\omega = \omega(\mathbf{p})$) depends on the vector $\mathbf{p}$ of relevant prices of competing products (internal and external).

2. The developed customer choice model is based on revealed preference data that are constructed based on automatic data collection processes from the perspective of a single carrier. Henceforth, the resulting models can be updated frequently without the need for expensive and time-consuming manual surveys, which allows the results to be adapted reasonably quickly to changing market dynamics.

3. A proprietary data basis composed of three different sources (internal and external to the analyzed airline) is used, which creates a unique and exhaustive view of the dependencies between latent demand (based on request logs), realized demand (based on booking records) and the pricing environment (based on fare skimming).

4. The conventional multinomial logit model is extended toward its universal representation to deal with the inevitable choice data restrictions and modeling challenges based on the automatically collected data.

5. The analyses are based on the full period spanning the last 60 days before actual flight departure, as $\approx 75\%$ of demand arrives within that period, what makes it the relevant study subject.

6. No manual interventions (e.g., for calendar effects) or data enhancements are employed to obtain a realistic and unbiased impression of possible computational results based on pure system-sourced data.

Nevertheless, the developed model inevitably yields forecasting errors, as not all external effects and customer characteristics can be included that, in reality, drive purchasing decisions.

Chapter 9

Discrete Customer Choice Analysis

The following chapter introduces the methodology used in this part of the work to formulate a model of the price sensitivity of customers within the low-cost travel market that is based on real-world data. Following many other works on choice analysis (see below Table 9.1), the method of discrete customer choice analysis is employed here to model and understand customer purchasing behavior at a disaggregated individual decision maker level.[1]

Discrete choice modeling is rooted in the works of McFadden (1974, 1975), for which he was eventually awarded the Nobel Prize in 2000 (see McFadden, 2000b). A brief introduction to the topic focusing on use-oriented aspects can be found in Gönsch et al. (2008a,b). For a thorough introduction beyond the specific model types covered in this chapter, the reader is referred to the introductions of Ben-Akiva and Lerman (1985), Koppelman and Bhat (2006) and Train (2003). Specific reviews on experimental design for customer choice analysis can be found in Bierlaire (1997), Carson et al. (1994) and Hensher and Button (2000).

Section 9.1 briefly discusses the background and scope of customer choice models in general. Thereafter, Section 9.2 explains the elements of any generic choice decision process, based on which Section 9.3 discusses the different underlying theoretical concepts. Finally, Section 9.4 introduces the multinomial logit model that forms the basis for the later model-building in Chapter 11.

---

[1] Throughout Part III, the decision maker is referred to as "she" with the sole purpose of distinguishing it from the analyst or modeler, respectively, who is correspondingly referred to as "he". Both descriptions are by no means intended to discriminate based on gender.

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6_9,

Choice subject	Literature
Destination	Train (1998) Bhadra (2003)
Route	Yai et al. (1997) Cascetta et al. (2002) Erhardt et al. (2003)
Air carrier	Proussaloglou and Koppelman (1995) Proussaloglou and Koppelman (1999) Coldren et al. (2003)
Car brand	Hensher et al. (1992) Bhat and Pulugurta (1998)
Consumer goods/brands	Timmermans et al. (1992) Krishnamurthi et al. (1995) Kalyanam and Putler (1997)
Housing/location	Waddell (1993) Sermons and Koppelman (1998)
Urban travel mode	Train (1978)

Table 9.1: Literature on customer choice models by choice subject (excerpt, chronological)

9.1 Fundamentals of Choice Modeling

The discussion of fundamental pricing models in Chapter 3 of Part I has already shown that the forecasting of *realized demand* is of essential importance to any dynamic pricing scheme. While Part II looked at forecasting the quantum of people *latently interested* in air travel on a specific route, here the subject matter is the *actual choice behavior* of each such individuals when faced with a predefined set of choice alternatives.

The ultimate objective of discrete choice modeling – as of most econometric models – is twofold (Koppelman and Bhat, 2006, Chap. 1):

1. Predict and *understand the decision making behavior* or process of a group of individuals when choosing from a discrete set of available alternatives.

2. *Determine the specific drivers* (and their respective importance) behind individual decision outcomes, be they individual characteristics or attributes of the available alternatives.

In general, there exist two potential modeling solutions for the above objective: The modeling of an aggregate share behavior based on group characteristics (e.g., socio-demographics) or attributes of the available alternatives. This so-called *aggregate approach* forms the basis for the common multivariate statistical analyses (e.g., linear regression). Alternatively, methods using the so-called *disaggregate approach* realize that the individuals' influences and differences may affect aggregate or group behavior (see Ben-Akiva and Lerman, 1985, Chap. 1).

Discrete choice analysis is part of the second group as it aims at modeling the behavior of single customers based on their individual characteristics and their resulting evaluation of a set of alternatives.

The disaggregate approach yields a range of significant advantages: First, it *explains the reasons* behind individual behavior, possibly based on specific personal characteristics. This in turn makes the analysis and its results potentially more *transferable* to other points in time or datasets. The final model is also better suited for *proactive policy analysis* by the seller, i.e., it provides an answer to the question of how changes in alternative configuration affect individual choice decisions. Additionally, disaggregate model estimates have proven to yield *higher efficiency* than aggregate models based on equally sized data samples. Disaggregate models reportedly also show *less biased results* (see Koppelman and Bhat, 2006, pp. 2).

In brief, the principle objective in discrete choice analysis as introduced in the following sections is the explanation of individual choice from a set of mutually exclusive and collectively exhaustive (MECE) alternatives. Its methodology is based on the econometric utility maximization principle, that is, a decision maker rationally chooses the specific option with the highest personal utility among all available options.

An operational model structure is therefore based on a parameterized utility function consisting of independent variables describing the decision maker or the available alternatives and unknown parts of utility that either cannot be observed or are unknown to the analyst. The parameters of such a model are estimated based on a representative sample of decision makers and their articulated choices. Following the random utility theory, the true utilities are taken to be stochastic variables, and henceforth, the probability that an alternative is chosen is given as the probability that it exhibits the highest utility among all assessed alternatives in the function model (see Ben-Akiva and Lerman, 1985, Sec. 1.2).

The next section takes a closer look at the founding elements of an individual's decision process in the above sense.

9.2 Elements of a Choice Decision Process

This section formally defines the elements of a choice decision process based on which choice behavior theories are later defined in Section 9.3.

According to Ben-Akiva and Lerman (1985, Chap. 3.2), a "choice" itself is defined as the outcome or result of an individual's sequential five-stage decision-making process, as depicted in Figure 9.1. The aim of customer choice analysis is to understand and model such discrete choice processes in order to be able to predict outcomes under varying input scenarios.

Independent of its specific economic foundation and the consideration of uncertainty in the process, any particular choice theory is defined by a concrete specification of four constitutional elements, which are discussed in the subsequent sections: The definition of the decision maker and the description of its characteristics (see Section 9.2.1), the detection of the full available choice set (see Section 9.2.2), the evaluation and description of these alternative choices by a set of attributes (see Section 9.2.3) and the final definition of a specific decision rule for the decision maker to choose among the available choices (see Section 9.2.4).

It is worth noting that all possible behavioral assumptions about a decision maker – even random behavior, following intuition or a supposed opinion leader – can be expressed by the introduced four elements when defined correctly (see, e.g., Train, 2003, Chap. 2).

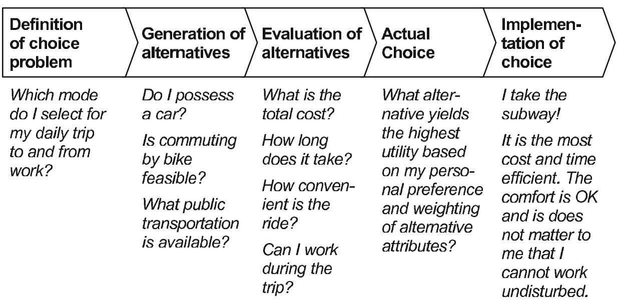


Figure 9.1: Choice decision process as framework for choice theories
Source: Visualization of Ben-Akiva and Lerman (1985, pp. 31)

9.2.1 Decision Maker and its Characteristics

The decision maker in each choice situation is the particular entity who is responsible for making the specific decision and whose characteristics, in turn, cardinally affect the decision process. The entity can hereby consist of a single individual (e.g., a traveler), a group of people (e.g., a household) or even an institution (e.g., a state or a company).

The selection of the appropriate decision maker entity depends on multiple criteria (see, e.g., Ben-Akiva and Lerman, 1985, Chap. 3 or Koppelman and Bhat, 2006, Chap. 2):

- **Affection:** As a first rule, the decision maker should reflect the entity that is responsible for making the decision and is subsequently also affected by it. For example, when selecting between different modes of transportation for the daily commute to work, the decision maker is typically the employee itself as she is the only one affected by the choice. On the other hand, when buying a new car, the decision maker could well be the family as a whole, because ultimately all members are affected by the decision – everybody will have to use the car they jointly decide to buy.

- **Characteristics bearer:** The decision maker should also be defined in such a way that the chosen entity carries the necessary characteristics to differentiate itself from others. Taking the car purchasing example from above, a potential determinant of choice is household income, which is linked to the family as a whole, but not to individual family members. Often it is possible to aggregate characteristics of single individuals within an entity to create overarching entity characteristics (as is the case with the mentioned variable of family income).

- **Level of abstraction:** Finally, the desired level of abstraction can help in defining the appropriate decision maker. At the end of the day, the actual decision maker will often be an individual person, possibly in interaction with others. However, in many cases the explicit decision process at the micro level (e.g., within a family or household) is not of interest and therefore should not be part of the model. Sometimes the level of available data also restricts the feasible level of abstraction.

In any case, it is essential to the modeling process to clearly define the appropriate entity under study early in the process. The next step, then, is to describe the set of discrete choices that is available to the decision maker, which is discussed in the next section.

9.2.2 Choice Set

In discrete choice analysis, the particular decision maker introduced in the above section is by definition supposed to make a choice from a nonempty, discrete set of predefined alternatives available to her (see Ben-Akiva and Lerman, 1985, Sec. 3.2).

This set of theoretically available choices can be restricted in multiple ways. Naturally, environmental limitations may exist; for example, taking the subway to work is only a valid alternative if the decision maker lives in a city large enough to possess a subway system. The choices remaining after consideration of possible external limitations form the *universal choice set*, which is by definition equal for all decision makers in the study.

However, individual characteristics of the decision maker may further limit the factually available choices. Legal constraints or physical handicaps may hinder her from exercising all options within the universal choice set. Consequently, the resulting choice set after deduction of such infeasible options is termed the *feasible choice set*.

Finally, when actually making a decision, an individual may not be aware of all alternatives available to and feasible for her. This could result from the decision maker not spending enough time on research or simply from the choice set being too large to understand and evaluate exhaustively. The final group of choices incorporated into the decision process is part of the *consideration choice set*. In many studies in the literature, the consideration choice set is simply referred to as *the choice set*, and this convention is followed here for brevity (see Koppelman and Bhat, 2006, Sec. 2.3).

The resulting final (consideration) choice set needs to exhibit two specific characteristics; that is, it has to be MECE (see Train, 2003, Sec. 2.2):

- **Mutually Exclusive:** The alternatives have to be defined in a way that necessarily excludes the choice of all remaining alternatives as soon as the decision maker selects a specific alternative from the choice set. That is, only one alternative at a time can be chosen by the decision maker under any potential circumstances.

- **Collectively Exhaustive:** Naturally, the choice set needs to contain all possible alternatives that the decision maker may choose from in reality. This necessarily includes the option to choose nothing (the so-called *no-buy option*). Any collection of alternatives can artificially be expanded to be collectively exhaustive by including an additional alternative defined to contain all other options. Needless to say, the collection of alternatives must also be finite.

After the definition of the choice set, the contained alternatives each need to be characterized, which is described in the following section.

9.2.3 Alternative Attributes

Each alternative in the choice set needs to be sufficiently characterized to be evaluated by the decision maker following her specific decision rule (see next Section 9.2.4). Therefore, a set of attributes has to be collected for each available alternative. In the example of choosing the mode of transportation to use for one's commute, possible attributes could be travel time, total cost or waiting time.

These characteristics are represented by a vector of attribute values, which can (sometimes partially) be homogeneous or generic in the sense that multiple alternatives share the same descriptive features (e.g., cost), but may also be heterogeneous or alternative-specific if specific characteristics only exist for selected alternatives (e.g., fuel consumption).

Similarly, the characteristics can sometimes partially be deterministic (e.g., train ticket cost) or stochastic (e.g., expected trip time). Stochastic characteristics are typically modeled as two separate attributes containing the moments of the probability distribution, i.e., the first attribute contains the expectancy of the outcome and the second one holds a measure for the stochastic variance (see Ben-Akiva and Lerman, 1985, Chap. 3).

Following the framework in Figure 9.1, the attractiveness of each alternative is evaluated by the decision maker based on the vector of its characteristics. Therefore, the model needs to include all relevant attributes based on a metric that allows for a stringent evaluation.

Similarly, it is important to include attributes that may be proactively influenced by a prospective policy maker (e.g., the pricing department of the urban transit system), so that the effects of possible policy changes can be simulated (see Koppelman and Bhat, 2006, Sec. 2.4).

The next section introduces the definition of decision rules for a decision maker based on the attributes of the available alternatives.

9.2.4 Decision Rule

Behavioral theory suggests that whenever an individual (or decision maker, respectively) truly chooses between two or more alternatives, some form of decision rule is employed. After the available information is evaluated, that rule suggests a specific choice of action.

Principally, this process may include irrational behavior, e.g., by random choice or the mentioned approach of following supposed opinion leaders in their decisions. However, for customer choice analysis to yield efficient and unbiased results, the deployed decision rule has to be rational to the extent that it satisfies the following two requirements of the so-called *bounded rationality* (see Ben-Akiva and Lerman, 1985, Sec. 3.3):

- **Consistency:** Repeated choices under identical circumstances, i.e., using the same choice set with the alternatives exhibiting fixed attribute values, would yield the same choice results.

  True random behavior violates this requirement, while following a leader satisfies it as long as the leader itself behaves consistently.

- **Transitivity:** The evaluation of alternatives results in a unique ordering according to their attractiveness; i.e., if A is more attractive than B and at the same time B is more attractive than C, then A is also more attractive than C.

  Random behavior does not satisfy this requirement, while a group dynamic behavior (individuals following the group's decision) would.

Although a number of other possible rational decision rules exist (see Ben-Akiva and Lerman, 1985, pp. 35), discrete choice theory (like most economic theories) rests on the *utility maximization principle*. That is, the attractiveness of an alternative is somehow contained in the specific values of its attributes, which may compensate one another but can finally be reduced to a scalar measure of utility. These one-dimensional utilities therefore contain the results of possible trade-off processes during an individual's decision making as she always chooses the alternative that in total yields the highest utility, i.e., that maximizes her specific utility (see Koppelman and Bhat, 2006, Sec. 2.4).

In the commute choice example of Figure 9.1, the decision maker may trade off the total cost for her commute with the elapsed trip time and comfort, whereby the utility values of cost and time may counterbalance based on personal preference. However, eventually, the process results in some specific utility order of available mode types.

After the above introduction of the overall choice decision process, the next section explains the theoretical principles behind the concept of utility-based choice.

9.3 Individual Choice Behavior

The aim of this section is to substantiate the theory of *customer choice* analysis within the economic *consumer theory*. To achieve this, Section 9.3.1 briefly recaps the basis of economic utility-based consumer theory. Based thereon, Sections 9.3.2 and 9.3.3 derive the deterministic and stochastic discrete choice concepts, respectively.

9.3.1 Economic Utility-based Consumer Theory

In micro-economic consumer theory, individual customers are assumed to choose a specific consumption bundle $Q = \{q_1, \ldots, q_L\}$ consisting of L different products or services. The bundle itself is defined by the different quantities q_l that are consumed from each respective product l. Naturally, these quantities are bound to be non-negative, $q_l \geqslant 0 \; \forall l$, and total consumption is limited by the so-called budget constraint $\sum_{l=1}^{L} q_l \cdot p_l \leqslant I$, where p_l is the unit price of product l and I is the individual's total budget or income.

In classic economic theory, the specific attributes or properties of the different products are not considered, and therefore, the simultaneous choice of the L different quantities q_l solely defines a single alternative from the decision maker's point of view.

The individual is assumed to be able to express preferences over such alternative bundles (often assuming perfect information): $Q_i \succeq Q_j$, i.e., bundle Q_i is at least as good as bundle Q_j. To formalize such ordinal relationships between bundles, decision makers are assumed to first internally evaluate a utility function $U : \mathbb{R}^L \to \mathbb{R}$ yielding a scalar utility value based on the chosen quantities $U = U(q_1, \ldots, q_L)$ and then to choose the particular bundle Q_k whose utility U_k at least matches the utility of all other possible bundles $Q_k \succeq Q_i \Leftrightarrow U_k \geqslant U_i \; \forall \; i$. This means that the decision maker is assumed to adjust the individual quantities so as to maximize the resulting utility (see, e.g., Varian, 1992), behaving henceforth rational in the sense of Section 9.2.4.

Raw economic theory assumes a principal substitutability of all goods, resulting in a single comprehensive utility function where all products are lastly related because of the joint budget constraint.

To restrict such heavy interdependence, Strotz (1957, 1959) proposes a structure called a *utility tree* that may arrange products or services in separable (or possibly nested) groups or branches in such a way that only products within the same branch are interrelated in an additive utility function. The total utility is then composed from the single sub-utilities or -groups

$$U = U\big(U_1(q_{11}, \dots, q_{1L}), U_2(q_{21}, \dots, q_{2L}), \dots, U_K(q_{K1}, \dots, q_{KL})\big), \qquad (9.1)$$

where

$$\begin{aligned} U_k(q_{k1}, \dots, q_{kL}) &: \text{utility of products in branch } k, \\ kL &: \text{the amount of products in branch } k, \\ K &: \text{the number of branches.} \end{aligned}$$

Ascending the described utility tree (in an abstract sense) toward the outer leaves eventually results in typical problems of discrete choice. As an example, the sub-utility U_{work} of the exemplary branch "commute to work" may be based on the discrete choice between three different mode types, where the possible quantities are restricted to be either one or zero and only a single mode can be selected at a time

$$U_{work} = U(q_1, q_2, q_3) \qquad (9.2)$$

with

$$q_i = \begin{cases} 1 & \text{if mode 1 is chosen,} \\ 0 & \text{otherwise} \end{cases}$$

$$\sum_{i=1}^{3} q_i = 1.$$

The above example marks the step from general *consumer theory* to the more specific *choice theory*, where choices are discreet and denumerable. In fact, (9.2) allows for only three different bundles or choices as outcomes: $U(1,0,0), U(0,1,0)$ or $U(0,0,1)$. The *discrete domain* for the actual choice in q_i renders the maximization of the utility function impossible using simple calculus (i.e., it is not differentiable with respect to the quantities q_i), unlike in the continuous economic consumer theory above (see Ben-Akiva and Lerman, 1985, Sec. 3.6). The appropriate analytical approach to utility maximization in such cases is discussed in Chapter 11.

The discretization allows the utility of each choice not only to depend on the chosen quantities (which are of binary domain here anyway), but to include attributes of the specific products evaluated by the decision maker as drivers for their respective utility (see Lancaster, 1966).

In the above example, the individual utilities U_{in} of the three modes i of commuting to work as seen by the decision maker n may each depend on a specific vector $\mathbf{z}_{in}$ of attributes describing the relevant properties of the corresponding alternative i, so that $U_{in} = U_i(\mathbf{z}_{in})$. Note that besides the attributes, the utility function itself (i.e., the evaluation of the attributes) also depends on the specific alternative chosen. For example, the (dis-)utility derived from travel time may differ by mode and henceforth may be valued differently between modes. Again, each decision maker is assumed to select the alternative k exhibiting the highest overall utility

$$U_k(\mathbf{z}_{kn}) \geqslant U_i(\mathbf{z}_{in}) \ \ \forall i \tag{9.3}$$

where

$U_i(\mathbf{z}_{in})$: utility of alternative i for decision maker n,

$\mathbf{z}_{in}$: attributes of alternative i as seen by decision maker n.

In choice theory, utility is derived from the properties of a single chosen alternative instead of the specific quantity composition in a consumption bundle. Correspondingly, the objective here is to derive the functional composition of such utility instead of general demand functions as in consumer theory.

The following sections now take a closer look at possible deterministic and stochastic compositions of these individual utility functions for choice alternatives.

9.3.2 Deterministic Choice Theory

The introduced utility maximization rule states that among a set of choice alternatives, each individual decision maker chooses the particular alternative that yields the highest utility to her, over all available alternatives. The existence of particular utility functions that evaluate to each customer's individual utility moreover implies that such utility is dependent on two parameter vectors that are external and internal to the decision maker, respectively

$$U_i(\mathbf{z}_{in}) = U_i(\mathbf{x}_i, \mathbf{s}_n) \ \ \forall i, n \tag{9.4}$$

where

$$\mathbf{x}_i : \text{attributes of alternative } i,$$
$$\mathbf{s}_n : \text{characteristics of decision maker } n.$$

Here, $\mathbf{x}_i$ explicitly represents the attributes and properties of alternative i (e.g., total travel time, cost, etc.) and $\mathbf{s}_n$ contains the characteristics of the decision maker (e.g., household income, age, etc.), the latter being identical for all utility functions irrespective of the alternative. Both inputs are considered deterministic, and therefore utility maximization in principle implies no uncertainty as a clear utility ranking of alternatives is always possible. The decision maker again chooses the alternative k with the highest utility (deterministic choice)

$$U_k(\mathbf{x}_k, \mathbf{s}_n) \geqslant U_i(\mathbf{x}_i, \mathbf{s}_n) \quad \forall i. \tag{9.5}$$

The deterministic choice resulting from a utility ranking as in (9.5) is illustrated in Figure 9.2 for the case of two alternatives. Six decision makers A–F evaluate the utility of two alternatives as shown in the grid. Whenever the result lies above the diagonal line of equal utility, alternative 1 is chosen; if it falls below the line, alternative 2 is chosen.

Intuitively, the absolute scale of utility does not affect the resulting choice, i.e., rescaling both dimensions in Figure 9.2 by any factor or the addition of

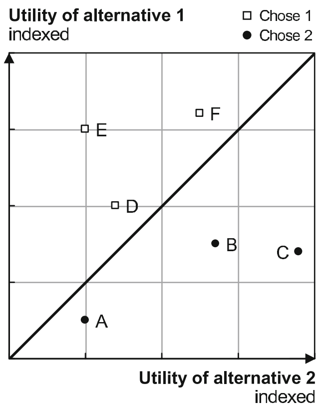


Figure 9.2: Deterministic choice grid
Source: Following Train (2003, pp. 16)

a constant value to both utilities would not change the outcome. Therefore, "any function that produces the same preference orderings can serve as a utility function and will give the same predictions of choice, regardless of the numerical values of the utilities assigned to individual alternatives" (Train, 2003, p. 15).

However, the difficulty in real life situations is the unobservability of the utility attributed by the individual decision maker, as only the actual choice can be recorded by an outside analyst. It is frequently observed that individuals supposedly exhibiting the same utility evaluations or hierarchies choose differently, not consistently choosing the alternative yielding the highest utility. Here, "the challenge is to develop a model structure that provides a reasonable representation of these unexplained variations" (Train, 2003, p. 16), which is tackled in the next section.

9.3.3 Probabilistic Choice Theory

Probabilistic choice theory, or more specifically the *random utility* approach developed by Manski (1977), accounts for unobservable (i.e., supposedly random) variation in customer utility and extends the deterministic theory from Section 9.3.2 above to incorporate stochastic effects.

Ben-Akiva and Lerman (1985) as well as Train (2003) report three basic sources of variation that may genuinely affect the deterministic utility definition in (9.4):

- **Unobserved attributes:** The observed vector of alternative attributes that supposedly affects utility is incomplete. Therefore, the true utility function $U_{in} = U_i(\mathbf{x}_i, \tilde{\mathbf{x}}_i, \mathbf{s}_n)$ includes an additional vector $\tilde{\mathbf{x}}_i$, which observationally can be seen as a random variable.

- **Unobserved taste variation:** The observed vector of decision maker characteristics is incomplete and varies between all decision makers. The true utility function $U_{in} = U_i(\mathbf{x}_i, \mathbf{s}_n, \tilde{\mathbf{s}}_n)$ contains an additional vector $\tilde{\mathbf{s}}_n$, which is again unobserved and therefore can be considered a random variable.

- **Measurement errors:** The observations of both, alternative attributes and decision maker characteristics may contain measurement errors, so that $\mathbf{x}_i = \tilde{\mathbf{x}}_i + \epsilon_i$ and $\mathbf{s}_n = \tilde{\mathbf{s}}_n + \epsilon_n$. This, in turn, leads the overall utility to be a random variable.

The above discussion explains why, in reality, individual decisions cannot be projected or explained with certainty. However, assuming that the decision maker herself knows the true utility of each alternative and acts rationally, random utility theory attempts to derive the *probability* that a decision maker chooses a specific alternative – considering that its utility itself can only be observed as a random variable.

Assuming an additive utility structure (as is typically done in the literature), the true unobservable utility U_{in} is composed of the observable and deterministic utility portion V_{in} (so-called systematic utility) and a stochastic error term ϵ_{in}

$$U_{in} = V_{in} + \epsilon_{in}. \tag{9.6}$$

Assuming a joint stochastic distribution for the random vector $\boldsymbol{\epsilon}_n = (\epsilon_{1n}, \ldots, \epsilon_{Cn})^T$ for a decision maker n across all available alternatives $i = 1, \ldots, C$, the probability p_{in} of her choosing a particular alternative i can be stated as

$$\begin{aligned} p_{in} &= P(U_{in} > U_{jn}\ \forall i \neq j) \\ &= P(V_{in} + \epsilon_{in} > V_{jn} + \epsilon_{jn}\ \forall i \neq j) \\ &= P(V_{in} - V_{jn} > \epsilon_{jn} - \epsilon_{in}\ \forall i \neq j). \end{aligned} \tag{9.7}$$

That is, based on estimates of the parameters of the systematic part of the utility function, random utility models for discrete customer choice provide the probability of a decision maker's choice as given by (9.7).

The calculation of such probabilities requires an explicit assumption about the distribution of the ϵ_{in}, and indeed, the various customer choice models distinguish themselves by their specific stochastic suppositions. Intuitively, the resulting integral for the evaluation of (9.7), which is dependent on these assumptions, may well not always exist in closed form (see Manski, 1977, for a comprehensive overview of random utility models).

Table 9.2 gives an overview of the most popular customer choice models together with their most important distinguishing assumptions about the distribution of error terms and utility parameters (where necessary). For the interested reader, the original literature sources are provided. However, the table is not intended to be complete, but rather to provide an informed selection with the aim of illustrating the importance of model selection based on justifiable assumptions about the particular choice situation being modeled. An introduction to the various models can be found in Koppelman and Bhat (2006, Chap. 12).

Model	Important assumptions	Original sources
· Probit	· Error terms are assumed to be normal distributed	· Hausman and Wise (1978) · Daganzo (1979)
· Generalized Extreme Value (GEV)	· Error terms are assumed to be extreme value/Gumbel distributed	· Luce (1959) · McFadden (1978)
· Multinomial Logit (MNL)	· Error terms are additionally assumed independent and identically distributed (IID) over *all* alternatives	· McFadden (1974) · McFadden (1978)
· Nested Logit (NL)	· Error terms are additionally assumed IID only over specific *subsets* (i.e., nests) of alternatives	· Williams (1977) · Daly and Zachary (1979)
· Cross-nested Logit (CNL)	· Like NL with specific alternatives being allowed in multiple nests (with errors not IID)	· Small (1987) · Vovsha (1997)
· Mixed Multinomial Logit (MMNL)	· *Coefficients* of utility function are considered random variables that are normal distributed	· Boyd and Mellman (1980) · Cardell and Dunbar (1980)
· ...		

Table 9.2: Overview of discrete choice models (excerpt, chronological)

This section has given a general introduction to individual customer choice theory. The following section takes a closer look at the multinomial logit model in particular, which is used in Chapter 11 to analyze customer choice behavior in the low-cost air travel market.

9.4 The Multinomial Logit Model

The most common and widely used model formulation to analyze customer choice behavior is the multinomial logit (MNL), which is advantageous to other more complex models in three ways: First of all, the formula for the evaluation of the resulting choice probability given by (9.7) takes a closed form, i.e., is *mathematically tractable*. Furthermore, the model structure and resulting probabilities are straightforwardly *interpretable*, and the results are *robust* in scenarios where the fairly restrictive underlying assumptions are slightly violated (see, e.g., Train, 2003, Sec. 3.1).

The following section derives and explains the functional form of MNL as well as the necessary assumptions. Based thereon, Section 9.4.2 highlights the resulting properties of the model before Section 9.4.3 explains the estimation approach used when employing such models. Finally, Section 9.4.4 discusses the necessary tests to detect violations of the assumptions and to compare differing model specifications.

9.4.1 Description and Functional Form

The multinomial logit model is based on the assumptions of the random utility theory introduced in Section 9.3.3: A decision maker n faces a set of C different alternatives or possible choices, respectively, that are MECE in the sense of Section 9.2.2.

Following (9.7), an alternative i's overall utility U_{in} as seen by the decision maker n is assumed to consist of a parameterized systematic utility function V_{in} that is considered deterministic and known by the analyst plus an unknown supposedly random part ϵ_{in} so that $U_{in} = V_{in} + \epsilon_{in}$.

The MNL now assumes that the ϵ_{in} are independent identically distributed (IID) extreme value or Gumbel. This implies that the unobserved parts of utility are uncorrelated between alternatives and exhibit the same variance (see Ben-Akiva and Lerman, 1985, p. 104), or "stated equivalently, the researcher has specified V_{in} sufficiently that the remaining, unobserved portion of utility is essentially 'white noise' " (Train, 2003, p. 39).

The assumption of extreme value distributed errors is empirically close to the normal distribution used in probit models as the cumulative distributions are "indistinguishable empirically" (Train, 2003, p. 39), as exemplary shown in Figure 9.3, with the extreme value distribution yielding significant analytical advantages (see below).

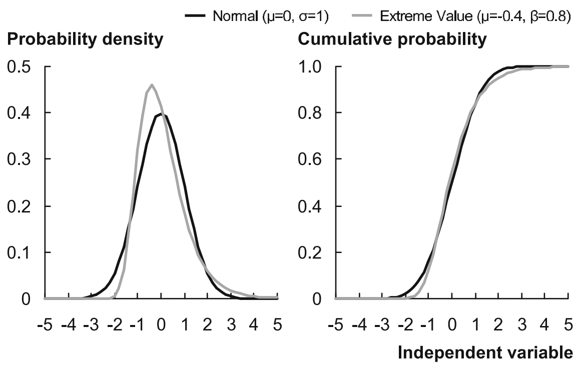


Figure 9.3: Normal vs. extreme value distribution
Source: Own design

The probability density and cumulative distribution function, respectively, of the standard extreme value distributed errors are

$$f(\epsilon_{in}) = e^{-\epsilon_{in}} e^{-e^{-\epsilon_{in}}} \quad \text{and} \quad F(\epsilon_{in}) = e^{-e^{-\epsilon_{in}}}. \tag{9.8}$$

Note that the mean of the standard extreme value distribution is not zero, but – as mentioned in Section 9.3.3 – only differences in utility matter, and therefore the actual mean is irrelevant for the choice probability

$$\begin{aligned} p_{in} &= P(\epsilon_{jn} - \epsilon_{in} < V_{in} - V_{jn} \ \forall j \neq i) \\ &= P(\epsilon_{jn} < V_{in} - V_{jn} + \epsilon_{in} \ \forall j \neq i). \end{aligned} \tag{9.9}$$

The explicit assumption of a standard extreme value distribution also implies a fixed variance of the ϵ_{in} of $\frac{\pi^2}{6}$, which automatically normalizes the scale of the model's overall utilities.[2]

Based on the IID assumption on the ϵ_{in}, and using the density and distribution in (9.8), the probability in (9.9) can be rewritten as

$$p_{in} = \int \left(\prod_{i \neq j} e^{-e^{-(V_{in} - V_{jn} + \epsilon_{in})}} \right) e^{-\epsilon in} e^{-e^{-\epsilon_{in}}} d\epsilon_{in}, \tag{9.10}$$

which can in turn be transferred through algebraic manipulation (for a detailed proof, see Train, 2003, Sec. 3.10) to

$$p_{in} = \frac{e^{V_{in}}}{\sum_{j \in C} e^{V_{jn}}}. \tag{9.11}$$

The reduced expression (9.11) is called the *logit choice probability* (see, e.g., Ben-Akiva and Lerman, 1985, Sec. 5.2). It defines the probability that the decision maker chooses alternative i if she assigns systematic utility V_{in} to that particular alternative and utilities V_{jn} to all available alternatives $j \in C$.

Obviously, the assumption of the stochastic part of utility to be extreme value IID leads to a fairly benign representation of choice probabilities. Moreover, for a linear specification of the deterministic part of the utility $V_{in} = \boldsymbol{\beta}_i^T \mathbf{z}_{in}$ (compare to Section 4.1), McFadden (1974) proves that the log-

[2] The normalization of the utility scale yields no negative implications for the model itself. However, the scale of utility and its parameters may vary between models. Therefore, no direct comparison of parameter values between models is possible (see Chapter 11).

likelihood function of (9.11) is globally concave in its parameters $\boldsymbol{\beta}_i$, which finally allows for the numerical maximization and estimation of V_{in}, as discussed in Section 9.4.3 below.

Before looking at estimation of the parameters in MNL models, the next section highlights the resulting specific properties of the multinomial logit choice probabilities along (9.11).

9.4.2 Specific Properties and Limitations

The aim of this section is to highlight the (mostly desirable) properties of the MNL and to discuss possible limitations that might follow from these.

General Properties The choice probabilities as defined by (9.11) satisfy the basic requirements for any stochastic probability.

First of all, the resulting probabilities are constructionally bound by zero and one. In addition, the actual choice probability of an alternative can reach neither bound: The exponential function does not evaluate to zero, and thus, if an alternative is believed to have exactly zero choice probability it should be removed from the choice set. In turn, an alternative can only achieve a definite chance of being selected if it is the only one in the choice set. Neither case limits the usability of the model.

Secondly, and also by construction, the sum of choice probabilities always equals one. Henceforth, a normalization or indexing of choice probabilities is not required. This property assures that the decision maker always chooses an alternative, while the requirement for the alternatives to be MECE automatically impedes the selection of multiple alternatives at the same time.

Sigmoid Shape Looking at how utility transforms to actual choice probability, the MNL model exhibits a sigmoid or S-shaped relationship (see Figure 9.4). Through this non-linear transformation, a change in utility yields the greatest increase in choice probability if the original probability was observed close to 0.5.

This property is reasonable in many real-life situations. Going back to the choice of commute mode example of Section 9.2: if, for a particular decision maker, using the bus has a very low choice probability because it is very inconvenient, it would take many major changes in utility to convince her to choose it. If, on the other hand, she is basically undecided between two modes, only a slight increase in utility on either side might tip the decision.

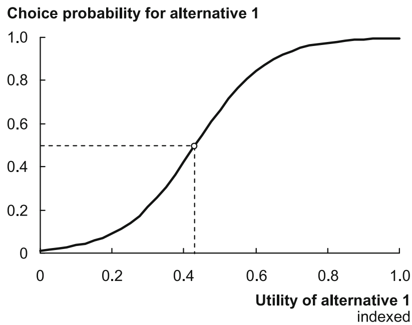


Figure 9.4: Sigmoid relationship of utility to choice probability
Source: Own design

Economic Meaning As mentioned in the previous section, the functional composition of systematic utility is assumed to be linear in alternative attributes $\mathbf{x}_i$ and decision maker characteristics $\mathbf{s}_n$ so that $V_{in} = \boldsymbol{\beta}_i^T \mathbf{z}_{in} = \boldsymbol{\beta}_{1i}^T \mathbf{x}_i + \boldsymbol{\beta}_{2i}^T \mathbf{s}_n$, which is a rather mild restriction as "under fairly general conditions, any function can be approximated arbitrary closely by one that is linear in parameters" (Train, 2003, p. 41).

However, the assumption of linearity gives economic meaning to the individual coefficients β_{im} of the parameter vector $\boldsymbol{\beta}_i$ as the ratio of two parameters represents the decision maker's trade-off in utility between the corresponding attributes that keeps the overall utility constant. This can be understood by taking the total derivative of overall utility with respect to one of the two attributes under consideration, setting it to zero and solving it for the change in the other attribute that keeps utility constant (see, e.g., Train, 2003, p. 43)

$$\frac{dV}{dx_1} = \frac{\partial V}{\partial x_1}\frac{dx_1}{dx_2} + \frac{\partial V}{\partial x_2}\frac{dx_2}{dx_1} \overset{!}{=} 0 \qquad \Leftrightarrow$$
$$\frac{dx_2}{dx_1} = -\frac{\frac{\partial V}{\partial x_1}}{\frac{\partial V}{\partial x_2}} = -\frac{\beta_1}{\beta_2}. \tag{9.12}$$

Equivalent Differences Property As mentioned before, the choice probabilities of the MNL only depend on the differences in systematic utility between alternatives and not on their actual level or scale. Koppelman and Bhat (2006, Sec. 4.1.2) show that the choice probability in (9.11) can be rewritten to

$$p_{in} = \frac{1}{1 + \sum_{j \neq i} e^{V_{jn} - V_{in}}}, \tag{9.13}$$

which formalizes the above statement, given that equivalent differences in utility yield the same choice probability.

Additionally, the property has a direct implication for the definition of systematic utility: The MNL is not identified if constants (coefficients with constant attributes) are added to the utility functions of all alternatives as the coefficients could be scaled by an arbitrary value, maintaining the differences in utility and with them keeping the choice probability constant.

Taste Variation The systematic part of a decision maker's choice behavior or *taste*, respectively, is supposedly contained in the utility function. However, diverse individuals will typically exhibit different choice behaviors regarding identical alternatives, based on the same observed attributes.

The MNL can by construction handle so-called *systematic taste variation*, which can be attributed to specific decision maker characteristics (typically socio-demographics). For example, the utility effect of price is often influenced by an individual's income, which can be incorporated by defining an interaction variable $z_{in} = \frac{price_i}{income_n}$ that enters the specification of V_{in} in the conventional linear way.

On the other side, *random taste variation* appears stochastic with respect to the observed decision maker characteristics and alternative attributes. That is, differences in utility between decision makers result from individuals assigning different valuations to alternative attributes or interaction variables despite exhibiting the same socio-demographic characteristics.

These taste variations can be modeled by assuming stochastic coefficients as is done in mixed multinomial models, but the standard MNL assumptions do not allow for such modifications. Therefore, random taste variation cannot be modeled, which may limit the explanatory value of MNL models. However, "as an approximation, logit might be able to capture the average tastes fairly well even when tastes are random, since the logit formula seems to be fairly robust to misspecifications" (Train, 2003, p. 48).

Independence of Irrelevant Alternatives The most prominent property and simultaneous limitation of the MNL can easily be derived by examining the ratio of choice probabilities for two alternatives, a and b (see, e.g., Koppelman and Bhat, 2006, Sec. 4.2)

$$\frac{p_{an}}{p_{bn}} = \frac{\frac{e^{V_{an}}}{\sum_{j\in C} e^{V_{jn}}}}{\frac{e^{V_{bn}}}{\sum_{j\in C} e^{V_{jn}}}} = \frac{e^{V_{an}}}{e^{V_{bn}}} = e^{V_{an}-V_{bn}}. \tag{9.14}$$

Obviously, the ratio of choice probabilities for alternatives a and b is fully independent of the utility of any other alternative, i.e., it is independent of (quasi-)irrelevant alternatives (IIA).

The implications of the IIA property are double-edged: On the one hand, even a partial MNL can still be correctly estimated based on a subset of alternatives, enabling scenarios where different decision makers face varying subsets of available alternatives. Moreover, the property also simplifies the estimation process as described in the next section.

On the other hand, the IIA implies that an increased choice probability for a specific alternative draws proportionally from the choice probabilities of all other alternatives, as the ratios between pairs of these do not change. While this confers no significant limitation on many occasions, sometimes it may violate structural assumptions of reality. The red bus/blue bus paradox is the most famous example of such a scenario (see, e.g., Ben-Akiva and Lerman, 1985, pp. 51).

Suppose an employee can choose between two modes of transportation for her trip to work, taking the car or taking a blue bus, which she values equally in terms of utility, so that her choice probabilities are $\frac{1}{2}$ each. If, now, a new bus service is introduced that employs red buses, but is equal in all other attributes (i.e, schedule, price, comfort, etc.), the equivalent differences property would suggest that choosing either of the two bus services is equally likely. At the same time – following the IIA – the new service draws proportionally from the old choice probabilities (of car and blue bus), resulting in new choice probabilities of $\frac{1}{3}$ for each of the three modes. This clearly opposes the intuition whereby the decision maker treats the two bus services as a single option, keeping the choice probability for commuting by car at $\frac{1}{2}$ and assigning $\frac{1}{4}$ to each of the bus lines. Here, the true decision process is most probably hierarchical, as pictured in Figure 9.5.

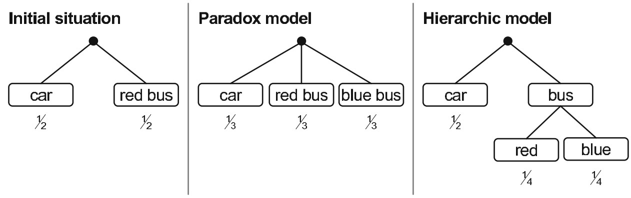


Figure 9.5: Example of a hierarchic choice structure
Source: Own design

Hierarchical choice structures do not meet the IIA assumption of the MNL and were the reason behind the development of more advanced *generalized extreme value* models like the nested logit (for literature notes on these models, see Table 9.2).

Having discussed their properties and limitations above, the next section now discusses the actual estimation of MNL models, incorporating those results to aid the numerical process.

9.4.3 Coefficient Estimation

This section explains the theory behind the coefficient estimation of the parameterized utility function of a MNL as introduced in Section 9.4.1.

First of all, it is important to note that the estimation methodology for choice models itself is dependent on the underlying sample of observations, which is supposed to be used to train the model. Typically, study design explicitly assures that an *exogenous* and *representative* sample is created, which is also the underlying assumption behind the theory explained below.

For scenarios where this ideal standard cannot be achieved, e.g., because the dataset has to be constructed based on sampling from actual choice groups (censored data), other more specific and advanced methods have to be used. Cosslett (1981) as well as Manski and McFadden (1981) list a variety of such methods under different sampling procedures.

Assuming a representative and exogenous sample is available for estimation, the overall probability $l_n(\boldsymbol{\beta})$ of a decision maker n choosing the alternative she was observed having chosen based on attribute valuation $\boldsymbol{\beta}$ in the sample can be stated as

$$l_n(\boldsymbol{\beta}) = \prod_{i \in C} \big(p_{in}(\boldsymbol{\beta}_i, \mathbf{z}_{in})\big)^{b_{in}}, \tag{9.15}$$

where

$$p_{in}(\boldsymbol{\beta}_i, \mathbf{z}_{in}) = \frac{e^{\boldsymbol{\beta}_i^T \mathbf{z}_{in}}}{\sum_{j \in C} e^{\boldsymbol{\beta}_j^T \mathbf{z}_{jn}}}$$

is the parameterized choice probability from (9.11) and

$$b_{in} = \begin{cases} 1 & \text{if decision maker } n \text{ actually chose } i, \\ 0 & \text{otherwise.} \end{cases}$$

Note that (9.15) is only a different formulation of (9.11) assuming a linear parametric formulation for $V_{in} = \boldsymbol{\beta}_i^T \mathbf{z}_{in}$.

Based on the assumption that in an exogenous sample, the decision makers made their choices independently of each other, the joined probability that each member of the sample made the actual choice that was observed is

$$L(\boldsymbol{\beta}) = \prod_{n=1}^{N} \prod_{i \in C} \big(p_{in}(\boldsymbol{\beta}_i, \mathbf{z}_{in})\big)^{b_{in}}. \tag{9.16}$$

The joined probability (9.16) results from the multiplication of N individual choice probabilities as in (9.15) above. The resulting probability $L(\boldsymbol{\beta})$ is called the *Likelihood function* and naturally depends on the coefficient vector $\boldsymbol{\beta}$ for all alternatives that, accordingly, shall be estimated so as to maximize the likelihood of the observations.

McFadden (1974) shows that the logarithmized, so-called *Log-Likelihood function*

$$LL(\boldsymbol{\beta}) = \sum_{n=1}^{N} \sum_{i \in C} \Big(b_{in} \cdot \ln\big(p_{in}(\boldsymbol{\beta}_i, \mathbf{z}_{in})\big)\Big) \tag{9.17}$$

is globally concave for a linear definition of systematic utility V_{in} (as in the above case). Henceforth, it is possible to numerically estimate the particular values $\hat{\boldsymbol{\beta}}$ that maximize $LL(\boldsymbol{\beta})$ – this is called the *Maximum Likelihood*

Software	Manufacturer	Probit	MNL	NL	MMNL
BIOGEME	Michel Bierlaire	✓	✓	✓	✓
Discrete Choice	Aptech Systems	✓	✓	✓	
NLOGIT	Econometric Software	✓	✓	✓	✓
ETS	SAS Institute	✓	✓	✓	✓
SPSS	SPSS		✓		
STATA	StataCorp	✓	✓	✓	

Table 9.3: Software packages for discrete choice modeling
Source: Summary of Gönsch et al. (2008b, p. 417)

estimator (for proof and discussion, see, e.g., Train, 2003, Sec. 3.7.1).

Under the assumption that the observed realizations of the decision makers' choices are just a sample based on a random experiment whose outcomes are supposedly normal distributed, the Maximum Likelihood estimator for (9.17) itself follows a normal distribution. That is, similar to for example conventional linear regression, the estimation results do not provide a point estimate, but rather the expectancy for the coefficient vector $\boldsymbol{\beta}$ together with a corresponding variance (for details on the numerical approach see, e.g., Ben-Akiva and Lerman, 1985, Sec. 5.5).

The obtained estimator is benign in the sense that it is consistent, asymptotic efficient and normally distributed.

While the described estimation process is tedious for many of the choice models in Table 9.2, for the MNL many software packages are available for estimation (see Table 9.3). Based on its performance and the fact that it is freely available, BIOGEME Ver. 1.8 is used throughout Part III of this work (see Bierlaire, 2003, 2009).

Now that the formal estimation process for multinomial logit models has been described, the next section identifies the relevant tests to assure that the underlying assumptions of the MNL are met and to verify the statistical relevance of the coefficient vector.

9.4.4 Tests of Model Specifications

This section introduces the necessary tests to evaluate the generated multinomial logit models based on their informal conformity with reality and intuition, their adherence to the model assumptions and the individual goodness of alternate specifications.

Informal Tests The first and most basic tests of a prospective model design are the informal tests of whether the results are rational and conform to the analyst's intuition or expectations. The model building process typically follows certain logic and rationale, in terms of what might drive customer choice probability, and the resulting model coefficients should reflect that initial thought.

As a basic requirement, the signs of the coefficients should correspond with the expected directions of the attributes' influences. For example, price, cost or waiting time should typically exhibit a negative sign, while comfort or convenience factors usually exhibit positive signs.

A similar rationality test can be applied to ratios of coefficients, which reflect the trade-off values the decision makers attach to specific characteristics (see Section 9.4.2). In case the expected trade-off ratio is very different from the observed value, the possible need for the addition or omission of variables or even their transformation has to be checked. For example, a logarithmic transformation of attributes can change an additive relationship into a multiplicative relationship, according to the logarithmic rules (see also Part II).

Formal Tests of Alternative Specifications Similar to linear regression (compare Section 4.1), a range of statistical hypothesis tests exist to verify the significance and goodness of fit for alternative model specifications. Naturally, these tests assume that the model type has been correctly selected, which can be verified using the appropriate tests in the next paragraph below.

A conventional *t-test* can be used to evaluate whether a specific coefficient β_m is significantly different from a predefined constant (typically zero). The application to MNL is straightforward following the definition in Section 4.1 and is henceforth not repeated here.

Directly derived from the t-statistic is the calculation of an *asymptotic confidence interval* for a single parameter: The $(1-\alpha)$ confidence interval for a single coefficient β_m is defined by

$$P\left(-t_{\frac{\alpha}{2}} \leqslant \frac{\hat{\beta}_m - \beta_m}{\sqrt{var(\hat{\beta}_m)}} \leqslant t_{\frac{\alpha}{2}}\right) = 1-\alpha,$$

which rearranges to

$$P\left(\hat{\beta}_m - t_{\frac{\alpha}{2}} \cdot \sqrt{var(\hat{\beta}_m)} \leqslant \beta_m \leqslant \hat{\beta}_m + t_{\frac{\alpha}{2}} \cdot \sqrt{var(\hat{\beta}_m)}\right) = 1-\alpha,$$

with $t_{\frac{\alpha}{2}}$ being the particular quantile of the normal distribution so that the t-ratio will exceed $t_{\frac{\alpha}{2}}$ only with probability $\frac{\alpha}{2}$.

The conventional t-test will not reject the null hypothesis that $\beta_m = 0$ to a significance level α whenever the corresponding confidence interval includes zero (see Ben-Akiva and Lerman, 1985, Sec. 7.4).

Using covariance information on $\boldsymbol{\beta}$ from the estimation process, it is also possible to run an *asymptotic t-test* on the possible *equality of two coefficients* β_l and β_m:

Hypothesis
$H_0 : \beta_l = \beta_m$.

Test statistic
$t_{l,m} = \frac{\hat{\beta}_l - \hat{\beta}_m}{\sqrt{var(\hat{\beta}_l - \hat{\beta}_m)}}$
with
$var(\hat{\beta}_l - \hat{\beta}_m) = var(\hat{\beta}_l) + var(\hat{\beta}_m) - 2cov(\hat{\beta}_l, \hat{\beta}_m)$.

Critical values
The null hypothesis H_0 is rejected if the absolute value of the test statistic is greater than the corresponding quantile of the standard normal distribution $|t_{l,m}| > t_{1-\frac{\alpha}{2}}$.

For more complex hypotheses on coefficient relationships, the *Likelihood Ratio test* can be employed for joint testing of several coefficient values. Here the null hypothesis imposes two or more linear restrictions on the original model and its coefficients, respectively. A typical linear restriction to test during model building is whether several coefficients are jointly zero, i.e., whether specific coefficients can be omitted from the model:

Hypothesis
$H_0 : \beta_l = \ldots = \beta_m = 0$.

Test statistic
$LR_{l,m} = -2\big(LL(\hat{\boldsymbol{\beta}}_R) - LL(\hat{\boldsymbol{\beta}}_U)\big)$,
where $\hat{\boldsymbol{\beta}}_R$ is the reduced coefficient vector of the restricted model (when the null hypothesis is true) and $\hat{\boldsymbol{\beta}}_U$ is the full coefficient vector of the unrestricted model.

Critical values
Under the null hypothesis, $LR_{l,m}$ is χ^2_r-distributed with degrees of freedom equal to the number of restrictions on the original model $r = |\hat{\boldsymbol{\beta}}_U| - |\hat{\boldsymbol{\beta}}_R|$. Accordingly, the null hypothesis is rejected if the test statistic is greater than the corresponding quantile of the χ^2-distribution $LR_{l,m} > \chi^2_{r,1-\frac{\alpha}{2}}$.

Note that it is essential for the restricted model to be nested based on the original model specification so that it can be obtained by enforcing a set of linear restrictions on selected coefficients.

The Likelihood Ratio test can be used in similar fashion to test whether coefficient values are not alternative-specific, but rather *generic*, that is, the coefficient values for a specific attribute (e.g., price) are equal over different alternatives.

Another useful application is to test for possible evidence that a non-linear parameter specification is beneficial (i.e., based on a transformation using basis functions as described in Section 4.1). For example, through a piecewise linear approximation by sections with length τ of the original variable space $x_i > 0$, a set of input variables x_{ik} with corresponding coefficients is created

$$x_{i,k} = \begin{cases} x_i & \text{if } (k-1)\cdot\tau < x_i \leqslant k\cdot\tau \\ & \text{with } k \in \{1,\ldots,K\}, \\ 0 & \text{else.} \end{cases} \tag{9.18}$$

Based on the estimation results of the unrestricted model (the coefficient of x_i is allowed to vary over the input variable's domain), the Likelihood Ratio test can be used to test the null hypothesis that the new coefficients are all equal (see, e.g., Ben-Akiva and Lerman, 1985, pp 174).

The *Likelihood Ratio Index* can be used to compare different models that have been estimated on the same dataset and an equal choice set. The index is defined based on the Log-Likelihood function (9.17) as

$$\rho^2 = 1 - \frac{LL(\hat{\boldsymbol{\beta}})}{LL(\mathbf{0})}, \tag{9.19}$$

where $LL(\mathbf{0})$ is the Log-Likelihood value of the artificial model where all coefficients are set to zero. That is, ρ^2 measures the increase in prediction accuracy compared to a model with no explanatory power. Therefore, the pure value of ρ^2 alone is only of limited help as even abstruse models might perform better than no model. However, it can help in comparing different models, where a higher ρ^2 indicates a better model.

Note that although ρ^2 ranges from zero to one, it cannot be interpreted in an intuitive way like the R^2 value in linear regression because it does *not* mirror the share of correctly explained variance in the data. This is also why two separate models estimated on different data or choice sets *cannot* be compared using the Likelihood Ratio Index (see Train, 2003, Sec. 3.8.1).

Formal Tests of Model Structure Similar to linear regression (see Section 4.1), the formal model assumptions of MNL – the random part of utility is IID – need to be tested.

If the assumption is correct, the resulting model will exhibit the IAA property introduced in Section 9.4.2, i.e., it is possible to estimate a coefficient vector $\hat{\boldsymbol{\beta}}_{\widetilde{C}}$ based on a limited choice set that is consistent with the corresponding sub-vector of $\hat{\boldsymbol{\beta}}_C$ belonging to the full choice set. The test of Hausman and McFadden (1984) leverages the IIA property to test for adherence to the model assumptions.

Hypothesis
$H_0 : \hat{\boldsymbol{\beta}}_C = \hat{\boldsymbol{\beta}}_{\widetilde{C}}$.

Test statistic
$\mathrm{HM}_{\widetilde{C}} = (\hat{\boldsymbol{\beta}}_{\widetilde{C}} - \hat{\boldsymbol{\beta}}_C)^T(\Sigma_{\hat{\beta}_{\widetilde{C}}} - \Sigma_{\hat{\beta}_C})^{-1}(\hat{\boldsymbol{\beta}}_{\widetilde{C}} - \hat{\boldsymbol{\beta}}_C)$,
where $\Sigma_{\hat{\beta}_{\widetilde{C}}}$ is the covariance matrix of the limited choice model and $\Sigma_{\hat{\beta}_C}$ is the appropriate sub-matrix of the full model.

Critical values
$\mathrm{HM}_{\widetilde{C}}$ is asymptotically $\chi^2_{\widetilde{K}}$-distributed with $\widetilde{K}$ degrees of freedom, where $\widetilde{K}$ is the number of coefficients in the restricted vector $\hat{\boldsymbol{\beta}}_{\widetilde{C}}$ that are identifiable from the restricted choice set model and data (see Ben-Akiva and Lerman, 1985, pp. 184). The null hypothesis is rejected if the test statistic is greater than the corresponding quantile of the χ^2-distribution $\mathrm{HM}_{\widetilde{C}} > \chi^2_{\widetilde{K},1-\frac{\alpha}{2}}$.

Information Criteria The comparison and evaluation of different model specifications that are not nested (which is a prerequisite for the applicability of the Likelihood Ratio test) requires the use of so-called information criteria. These measure the goodness of the overall model's fit to the observations in relationship to the complexity of the model, the latter being measured in number of input parameters used. Similar to the adjusted R^2 of linear regression, the count of coefficients is considered through an according penalty.

The *Akaike Information Criterion* (AIC) is based on the Log-Likelihood value of the estimated model $LL(\hat{\boldsymbol{\beta}})$ and the total number of parameters M in its utility functions (Akaike, 1973)

$$\text{AIC} = -2LL(\hat{\boldsymbol{\beta}}) + 2M.$$

Similar to the Likelihood Ratio Index, the AIC itself has no explanatory value. It derives its value from the comparison of two AIC that stem from different models, but were estimated based on the same data. Therefore, only differences Δ in AIC matter, for whose evaluation Burnham and Anderson (2004, p. 271) offer a rule of thumb: "Models having $\Delta \leqslant 2$ have substantial support (evidence), those in which $4 \leqslant \Delta \leqslant 7$ have considerably less support, and models having $\Delta > 10$ have essentially no support."

The authors also point out that differencing removes scaling constants and the effects of sample size, and thus the above rule retains its validity even for higher absolute levels of AIC.

In contrast to the *Bayesian Information Criteria* (BIC) (Schwarz, 1979), the AIC implicitly considers that based on a sufficiently large data basis, even models with numerous parameters can still be estimated reliably. This is also the reason why the BIC is not considered in this work.

Note that for very small sample sizes or very large numbers of parameters ($\frac{N}{M} < 40$), the AIC exhibits a bias. Therefore, in such situations (which is not the case in this work), a corrected AIC_C should be used, defined as

$$\text{AIC}_C = -2LL(\hat{\boldsymbol{\beta}}) + 2M + \frac{2M(M+1)}{N-M-1}.$$

In sum, this chapter has introduced the general theory behind customer choice modeling. The next Chapter 10 describes and analyzes the dataset, which is later used for the construction of the MNL model in Chapter 11. In addition, it highlights the supposedly relevant drivers behind customer decision making that must in turn be analyzed for their explanatory value throughout the later model building process.

Chapter 10

Choice Situation in Low-Cost Markets

This chapter presents and descriptively analyzes a proprietary dataset, which contains information on the choice situation of air travel customers in low-cost markets. The analysis is based on the structure and characteristics of exclusively collected real-world data that later also form the basis for the development of a corresponding discrete choice model in Chapter 11.

The objectives of this chapter are to identify possible demand drivers and to discover potentially relevant attributes that might influence customer choice behavior. The specific significances and coefficient weights of these drivers are later computed and tested thoroughly in Chapter 11 using the appropriate methods.

Below, Section 10.1 first describes the compilation of the complete dataset from selected individual data sources and the resulting limitations. Based on those data, an overview of market conduct and participants' behavior is given in Section 10.2. Then, Section 10.3 reports the observed schedule and booking preferences, together with hints on the inherent price sensitivity, before the implications of these findings on the prospective choice model are finally given in Section 10.4.

10.1 Experimental Data Set

This section describes the dataset that is used in the analyses in the following chapters. As the overall objective of Part III is to understand the choice behavior of potential customers having expressed a latent flight demand, the data employed here are naturally an extension of the data basis of Part II.

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6_10,

The data preparation and processing, i.e., the combination of records from the different sources, the coding and transformation to the desired input variables as well as the formating and generation of input data files for BIOGEME has been accomplished using a deliberately coded proprietary data management system on the basis of Microsoft's Visual Basic for Applications on top of a Microsoft Access database.

To undertake a customer choice analysis by the means described in Chapter 9, three different types or sources of data are required and must be combined into a single data set:

- **Customer sample with characteristics:** A sample of the examined decision makers and their individual characteristics that is representative for the entire population under analysis in the sense of Section 9.2.1.
- **Alternative set with attributes:** The complete set of alternatives available to decision makers, which is MECE in the sense of Section 9.2.2 and for which a sufficiently large set of descriptive attributes is available in the sense of Section 9.2.3.
- **Choice decision:** An observation of the final, discrete purchasing decision for each decision maker within the observation sample in the sense of Section 9.2.4.

The data set employed in Part II contains a representative customer sample of decision makers in which all potential customers that request a fare quote from the air carrier under observation are contained. The data can also be considered representative for the full population of potential travelers as 99% of customers report being aware of the monitored offering and henceforth are part of the data (see Section 5.1).

To meet the full data requirements expressed above, the plain observation sample from Part II is amended with data from three additional independent data sources (see Figure 10.1 for an illustration):

① The full *alternative set* can be constructed by using public sources like the proprietary *OAG schedule database*, which "offers daily updated schedules, available in print and electronic formats" (OAG, 2009), or simply by manual examination of affected airport schedules. These sources also already contain several of the relevant alternative-specific

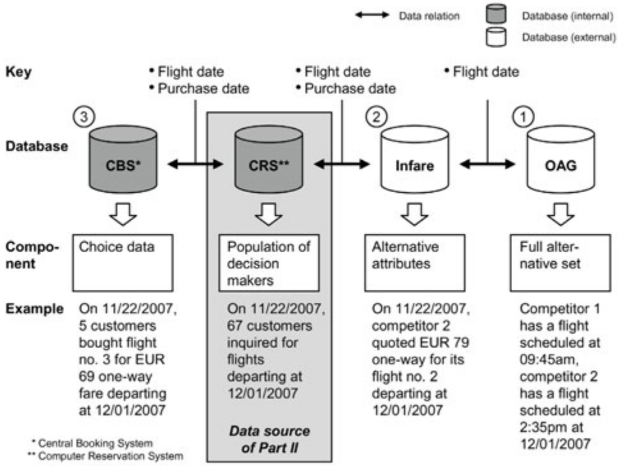


Figure 10.1: Construction of complete data set incorporating three additional data sources
Source: Own design

attributes (e.g., operating carrier, departure time, flight time, aircraft type, etc.). However, this alternative-specific information needs to be amended with flight availability and fare data to complete the necessary set of descriptive attributes, which requires a further additional data source.

② A second proprietary database containing additional *alternative-specific attributes* like flight availability and fare data for all flights was provided by *Infare Solutions* (Infare, 2009) for exclusive use in this work. Dynamic flight status-dependent information is collected by Internet bots on a daily basis from the booking websites of all relevant competitors. An alternative manual data collection would have considerably limited the amount of data that could reasonably have been collected and would have also been prone to error.

Fare data and flight availability were scanned automatically by Infare's system at night (midnight to 5:00 AM) between September 15, 2007, and February 15, 2008 (see Figure 10.2). For all specified carriers, one-way fares were collected for each in- and outbound flight on a selected short-haul route within a time window of 60 days before departure. Additionally, for the competing flag carrier, round-trip fares were collected, assuming a seven-day stay as traditional pricing promotes round-trip travel and purchase.

③ The actual *choice decision* of customers can technically be tracked based on actual bookings within each carrier's *central booking system* (CBS), which is not publicly available to all market participants. As this work takes the perspective of a single carrier, only its own customers' choice information is available, i.e., whether or not a booking was made at the considered carrier – not differentiating customers that did not purchase a ticket at all and those that did buy, but from a competing carrier.[1]

Section 11.1 shows that this data limitation restricts the possible model structures, but still allows the construction of insightful models.

The necessary combination of data sources unfortunately leads to a shortened maximum analysis horizon, as illustrated in Figure 10.2, because data availability inevitably varies between sources. While the data from the internal sources (CRS and CBS) span the entire analysis horizon as used in Part II (see Figure 5.4) and the OAG database claims to be complete over any recent time window, the automatic fare collection by Infare was limited to a shorter time frame (as reported above) because structural changes in the websites of the monitored competitors after February 15, 2008, required a reconfiguration of the bots, leading to a large amount of missing data, rendering the information collected later unusable.

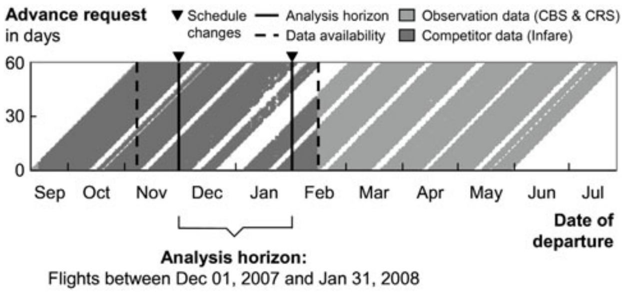


Figure 10.2: Analysis horizon based on data availability
Source: Own design, based on collected data

[1] As the examined market is a short-haul leisure destination, only single-class products (economy class only) are operated and considered in the analysis.

Additionally, major schedule changes were performed twice within the remaining potential analysis horizon: completion of the switch to the 2007/2008 winter schedule and fundamental schedule adjustments towards the end of January 2008. To obtain a consistent and complete dataset, the finally considered analysis horizon has therefore been limited to December 1, 2007 to January 31, 2008 (see Figure 10.2), yielding 4,567 different observation points with a total of 97,637 potential customers logged in the CRS.

The usage of the described compound data source results in some major advantages compared to most other research performed in this area: All data are collected consistently and automatically over time, i.e., a full sample of the truly relevant population is generated online without the need to conduct expensive and time-consuming manual surveys with potential travelers. Also, the obtained data reveal the actual choices (*revealed preferences*) of the respondents, which renders the data less biased compared to results of manual surveys (*stated preferences*) (see, e.g., Gönsch et al., 2008b, p. 415).

Additionally, the data collection could potentially be run in real time within a future productive environment. Analyses based on this approach would be able to cope with the highly dynamic and changing environment of the LCC market (see Section 1.2). Traditional survey-based approaches inevitably explain characteristics of markets that may already have ceased to exist once the data are available.

The exclusive usage of online data naturally also implies some disadvantages that stem from the restricted amount of data that is legally allowed to be collected by companies: No unique respondent identification is possible in all of the data sources, making it impossible to link single entries in multiple databases to identify individuals that appear multiple times within parts of the data and to attach personal characteristics (such as gender or income) to the individual records.

For the analysis in Chapter 11, this means that the population may potentially contain multiple entries for single customers who searched the Internet extensively for cheap airfares, i.e., the system cannot detect that the same individual searched multiple times for different flight alternatives but made only one final decision instead of opting not to buy a ticket multiple times. Also, so-called random taste variation (see Section 9.4.2) might bias the results as no truly personal characteristics can be leveraged to explain the inevitable choice variation of different customers when faced with identical options.

For the same reason, the basic metric here is restricted to transactions or bookings without consideration of the number of actual passengers that are contained in a booking (so-called passenger name record – PNR), which can be different. The number of passengers can be read from the CBS, but cannot be linked to the CRS records due to the lack of a unique record identifier.

The following section aggregates the available data to a market overview in terms of market participants, schedules and pricing behavior.

10.2 Market Overview

This analysis and the model defined in Section 11 are based on the same intra-European short-haul leisure market that was used in Part II. The market has been selected for this work based on its high level of competition and strong customer demand. It is especially appropriate for the analysis of inherent price sensitivity as charter, low-cost and flag carriers compete in this market – mostly via price.

The next sections give an overview of market participants and schedules (Section 10.2.1) as well as the recorded fare levels and supposed market efficiency (Section 10.2.2).

10.2.1 Market Participants and Supply

The analyzed market consists of a short-haul leisure-focused route from mainland Europe to a southern European island. Throughout the considered period, three airlines competed on the route, offering a total of 21 weekly frequencies (each consisting of an out- and inbound flight – seen from the mainland origin). The largest share of flights is thereby offered by a charter and a LCC carrier, which both operate a daily connection early in the morning (outbound) and at midday (inbound) in conjunction with an additional later flight on selected weekdays. The flag carrier operates a single daily flight on weekends only. The entire schedule is illustrated in Figure 10.3.

All carriers operate similarly sized equipment holding 144 seats (A319/ Boeing 737-700) or 174/186 seats (A320/Boeing 737-800). The majority of carrying capacity is offered by the two charter/LCC carriers at similar flight times early in the morning (outbound) and in the afternoon (inbound).

Choice behavior in air travel is typically at least partially driven by flight schedule in relation to an individual's ideal departure time (see Figure 10.4 or Proussaloglou and Koppelman, 1999, directly). As such individual preferred departure times cannot be drawn from the source data (see above discussion),

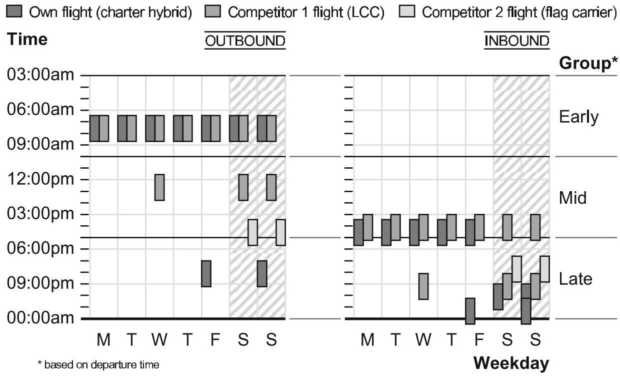


Figure 10.3: Flight schedule on considered route – outbound vs. inbound
Source: Own design

the definition of an exact departure time preference variable is not possible. Additionally, the specific departure times of the charter and LCC flights tend to vary based on operational schedule changes – even across flight numbers. As a result, throughout the following chapters, flight times are considered based on three departure time segments as defined in Figure 10.3.

Overall, the market is an extremely high-performing leisure destination – especially as Figure 10.3 only shows the *winter schedule*. Accordingly, most flights are timed to leave at convenient and competitive times, i.e., early in the morning for outbound flights to allow for a full day at the beach or in the afternoon to allow for a nearly full working day at home; the inbound flights do not depart before afternoon to allow for sufficient checkout time at the hotel and a relaxed transfer to the airport.

The next section examines the pricing environment in a market where multiple players compete head-to-head within tightly aligned schedules from the analyzed carrier's perspective.

10.2.2 Pricing Environment and Behavior

This section provides a short overview of the pricing environment in which the market participants are situated. In particular, insight is generated on fare premiums for selected carriers, flight times or departure dates.

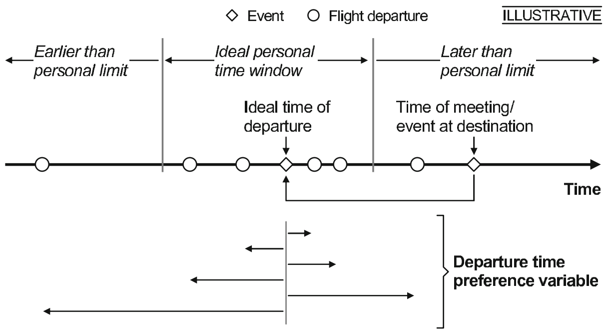


Figure 10.4: Definition of a departure time preference variable
Source: Following Proussaloglou and Koppelman (1999, p. 200)

Note that all fares in Part III are shown as gross fares, including service charges, taxes and applicable fees. For the charter and low-cost carrier, the *lowest available one-way fare* for each departure date is considered (independent of terms that may restrict refund options). For the flag carrier, both one-way and round-trip fares were collected, but the minimum of the lowest available one-way fare and *half of the lowest available round-trip fare* is considered, to account for reduced fares when booking return tickets (for more information on conventional round-trip ticket pricing of incumbent airlines, see, e.g., Pompl, 2006, Chap. 6).

Close examination of Figure 10.5, showing the distribution of the median of the lowest available fares over all market carriers, reveals some interesting facts about the market and the participants' pricing behavior:

- **Low-season pricing:** The overall fare level within the last 60 days before departure is rather low, averaging between 68 and 72 EUR for outbound and inbound flights, respectively (excluding the holiday season). The same observation holds for the fare dynamics (i.e., variation over the 60-day period) as the depicted standard deviation of the median fares is also small (15 to 18 EUR).

- **Outbound focus:** The variation over macro-seasonalities is heavily outbound focused. Besides a steep fare level increase for returning holiday travel following New Year, the average fare level is rather flat on the inbound side while exhibiting strong micro-seasonalities on the outbound side.

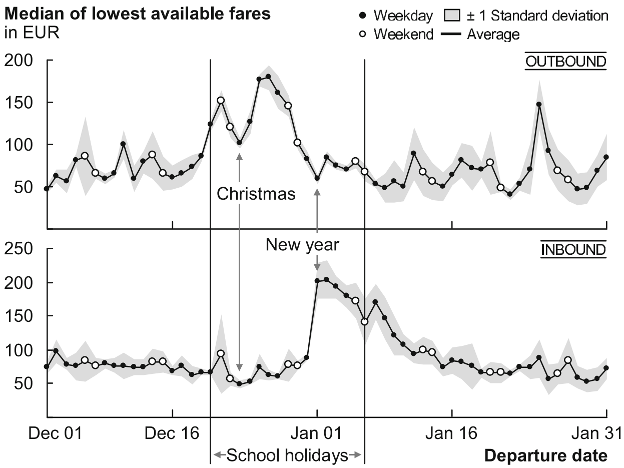


Figure 10.5: Median fares on considered route – outbound vs. inbound
Source: Own design, based on collected data

- **Micro-seasonality:** Mostly, the outbound chart exhibits strong fare level variations based on the prevalent micro-seasonality, e.g., departures towards the weekend appear to be genuinely higher priced. The same holds true for departures between Christmas and New Year's Day.

- **Holiday premium:** Flight departures for Christmas vacation show a hefty price premium for outbound flights *before* New Year's Day and returns *thereafter*. Interestingly, the return price premium does not expire until one week *after* school resumes.

- **Varying dynamics:** While the overall fare fluctuations over the 60 days before departure seem low, as indicated by a low standard deviation, the high-priced outbound flights during the holidays exhibit extremely low variation (nearly no visible variance), indicating that fares were already high early on in the booking period.

To complete the overview of pricing behavior, Figure 10.6 illustrates the individual deviation of each carrier's lowest available fare from the market median fare for each departure time segment. The figure shows slightly modified box-plots, indicating only the 5^{th} and 95^{th} percentiles at the outer ends to adjust for distant fare outliers.

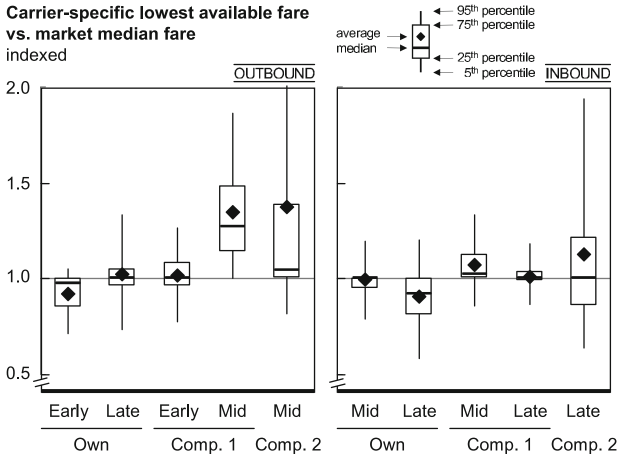


Figure 10.6: Fare distribution on considered route – outbound vs. inbound
Source: Own design, based on collected data

Again, it is apparent that the fare variation for the outbound flights is greater than for the inbound flights. Competitor 2 (the flag carrier) exhibits the most variation and a rather high average deviation from the market mean, which results from punctual high fare levels based on strong general network demand (e.g., via hub connections to overseas destinations).

Overall, the market seems quite efficient, with most fares close to the market mean. Competitor 1 seems to collect a fare premium over the analyzed carrier for flights within the same time segment. Moreover, competitor 1 extracts a considerable fare premium on its midday departures, especially on the outbound direction. Besides a simple brand premium, this could also be caused by schedule differences (mainly weekend departures on the midday outbound flight).

Following this brief overview of the market environment, the next section describes the resulting demand behavior as observed by the considered carrier based on the newly constructed dataset.

10.3 Observed Demand Behavior

This section reports the observed revealed preferences of potential customers whose fare requests arrive in the CRS of the airline under study. The relevant metric by which to measure customer behavior is the so-called *book-to-look ratio* ω, which is the fraction of latent demand that actually results in a buying decision (see Section 3.4 in Part II)

$$\omega = \frac{d}{D}. \tag{10.1}$$

As a reminder, the objective of this part (in particular, Chapter 11) is to develop a model $\tilde{g}(\cdot)$ that approximates the true functional form of ω

$$\omega = g(\cdot) \approx \tilde{g}(\cdot). \tag{10.2}$$

The individual sections below examine potential one-dimensional drivers for $\tilde{g}(\cdot)$ that can be directly observed by a single carrier operating in the market. The objective here is to discover promising attributes for the final multinomial logit model later developed in Chapter 11, which then computes the full composition of these drivers to derive the book-to-look ratio ω.

First, Section 10.3.1 examines the expectable single-dimensional price sensitivity of prospective customers in relation to the airfares of different carriers. Section 10.3.2 then discusses the supplemental schedule effects of different available departure times and the micro-seasonal effects of different departure weekdays. Finally, Section 10.3.3 examines the variation in customer behavior based on the specific weekdays of fare request and booking.

10.3.1 Price Sensitivity

This section takes a brief look at the most obvious and intuitive driver behind the book-to-look ratio – the price level of the lowest available fare. The analysis is neither intended to be exhaustive nor to actually specify the influence of a particular carrier's ticket price alone, but rather to reveal evidence for its impact and to highlight the interdependencies of prices in this competitive market setting.

Standard microeconomic concepts (see, e.g., Varian, 1992) would suggest a demand curve with negative slope for the book-to-look ratio over increasing levels of lowest available fare for a specific flight: following the "law of demand", realized demand of a specific flight drops as its fare level increases.

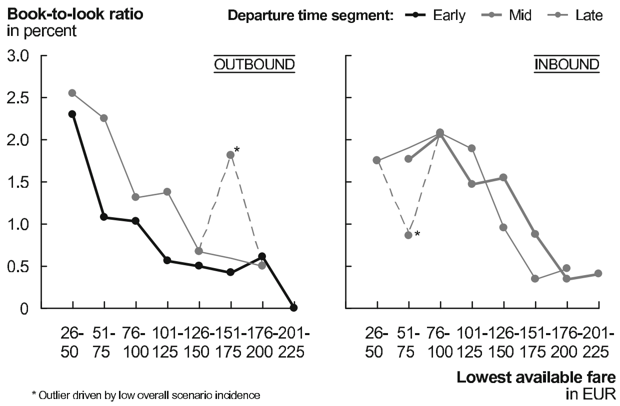


Figure 10.7: Demand curves for own flights – outbound vs. inbound
Source: Own design, based on collected data

Naturally, such a demand curve can only be constructed for products or services where the purchasing behavior is known or can be observed. Accordingly, Figure 10.7 only depicts the demand curves for the flights offered by the analyzed carrier.[2]

First of all, the demand curves exhibit the expected (downward sloping) shape. However, the charts differ greatly between outbound and inbound travel and relay additional information about the specific relationship between book-to-look ratio and fare level that may prove important for the later model development in Chapter 11:

- **Outbound focus:** In accordance with the findings in Figure 10.5, the price sensitivity and reaction of customers is highest in the outbound direction. First, book-to-look ratios decrease steeply on the outbound side as soon as fares exceed bargain levels. Second, book-to-look ratios are much lower in general on the inbound side, indicating higher browsing rates, especially as the total count of bookings is rather stable while the number of fare inquiries is high (not shown in the figure). In conjunction with the findings from Figure 10.5 (i.e., the generally higher level of inbound fares), this explains the third insight, namely that the decline of inbound price sensitivity starts at slightly higher fare levels than is the case for the outbound flights.

[2] To enhance readability, book-to-look ratios are depicted over fare groups of EUR 25 and results are shown as line plots.

- **Decreasing price sensitivity:** In both flight directions, price sensitivity does not seem to decrease linearly. The rate of decrease in book-to-look ratio is steepest between low fare level groups and flattens toward the truly expensive tickets. This pattern is in line with intuition and the findings of many works that a 50 EUR price increase triggers different customer reactions when launched on top of a 50 EUR base price than on a 200 EUR base fare.

- **Time preference:** The observed purchasing behavior clearly reveals preferences for departure time segments: While on outbound flights, the late departure is favored over all fare levels, in the inbound direction, midday departures are preferred, especially during peak times when fares are generally high.

In the considered competitive setting (see Section 10.2.1), the discussed demand curves depict the average customer behavior without considering the influences of specific competitive pricing measures, i.e., the observed average book-to-look ratio at a selected fare supposedly disintegrates into a lower rate if the competition's fares are cheaper and a higher rate if competition's fares are more expensive. The magnitude of a specific competitor's influence depends on a variety of characteristics, such as brand, schedule, etc., which are external to the analyzed carrier, but need to be reflected later in the model (see Chapter 11).

These cross-influences are multi-dimensional and cannot be illustrated meaningfully, even in the rather well-arranged market setting here. However, Figure 10.8 provides an example of the cross-effect of a competitor's early and midday departure fares (y-axis) on the book-to-look ratio depending on the analyzed carrier's early departure fares (x-axis).

While the graph shows only a two-dimensional example of how purchasing behavior is affected under different fare combinations of the considered airline and one of its competitors, some interesting insights with relevance for the later model in Chapter 11 can still be derived:

- **Efficient market:** In the two examples in Figure 10.8, dots and bubbles indicate the fare combinations that were actually observed in the data. All observed combinations reside along the diagonal line of equal fare levels (besides the situation when no competing flight is available). While the comparison of the early departures exhibits varying levels around both sides of the parity, the midday departure of competitor 1 consistently exhibits a premium, supposedly based on the preferential departure time.

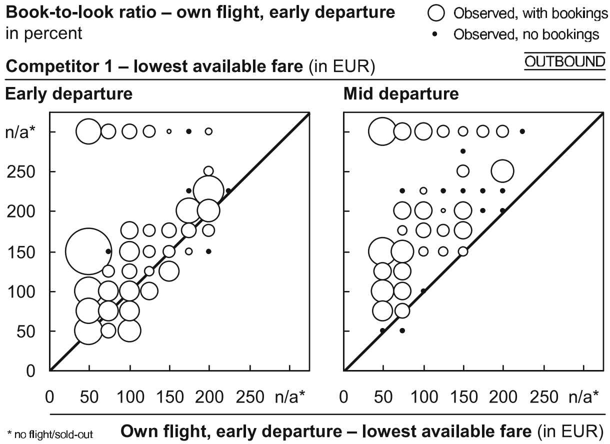


Figure 10.8: Demand sensitivity under competition
Source: Own design, based on collected data

- **Limited premiums:** The observed premiums are limited in magnitude in both examples. While competitor 1 apparently tries to achieve a sizable premium for its midday departure, the feasible deviations from the parity line are more limited on the aligned early departure times. Here, both competitors mirror their pricing more closely.

- **Varying relevant competition:** The alignment of the fares and the possible premiums also gives an indication of the relevant competition in terms of carrier and departure time segment. For the early departure time segment, the fares of competitor 1 seem relevant to the customers of the considered airline as the book-to-look ratio varies greatly over combinations. On the other hand, the fares for the midday departure of competitor 1 seem to have no visible effect on booking rates, despite being considerably higher.

This section has shown that besides the analyzed carrier's own fares, such of relevant competitors are promising candidates for the later development of the decision model in Chapter 11. Their influence may vary depending on the competitiveness of the departure time segments as well as the overall fare level in the market.

The next sections examine schedule and booking day preferences that may amend the reported price influence.

10.3.2 Schedule Preference

Section 10.3.1 above sheds light on the most intuitive driver behind purchasing behavior – the level of the lowest available fare. While in price-sensitive markets, like low-cost travel, price is typically a major factor, customers also tend to have specific preferences for travel dates depending on the expected duration of their stay (e.g., two-week vacation vs. weekend stay) and trip occasion (vacation vs. visiting friends and relatives).

Differences in schedule preference may manifest in macro- and micro-seasonal variances of the book-to-look ratio. Part II has shown that overall demand already exhibits macro-seasonal variation, resulting in higher traffic during vacations, summer or special events (see Figure 10.9).

The generally higher latent demand during macro-seasonal peaks affects the resulting book-to-look ratio in two ways: First, as total capacity is fixed, book-to-look ratios become limited when flights reach capacity. Second, in an efficient market, the available fare (not shown in Figure 10.9) reflects the prevalent latent demand or the pricing analyst's expectation about it and tends to be higher in peak times, pushing down book-to-look ratios.

Henceforth, in later model development, the effective need for an additional macro-seasonal variable (like departure date or season) has to be carefully evaluated, as it may already be contained in the price variables.

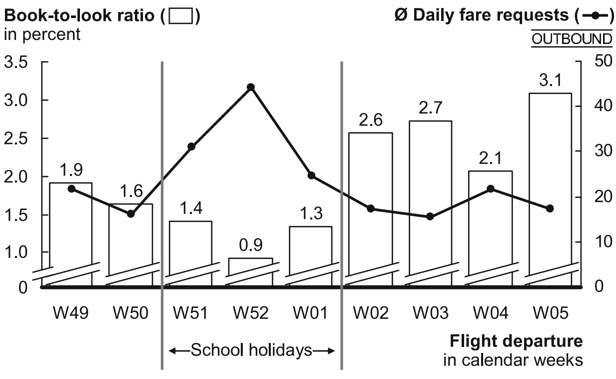


Figure 10.9: Macro-seasonal influences on book-to-look ratio
Source: Own design, based on collected data

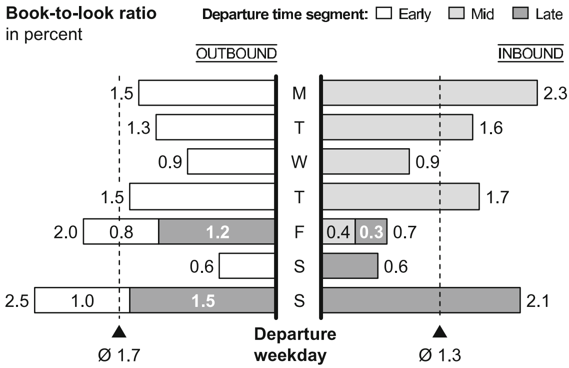


Figure 10.10: Micro-seasonal influences on book-to-look ratio
Source: Own design, based on collected data

In addition, it was established in Part II that micro-seasonal demand variations lead to higher latent demand on attractive weekdays (e.g., Fridays for outbound flights and Sundays for returns). Figure 10.10 depicts the resulting micro-seasonal comparison of book-to-look ratios for departure weekdays.

Here, the ratios are visibly elevated on the supposedly attractive days. Additional large differences between adjacent weekdays make it unlikely that the entire effect can be contained and steered through pricing measures alone. Especially early on in the booking period, entry fares may not be differentiated enough to drive such strong differences between weekdays.

Obviously, the mere presence of a second flight operated by the analyzed carrier considerably limits the book-to-look ratio of each individual flight (especially when flight departure times are perceived as less convenient). Only if there is considerable additional demand (as on Fridays and Sundays in the outbound direction), this effect can be mitigated.

Similar to the mentioned macro-seasonalities, the model design process in Chapter 11 must carefully evaluate whether additional variables representing the micro-seasonal variations are necessary. The observed effect again may well be only partially contained in fare levels.

After the analysis of choice variations over departure date seasonality, the next section investigates additional micro-seasonal effects based on potential booking day preferences.

10.3.3 Booking Day Preference

The last dimension examined in this part that may affect customer choice behavior is the booking weekday. While the possibility to book a flight on a specific weekday may not directly induce additional utility for a customer, particular customer segments (exhibiting a general differential utility for the same flight) may be prone to specific booking days. At the same time, there exist no reports of airlines capitalizing on that potential effect by varying their fares over different weekdays.

Figure 10.11 shows the book-to-look ratios over different booking weekdays, which fluctuate heavily for both flight directions. Section 5.3.3 in Part II has already reported that lower levels of latent demand are expressed on weekends. For the outbound direction, this apparently gets expedited by a lower book-to-look ratio to lower realized demand, i.e., the few customers that shop on weekends also tend to book less. For the inbound direction, the weekdays with low conversion rates seem to be shifted toward the beginning of the week (Sunday to Tuesday).

The observed effect could possibly be attributed to varying fare levels and the resulting competitiveness throughout the week. In this scenario, pricing analysts would monitor outbound flights early in the week, adjusting fares to competitive levels that drift off (e.g., driven by received bookings and fare adjustments by the competition) toward the end of the week. Inbound

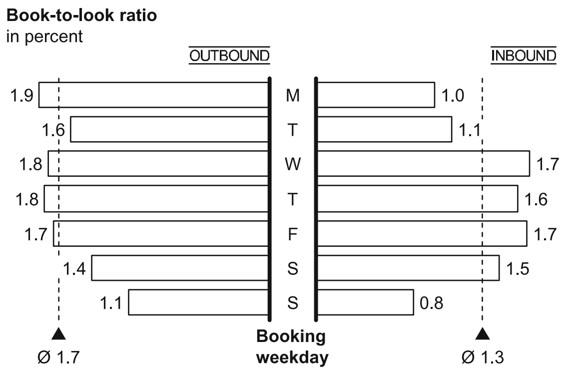


Figure 10.11: Booking day influence on book-to-look ratio
Source: Own design, based on collected data

flights would accordingly be monitored at the end of the work week, yielding increasingly competitive fare levels toward the weekend, with declining competitiveness toward the start of the next week. This could also explain the gradual decline with sharp rebound on a specific weekday.

As the observed effect in Figure 10.11 could be triggered by various other drivers (besides the mentioned fare adjustment schedule); it cannot be attributed to a specific cause at this stage, but must be evaluated more closely in the next chapter when the model is composed and reviewed.

In sum, a series of potential drivers for realized demand (measured by book-to-look ratio) has been analyzed above. The next section summarizes their possible implications on the model building process and the final model itself.

10.4 Implications for Choice Model

This section summarizes the findings on the choice data's underlying characteristics and potential demand drivers, which have been developed in Sections 10.2 and 10.3, with the aim to outline implications for the construction of the appropriate discrete choice model in Chapter 11.

Accordingly, the following guidelines and limitations can be derived for the creation of the choice model based on the analyses discussed above:

- **Transaction-based metric:** The relevant metrics to record observations and corresponding choices of customers are limited to transactions or bookings. Actual passenger numbers cannot be attributed to observations that do not result in completed bookings and henceforth cannot be used consistently throughout the dataset. Based on the analyzed market, only publicly available economy class tickets are considered.[3]

- **Directional model differentiation:** Many observed effects and potential driver variables vary widely between out- and inbound flights. Thus, separate models should be created to evaluate possible differences in the coefficient weights of joint drivers and to allow for the evaluation of potentially differing models based on their variable compositions.

- **Segmented departure time variable:** Flight departure times should be reflected in a segmentation by departure time according to the in-

[3] Business class seats are not available on the chosen route, and special fares and conditions for employers are not included in the analysis.

troduced time segments *early*, *midday* and *late* (see Figure 10.3), as any particular flight number is not the true source of utility and may also not be consistently bound to a fixed departure time.

This construction can also bridge the gap created by the lack of data on the customers' preferred or ideal departure times by directing their preferences to a limited number of segments instead. The utility of individual alternatives may then well prove to be unaffected by attributes that are bound to selected departure time segments.

- **Varying brand relevance:** The specific utility of the three different brands may also vary between flight directions and time segments. As the pure brand value does not lend itself to a specific metric, differences should again be reflected in an appropriate segmentation where brand value could, for instance, be reflected in fare coefficients that are carrier-specific.

- **Non-linear price effect:** The undeniably important fare-based utility drivers may exhibit a non-linear relationship with increases from lower base fares, inducing a stronger utility shift than from initially high fares. The potential effect should be tested and, if found to be significant, incorporated into the model through an adequate transformation of the original source variable.

- **Micro- vs. macro-seasonal drivers:** A range of micro-seasonal variables (i.e., weekday of flight departure and weekday of booking) must be evaluated for individual significance and whether their expected influences are mitigated by varying fare levels. The effect of holiday departures may well be contained in generally higher fare levels throughout the booking period.

  Whether the overall macro-seasonal variation in choice behavior must be captured by a separate driver has also to be tested. The potential variable should be based on an adequate transformation of the overall latent demand described in Part II, which already captures macro-seasonal demand variations.

- **Limited customer characteristics:** Individual customer characteristics cannot be included based on the available data. Therefore, the developed model does not allow for the consideration of such effects, but should be evaluated against the possible error or bias induced by this limitation. Future work should also try to augment the employed datasets with such individual attributes as exhaustively as possible.

The following Chapter 11 develops a proprietary discrete choice model based on the multinomial logit explained in Chapter 9 that draws from the discussed structural findings of this chapter.

Chapter 11

Multinomial Logit Model for Low-Cost Travel Choice

he objective of this chapter is to develop a multinomial logit model, as introduced in Chapter 9. The model simulates low-cost travel choice based on the market's demand specifics as identified in Chapter 10.

In his Nobel lecture, McFadden (2000b) illustrated the contemporary view of the different elements of a choice process (see Figure 11.1) based on both, the traditional economic view that is rooted in *rational choice behavior* following the evaluation of attributes and less stringent *psychological factors*

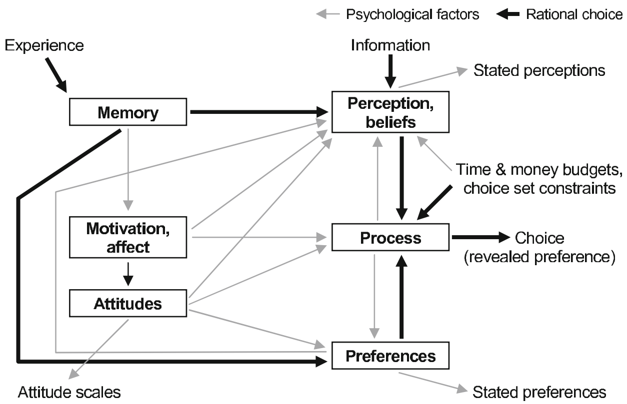


Figure 11.1: Elements in a contemporary view of the theory of choice
Source: Redrawn from McFadden (2000b, p. 336)

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6_11,

that also drive decision making, but in a more non-linear way. The picture highlights that choice or *revealed preference* may differ from what is *stated* based on experience or perception. For this reason, meaningful choice models should be based on revealed preference data – as is the case in this work.

Additionally, the apparent complexity of the decision process typically prohibits the immediate identification of all relevant decision drivers and their assembly into a meaningful model. Therefore, the development process of customer choice models is supposed to be twofold cyclical, as shown in Figure 11.2.

Naturally, a model's adherence to its underlying assumptions and the specifics of the employed methodology must be tested. Additionally, intermediate results and conclusions might lead to the adjustment of such initial assumptions and expectations about customer choice behavior.

Before delving into model design, the next section discusses the modeling constraints that arise from the data availability particular to this work (see Section 10.1), in the sense that only part of the overall decision process as depicted in Figure 11.1 can be observed. Then, Section 11.2 reflects on the iterative model building process as shown in Figure 11.2, showing the intermediate steps and reasoning behind the chosen specifications.

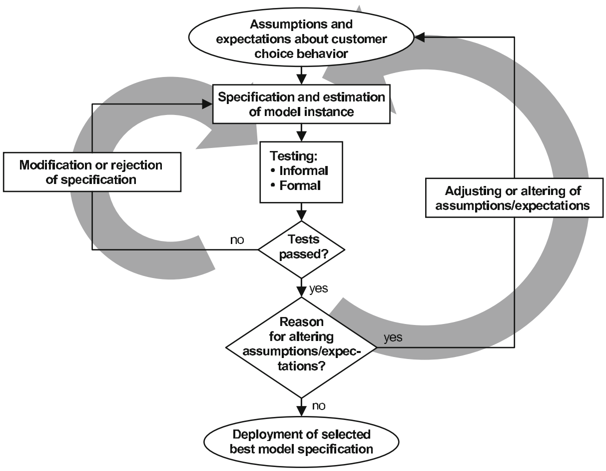


Figure 11.2: Choice model development process
Source: Adapted from Gönsch et al. (2008b, p. 413), originally based on Ben-Akiva and Lerman (1985, pp. 154)

11.1 Modeling Constraints and Specifics

This section describes the modeling constraints that arise from the limitations of the available dataset, which has been introduced in Section 10.1.

Section 10.1 highlights that the available data sample contains revealed preference data (i.e., is unconstrained) of air travel customers exhibiting a latent flight demand together with their actual choices when a flight was booked with the analyzed carrier. As the data were collected automatically from existing systems, its granularity is limited to transactions or bookings (i.e., the number of passengers in a PNR is not known). Additionally, requests and bookings cannot be directly related to one another on an individual basis, but only on the "aggregate" level of single days. Data were collected from the perspective of a single carrier, and therefore no choice information about the competition is available; nor are individual customer characteristics (e.g., socio-demographics or preferred travel time). Based on the route chosen for the analysis, data attributes are based on one-way travel in economy class.

Figure 11.3 illustrates the described data availability based on a typical view of the decision maker on the available choice set: A prospective customer queries the website of the considered carrier for flight availability and fares on a specific route. The computer reservation system (CRS) logs the request

Number of transactions (CRS)/ number of bookings (CBS)*

Data availability: Full (choices and attributes); Partial (attributes only)

OUTBOUND

Data type	Tracking system	Choice set		Departure date 11/13	11/14	11/15	11/16	11/17	11/18	11/19
Observation	CRS			116	121	120	134	128	114	115
Choice	CBS	Consid. carrier	Early	16	0	12	15	9	16	11
			Late	–	–	–	17	–	22	–
		Comp. carrier 1	Early	?	?	?	?	?	?	?
			Mid	–	?	–	–	?	?	–
		Comp. carrier 2	Late	–	–	–	–	?	?	–
		Other/no-buy		?	?	?	?	?	?	?

* Combinations marked with "–" do not exist/are not available

Figure 11.3: Data availability along customer choice set view
Source: Own design

with the corresponding desired departure date (here, November 16). Up to this point, no information about preferred departure time or the size of the traveling party is available. The decision maker can then view the available flights in the so-called fare display, which typically also includes adjacent departure dates. As she is assumed to also query the websites of both competitors (which is realistic based on their strong brand awareness), the full choice set of five different flights on the requested date, including associated fares, is available to her.

If a booking is made based on the initial query, it is recorded in the central booking system (CBS). However, as the latter is not directly linked to the CRS, the transfer cannot be traced, i.e., a query for a specific date could potentially result in a booking for other displayed flight departures. Also, if the booking is instead made at one of the competitors' websites, it is naturally not recorded in the considered carrier's own CBS and is thus invisible to it. That is, for requests in the CRS that do not materialize as bookings in the carrier's own CBS, it cannot be said whether these were lost to competition or if the decision maker simply did not purchase a ticket at all.

At the same time, a large range of alternative attributes that could possibly be relevant to the decision maker (see Table 11.1) can be observed by the considered airline, either through publicly available sources like the OAG database (OAG, 2009, for schedule and equipment data) or through service firms that provide customized proprietary information (Infare, 2009, for airline fare data).

	Modeling options			
Variable	Alternative definition	Alternative attribute	Decision maker characteristic	Segmentation of alternatives
Brand	✓			
Punctuality	✓			
Ticket fare		✓		
Advance purchase		(✓)	✓	
Departure weekday		(✓)	✓	
Booking weekday		(✓)	✓	
Departure time				✓
Flight time	not relevant on the considered route, as fundamentally equal across alternatives – independent of carrier.			
Aircraft type				
Stops				

Table 11.1: Utility drivers with possible modeling options

Table 11.1 lists the utility drivers that are considered in this work based on the available data, together with the choices of how they could be modeled:

- **Alternative definition:** Airline brand and punctuality are not modeled as separate attributes (not least because both are difficult to quantify meaningfully on a numeric scale), but are represented indirectly through the definition of the individual alternatives. That is, the alternative of choosing a specific flight of a competing carrier implicitly contains the utility effect of its brand and punctuality.

- **Alternative attributes:** Ticket fare and departure weekday are obvious attributes of a specific choice alternative, i.e., both describe a specific flight in more detail, what might affect the choice decision. However, the departure weekday could also be regarded as a decision maker characteristic depending on whether the prospective customer prefers that particular day, i.e., carries it as a personal preference attribute.

- **Decision maker characteristics:** Advance purchase time and booking weekday describe the decision maker in terms of her preferred booking and choice behavior. However, depending on the specific definitions of alternatives, both could also be regarded as alternative attributes, describing the choice to book a specific flight on a specific day a selected number of days in advance.

- **Segmentation of alternatives:** As explained in Section 10.2.1, the actual departure times of flights can only be modeled based on time segments because the actual preferred departure time of a customer cannot be known and departure times often vary slightly in the considered market, the latter being considered irrelevant to the overall choice.

Based on the described choice scenario (see Figure 11.3) and the prospective utility drivers (see Table 11.1), an intuitive and feasible choice set definition would consist of two alternative choices for the considered carrier (segmented by time) plus an additional option containing both, all other choices that cannot be monitored separately and the obligatory no-buy decision (see Figure 11.4).

As illustrated in the figure, the individual fares would be assigned as alternative attributes according to the choice set definition, leaving the "other/no-buy"-option with three attached fare attributes, possibly exhibiting differing weights. The mappings of departure weekday, booking weekday and advance purchase as decision maker characteristics would depend on their actual effects on the choice probabilities and thus would have to be tested for particular significance.

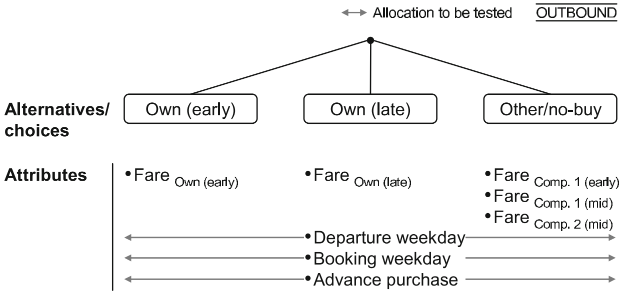


Figure 11.4: Intuitive definition of choice set structure
Source: Own design

Modeling attempts and corresponding tests reveal the obvious at second sight: The pictured structure cannot be in line with the IIA property as described in Section 9.4.2. The IIA directly implies that an increased choice probability of one alternative proportionally draws from all other alternatives as their choice ratio must remain constant, independent of the "irrelevant" alternative that has changed.

However, in the considered scenario, if competitor 1 lowers its fare on the early departure, supposedly leading to a higher choice probability of the "other/no-buy"-option, the choice probabilities of the considered carrier's departures would most likely not decrease proportionally, but rather the early departure would be more affected, being a simultaneous flight.

Following Timmermans et al. (1992, pp. 180), the IIA property of the conventional MNL configuration may generally be violated if at least one of the following assumptions is disregarded:

1. The error terms of the individual utility functions are IID.

2. The choice alternatives' systematic utilities are a function of their own attributes and those of the decision maker only.

3. Decision makers process and evaluate the attributes of available alternatives simultaneously and not in a sequential or hierarchical form.

In the considered case, assumptions 2 and 3 may not hold true: As described, utility of the considered carrier's own early departure may vary overproportionally with the fare of the early departure of competitor 1 contained in the "other/no-buy"-alternative. Also in reality, the choice process might likely be hierarchic or nested (as depicted in Figure 11.5). That is, the decision maker potentially has a clear preference for a departure time segment based on personal circumstances or characteristics (e.g., whether she wants to travel at the end of the work day or instead spend a full day at the beach). Within that time segment, she chooses between available alternatives (i.e., carriers and flights), and at that lower level, the IAA property may well hold: if a certain carrier lowers its fare on the early departure, its increasing choice probability proportionally affects all other departures in the same time segment. Similar argumentation holds for the higher level: a genuinely more attractive early departure segment might proportionally attract travelers from others.

As indicated in Figure 11.5, the available choice data unfortunately do not allow for modeling and testing the potentially most appropriate nested logit (NL) model structure (see Train et al., 1989, Sec. III).

Standard MNL construction rules, as introduced in Chapter 9, limit the attribute vector $\mathbf{x}_i$ of an alternative i to depend *only* on attributes of itself as well as to possibly interact with decision maker characteristics $\mathbf{s}_n$. That is, $\mathbf{x}_i$ does not depend on any attributes of alternatives other than i (see McFadden, 1984, Sec. 3.5).

However, when the vector of alternative attributes is allowed to be a function of *all* available alternatives, the resulting more "general MNL form is sufficiently flexible to approximate any continuous positive response probability model on a compact set of the explanatory variables" (McFadden, 1984, p. 1414). This approximation has been termed the *universal* or *mother logit* (see, e.g., Koppelman and Sethi, 2000, Sec. 3.3) and was initially developed in McFadden (1975) – for details, see McFadden (1981). In practice, the

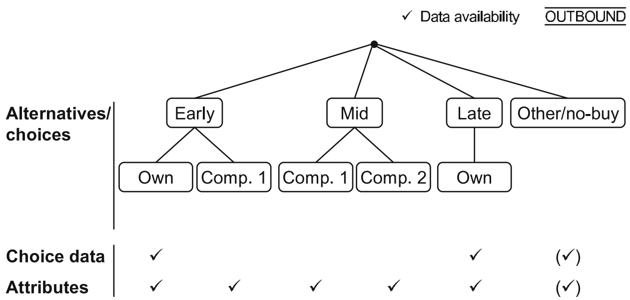


Figure 11.5: Data availability in nested choice set structure
Source: Own design

application of the universal logit may be computationally infeasible or inefficient as the increased number of variables may leave the model unidentified.[1] Naturally, through the bypass of the IIA property, universal logit models can also not be efficiently estimated based on alternative subsets (see McFadden, 1984, Sec. 3.5).

The actual behavior of a model specification depends on the particular definition of the alternatives' representative utility. "Thus, the IIA property might be valid for one specification of representative utility and not for another, even though both specifications relate to the same choice situation" (McFadden et al., 1977, p. 41).

Figure 11.6 depicts the revised universal model structure used in this work. Here, the competitor fares can potentially be included in any of the systematic utilities of the considered carrier's two departure alternatives, but the construction is clearly identified. However, the factual mapping must be evaluated and tested during model building. Intuitively, the fares will likely affect more strongly the particular choice that falls into the same departure time segment.

Note that the competitor fares cannot be included in the systematic utility of all choice alternatives, as this again would leave the model unidentified (see above).

Alternatively, the fares may affect both alternatives of the considered carrier, but to a nonuniform extent; that is, competitors' fares for flights within

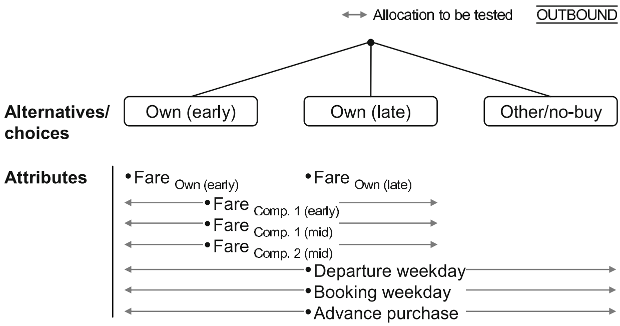


Figure 11.6: Revised universal definition of choice set structure
Source: Own design

[1] Naturally, the inclusion of all alternative attributes into the systematic utility of all other alternatives yields an unidentified model with a heavily increased variable count.

the same departure time segment may affect utility more than those in others. For example, consumer goods marketing studies have shown that such *cross-price asymmetries* do also exist between national value brands and private label brands (see, e.g., Blattberg and Wisniewski, 1989) or between different price tiers, i.e., low priced vs. premium products (see, e.g., Allenby, 1989).

Evaluations and tests based on the revised approach in Figure 11.6 have produced promising results, which are described and discussed exhaustively in the following section.

11.2 Model Building and Goodness of Fit

This section describes the model building process, including the necessary informal and formal tests depicted in above Figure 11.2. Where appropriate, the impact of possible variable transformations (i.e., ticket fare and advance request time) is evaluated as well. The described models are numbered chronologically to ease comparison and testing.

All multinomial logit models in this work have been estimated and evaluated using Michel Bierlaire's free estimation package for generalized extreme value models, BIOGEME Ver. 1.8 (for download and an introduction, see Bierlaire, 2003, 2009).

The overall model evaluation follows a three-step approach: First, informal tests (magnitude and directional sign of coefficients) are performed to assure the model's basic conformity with intuition and reality. Second, standard t-tests and asymptotic confidence intervals (see Section 9.4.4) are used to evaluate the explanatory value of single attributes or characteristics. Finally, the Likelihood Ratio test is employed to test differences in the goodness of fit between nested models, while the Akaike Information Criterion (AIC) is genuinely used to confirm these results and to compare models that are not nested by linear restrictions (as required for the Likelihood Ratio test).

The model development process is discussed in detail based on the outbound direction of flights as these exhibit more variation in the data (see Section 10.2). For the inbound direction, only those intermediate steps that exhibit divergent results in Section 11.2.4 are presented.

Modeling starts with internal choice drivers that can be controlled by the considered carrier in Section 11.2.1, then includes those variables that are modeled as decision maker characteristics in Section 11.2.2 and finally includes the appropriate allocation of externally controlled drivers (i.e., those of the competition) in Sections 11.2.3 and 11.2.4.

11.2.1 Internal Choice Drivers

This section examines the explanatory value of own ticket fares for customer choice decisions. The starting point is a model based on alternative specific constants only, which represents the average utility of each alternative.

Note that throughout the following sections, the modeled alternatives are indexed with e and l for the early and late departures of the considered carrier, respectively. The index o marks the third alternative joining the available other/no-buy choices, so that $C = \{e, l, o\}$. The index m is used to identify variables belonging to the competitor's midday flights.

Alternative Specific Constants Only

$$\begin{aligned} V_n^o &= \alpha^o \\ V_n^e &= \alpha^e \\ V_n^l &= \alpha^l \end{aligned} \qquad \text{(Model 01)}$$

Model 01 above consists solely of alternative specific constants and henceforth mirrors the average choice behavior in case all alternatives are available to the decision maker, i.e., it resembles the overall market share of choices.

To yield an identified and assessable model, one of the constants α^i has to be fixed to a predefined value or scale.[2] Throughout this work, the alternative specific constant for the third alternative is fixed to zero $\alpha^o = 0$. Besides scaling the model, this also fosters interpretation of the results as the utilities of the remaining alternatives then describe the utility gain (or loss) when purchasing a ticket from the considered carrier in comparison to buying from the competition or not traveling at all. As the coefficient is genuinely zero, it is not reported for brevity.

The following tables reporting the estimation results all show the estimates (i.e., the expectancy) for the named coefficients together with their standard deviation. Under the normal distribution assumption, this entails

[2] Otherwise, the overall level of representative utility could be raised and lowered arbitrarily, which would not allow for an unambiguous estimation of the coefficients, i.e., the model would not be identified.

a significance measure on how certain the coefficient is different from zero (i.e., is in fact of explanatory value).[3]

Early departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^e	−2.750000	0.015200	0.00	α^l	−2.350000	0.021600	0.00

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.2: Estimation results for Model 01 (alternative specific constants)

The estimation results in Table 11.2 illustrate that the joint utility of buying a ticket at a competitor or not buying a ticket at all is highest, yielding a choice probability of $\approx 86\%$ if all alternatives are available. Similarly, when available, the late flight departure, with a choice probability of $\approx 8\%$, seems more attractive compared to the early departure ($\approx 6\%$).[4]

The Log-Likelihood value of this very basic model is $-25,468$ based on $97,637$ observations and two estimated parameters. Caused by the high choice probability of the "other/no-buy"-alternative, the Likelihood Ratio Index ρ^2 (see Section 9.4.4) always yields a comparably high level, and therefore it is not further evaluated as its explanatory value is low. Subsequent model comparisons are therefore based on the Likelihood Ratio test (on nested models) and the AIC (on all models).

The consequent next step is to understand the impact of the considered carrier's own ticket fares on the utility of its flights. The fares are expected to negatively affect utility, and the magnitudes of the coefficients may well vary between departure time segments.

Inclusion of Own Fares

$$\begin{aligned} V_n^o &= \alpha^o \\ V_n^e &= \alpha^e + \beta_{p_e}^e \cdot p_e \\ V_n^l &= \alpha^l + \beta_{p_l}^l \cdot p_l \end{aligned} \qquad \text{(Model 02)}$$

[3] The reported significance measure indicates the level of α up to which the estimate can be considered significantly different from zero. For example, a significance of 0.24 denotes that for any $\alpha < 0.24$ the coefficient is *not* significantly different from zero.

[4] The choice probabilities can be reproduced by substituting the coefficients in Model 01 with the estimates in Table 11.2 and using the resulting representative utility to calculate the MNL choice probability according to (9.11).

The considered carrier's own ticket fares are included as p_i with individual coefficients $\beta^i_{p_i}$ for the corresponding choice alternatives $i \in \{e, l\}$. Table 11.3 shows the estimation results for Model 02.

Early departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^e	-2.070000	0.037000	0.00	α^l	-1.890000	0.055600	0.00
$\beta^e_{p_e}$	-0.009620	0.000511	0.00	$\beta^l_{p_l}$	-0.005350	0.000614	0.00

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.3: Estimation results for Model 02 (inclusion of own fares)

The new coefficients clearly exhibit the expected negative sign and also a larger effect of fare increases on the early departure $\beta^e_{p_e} < \beta^l_{p_l}$, which is in line with the observed choice probabilities in Section 10.3.1.

The reduction in the alternative specific constants indicates that some of the formerly average utility of an alternative is now explained by fare variation and is therefore contained in $\beta^i_{p_i}$ rather than in α^i.

The new model Model 02 also exhibits improved statistical performance over Model 01, which comes as no surprise (see Table 11.4). That is, both the Likelihood Ratio test and AIC suggest that including the fares significantly (at a level of $\alpha = 5\%$) improves the model's accuracy.

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{2,0.975}$	Evaluation
02	$-25,220$	4	$50,449$	Model 02 clearly better	494.676	7.378	Model 02 clearly better
01	$-25,468$	2	$50,940$				

Table 11.4: Comparison Model 02 vs. Model 01

However, Model 02 assumes that fare changes affect the utility linearly, i.e., an increase of 10 EUR for the early departure (all else equal) reduces a customer's specific utility by 0.0962 points, independent of the prevalent fare level. With typical fares ranging from 25 – 200 EUR on the considered routes, this assumption does not seem realistic.

Intuition would suggest the disutility of fare changes to decline in absolute fare level. That is, on attractive and highly priced flights, a fare increase of 10 EUR yields less disutility than on those sold at bargain fares. The described effect should result in a non-linear influence of p_i.

If fare changes do in fact interact with utility non-linearly, the encountered functional dependence can be linearly approximated by replacing the original fare variables p_i with a set of separating variables (see Section 9.4.4 or Ben-Akiva and Lerman, 1985, pp. 174), which replace the single fare variable in Model 03 with a piecewise linear substitution

$$p_{i_k} = \begin{cases} p_i & \text{if } (k-1)\cdot 25 < p_i \leqslant k \cdot 25 \\ & \text{with } k \in \{1,\ldots,8\}, \\ 0 & \text{otherwise.} \end{cases} \qquad i \in \{e, l\} \tag{11.1}$$

Linearly Approximated Fares

$$\begin{aligned} V_n^o &= \alpha^o \\ V_n^e &= \alpha^e + \sum_{k=1}^{8} \beta^e_{p_{e_k}} \cdot p_{e_k} \\ V_n^l &= \alpha^l + \sum_{k=1}^{8} \beta^l_{p_{l_k}} \cdot p_{l_k} \end{aligned} \qquad \text{(Model 03)}$$

The adjusted model now includes the full set of newly defined fare variables from (11.1), instead of a single uniform fare variable. If the assumption of a non-linear fare effect is correct, the coefficient values should vary and tend toward zero for higher fare levels.

The estimation results, which are shown in Figure 11.7, confirm the expected effect, which also appears to be stronger on the genuinely more price-sensitive early departure.

The graphic depicts the estimates for $\beta^i_{p_{i_k}}$ together with the coefficient's confidence interval, which allows a direct interpretation of the standard deviation (i.e., the interval width directly depends on the standard deviation and the chosen α – see Section 9.4.4) as well as lending intuitive access to the level of significance (i.e., if a particular confidence interval includes zero, the estimate is not significant). All but one of the coefficients are significant at $\alpha = 5\%$, with the large confidence interval of p_{e_8} (fares $>$ 200 EUR) on the early departure being due to the low incidence.

Also, the statistical performance of Model 03 has improved over Model 02,

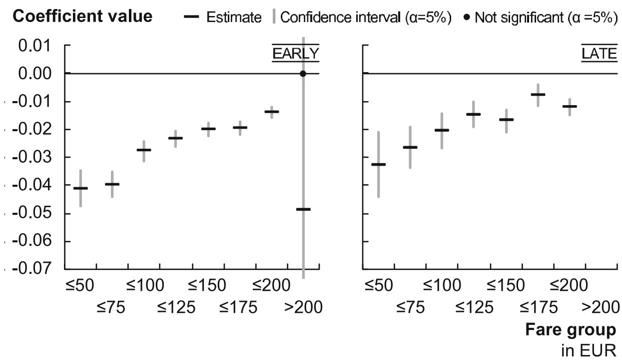


Figure 11.7: Estimation results for Model 03 (linearly approximated fares)
Source: Own design, based on estimates

although 14 additional parameters have been used (16 new based on (11.1), minus the two former fare variables), which penalizes both test statistics (see Table 11.5).

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{14,0.975}$	Evaluation
03	$-25,007$	18	50,050	Model 03 clearly better	426.848	26.119	Model 03 clearly better
02	$-25,220$	4	50,449				

Table 11.5: Comparison Model 03 vs. Model 02

While the coefficient estimates in Figure 11.7 already give a clear indication of a non-linear relationship between fare and utility, the values alone do not directly indicate the type of appropriate transformation. For this purpose, Figure 11.8 depicts the resulting absolute (dis-)utility for both flight alternatives over the available fare level, disregarding the utility contained in the alternative specific constants. The shaded areas depict a fitted channel of width 0.2 with logarithmically decreasing utility, which apparently covers the measured utility extremely well.

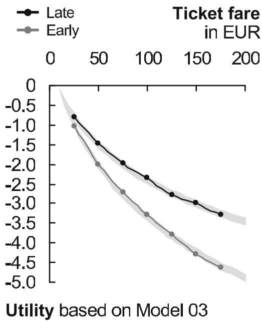


Figure 11.8: Logarithmic utility development over linearly approximated fares
Source: Own design, based on estimates

Based on the above findings, an alternative specification to the linear Model 02 containing logarithmized fares is estimated and evaluated below.

Logarithmized Fares

$$\begin{aligned} V_n^o &= \alpha^o \\ V_n^e &= \alpha^e + \beta^e_{\ln(p_e)} \cdot \ln(p_e) \\ V_n^l &= \alpha^l + \beta^l_{\ln(p_l)} \cdot \ln(p_l) \end{aligned} \qquad \text{(Model 04)}$$

Again, in the model with logarithmized fares, the coefficients for the transformed fare attributes $\beta^i_{\ln(p_i)}$ adhere to the expectations in sign and magnitude. The latter is considerably higher as in Model 02 and Model 03, which stems from the lower level of correspondingly transformed $\mathbf{x}_i$. The new coefficients for the fares are again significant at $\alpha = 5\%$ (see Table 11.6).

Early departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^e	1.020000	0.164000	0.00	α^l	0.015800	0.228000	0.94
$\beta^e_{\ln(p_e)}$	−0.908000	0.040000	0.00	$\beta^l_{\ln(p_l)}$	−0.543000	0.052500	0.00

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.6: Estimation results for Model 04 (logarithmized fares)

At this stage, the introduction of logarithmized fares has induced so much explanatory value for the late departure that the average utility contained in the alternative specific constant is close to zero and has lost its significance, which does not negatively affect the model.

Due to the transformation of fares, the original Model 02 and the improved Model 04 are not linearly nested and can therefore not be compared using the Likelihood Ratio test (see Section 9.4.4). However, besides the pure informal evaluation of the parameters and testing for their significance, the AIC can still be used for comparison (see Table 11.7).

Obviously, Model 04 yields an AIC that is substantially smaller than the fare-linear Model 02 and is henceforth deemed to report sufficiently better results.

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{0,0.975}$	Evaluation
04	−25,135	4	50,277	Model 04 clearly better	n/a	n/a	n/a
02	−25,220	4	50,449				

Table 11.7: Comparison Model 04 vs. Model 02

This section has included the internal choice drivers in the model. The following section now takes a look at characteristics of the decision maker when observing the choice set, before finally the effects of external choice drivers controlled by the competition are evaluated in Sections 11.2.3 and 11.2.4.

11.2.2 Decision Maker Characteristics

This section discusses the inclusion of decision maker characteristics into the choice model that has been developed in the previous section. In the sense of this work, the decision maker can carry specific advance request time preferences as well as departure and booking weekday needs as individual characteristics.

Similar to the demonstrated level-dependent effect of fares on a flight's utility, which is included based on a logarithmic transformation, the remaining time until departure at the time of booking may drive a decision maker's utility non-linearly – all else equal. That is, a flight that can be booked on short notice but still has the same price as before may exhibit a greatly increased utility to the decision maker.

A plot of book-to-look ratios for the considered flights over the remaining time to actual departure clearly exhibits a positive and possibly exponential

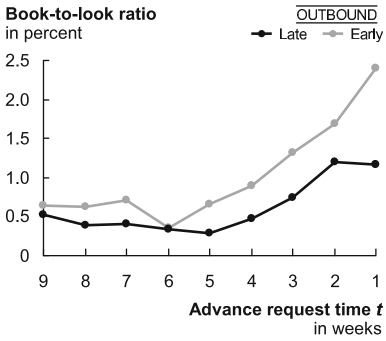


Figure 11.9: Book-to-look ratio development over advance request time
Source: Own design, based on collected data

trend (see Figure 11.9). Note that the shown trend depicts the average development with respect to all other drivers, so that the dip for the late flight in the last week before departure could still be entirely due to fare changes (i.e., raises), rather than being a specific effect of advance purchase time.

Time to departure is included in both utility functions separately as its coefficient weight again may differ for early and late flights. Note that in Part II, the variable for time to departure t has been modeled as decreasing $t \in \{60, \ldots, 1\}$ to represent the remaining days until departure, whereas here time is modeled as an increasing variable $s = 61 - t$ (i.e., $s \in \{1, \ldots, 60\}$) to allow for the testing of possibly exponential utility effects below.

Advance Request Time

$$\begin{aligned} V_n^o &= \alpha^o \\ V_n^e &= \alpha^e + \beta^e_{\ln(p_e)} \cdot \ln(p_e) + \beta^e_s \cdot s \qquad \text{(Model 05)} \\ V_n^l &= \alpha^l + \beta^l_{\ln(p_l)} \cdot \ln(p_l) + \beta^l_s \cdot s \end{aligned}$$

Estimation results for Model 05 in Table 11.8 show the expected results on the established coefficients of the fare effect $\beta^i_{\ln(p_i)}$, which have both declined in magnitude as part of the increased purchasing probability close to departure can now – supposedly correctly – be attributed to the time variable s. The effect is most obvious for the early departure, for which Figure 11.9 already indicated the strongest increase. Also, the signs of all coefficients are in line with expectations: positive for the time variable and still negative for the ticket fares. All else equal, a purchase made closer to departure yields a higher utility for both early and late departures.

Early departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^e	−0.723000	0.166000	0.00	α^l	−0.653000	0.232000	0.00
$\beta^e_{\ln(p_e)}$	−0.750000	0.038600	0.00	$\beta^l_{\ln(p_l)}$	−0.528000	0.052100	0.00
β^e_s	0.033300	0.000927	0.00	β^l_s	0.019100	0.001220	0.00

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.8: Estimation results for Model 05 (advance request time)

Comparing the results from Tables 11.6 and 11.8, it becomes apparent that the disutility attributed to a rising fare level can be offset partially when the booking happens closer to departure.

To demonstrate and calculate the actual effect, the marginal rate of substitution (MRS) between advance request time s and fare level p_i for each alternative can be calculated following Section 9.4.4 to

$$MRS^i_{s,p_i} = -\frac{\beta^i_s}{\beta^i_{\ln(p_i)}} \cdot p_i. \tag{11.2}$$

Owing to the logarithmized fare representation, (11.2) depends not only on the affected coefficients, but also on the prevalent base fare level. That is, with rising overall fare level, the decision makers will trade a larger fare increase for an additional day of advance booking. However, as fare enters linearly into the MRS, it simply results in a fixed *percentage* fare increase to keep utility constant. Without logarithmic transformation, the required fare increase would have been fixed in *absolute* terms.

Considering the estimation results in Table 11.8, decision makers would substitute a 4.44% fare increase for an additional day of advance booking on the early departure and a 3.62% increase on the late departure to keep their overall utility constant, i.e., without altering their choice probabilities.

The consideration of advance purchase time has considerably improved the statistical performance of Model 05 compared to Model 04. As the two are nested by linear restriction on β^i_s, the comparison can resort to the Likelihood Ratio test (see Table 11.9).

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{2,0.975}$	Evaluation
05	$-24,335$	6	$48,682$	Model 05 clearly better	1,599.430	7.378	Model 05 clearly better
04	$-25,135$	4	$50,277$				

Table 11.9: Comparison Model 05 vs. Model 04

By model construction so far, the marginal rate of substitution of advance purchase time for fare does not depend on time until actual flight departure, which is in line neither with intuition nor with the observations from Figure 11.9. In reality, customers would probably value booking time close to departure higher than they would early in the booking process.

To test this assumption, an exponential (here, quadratic) transformation of time is included in the model below.

Exponential Advance Request Time

$$\begin{aligned} V_n^o &= \alpha^o \\ V_n^e &= \alpha^e + \beta^e_{\ln(p_e)} \cdot \ln(p_e) + \beta^e_{s^2} \cdot s^2 \\ V_n^l &= \alpha^l + \beta^l_{\ln(p_l)} \cdot \ln(p_l) + \beta^l_{s^2} \cdot s^2 \end{aligned} \qquad \text{(Model 06)}$$

Model 06 now allows the time variable to drive a strong increase in utility toward departure independent of fare changes. Naturally, the model is not a direct extension of (11.2), as it is not linearly nested.

The new estimation results are recorded in Table 11.10. All coefficients are again in line with intuitions regarding sign and magnitude. The much smaller values of the time coefficients $\beta^i_{s^2}$ are due to the increased base level of the time attribute. Notably, coefficients for the alternative specific variables and the logarithmized fares vary only fractionally.

Based on the new Model 06, the marginal rate of substitution of advance purchase time to fare changes to

$$MRS^i_{s,p_i} = -\frac{\beta^i_s}{\beta^i_{\ln(p_i)}} \cdot 2p_i s. \qquad (11.3)$$

Based on (11.3), the decision maker's willingness to trade off a day of additional booking time against fare raises now depends on both the prevalent fare level and the actual time until flight departure.

Early departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^e	−0.580000	0.164000	0.00	α^l	−0.658000	0.231000	0.00
$\beta^e_{\ln(p_e)}$	−0.695000	0.038600	0.00	$\beta^l_{\ln(p_l)}$	−0.486000	0.052100	0.00
$\beta^e_{s^2}$	0.000518	0.000013	0.00	$\beta^l_{s^2}$	0.000319	0.000018	0.00

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.10: Estimation results for Model 06 (exponential advance request time)

According to the coefficient estimates in Table 11.10, at 60 days prior to departure, customers value an incremental day in advance purchase time at only 0.15% of the fare level for the early departure (0.13% for the late one), but at 8.94% and 7.88%, respectively, on the last day of the booking period.

Figure 11.10 illustrates the results on the marginal rates of substitution of (11.2) and (11.3) for both early and late departures. The chart shows the relative fare increase that customers would trade off for an additional day of advance purchase time. Based on the logarithmic transformation, this rate is constant over initial fare levels.

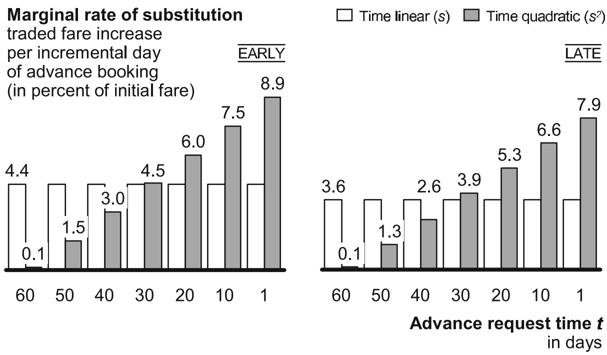


Figure 11.10: Marginal rate of substitution – linear vs. exponential transformation of time variable
Source: Own design, based on estimates

As seen before, the utility for the early departure reacts more strongly to fare changes, which is in line with expectations. However, the marginal rates of substitution for early and late departures are considerably closer than expected from Figure 11.9. This illustrates that the increased booking activity on the early flight closer to departure may not be driven solely by a higher utility of time, but rather by generally lower prices (see also Figure 10.5).

Additionally, this is reflected in the raw data: whenever offered, the later flight seems to be more popular than the early departure. This quickly leads to a lower number of free seats on the late departure and corresponding higher prices, which reduce flight utility and trigger increased booking on the early flight (so-called demand rolling).

The new model formulation, including exponential fares, also performs considerably better statistically than the model with linear fares (see Table 11.11). As the models are not linearly nested, the AIC is used exclusively to compare the models.

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{0,0.975}$	Evaluation
06	$-24,266$	6	$48,543$	Model 06 clearly better	n/a	n/a	n/a
05	$-24,335$	6	$48,682$				

Table 11.11: Comparison Model 06 vs. Model 05

Based on the above results, going forward the model is based on logarithmized fares and a quadratic time variable. After these two obvious drivers have been included, the next two models evaluate the influences of departure and booking weekdays.

Flight Departure Weekdays

Section 10.3.2 has indicated that experienced utility may vary by weekday of flight departure. While this variation may well already be contained or offset by differing fare levels, Model 07 now incorporates an additional set of variables to test for specific utility effects not covered by price alone.

Similar to Part II, the differentiation of flight departure weekdays is included based on a set of binary dummy variables:[5]

$$f_d = \begin{cases} 1 & \text{if } weekday(flight\ date) = d \\ & \text{with } d \in \{1, \ldots, 7\}, \\ 0 & \text{otherwise.} \end{cases} \tag{11.4}$$

[5] Weekdays are assumed to be numbered from 1 to 7, with 1 being the Monday.

Note that the variables f_d are independent of the specific alternative i, as the decision maker is assumed to choose only between flights departing on the same date. However, the coefficients may well vary between departure time segments, not least because late departures in the outbound direction are only scheduled on Friday and Sunday (see Figure 10.3).

To keep the model identified, all but one weekday can be included in the model so that the alternative specific constant covers the part of the absent variable (here chosen to be Sunday).

$$\begin{aligned}
V_n^o &= \alpha^o \\
V_n^e &= \alpha^e + \beta^e_{\ln(p_e)} \cdot \ln(p_e) + \beta^e_{s^2} \cdot s^2 + \sum_{d=1}^{6} \beta^e_{f_d} \cdot f_d \qquad \text{(Model 07)} \\
V_n^l &= \alpha^l + \beta^l_{\ln(p_l)} \cdot \ln(p_l) + \beta^l_{s^2} \cdot s^2 + \beta^l_{f_5} \cdot f_5
\end{aligned}$$

Table 11.12 reports the estimates for Model 07, which reflect the observed variation of choice behavior among departure weekdays in Section 10.3.2 (see Figure 10.10). Although the coefficients $\beta^i_{f_d}$ adjust the model for corresponding variations in utility that are not reflected in the fares of the analyzed carrier, coefficients for pre-existing variables react in a fairly stable fashion.

Predictably, the effect of advance purchase time $\beta^i_{s^2}$ has barely changed, while shifts in the fare coefficients $\beta^i_{\ln(p_i)}$ occurred, especially for the late departure, where the single coefficient for departure weekday can clearly separate the effects of the strong Sunday departures vs. the weaker Friday departures.

Some coefficients of particular departure weekday variables are far from significant at the chosen level of $\alpha = 5\%$. However, the full coefficient set is kept for the moment in order to check for possible interactions with external choice drivers in Section 11.2.3 below. That is, the fares of competitors may well vary by departure weekday, which might induce additional variation in utility that could then be explained by the above variables.

Early departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^e	−0.838000	0.174000	0.00	α^l	−1.080000	0.262000	0.00
$\beta^e_{\ln(p_e)}$	−0.650000	0.041600	0.00	$\beta^l_{\ln(p_l)}$	−0.371000	0.062200	0.00
$\beta^e_{s^2}$	0.000523	0.000013	0.00	$\beta^l_{s^2}$	0.000320	0.000019	0.00
$\beta^e_{f_1}$	0.007470	0.051300	0.88				
$\beta^e_{f_2}$	0.297000	0.054800	0.00				
$\beta^e_{f_3}$	−0.035600	0.064400	0.58				
$\beta^e_{f_4}$	0.284000	0.051000	0.00				
$\beta^e_{f_5}$	−0.056000	0.054100	0.30	$\beta^l_{f_5}$	−0.171000	0.052200	0.00
$\beta^e_{f_6}$	−0.102000	0.070800	0.15				

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.12: Estimation results for Model 07 (flight departure weekdays)

Table 11.13 reveals that the model has again statistically benefited from adding the described variables, although the number of coefficients to be estimated has doubled – leading to an increasing penalty for both the Likelihood Ratio test and the AIC.

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{7,0.975}$	Evaluation
07	−24,215	13	48,456	Model 07 clearly better	100.486	16.013	Model 07 clearly better
06	−24,266	6	48,543				

Table 11.13: Comparison Model 07 vs. Model 06

Before finally adding external choice drivers in terms of competitor fares, the next model extension adds another set of dummy variables to include the effect of varying booking weekdays.

Advance Booking Weekdays

Although not immediately comprehensible, Section 10.3.2 has indicated that the decision maker's utility may be affected by the specific weekday on which she actually books her ticket. A second look suggests that depending on the flight direction, specific booking weekdays naturally are preferential. For example, last-minute decisions to take weekend trips can simply not be made on that same weekend.

Similar to above, a further set of dummy variables is defined, indicating the weekday of the advance request day of the recorded booking or transaction (if no booking has been made)

$$a_d = \begin{cases} 1 & \text{if } \mathit{weekday}(\mathit{advance\ request\ date}) = d \\ & \text{with } d \in \{1, \ldots, 7\}, \\ 0 & \text{otherwise.} \end{cases} \tag{11.5}$$

Based on this definition, Model 07 is extended by a total of twelve new variables. As both alternatives (early and late) can be booked on either advance request weekday, a theoretical maximum of seven new coefficients would have to be added. To assure identification of the model, the Sunday variables are again excluded. The new variables are independent of the specific alternatives as the decision maker defines them on the same occasion.

$$\begin{aligned} V_n^o &= \alpha^o \\ V_n^e &= \alpha^e + \beta^e_{\ln(p_e)} \cdot \ln(p_e) + \beta^e_{s^2} \cdot s^2 + \sum_{d=1}^{6} \beta^e_{f_d} \cdot f_d + \sum_{d=1}^{6} \beta^e_{a_d} \cdot a_d \\ V_n^l &= \alpha^l + \beta^l_{\ln(p_l)} \cdot \ln(p_l) + \beta^l_{s^2} \cdot s^2 + \beta^l_{f_5} \cdot f_5 + \sum_{d=1}^{6} \beta^l_{a_d} \cdot a_d \end{aligned} \qquad \text{(Model 08)}$$

Table 11.14 reports the estimation results on the larger model. The existing coefficients have remained fairly stable, especially the pairs measuring the effects of own fares and advance booking time. The alternative specific constants have changed, indicating that additional utility inherent to the two alternatives (other than the "other/no-buy"-option) remains.

Still, some of the coefficients for the departure weekday effect remain insignificant at that stage. Those newly added for the advance request weekdays seem significant only for workdays. Again, these are nevertheless kept until the effects of competitor fares have been added in Section 11.2.3.

Despite the addition of twelve new variables for advance request weekday variations, the model has again improved its statistical performance, which is reported in Table 11.15.

The current model now includes the relevant drivers that can be controlled *internally* (i.e., through fare changes depending on a specific flight's departure time segment and timely distance to departure) as well as decision maker characteristics in terms of when customers want to fly and when they want to book their flights. Up to this point, the model also follows standard MNL modeling rules (i.e., it is not universal).

Early departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^e	−1.020000	0.180000	0.00	α^l	−1.660000	0.276000	0.00
$\beta^e_{\ln(p_e)}$	−0.667000	0.041800	0.00	$\beta^l_{\ln(p_l)}$	−0.353000	0.062900	0.00
$\beta^e_{s^2}$	0.000528	0.000014	0.00	$\beta^l_{s^2}$	0.000319	0.000019	0.00
$\beta^e_{f_1}$	0.008680	0.051400	0.87				
$\beta^e_{f_2}$	0.313000	0.054900	0.00				
$\beta^e_{f_3}$	−0.015300	0.064500	0.81				
$\beta^e_{f_4}$	0.296000	0.051200	0.00				
$\beta^e_{f_5}$	−0.058200	0.054300	0.28	$\beta^l_{f_5}$	−0.160000	0.052600	0.00
$\beta^e_{f_6}$	−0.106000	0.070900	0.14				
$\beta^e_{a_1}$	0.293000	0.061200	0.00	$\beta^l_{a_1}$	0.487000	0.095200	0.00
$\beta^e_{a_2}$	0.151000	0.061000	0.01	$\beta^l_{a_2}$	0.746000	0.090100	0.00
$\beta^e_{a_3}$	0.381000	0.059000	0.00	$\beta^l_{a_3}$	0.327000	0.094700	0.00
$\beta^e_{a_4}$	0.538000	0.062100	0.00	$\beta^l_{a_4}$	0.963000	0.094200	0.00
$\beta^e_{a_5}$	0.130000	0.064400	0.04	$\beta^l_{a_5}$	0.691000	0.095200	0.00
$\beta^e_{a_6}$	0.059800	0.066000	0.37	$\beta^l_{a_6}$	−0.140000	0.117000	0.23

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.14: Estimation results for Model 08 (advance booking weekdays)

Model estimation and corresponding tests have shown that the reported findings concerning internal choice drivers and the considered decision maker characteristics can be directly transferred to the inbound directional data – naturally with different coefficient estimates. However, as the influence of competitor fares likely differs, Sections 11.2.3 and 11.2.4 discuss the results separately by flight direction.

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{12,0.975}$	Evaluation
08	−24,044	25	48,137	Model 08 clearly better	343.430	23.337	Model 08 clearly better
07	−24,215	13	48,456				

Table 11.15: Comparison Model 08 vs. Model 07

For easier comparison with the directional model results below as well as allowing for a more intuitive interpretation, the current model's coefficient estimates together with their confidence intervals for $\alpha = 5\%$ are depicted graphically in Figure 11.11.

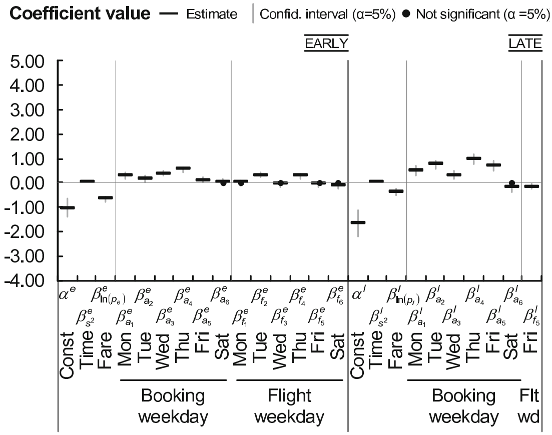


Figure 11.11: Estimation results for Model 08 (internal drivers and decision maker characteristics)
Source: Own design, based on estimates

11.2.3 External Outbound Choice Drivers

The objective of this section is to model the utility effect of competitor fares on the choice probabilities for the two outbound flight alternatives of the considered carrier (early and late departure).

Based on the limited data availability and general model structure (see Section 11.1), from here on the approach borrows from universal logit to avoid interference with the IAA property of the conventional MNL formulation (see Section 9.4.2).

Individual fare attributes of competitor flights can potentially affect both alternatives' representative utility of the considered carrier, so the appropriate allocation – possibly with individual coefficients – has to be tested. The magnitudes of individual influences of such fares may also depend on other fares of the competition, as naturally any competitor competes not just with the analyzed carrier, but also with the remaining competition. Hence, the external fare data are not included incrementally into the model in a forward directional approach as in Sections 11.2.1 and 11.2.2; instead, all drivers are included simultaneously. Individual variables that do not seem to be significant for a particular alternative are then eliminated in a *backward directional* manner.

The modeling approach based on the universal logit mindset demands specific treatment of cases where competitor flights are not available. In conventional MNL modeling, in case an explicitly modeled alternative is not available to the decision maker, it is not included in its particular choice set and the representative utility is naturally forced to be zero – independent of its individual attributes (i.e., the latter do not affect the estimation results).

In the employed universal formulation, the attributes of competing alternatives are entered into the model as long as the considered carrier's flights are scheduled. In our specific case, this naturally triggers the question of how to treat missing fare values of competitor alternatives in case those are not available to the decision maker. That is, if a competing flight is sold out, its fare attribute will be missing from the dataset; however, it has to be included in the estimation process.

The first option to solve this problem would be to replace such missing fare values by a sufficiently high compensating figure representing a prohibitive fare that would render the particular flight not feasible to the decision maker; the disadvantage being that the number would have to be arbitrarily selected. The second option would be to add a separate binary variable indicating whether a particular flight is available. This way, the coefficient of the latter would represent the specific utility induced by the sheer availability of a flight (independent of its fare) and would not have to be arbitrarily chosen, but could be calculated using the conventional estimation process.

In this work, the second option is selected based on its adequate accuracy and the fact that it integrates smoothly into the estimation procedure. Accordingly, for each of the three competitor flights, two variables (fare and availability) are included in each of the considered carrier's alternatives. Similar to Section 11.2.1, the fares are considered logarithmized as $\ln(p_{j,e|m|l})$ and the binary variable for flight availability is defined as

$$b_{j,u} = \begin{cases} 1 & \text{if competitor } j \text{ offers a flight in time segment } u, \\ & \text{with } u \in \{e = early, m = mid, l = late\}, \\ 0 & \text{otherwise.} \end{cases} \tag{11.6}$$

Note that for brevity the already specified systematic utility for the alternatives in Model 08 is replaced by $\widetilde{V}_n^i$ in the models below. However, estimates for all coefficient values are still reported to illustrate possible changes.

Maximum Model

Considering the above, the full model including all competitor fares with corresponding binary availability variables in both alternatives reads as

$$
\begin{aligned}
V_n^o &= \widetilde{V}_n^o \\
V_n^e &= \widetilde{V}_n^e + \beta^e_{b_{1,e}} \cdot b_{1,e} + \beta^e_{\ln(p_{1,e})} \cdot \ln(p_{1,e}) && \text{(comp. 1, early dep.)} \\
&\quad + \beta^e_{b_{1,m}} \cdot b_{1,m} + \beta^e_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) && \text{(comp. 1, mid dep.)} \\
&\quad + \beta^e_{b_{2,m}} \cdot b_{2,m} + \beta^e_{\ln(p_{2,m})} \cdot \ln(p_{2,m}) && \text{(comp. 2, mid dep.)} \\
V_n^l &= \widetilde{V}_n^l + \beta^l_{b_{1,e}} \cdot b_{1,e} + \beta^l_{\ln(p_{1,e})} \cdot \ln(p_{1,e}) && \text{(comp. 1, early dep.)} \\
&\quad + \beta^l_{b_{1,m}} \cdot b_{1,m} + \beta^l_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) && \text{(comp. 1, mid dep.)} \\
&\quad + \beta^l_{b_{2,m}} \cdot b_{2,m} + \beta^l_{\ln(p_{2,m})} \cdot \ln(p_{2,m}) && \text{(comp. 2, mid dep.)} \qquad \text{(Model 09)}
\end{aligned}
$$

Model 09 now allows for the desired cross-effects between utility of the considered carrier's own flights and those of competitors (see Section 11.1); however, it is not a full universal logit representation as cross-alternative characteristics do not enter the systematic utility of all choices, and moreover, competitive alternatives are not included separately.

The estimation results for the fully-fledged Model 09 in Table 11.16 exhibit the expected influence of the newly added competition attributes on the dummy variables differentiating the flight departure weekday as well as the fares of the considered carrier. Based on the newly included variables, the overall scale of utility has shifted slightly; otherwise, the model estimates prove fairly stable.

Figure 11.12 provides a graphic illustration of the coefficients that allows for quick spotting of issues and an intuitive comparison with the estimates of Model 08 in Figure 11.11.

Looking at the new coefficient estimates for the attributes of competitor alternatives, a few issues become apparent: Quite a few coefficients are not significant at the chosen level of $\alpha = 5\%$, which could be attributed to the fact that the availability or fare level of particular competitor flights might simply not affect the choice probability of the corresponding alternative flights of the considered carrier.

However, some of the estimated results also do not exhibit the expected signs. The new binary variable indicating the *availability* of a competitor flight should be negative, resulting in a genuinely reduced utility of the considered carrier's own flights if a particular competitor flight is available. Correspondingly, the *fare* coefficients should have a positive sign, that is, utility of the considered carrier's flights increase with rising competitor fares.

Early departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^e	1.140000	0.260000	0.00	α^l	−1.170000	0.359000	0.00
$\beta^e_{\ln(p_e)}$	−1.110000	0.060100	0.00	$\beta^l_{\ln(p_l)}$	−0.456000	0.079500	0.00
$\beta^e_{s^2}$	0.000487	0.000014	0.00	$\beta^l_{s^2}$	0.000346	0.000020	0.00
$\beta^e_{f_1}$	−0.533000	0.073700	0.00				
$\beta^e_{f_2}$	−0.210000	0.076100	0.01				
$\beta^e_{f_3}$	−0.185000	0.071700	0.01				
$\beta^e_{f_4}$	−0.225000	0.071500	0.00				
$\beta^e_{f_5}$	−0.625000	0.074300	0.00	$\beta^l_{f_5}$	−0.379000	0.081900	0.00
$\beta^e_{f_6}$	−0.074400	0.072100	0.30				
$\beta^e_{a_1}$	0.274000	0.061400	0.00	$\beta^l_{a_1}$	0.536000	0.095900	0.00
$\beta^e_{a_2}$	0.167000	0.061300	0.01	$\beta^l_{a_2}$	0.813000	0.090800	0.00
$\beta^e_{a_3}$	0.345000	0.060700	0.00	$\beta^l_{a_3}$	0.349000	0.096900	0.00
$\beta^e_{a_4}$	0.531000	0.063300	0.00	$\beta^l_{a_4}$	0.937000	0.095500	0.00
$\beta^e_{a_5}$	0.143000	0.064700	0.03	$\beta^l_{a_5}$	0.689000	0.095800	0.00
$\beta^e_{a_6}$	0.070600	0.066100	0.29	$\beta^l_{a_6}$	−0.151000	0.117000	0.20
$\beta^e_{b_{1,e}}$	−2.470000	0.337000	0.00	$\beta^l_{b_{1,e}}$	0.287000	0.465000	0.54
$\beta^e_{b_{1,m}}$	−0.360000	0.586000	0.54	$\beta^l_{b_{1,m}}$	−2.270000	0.764000	0.00
$\beta^e_{b_{2,m}}$	−1.350000	0.389000	0.00	$\beta^l_{b_{2,m}}$	2.050000	0.416000	0.00
$\beta^e_{\ln(p_{1,e})}$	0.674000	0.078000	0.00	$\beta^l_{\ln(p_{1,e})}$	−0.018500	0.101000	0.85
$\beta^e_{\ln(p_{1,m})}$	−0.058100	0.123000	0.64	$\beta^l_{\ln(p_{1,m})}$	0.420000	0.159000	0.01
$\beta^e_{\ln(p_{2,m})}$	0.228000	0.081700	0.01	$\beta^l_{\ln(p_{2,m})}$	−0.482000	0.088900	0.00

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.16: Estimation results for Model 09 (maximum model)

A counterintuitive effect is most visible for the midday departure of competitor 2 ($\beta^l_{b_{2,m}}$ and $\beta^l_{\ln(p_{2,m})}$) and can also be seen, albeit less pronounced, for the early departure of competitor 1 ($\beta^l_{b_{1,e}}$ and $\beta^l_{\ln(p_{1,e})}$), both regarding the utility of the considered carrier's late flight departure.

The counterintuitive effect may be induced by varying the search behavior of decision makers. As described in Section 10.1, the CRS log does not allow us to differentiate between customers who specifically look for a flight, which then quickly results in a booking or a lost sale (serious demand), and customers who repeatedly query the system in search of bargain fares without a definite purchase intention (bargain demand). Obviously, the purchase probability is higher when the decision maker expresses serious demand – all else equal.

Unfortunately, the data do not directly allow a differentiation of customers into these two segments. However, assuming a steady base level of

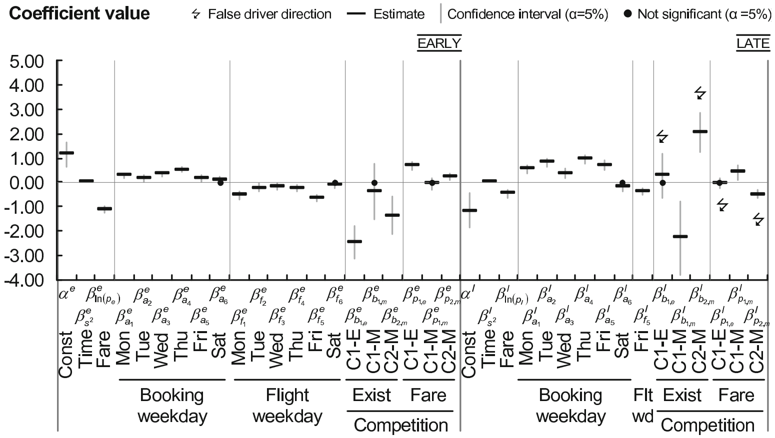


Figure 11.12: Estimation results for Model 09 (maximum model)
Source: Own design, based on estimates

serious demand in the data, purchase probability declines in absolute latent demand levels. Put differently, the share of serious demand customers is higher if the overall level of demand is low, because then most decision makers assign a serious utility to the choice alternative for which they search.

As the late departure is genuinely more attractive than the early departure and is moreover available only on Fridays and Sundays, which obviously are the most attractive departure weekdays for outbound flights, the described effect may manifest in the following scenario: Overall demand is rather low, yielding an elevated choice probability based on a high share of serious demand and the sheer mathematical fact that ratios tend to be higher on small base levels. At the same time, as demand is genuinely low, competing flights most probably have space available (and vice versa). In this case, the higher choice probability is not driven by the sheer fact that a competing flight is available, but rather depends on the overall latent demand level and the share of serious demand. That is, the availability of competing flights is an indirect effect of a different true underlying driver.

The described effect is expected to be present in the data, resulting in distortions of the binary variables as shown in Figure 11.12. To adjust and test the model for the described particularity, an additional variable is included below containing an approximate measure for the overall share of serious demand.

Note that as the backward approach of variable elimination is used in this section, the current model is not statistically compared to the last model in the above section; however, this will be accomplished with the final model at the end of this section.

Full Model with Demand Type

To control for the effect described above, an additional variable $\widetilde{D}$ is introduced into the model that depends on the overall level of latent demand D (see Part II), which is supposed to indicate the seriousness or utility that individual decision makers assign to the searched-for flight as

$$\widetilde{D} = \frac{1}{D}. \tag{11.7}$$

Note that $\widetilde{D}$ can only serve as an approximation of the true decision maker or demand type. Naturally, the explicit knowledge of this characteristic would be more definite but is not possible here due to data limitations (see Section 10.1).

$$\begin{aligned}
V_n^o &= \widetilde{V}_n^o \\
V_n^e &= \widetilde{V}_n^e + \beta^e_{\widetilde{D}} \cdot \widetilde{D} + \beta^e_{b_{1,e}} \cdot b_{1,e} + \beta^e_{\ln(p_{1,e})} \cdot \ln(p_{1,e}) \\
&\quad + \beta^e_{b_{1,m}} \cdot b_{1,m} + \beta^e_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) \\
&\quad + \beta^e_{b_{2,m}} \cdot b_{2,m} + \beta^e_{\ln(p_{2,m})} \cdot \ln(p_{2,m}) \\
V_n^l &= \widetilde{V}_n^l + \beta^l_{\widetilde{D}} \cdot \widetilde{D} + \beta^l_{b_{1,e}} \cdot b_{1,e} + \beta^l_{\ln(p_{1,e})} \cdot \ln(p_{1,e}) \\
&\quad + \beta^l_{b_{1,m}} \cdot b_{1,m} + \beta^l_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) \\
&\quad + \beta^l_{b_{2,m}} \cdot b_{2,m} + \beta^l_{\ln(p_{2,m})} \cdot \ln(p_{2,m})
\end{aligned} \tag{Model 10}$$

As expected and intended, the inclusion of the inverted demand level keeps the majority of the coefficients stable in their relations but changes the overall level of utility and with it the binary variables indicating the availability of competitor flights (see Table 11.17).

The inclusion of the demand type variable has additionally improved the significance of quite a few of the binary variables that indicate the utility effect of booking and departure weekdays as it partially contains the macro-seasonal variations (attractive vs. less attractive flights), based upon which the micro-seasonal fluctuations exhibit a more significant effect.

Early departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^e	0.753000	0.264000	0.00	α^l	−2.020000	0.360000	0.00
$\beta^e_{\ln(p_e)}$	−1.070000	0.060600	0.00	$\beta^l_{\ln(p_l)}$	−0.386000	0.079300	0.00
$\beta^e_{s^2}$	0.000501	0.000014	0.00	$\beta^l_{s^2}$	0.000381	0.000021	0.00
$\beta^e_{f_1}$	−0.561000	0.073800	0.00				
$\beta^e_{f_2}$	−0.231000	0.076100	0.00				
$\beta^e_{f_3}$	−0.260000	0.072100	0.00				
$\beta^e_{f_4}$	−0.263000	0.071700	0.00				
$\beta^e_{f_5}$	−0.624000	0.074500	0.00	$\beta^l_{f_5}$	−0.292000	0.083800	0.00
$\beta^e_{f_6}$	−0.143000	0.072400	0.05				
$\beta^e_{a_1}$	0.380000	0.062500	0.00	$\beta^l_{a_1}$	0.747000	0.097900	0.00
$\beta^e_{a_2}$	0.266000	0.062300	0.00	$\beta^l_{a_2}$	1.000000	0.092700	0.00
$\beta^e_{a_3}$	0.388000	0.061200	0.00	$\beta^l_{a_3}$	0.443000	0.098500	0.00
$\beta^e_{a_4}$	0.606000	0.063900	0.00	$\beta^l_{a_4}$	1.020000	0.097200	0.00
$\beta^e_{a_5}$	0.198000	0.065100	0.00	$\beta^l_{a_5}$	0.837000	0.097300	0.00
$\beta^e_{a_6}$	0.104000	0.066300	0.12	$\beta^l_{a_6}$	−0.094200	0.118000	0.42
$\beta^e_{\widetilde{D}}$	1.840000	0.168000	0.00	$\beta^l_{\widetilde{D}}$	4.460000	0.237000	0.00
$\beta^e_{b_{1,e}}$	−2.750000	0.338000	0.00	$\beta^l_{b_{1,e}}$	−1.010000	0.462000	0.03
$\beta^e_{b_{1,m}}$	−0.552000	0.583000	0.34	$\beta^l_{b_{1,m}}$	−3.050000	0.747000	0.00
$\beta^e_{b_{2,m}}$	−1.620000	0.391000	0.00	$\beta^l_{b_{2,m}}$	0.885000	0.423000	0.04
$\beta^e_{\ln(p_{1,e})}$	0.742000	0.078400	0.00	$\beta^l_{\ln(p_{1,e})}$	0.263000	0.099800	0.01
$\beta^e_{\ln(p_{1,m})}$	−0.014300	0.122000	0.91	$\beta^l_{\ln(p_{1,m})}$	0.612000	0.155000	0.00
$\beta^e_{\ln(p_{2,m})}$	0.290000	0.082000	0.00	$\beta^l_{\ln(p_{2,m})}$	−0.203000	0.090200	0.02

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.17: Estimation results for Model 10 (full model with demand type)

Figure 11.13 depicts that the pseudo-utility effect, which was contained in the binary variables has been erased entirely for the early flight of competitor 1, whereas the midday departure of competitor 2 still exhibits the unforeseen behavior when competing against the late departure of the considered carrier – albeit at a lower level.

The remaining counterintuitive signs of the coefficients for competitor 2 on the late departure alternative can be traced back to some peculiarities of the schedule: The considered carrier's late departing flight is offered only on Fridays and Sundays, while the midday departure of competitor 2 is offered only on weekends. Consequently, the binary variable $\beta^l_{b_{2,m}}$ does not simply indicate the availability of seats on competitor flights, but also unintentionally acts as a flag for Sunday departures, which is the only day on which these two carriers compete head-on with their midday and late flights.

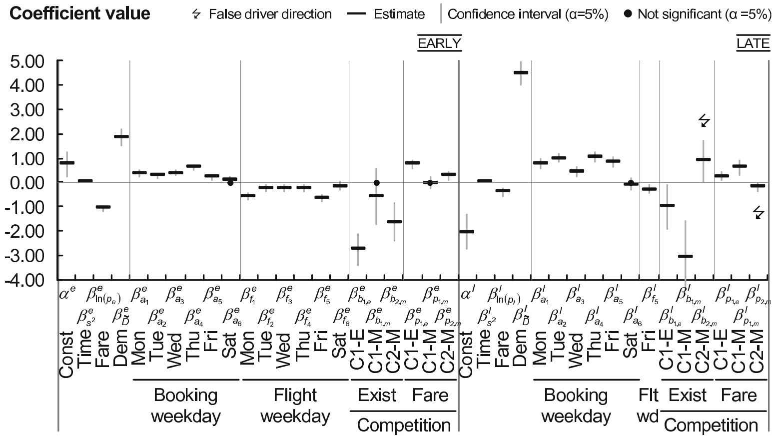


Figure 11.13: Estimation results for Model 10 (full model with demand type)
Source: Own design, based on estimates

At the same time, the overall micro-seasonal book-to-look ratios (see Figure 10.10) suggest (also when compared to the early departure alternative) that the higher choice probabilities on Sundays are due to the weekday effect rather than the mere existence of a flight offering by competitor 2. Moreover, presumably the competitor does offer flights on Sunday simply because of the higher demand.

Before presenting a solution to the second observed pseudo-utility driver, it must be noted that besides adhering more closely to intuition and expectations, Model 10 also performs significantly better than the starting point in Model 09, which is evidenced in Table 11.18. The models are linearly nested as two additional parameters have simply been added and can therefore also be compared using the Likelihood Ratio test.

Obviously, schedule effects regarding the flight plan of competitor 2 are captured by its flight availability variable $\beta^l_{b_{2,m}}$. While this first leads to a different interpretation of the variable, it can also be argued that the model already contains a dedicated variable for such schedule effects ($\beta^l_{f_5}$) that could carry the effect and is possibly biased in the current estimates.

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{2,0.975}$	Evaluation
10	−23, 697	39	47, 473	Model 10 clearly better	398.904	7.378	Model 10 clearly better
09	−23, 897	37	47, 868				

Table 11.18: Comparison Model 10 vs. Model 09

The following model therefore eliminates the supposedly redundant variable $\beta^l_{b_{2,m}}$ from the second alternative to observe the changes to the remaining coefficients (especially $\beta^l_{f_5}$) and to see whether the statistical performance of the model improves.

Exclude Competitor 2 Availability

The specification of Model 11 below eliminates the possibly biased availability variable of competitor 2 from the second alternative.

$$
\begin{aligned}
V^o_n &= \widetilde{V}^o_n \\
V^e_n &= \widetilde{V}^e_n + \beta^e_{\widetilde{D}} \cdot \widetilde{D} + \beta^e_{b_{1,e}} \cdot b_{1,e} + \beta^e_{\ln(p_{1,e})} \cdot \ln(p_{1,e}) \\
&\quad + \beta^e_{b_{1,m}} \cdot b_{1,m} + \beta^e_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) \\
&\quad + \beta^e_{b_{2,m}} \cdot b_{2,m} + \beta^e_{\ln(p_{2,m})} \cdot \ln(p_{2,m}) \\
V^l_n &= \widetilde{V}^l_n + \beta^l_{\widetilde{D}} \cdot \widetilde{D} + \beta^l_{b_{1,e}} \cdot b_{1,e} + \beta^l_{\ln(p_{1,e})} \cdot \ln(p_{1,e}) \\
&\quad + \beta^l_{b_{1,m}} \cdot b_{1,m} + \beta^l_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) \\
&\quad + \beta^l_{\ln(p_{2,m})} \cdot \ln(p_{2,m}) \qquad \text{(Model 11)}
\end{aligned}
$$

The estimates of the new specification remove the discussed effect on the availability variable of competitor 2 (see Table 11.19). Driven by the erroneous positive sign of $\beta^l_{b_{2,m}}$, the fare variable $\beta^l_{\ln(p_{2,m})}$ has been "flipped" in magnitude in Model 10. The sign is still negative (as is the one of $\beta^e_{\ln(p_{1,m})}$), but neither is significantly different from zero, which explains the particular sign as somewhat random.

Comparison of the estimation results in Table 11.19 with those in Table 11.17 highlights the changes: The coefficients for the early alternative prove fairly stable (as expected), while for the late flight alternative the constant, the binary variable for the departure date and the binary variables capturing the availability of competitor flights are affected.

Figure 11.14 depicts the resulting coefficient landscape, which is absolutely in line with expectations in terms of coefficient sign and magnitude.

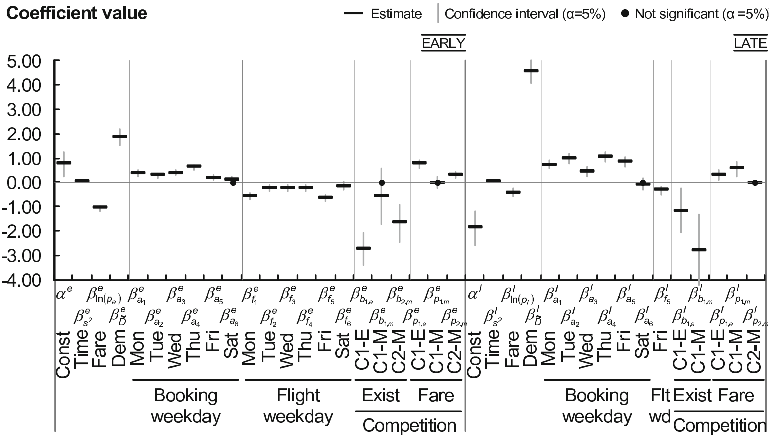


Figure 11.14: Estimation results for Model 11 (exclude competitor 2 availability)
Source: Own design, based on estimates

Also, the variables that seem not to be significant at level $\alpha = 5\%$ may now lend themselves to a concise explanation.

Two interesting observations can be drawn from Figure 11.14: To begin with, the posted fare for the midday departure of competitor 2 seems not to have a significant effect on utility for the late departure of the considered carrier. Three possible explanations for that result exist: First, based on the earlier departure time segment of the competitor flight, decision makers may simply not regard it as a reasonable alternative to the late departure of the considered carrier, independent of its price. Second, Figure 10.6 depicts that the average fare level of competitor 2 is $\approx 20\%$ higher than that of the considered carrier's late departure, and so competitor 2's midday departure would only affect price-insensitive customers anyway. Third, competitor 2 may genuinely not be regarded as a valid alternative for decision makers considering whether to take the considered carrier's late flight.

Additionally, the fare of the other competitor's midday departure seems not to affect choice for the considered carrier's early departure. As competitor 1 is the main carrier on the analyzed route and its early departure exhibits heavy impact on the model, the rationale here can only be either the high fare level or, again, the differing departure time segments. In fact,

Early departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^e	0.754000	0.264000	0.00	α^l	−1.870000	0.351000	0.00
$\beta^e_{\ln(p_e)}$	−1.070000	0.060600	0.00	$\beta^l_{\ln(p_l)}$	−0.408000	0.078100	0.00
$\beta^e_{s^2}$	0.000501	0.000014	0.00	$\beta^l_{s^2}$	0.000367	0.000020	0.00
$\beta^e_{f_1}$	−0.562000	0.073800	0.00				
$\beta^e_{f_2}$	−0.232000	0.076100	0.00				
$\beta^e_{f_3}$	−0.260000	0.072100	0.00				
$\beta^e_{f_4}$	−0.264000	0.071800	0.00				
$\beta^e_{f_5}$	−0.625000	0.074500	0.00	$\beta^l_{f_5}$	−0.314000	0.082700	0.00
$\beta^e_{f_6}$	−0.145000	0.072400	0.05				
$\beta^e_{a_1}$	0.379000	0.062500	0.00	$\beta^l_{a_1}$	0.727000	0.097400	0.00
$\beta^e_{a_2}$	0.265000	0.062300	0.00	$\beta^l_{a_2}$	0.983000	0.092300	0.00
$\beta^e_{a_3}$	0.387000	0.061200	0.00	$\beta^l_{a_3}$	0.429000	0.098200	0.00
$\beta^e_{a_4}$	0.606000	0.063900	0.00	$\beta^l_{a_4}$	1.020000	0.097100	0.00
$\beta^e_{a_5}$	0.197000	0.065100	0.00	$\beta^l_{a_5}$	0.831000	0.097200	0.00
$\beta^e_{a_6}$	0.103000	0.066300	0.12	$\beta^l_{a_6}$	−0.100000	0.118000	0.40
$\beta^e_{\tilde{D}}$	1.850000	0.168000	0.00	$\beta^l_{\tilde{D}}$	4.540000	0.233000	0.00
$\beta^e_{b_{1,e}}$	−2.740000	0.338000	0.00	$\beta^l_{b_{1,e}}$	−1.150000	0.456000	0.01
$\beta^e_{b_{1,m}}$	−0.584000	0.583000	0.32	$\beta^l_{b_{1,m}}$	−2.770000	0.736000	0.00
$\beta^e_{b_{2,m}}$	−1.670000	0.390000	0.00				
$\beta^e_{\ln(p_{1,e})}$	0.741000	0.078400	0.00	$\beta^l_{\ln(p_{1,e})}$	0.294000	0.098500	0.00
$\beta^e_{\ln(p_{1,m})}$	−0.007720	0.122000	0.95	$\beta^l_{\ln(p_{1,m})}$	0.550000	0.153000	0.00
$\beta^e_{\ln(p_{2,m})}$	0.303000	0.081800	0.00	$\beta^l_{\ln(p_{2,m})}$	−0.017000	0.015200	0.26

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.19: Estimation results for Model 11 (exclude competitor 2 availability)

Figure 10.6 reports an even higher fare premium for competitor 1 ($\approx 35\%$) on the considered carrier's early departure.

Table 11.20 reports the model's overall statistics compared to Model 10, which shows that Model 11 has not significantly suffered from the exclusion of the binary variable.

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{1,0.975}$	Evaluation
11	−23,700	38	47,475	no significant difference	4.488	5.024	no significant difference
10	−23,697	39	47,473				

Table 11.20: Comparison Model 11 vs. Model 10

The consequent next step is to test whether the omission of the insignificant competitive fare variables has a measurable impact on Model 11.

Exclude Insignificant Competitor Fares

$$
\begin{aligned}
V_n^o &= \widetilde{V}_n^o \\
V_n^e &= \widetilde{V}_n^e + \beta^e_{\widetilde{D}} \cdot \widetilde{D} + \beta^e_{b_{1,e}} \cdot b_{1,e} + \beta^e_{\ln(p_{1,e})} \cdot \ln(p_{1,e}) \\
&\qquad + \beta^e_{b_{1,m}} \cdot b_{1,m} \\
&\qquad + \beta^e_{b_{2,m}} \cdot b_{2,m} + \beta^e_{\ln(p_{2,m})} \cdot \ln(p_{2,m}) \\
V_n^l &= \widetilde{V}_n^l + \beta^l_{\widetilde{D}} \cdot \widetilde{D} + \beta^l_{b_{1,e}} \cdot b_{1,e} + \beta^l_{\ln(p_{1,e})} \cdot \ln(p_{1,e}) \\
&\qquad + \beta^l_{b_{1,m}} \cdot b_{1,m} + \beta^l_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) \qquad \text{(Model 12)}
\end{aligned}
$$

The new model now excludes the insignificant competitor fares from Model 11, leaving the representative utility of the late departing alternative V_n^l totally unaffected by the flight offer of competitor 2 (incl. price).[6]

Looking at the results in Table 11.21, one major change to the coefficient

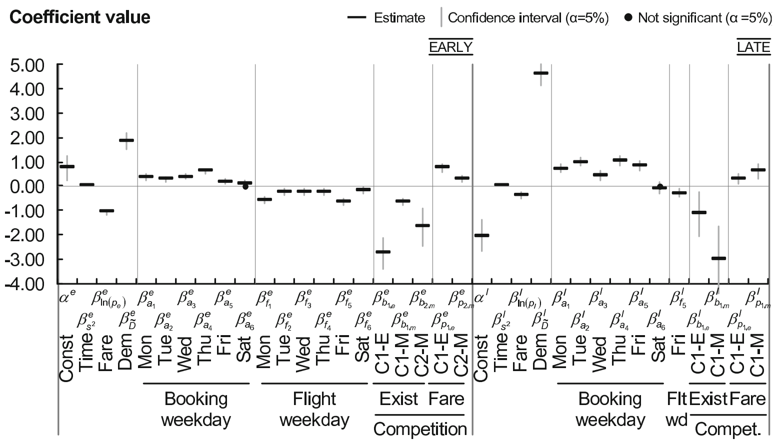


Figure 11.15: Estimation results for Model 12 (exclude insignificant competitor fares)
Source: Own design, based on estimates

[6] The exclusion of the specified competitor fares has also been tested individually, yielding the same results for the overall model. The intermediate incremental results are not reported here for brevity.

Early departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^e	0.746000	0.264000	0.00	α^l	−2.020000	0.326000	0.00
$\beta^e_{\ln(p_e)}$	−1.070000	0.060600	0.00	$\beta^l_{\ln(p_l)}$	−0.385000	0.075400	0.00
$\beta^e_{s^2}$	0.000501	0.000014	0.00	$\beta^l_{s^2}$	0.000369	0.000020	0.00
$\beta^e_{f_1}$	−0.558000	0.073800	0.00				
$\beta^e_{f_2}$	−0.228000	0.076100	0.00				
$\beta^e_{f_3}$	−0.257000	0.072000	0.00				
$\beta^e_{f_4}$	−0.260000	0.071600	0.00				
$\beta^e_{f_5}$	−0.621000	0.074000	0.00	$\beta^l_{f_5}$	−0.281000	0.077700	0.00
$\beta^e_{f_6}$	−0.145000	0.071700	0.04				
$\beta^e_{a_1}$	0.379000	0.062500	0.00	$\beta^l_{a_1}$	0.733000	0.097300	0.00
$\beta^e_{a_2}$	0.265000	0.062300	0.00	$\beta^l_{a_2}$	0.984000	0.092300	0.00
$\beta^e_{a_3}$	0.387000	0.061200	0.00	$\beta^l_{a_3}$	0.432000	0.098200	0.00
$\beta^e_{a_4}$	0.606000	0.063900	0.00	$\beta^l_{a_4}$	1.030000	0.096900	0.00
$\beta^e_{a_5}$	0.197000	0.065100	0.00	$\beta^l_{a_5}$	0.839000	0.096900	0.00
$\beta^e_{a_6}$	0.104000	0.066300	0.12	$\beta^l_{a_6}$	−0.092400	0.118000	0.43
$\beta^e_{\widetilde{D}}$	1.850000	0.167000	0.00	$\beta^l_{\widetilde{D}}$	4.580000	0.231000	0.00
$\beta^e_{b_{1,e}}$	−2.740000	0.324000	0.00	$\beta^l_{b_{1,e}}$	−1.130000	0.456000	0.01
$\beta^e_{b_{1,m}}$	−0.620000	0.065700	0.00	$\beta^l_{b_{1,m}}$	−3.000000	0.706000	0.00
$\beta^e_{b_{2,m}}$	−1.670000	0.385000	0.00				
$\beta^e_{\ln(p_{1,e})}$	0.739000	0.075300	0.00	$\beta^l_{\ln(p_{1,e})}$	0.291000	0.098400	0.00
				$\beta^l_{\ln(p_{1,m})}$	0.601000	0.146000	0.00
$\beta^e_{\ln(p_{2,m})}$	0.304000	0.081300	0.00				

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.21: Estimation results for Model 12 (exclude insignificant competitor fares)

estimates is apparent: As the fare of competitor 1's midday departure has been excluded, the negative utility effect of the sheer availability of that flight is now a significant driver in the model. That is, the utility of the considered carrier's early flight departure is negatively affected by seat availability on the midday flight of competitor 1 – independent of its prevalent fare level.

Figure 11.15 additionally depicts the estimation results from Table 11.21, wherein it is easy to see that all of the remaining coefficients now have reasonable signs and lend themselves to concise explanations. However, two of the binary dummy variables modeling fare inquiry or booking on Saturdays still do not exhibit significant deviations from zero (at $\alpha = 5\%$).

Before finally removing the insignificant weekday dummy variables from Model 12, the statistics shown in Table 11.22 indicate that the model again has not suffered to a significant extent from the omission of the two competitive fare variables.

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{2,0.975}$	Evaluation
12	$-23,700$	36	$47,472$	no significant difference	1.262	7.378	no significant difference
11	$-23,700$	38	$47,475$				

Table 11.22: Comparison Model 12 vs. Model 11

Specifically, the Log-Likelihood has only changed in the position after the decimal point, with the restricted Model 12 that dropped two binary variables yielding the better AIC.

The model below finally excludes the remaining non-significant dummy variables for request and booking on Saturday and compares its statistical performance against previous Model 08, which did not include any external fare variables whatsoever.

Final Model

This paragraph reports on the final outbound model, which only includes significant utility drivers (at $\alpha = 5\%$) in a meaningful and statistically sound composition. For lucidity, the model is notated in extenso below.

$$
\begin{aligned}
V_n^o &= \alpha^o \\
V_n^e &= \alpha^e + \beta^e_{\ln(p_e)} \cdot \ln(p_e) + \beta^e_{s^2} \cdot s^2 \\
&\quad + \sum_{d=1}^{6} \beta^e_{f_d} \cdot f_d + \sum_{d=1}^{5} \beta^e_{a_d} \cdot a_d + \beta^e_{\widetilde{D}} \cdot \widetilde{D} \\
&\quad + \beta^e_{b_{1,e}} \cdot b_{1,e} + \beta^e_{\ln(p_{1,e})} \cdot \ln(p_{1,e}) \\
&\quad + \beta^e_{b_{1,m}} \cdot b_{1,m} \\
&\quad + \beta^e_{b_{2,m}} \cdot b_{2,m} + \beta^e_{\ln(p_{2,m})} \cdot \ln(p_{2,m}) \\
V_n^l &= \alpha^l + \beta^l_{\ln(p_l)} \cdot \ln(p_l) + \beta^l_{s^2} \cdot s^2 \\
&\quad + \beta^l_{f_5} \cdot f_5 + \sum_{d=1}^{5} \beta^l_{a_d} \cdot a_d + \beta^l_{\widetilde{D}} \cdot \widetilde{D} \\
&\quad + \beta^l_{b_{1,e}} \cdot b_{1,e} + \beta^l_{\ln(p_{1,e})} \cdot \ln(p_{1,e}) \\
&\quad + \beta^l_{b_{1,m}} \cdot b_{1,m} + \beta^l_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) \qquad \text{(Model 13)}
\end{aligned}
$$

Early departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^e	0.797000	0.262000	0.00	α^l	−2.050000	0.323000	0.00
$\beta^e_{\ln(p_e)}$	−1.070000	0.060600	0.00	$\beta^l_{\ln(p_l)}$	−0.386000	0.075400	0.00
$\beta^e_{s^2}$	0.000501	0.000014	0.00	$\beta^l_{s^2}$	0.000369	0.000020	0.00
$\beta^e_{f_1}$	−0.558000	0.073800	0.00				
$\beta^e_{f_2}$	−0.230000	0.076100	0.00				
$\beta^e_{f_3}$	−0.258000	0.072000	0.00				
$\beta^e_{f_4}$	−0.261000	0.071600	0.00				
$\beta^e_{f_5}$	−0.623000	0.074000	0.00	$\beta^l_{f_5}$	−0.282000	0.077700	0.00
$\beta^e_{f_6}$	−0.145000	0.071700	0.04				
$\beta^e_{a_1}$	0.327000	0.052400	0.00	$\beta^l_{a_1}$	0.774000	0.082200	0.00
$\beta^e_{a_2}$	0.213000	0.052100	0.00	$\beta^l_{a_2}$	1.030000	0.076200	0.00
$\beta^e_{a_3}$	0.336000	0.050800	0.00	$\beta^l_{a_3}$	0.474000	0.083300	0.00
$\beta^e_{a_4}$	0.553000	0.053800	0.00	$\beta^l_{a_4}$	1.070000	0.081700	0.00
$\beta^e_{a_5}$	0.145000	0.055500	0.01	$\beta^l_{a_5}$	0.880000	0.081800	0.00
$\beta^e_{\tilde{D}}$	1.830000	0.167000	0.00	$\beta^l_{\tilde{D}}$	4.580000	0.231000	0.00
$\beta^e_{b_{1,e}}$	−2.720000	0.324000	0.00	$\beta^l_{b_{1,e}}$	−1.160000	0.456000	0.01
$\beta^e_{b_{1,m}}$	−0.621000	0.065700	0.00	$\beta^l_{b_{1,m}}$	−2.950000	0.706000	0.00
$\beta^e_{b_{2,m}}$	−1.650000	0.385000	0.00				
$\beta^e_{\ln(p_{1,e})}$	0.736000	0.075200	0.00	$\beta^l_{\ln(p_{1,e})}$	0.296000	0.098400	0.00
				$\beta^l_{\ln(p_{1,m})}$	0.590000	0.146000	0.00
$\beta^e_{\ln(p_{2,m})}$	0.299000	0.081200	0.00				

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.23: Estimation results for Model 13 (final outbound model)

Table 11.23 reports the merely fractionally altered estimates on the final model. Only the remaining coefficients for the dummy variables indicating the weekday on which a particular request was submitted have slightly changed to adjust for the missing influence of Saturdays. Note that all remaining coefficients are clearly significant at the chosen level of $\alpha = 5\%$ and that the coefficients all lend themselves to a reasonable explanation in terms of sign and magnitude.

The coefficients for Model 13 are also depicted in Figure 11.16, which allows for a straightforward qualitative interpretation of the model in anticipation of Section 12.2. Both alternatives exhibit an increased utility when booked closer to departure, the effect being most prominent for the early departure. A similar observation can be made for the fare level, where higher fares naturally yield lower utility especially on the early departure where customers seem extremely price-sensitive.

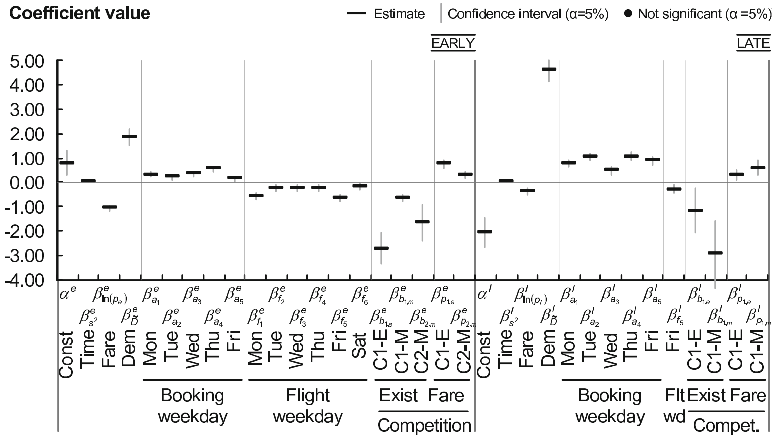


Figure 11.16: Estimation results for Model 13 (final model)
Source: Own design, based on estimates

Regarding competitive dynamics, decision makers respond to the sheer availability of alternatives. However, the effect depends on the closeness of the flights in departure time. The price sensitivity depends on the overall fare level of the considered carrier's alternatives, i.e., it is higher with the bargain fares on the early departures (for a more thorough discussion of the results, see Chapter 12).

The comparative statistics in Table 11.24 highlight that the final model again has not suffered considerably from the exclusion of non-significant variables. Additionally, Table 11.25 shows that the intermediate Model 08, which does not contain any external fare and availability variables, in comparison is significantly less adequate for the data.

Before Chapter 12 reports the computational results for the model in terms of overall fit, elasticities and substitutional patterns, the following section describes the model differences in terms of external choice drivers for

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{2,0.975}$	Evaluation
13	−23,702	34	47,472	no significant difference	3.128	7.378	no significant difference
12	−23,700	36	47,472				

Table 11.24: Comparison Model 13 vs. Model 12

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{0,0.975}$	Evaluation
13	$-23,072$	34	$47,742$	Model 13 clearly better	n/a	n/a	n/a
08	$-24,044$	25	$48,137$				

Table 11.25: Comparison Model 13 vs. Model 08

the inbound directional estimates.

11.2.4 External Inbound Choice Drivers

This section highlights the differences in the definition of the inbound directional model in comparison to the above finalized Model 13 for the outbound direction.

As the definition regarding internal choice drivers and decision maker characteristics given in Sections 11.2.1 and 11.2.2 is straightforward and also yields similar results (naturally with differing coefficient values), the intermediate steps are omitted and the full model including the overall latent demand driver is introduced below.

Based on the differences in schedule (see Figure 10.3), the flight alternatives of the considered carrier belong to the midday and late departure segments, while competitor 1 has also scheduled departures in the midday and late segments and competitor 2 has a scheduled departure within the late segment only. Note that the latter competitor's flights only compete with the late departing flights of the considered carrier as it does not operate midday flights on the weekend.

Depending on the differences in competitor schedules, the inbound model is rooted in the Friday departures. That is, the binary dummy variables indicating a Friday departure are omitted to yield an identifiable model.

Full Inbound Model

$$
\begin{aligned}
V_n^o &= \widetilde{V}_n^o \\
V_n^m &= \widetilde{V}_n^m + \beta^m_{\widetilde{D}} \cdot \widetilde{D} + \beta^m_{b_{1,m}} \cdot b_{1,m} + \beta^m_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) \\
&\qquad + \beta^m_{b_{1,l}} \cdot b_{1,l} + \beta^m_{\ln(p_{1,l})} \cdot \ln(p_{1,l}) \\
V_n^l &= \widetilde{V}_n^l + \beta^l_{\widetilde{D}} \cdot \widetilde{D} + \beta^l_{b_{1,m}} \cdot b_{1,m} + \beta^l_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) \\
&\qquad + \beta^l_{b_{1,l}} \cdot b_{1,l} + \beta^l_{\ln(p_{1,l})} \cdot \ln(p_{1,l})
\end{aligned}
$$

$$+ \beta^l_{b_{2,l}} \cdot b_{2,l} + \beta^l_{\ln(p_{2,l})} \cdot \ln(p_{2,l}) \qquad \text{(Model 14)}$$

Table 11.26 reports the estimation results for the full inbound model, yielding expectable results on the binary dummy variables for the request and booking weekdays as well as the squared advance purchase time and logarithmized fare variables.[7]

The illustration of estimation results in Figure 11.17 again lends an intuitive view of the model. Obviously, the coefficients do not suffer from any particularities that manifest in unanticipated signs. However, a few coefficients seem not to be significant drivers at the chosen level of $\alpha = 5\%$.

Specifically, two pairs of binary dummy variables indicating the existence

Mid day departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^m	−1.030000	0.287000	0.00	α^l	−3.190000	0.378000	0.00
$\beta^m_{s^2}$	0.000688	0.000015	0.00	$\beta^l_{s^2}$	0.000443	0.000021	0.00
$\beta^m_{\ln(p_m)}$	−0.847000	0.062800	0.00	$\beta^l_{\ln(p_l)}$	−0.322000	0.084000	0.00
$\beta^m_{\tilde{D}}$	3.690000	0.146000	0.00	$\beta^l_{\tilde{D}}$	1.540000	0.265000	0.00
$\beta^m_{a_1}$	0.717000	0.069900	0.00	$\beta^l_{a_1}$	0.699000	0.091300	0.00
$\beta^m_{a_2}$	0.658000	0.070400	0.00	$\beta^l_{a_2}$	0.855000	0.088700	0.00
$\beta^m_{a_3}$	0.607000	0.069800	0.00	$\beta^l_{a_3}$	0.618000	0.089000	0.00
$\beta^m_{a_4}$	0.753000	0.073000	0.00	$\beta^l_{a_4}$	0.594000	0.096500	0.00
$\beta^m_{a_5}$	0.568000	0.072000	0.00	$\beta^l_{a_5}$	0.567000	0.094600	0.00
$\beta^m_{a_6}$	0.274000	0.072800	0.00	$\beta^l_{a_6}$	0.420000	0.096800	0.00
$\beta^m_{f_1}$	0.302000	0.051500	0.00				
$\beta^m_{f_2}$	0.503000	0.056700	0.00				
$\beta^m_{f_3}$	0.311000	0.093200	0.00				
$\beta^m_{f_4}$	0.413000	0.050900	0.00				
				$\beta^l_{f_6}$	0.214000	0.111000	0.05
				$\beta^l_{f_7}$	1.060000	0.094300	0.00
$\beta^m_{b_{1,m}}$	−1.450000	0.357000	0.00	$\beta^l_{b_{1,m}}$	−0.096900	0.518000	0.85
$\beta^m_{b_{1,l}}$	−0.403000	0.660000	0.54	$\beta^l_{b_{1,l}}$	−1.860000	0.615000	0.00
				$\beta^l_{b_{2,l}}$	−0.923000	0.344000	0.01
$\beta^m_{\ln(p_{1,m})}$	0.309000	0.079700	0.00	$\beta^l_{\ln(p_{1,m})}$	0.125000	0.114000	0.27
$\beta^m_{\ln(p_{1,l})}$	0.138000	0.151000	0.36	$\beta^l_{\ln(p_{1,l})}$	0.350000	0.137000	0.01
				$\beta^l_{\ln(p_{2,l})}$	0.223000	0.074600	0.00

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.26: Estimation results for Model 14 (full model)

[7] The necessity for an exponential and logarithmic transformation, respectively, can be shown for the inbound direction in a similar manner as in Section 11.2.1 for the outbound direction, but is omitted here for brevity.

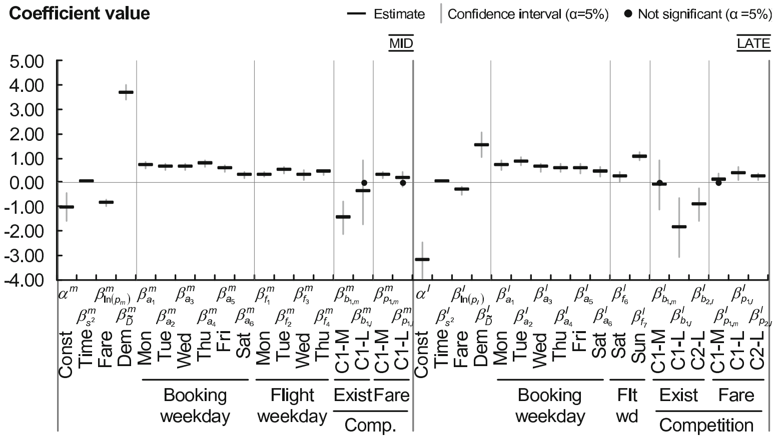


Figure 11.17: Estimation results for Model 14 (full inbound model)
Source: Own design, based on estimates

of competitor flights and their corresponding fare variables are not significant at the chosen level, one in each of the considered carrier's alternatives.

In both cases, the particular competitor flight that does not depart in the same but in an adjacent time segment is affected. At any one time, the binary variable exhibits a high variance yielding a comparably low significance.

Accordingly, both binary variables are removed below to examine the impact on the overall goodness and the remaining coefficients' significances.

Exclusion of Insignificant Binary Variables

$$
\begin{aligned}
V_n^o &= \tilde{V}_n^o \\
V_n^m &= \tilde{V}_n^m + \beta^m_{\widetilde{D}} \cdot \widetilde{D} + \beta^m_{b_{1,m}} \cdot b_{1,m} + \beta^m_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) \\
&\qquad + \beta^m_{\ln(p_{1,l})} \cdot \ln(p_{1,l}) \\
V_n^l &= \tilde{V}_n^l + \beta^l_{\widetilde{D}} \cdot \widetilde{D} \qquad + \beta^l_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) \\
&\qquad + \beta^l_{b_{1,l}} \cdot b_{1,l} + \beta^l_{\ln(p_{1,l})} \cdot \ln(p_{1,l}) \\
&\qquad + \beta^l_{b_{2,l}} \cdot b_{2,l} + \beta^l_{\ln(p_{2,l})} \cdot \ln(p_{2,l}) \qquad \text{(Model 15)}
\end{aligned}
$$

Model 15 now excludes the respective binary variables indicating whether a competing flight is available or sold out in the adjacent time segment of the considered carrier's own flights (see Table 11.27 for estimation results). The reasoning behind the new model specification is as follows: The decision maker may be specifically looking for a particular departure time segment and is initially unaffected by available flights in adjacent departure time segments. However, with declining fares of these flights, the decision maker may switch, but she remains uninfluenced if such flights become sold out.

Figure 11.18 illustrates the model coefficients from Table 11.27, showing two major developments: Most importantly, the exclusion of the binary variables for availability has led the corresponding fare variables to be significant at the chosen level. Additionally, Saturday as a departure weekday for the late inbound flight seems to no longer induce significantly different utility than the base day of Friday; that is, only Sunday departures carry

Mid day departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^m	−1.040000	0.287000	0.00	α^l	−3.230000	0.340000	0.00
$\beta^m_{s^2}$	0.000688	0.000015	0.00	$\beta^l_{s^2}$	0.000444	0.000020	0.00
$\beta^m_{\ln(p_m)}$	−0.845000	0.062800	0.00	$\beta^l_{\ln(p_l)}$	−0.314000	0.076300	0.00
$\beta^m_{\widetilde{D}}$	3.690000	0.146000	0.00	$\beta^l_{\widetilde{D}}$	1.550000	0.265000	0.00
$\beta^m_{a_1}$	0.718000	0.069900	0.00	$\beta^l_{a_1}$	0.700000	0.091400	0.00
$\beta^m_{a_2}$	0.658000	0.070400	0.00	$\beta^l_{a_2}$	0.856000	0.088700	0.00
$\beta^m_{a_3}$	0.606000	0.069800	0.00	$\beta^l_{a_3}$	0.619000	0.089100	0.00
$\beta^m_{a_4}$	0.752000	0.073000	0.00	$\beta^l_{a_4}$	0.595000	0.096600	0.00
$\beta^m_{a_5}$	0.567000	0.072000	0.00	$\beta^l_{a_5}$	0.568000	0.094600	0.00
$\beta^m_{a_6}$	0.274000	0.072800	0.00	$\beta^l_{a_6}$	0.420000	0.096800	0.00
$\beta^m_{f_1}$	0.304000	0.051400	0.00				
$\beta^m_{f_2}$	0.505000	0.056600	0.00				
$\beta^m_{f_3}$	0.305000	0.092800	0.00				
$\beta^m_{f_4}$	0.413000	0.050900	0.00				
				$\beta^l_{f_6}$	0.211000	0.111000	0.06
				$\beta^l_{f_7}$	1.060000	0.093900	0.00
$\beta^m_{b_{1,m}}$	−1.530000	0.336000	0.00				
				$\beta^l_{b_{1,l}}$	−1.940000	0.432000	0.00
				$\beta^l_{b_{2,l}}$	−0.921000	0.342000	0.01
$\beta^m_{\ln(p_{1,m})}$	0.325000	0.075000	0.00	$\beta^l_{\ln(p_{1,m})}$	0.104000	0.019800	0.00
$\beta^m_{\ln(p_{1,l})}$	0.138000	0.151000	0.36	$\beta^l_{\ln(p_{1,l})}$	0.366000	0.098200	0.00
				$\beta^l_{\ln(p_{2,l})}$	0.223000	0.074400	0.00

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.27: Estimation results for Model 15 (exclusion of insignificant binary variables)

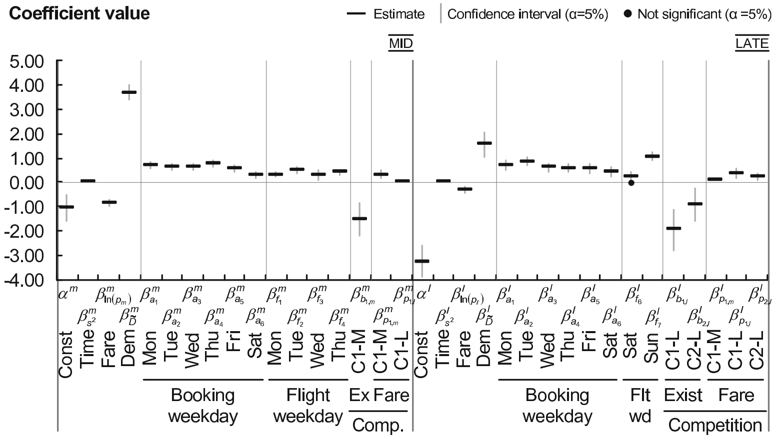


Figure 11.18: Estimation results for Model 15 (exclusion of insignificant binary variables)
Source: Own design, based on estimates

additional utility.

A comparison of Tables 11.26 and 11.27 shows that the coefficients have only changed marginally. Mostly, such fares of competitor flights have varied in influence where the binary existence variable has been removed.

Table 11.28 reports the statistical test results on whether Model 15 has suffered a significant loss of explanatory value, which is obviously not the case here. As Model 15 is a linearly restricted version of Model 14, the Likelihood Ratio test can be used for comparison.

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{1,0.975}$	Evaluation
15	−22,176	34	44,421	no significant difference	0.376	5.024	no significant difference
14	−22,176	35	44,422				

Table 11.28: Comparison Model 15 vs. Model 14

Similar to Section 11.2.3, the final step for the inbound model now excludes the single non-significant driver indicating whether the late inbound flight departure happens to be on a Saturday.

Final Inbound Model

$$
\begin{aligned}
V_n^o &= \alpha_n^o \\
V_n^m &= \alpha^m + \beta^m_{\ln(p_m)} \cdot \ln(p_m) + \beta^m_{s^2} \cdot s^2 \\
&\quad + \sum_{d=1}^{4} \beta^m_{f_d} \cdot f_d + \sum_{d=1}^{6} \beta^m_{a_d} \cdot a_d + \beta^m_{\widetilde{D}} \cdot \widetilde{D} \\
&\quad + \beta^m_{b_{1,m}} \cdot b_{1,m} + \beta^m_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) \\
&\qquad\qquad + \beta^m_{\ln(p_{1,l})} \cdot \ln(p_{1,l}) \\
V_n^l &= \alpha^l + \beta^l_{\ln(p_l)} \cdot \ln(p_l) + \beta^l_{s^2} \cdot s^2 \\
&\quad + \beta^l_{f_{7_l}} \cdot f_7 + \sum_{d=1}^{6} \beta^l_{a_d} \cdot a_d + \beta^l_{\widetilde{D}} \cdot \widetilde{D} \\
&\qquad\qquad + \beta^l_{\ln(p_{1,m})} \cdot \ln(p_{1,m}) \\
&\quad + \beta^l_{b_{1,l}} \cdot b_{1,l} + \beta^l_{\ln(p_{1,l})} \cdot \ln(p_{1,l}) \\
&\quad + \beta^l_{b_{2,l}} \cdot b_{2,l} + \beta^l_{\ln(p_{2,l})} \cdot \ln(p_{2,l}) \qquad \text{(Model 16)}
\end{aligned}
$$

Looking at the final model specification for the inbound directional flights, two major differences compared to the outbound version in Model 13 get apparent. First, the weekday on which a request or booking is submitted is of importance for all specific days. As return flights are typically not booked from the holiday destination, the reported last minute effects from Section 11.2.3 are not of any consequence here; that is, all weekdays are significant drivers.

The second observation refers to the impact of competitor availability and fares. The influence seems to be limited to flights and fares that belong to the same time segment. This may be due to decision makers having more specific time preferences when choosing their return flight than when looking for outbound travel.

The final estimates for Model 16 are reported in Table 11.29, whereas Figure 11.19 illustrates the same results for graphic interpretation. Besides the variation based on departure weekday and fare request day, utility seems to be most affected by competitor flights within the same departure time segment. As discussed above, only the fares of adjacent departure times (not their availability variables) affect utility, and even this impact is rather small – albeit significant at the chosen level of $\alpha = 5\%$.

Mid day departure				Late departure			
Coefficient	Estimate	Standard Deviation	Significance[a]	Coefficient	Estimate	Standard Deviation	Significance
α^m	−1.050000	0.287000	0.00	α^l	−3.300000	0.337000	0.00
$\beta^m_{s^2}$	0.000688	0.000015	0.00	$\beta^l_{s^2}$	0.000447	0.000020	0.00
$\beta^m_{\ln(p_m)}$	−0.843000	0.062700	0.00	$\beta^l_{\ln(p_l)}$	−0.276000	0.073500	0.00
$\beta^m_{\widetilde{D}}$	3.690000	0.146000	0.00	$\beta^l_{\widetilde{D}}$	1.610000	0.263000	0.00
$\beta^m_{a_1}$	0.719000	0.069900	0.00	$\beta^l_{a_1}$	0.700000	0.091300	0.00
$\beta^m_{a_2}$	0.658000	0.070400	0.00	$\beta^l_{a_2}$	0.856000	0.088600	0.00
$\beta^m_{a_3}$	0.606000	0.069800	0.00	$\beta^l_{a_3}$	0.616000	0.089000	0.00
$\beta^m_{a_4}$	0.752000	0.073000	0.00	$\beta^l_{a_4}$	0.595000	0.096500	0.00
$\beta^m_{a_5}$	0.568000	0.072000	0.00	$\beta^l_{a_5}$	0.564000	0.094600	0.00
$\beta^m_{a_6}$	0.275000	0.072800	0.00	$\beta^l_{a_6}$	0.412000	0.096700	0.00
$\beta^m_{f_1}$	0.301000	0.051400	0.00				
$\beta^m_{f_2}$	0.503000	0.056600	0.00				
$\beta^m_{f_3}$	0.303000	0.092800	0.00				
$\beta^m_{f_4}$	0.411000	0.050900	0.00				
				$\beta^l_{f_7}$	0.924000	0.058300	0.00
$\beta^m_{b_{1,m}}$	−1.500000	0.336000	0.00				
				$\beta^l_{b_{1,l}}$	−1.820000	0.429000	0.00
				$\beta^l_{b_{2,l}}$	−0.950000	0.344000	0.01
$\beta^m_{\ln(p_{1,m})}$	0.319000	0.075000	0.00	$\beta^l_{\ln(p_{1,m})}$	0.082800	0.016000	0.00
$\beta^m_{\ln(p_{1,l})}$	0.046300	0.023900	0.04	$\beta^l_{\ln(p_{1,l})}$	0.370000	0.098400	0.00
				$\beta^l_{\ln(p_{2,l})}$	0.237000	0.074300	0.00

[a] Level of α up to which the estimated coefficient value is significant.

Table 11.29: Estimation results for Model 16 (final inbound model)

Finally, the statistical goodness of the restricted Model 16 relative to the previous Model 15 is reported in Table 11.30. Obviously, the models do not exhibit a statistically significant difference.

			AIC criterion		Likelihood Ratio test		
Model	Log-likelih.	Param.	AIC	Evaluation	LR	$\chi^2_{1,0.975}$	Evaluation
16	−22,178	33	44,422	no significant difference	3.722	5.024	no significant difference
15	−22,176	34	44,421				

Table 11.30: Comparison Model 16 vs. Model 15

In sum, two multinomial logit models have been developed and tested based on informal and formal criteria in the above sections. The following chapter takes a look at the predictive performance of the outcomes and reports the computational results that can be derived from the specification, which may help in day-to-day pricing execution.

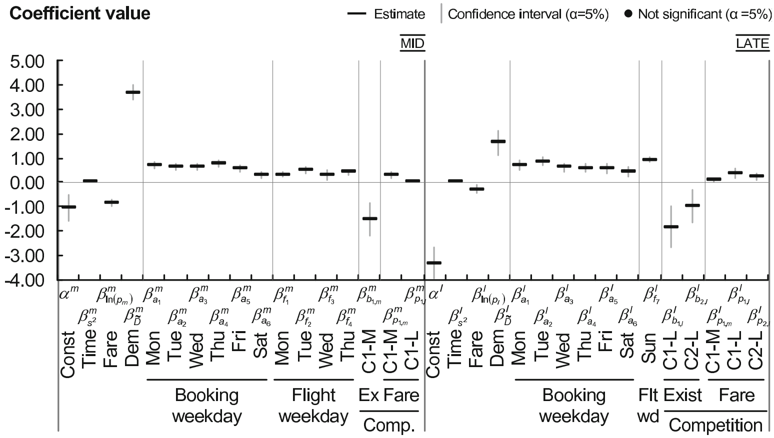


Figure 11.19: Estimation results for Model 16 (final inbound model)
Source: Own design, based on estimates

Chapter 12

Computational Results and Evaluation

wo statistically sound choice models have been developed in Chapter 11. This chapter examines the overall predictive performance of these models in the first Section 12.1. Additionally, specific elasticities and substitutional patterns are inferred in Section 12.2, and finally Section 12.3 draws targeted conclusions on the usage of the reported results in actual airfare pricing.

12.1 Predictive Model Performance

The data sample or observation choice set, whereon the universal multinomial logit model of Chapter 11 is based, has been constructed from a total of four different data sources (see Figure 10.1), which naturally has limited data availability considerably (see Figure 10.2). Nevertheless, it is "conventional statistical folklore" (Picard and Cook, 1984, p. 575) that model building and validation need to be based on different datasets to account for possible over-fitting as "testing the procedure on the data that gave it birth is almost certain to overestimate performance" (Mosteller and Tukey, 1977, p. 37).

The most widely used method for estimating prediction error while preventing measurement bias due to over-fitting the model to the training data is called *cross-validation*, or more precisely *k-fold cross-validation* (see, e.g., Hastie et al., 2001, Sec. 7.10).

In an ideal world of model validation, the initially available data are split into two groups, the *training set* and the *validation set*, where the first is used exclusively for developing and training the model, while the second is kept

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6_12,

aside for later testing of the predictive performance of the trained model on that independent test sample. Obviously, the required amount of data roughly doubles in such a scenario (see Bishop, 2006, Sec. 1.3).

Cross-validation bypasses this issue by using the complete set of available data for both training and validation while assuring that the model's performance is not evaluated on the same data subsets that were used for training the model initially.

In *k-fold cross validation*, the available N observations are split into K mutually exclusive, collectively exhaustive subsets by a random indexing function $\kappa : \{1, \dots, N\} \to \{1, \dots, K\}$. Thus, κ is uniformly distributed over K, in expectancy yielding equally sized sets. The model is then estimated or trained based on observations that belong to only $K - 1$ of these subsets, i.e., where $\kappa(n) \neq k$, while the resulting model parameterization is finally evaluated against the observations in the remaining subset, i.e., where $\kappa(n) = k$ (see Hastie et al., 2001, Sec. 7.10).

To fully utilize the available data and to obtain an unbiased estimate of the overall model goodness of fit or predictive error, respectively, the procedure is repeated K times so that each of the subsets is used for validation exactly once. The performance of the model is finally evaluated using the K different model predictions for the K subsets that were used for validation together with the corresponding original observations of the dependent variable (see Bishop, 2006, Sec. 1.3). The method is illustrated for the most commonly used form of ten-fold cross validation in Figure 12.1.

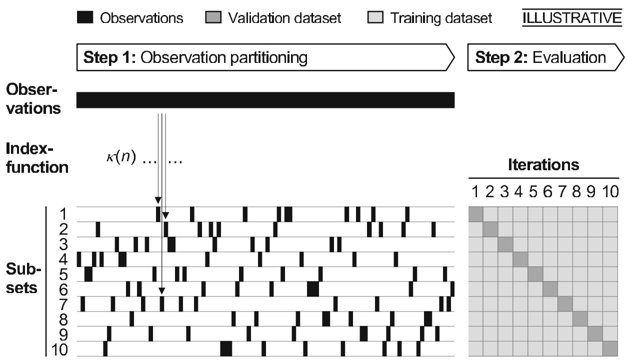


Figure 12.1: Illustration of 10-fold cross-validation
Source: Own design

For scenarios in which data are extremely scarce, *leave-one-out cross-validation* can be used, where $K = N$ and $\kappa(n) = n$. However, for the validation of the two final choice models – Model 13 for the outbound and Model 16 for the inbound direction – ten-fold cross-validation is chosen.

In a first step, the two available observation sets (separately for the outbound and inbound directions) have been partitioned into ten subsets, each based on a uniform distribution. In an iterative process, at any one time, nine of the resulting data sets have been used to train the final models according to their parameterizations in Chapter 11, each time leaving a single subset for later measurement of prediction error.

The estimation results for the ten subsamples can first be compared to those obtained from the initial model specifications above. For the outbound direction, Figure 12.2 depicts the familiar chart exhibiting the model estimates and confidence intervals from Figure 11.16 in Chapter 11 together with the respective results from the ten estimation runs based on ten-fold cross-validation.

While most subsample estimates lie within the 5%-confidence intervals, a few estimates for the inverted demand level ($\beta^i_{\tilde{D}}$) scatter outside the latter. This may be due to the differing macro-seasonailities already observed in Section 10.3.2, possibly reflected to a different degree within the ten subsets.

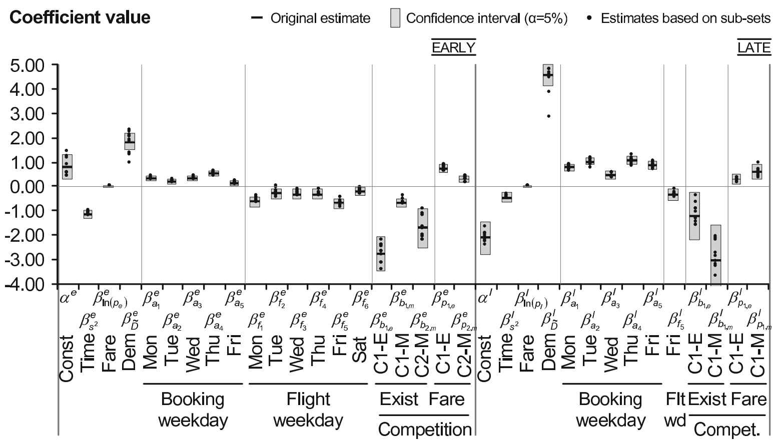


Figure 12.2: Estimation results for final model based on 10-fold cross-validation – outbound

Source: Own design, based on estimates

However, based on the depicted results, Model 13 seems sufficiently stable in coefficient estimations, especially for the many parameters exhibiting extremely small confidence intervals.

Figure 12.3 shows similar results for the inbound Model 16 whose original estimates from model building are taken from Figure 11.19. Again, only a few estimates for the inverted demand, if any, lie outside the provided confidence intervals. Additionally, the effect of Saturday requests for the midday departures exhibits a somewhat high variance.

Although a few estimates lie outside the confidence intervals obtained from the model creation process in Chapter 11, the results can be used to evaluate the model as the deviations are not deemed to falsify the overall model definition – especially as none of them has changed in sign.

In any case, based on the frequently reported limitations to the data (especially the absence of socio-demographic information about the decision makers), the model can only be considered an approximation of true choice behavior at this stage.

The obtained estimates from the ten-fold cross-validation runs are now used to actually predict the choice behavior for the observations contained in the data samples that were not used for training the cross-validation models.

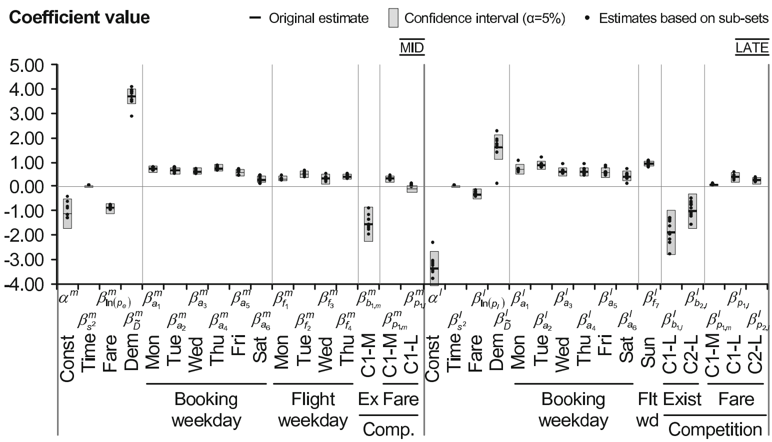


Figure 12.3: Estimation results for final model based on 10-fold cross-validation – inbound

Source: Own design, based on estimates

When combined, the prediction and validation datasets resemble the original full data range, containing the actual choices of the decision makers together with the predictions from the validation models.

As reported above, in all cases, the third "other/no-buy"-alternative yields the highest utility for the individual decision makers and with it the highest choice probability, which is in line with the recorded overall shares (see Section 11.2). Raw econometric theory suggests that the decision maker always chooses the alternative that yields the highest individual utility. However, the explicit acknowledgment of uncertainty within the multinomial logit model derives choice probabilities that, in expectancy, represent the share of decision makers who will choose a particular alternative if the number of observations chosen is sufficiently high.

To evaluate the predictive performance of the two developed models, this expected book-to-look ratio based on the utility calculations is compared to the actually observed ratio within the data. Naturally, the number of individual bookings fluctuates heavily based on a daily observation granularity per departure and booking date. Therefore, the book-to-look ratios are compared over the full booking period of individual flight departures, and the error – that is, the share of the total passengers on a flight that has been mispredicted – is again evaluated using the total absolute percentage error (TAPE), similar to Part II.

Figure 12.4 depicts the observed and predicted book-to-look ratios for the outbound directed flights separately for the early and late departure.[1] Notably, the predictions retrace the variations of the observed rate quite closely. However, the model is obviously not capable of capturing extreme spikes caused by specific calendar effects.

As the developed model is explicitly based on automatically collected data, it does not consider the depicted demand effects based on vacations or carnival (see Figure 12.4). Nevertheless, the results could potentially be adjusted manually to account for such events, which is common practice at most airlines.

Interestingly, the highest relative deviations from the observed actual choice as reported in Figure 12.5 are not triggered by such peak events, but rather by unusually low booking rates (① – ③) that are eventually overestimated (compare Figures 12.4 and 12.5).

---

1 For readability, the individual values are connected by a solid line, although strictly speaking they are discrete points.

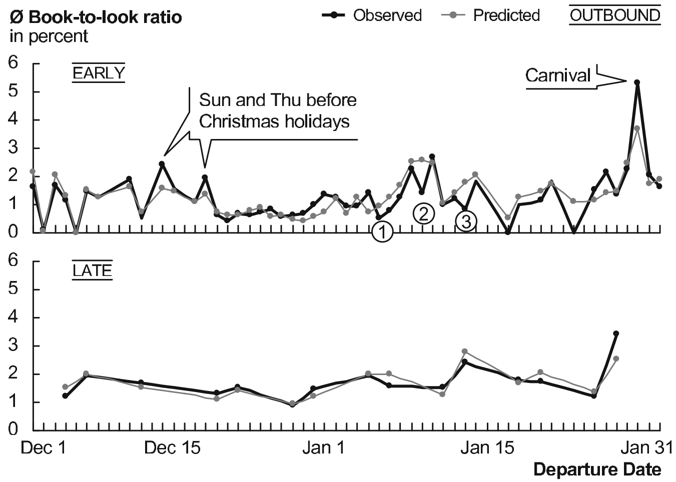


Figure 12.4: Observed and predicted book-to-look ratios – outbound
Source: Own design, based on results

The results in Figure 12.5 also show that the model's performance seems unaffected by the overall level of requests or observations. That is, the peak-demand season around Christmas is equally well predicted as the low-demand season in the end of January.

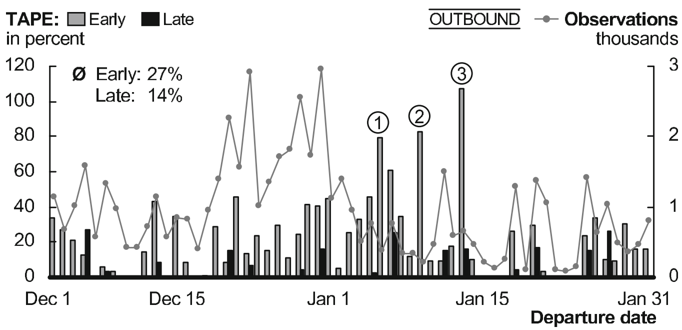


Figure 12.5: Customer choice prediction errors – outbound
Source: Own design, based on results

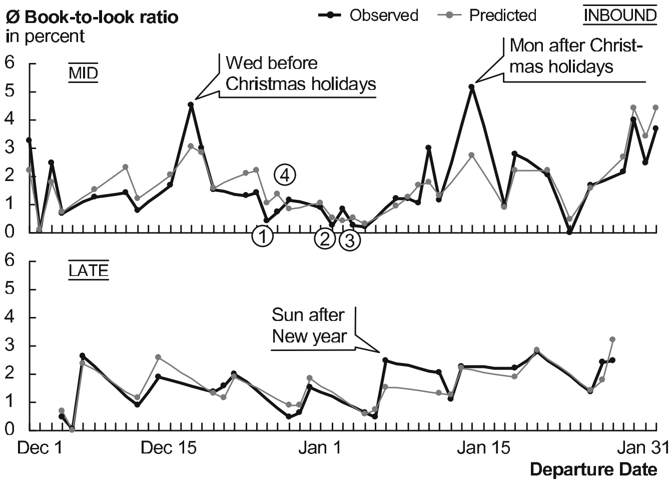


Figure 12.6: Observed and predicted book-to-look ratios – inbound
Source: Own design, based on results

In summary, the model's predictive performance is satisfactory, yielding TAPEs of 27% and 14% for the early and late flight departures, respectively. To calibrate the results, it should be noted again that the reported predictive results are based on raw data from automated systems that do not contain any specific sociodemographic information about the decision makers, which would likely further improve the results substantially.

The comparison of observed and predicted book-to-look ratios for the inbound direction in Figure 12.6 shows similar results to those obtained above. Here, the deviations seem higher as the overall variation between neighboring flight departure dates is severe (especially for the late departure at the bottom). Again, the model in the current specification and based on the available data basis is not capable of capturing singular calendar effects, e.g., passengers explicitly returning before or after Christmas.

The highest relative deviations from the observed choice can again be observed for departure days exhibiting unusually low booking levels (① – ③) that are eventually overestimated (compare Figures 12.6 and 12.7). The inbound errors are also high for the unforeseen return traffic shortly after Christmas (④), which exhibits an atypical pattern based on the underlying weekdays.

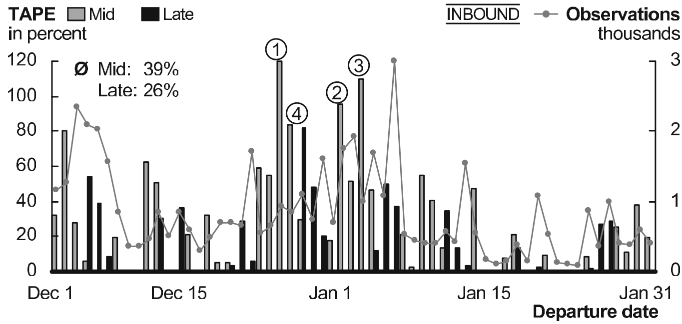


Figure 12.7: Customer choice prediction errors – inbound
Source: Own design, based on results

The overall error level is higher compared to the outbound direction, averaging at TAPEs of 39% and 26% for the midday and late departing flights, respectively. The worse performance of the inbound model can potentially be attributed to the stronger presence of peculiarities and singular effects in the data.

The above reported predictive performance of the developed models demonstrates that it is difficult to construct an accurate model of customer choice in a competitive market based on limited data, but nevertheless also achievable. Various other works in this area have obtained sound results when estimated based on explicitly collected survey results that naturally provide the full necessary data basis for more thorough model designs (see Warburg et al., 2006, Sec. 2.1, for a recent overview). However, aiming at continuous usage and calculation of purchase probabilities for dynamic pricing purposes, such approaches are not feasible; hence, the analyst has to rely on the available real-time and automated data sources.

Besides yielding pure choice probabilities under given scenarios, the power of discrete choice models lies in the analysis of specific choice drivers, which can give insights beyond plain book-to-look ratios – especially for manual intervention in dynamic pricing departments. The next section examines such additional insights that can be derived from the developed models.

12.2 Choice Elasticities of Fare Changes

The predicted results from Section 12.1 above describe the expected customer choice probabilities as a parameterized function of the observed input variables. Consequently, it is of particular interest to understand the responses of such choice probabilities to changes in the underlying functional drivers (see, e.g., Koppelman and Bhat, 2006, Sec. 4.4.2). This section lends a purposive view on specific drivers that can be actively influenced by the market participants.

The developed models contain two types of influenceable variables with direct effects on the customer choice probabilities: the considered carrier's *own fares*, which can be proactively altered by the airline's pricing analyst, and the relevant *competitor fares*, which reflect the reactions or similar proactive actions of the competition.

The appropriate economic measures for market reactions based on such changes are *elasticities* that – all else equal – define the ratio of the percentage change in the dependent variable y based on an infinitesimal percentage change of the independent variable x (see, e.g., Varian, 1992)

$$E_{y,x} = \frac{\partial y}{\partial x}\frac{x}{y}. \tag{12.1}$$

Based on the models developed in Chapter 11, three different scenarios for a market participant's action can be defined, resulting in three different elasticities to be observed for the **choice probability P_i of a particular flight alternative i** of the considered carrier with respect to a...

- **Change in fare p_i of the same alternative i:** Solely the fare of the considered flight alternative i is adjusted, resulting in a direct effect on its choice probability P_i (see, e.g., Train, 2003, p. 63):

$$\begin{aligned}
E_{P_i,p_i} = \frac{\partial P_i}{\partial p_i}\frac{p_i}{P_i} &= \frac{\partial\left(\frac{e^{V_i}}{\sum_j e^{V_j}}\right)}{\partial p_i}\frac{p_i}{P_i} \\
&= \frac{\frac{\partial e^{V_i}}{\partial p_i}\cdot\sum_j e^{V_j} - \frac{\partial \sum_j e^{V_j}}{\partial p_i}e^{V_i}}{(\sum_j e^{V_j})^2}\frac{p_i}{P_i} \\
&= \left[\frac{\partial V_i}{\partial p_i}\frac{e^{V_i}}{\sum_j e^{V_j}} - \frac{\partial V_i}{\partial p_i}\frac{e^{V_i}e^{V_i}}{(\sum_j e^{V_j})^2}\right]\frac{p_i}{P_i}
\end{aligned}$$

$$= \frac{\partial V_i}{\partial p_i}\big(P_i - P_i^2\big)\frac{p_i}{P_i}$$

$$= \frac{\partial V_i}{\partial p_i} p_i \big(1 - P_i\big), \tag{12.2}$$

which based on the logarithmized fares $\ln(p_i)$ with coefficients $\beta^i_{\ln(p_i)}$ in V_i can be reduced to

$$= \beta^i_{\ln(p_i)} \frac{1}{p_i} \cdot p_i \big(1 - P_i\big)$$

$$= \beta^i_{\ln(p_i)} \big(1 - P_i\big). \tag{12.3}$$

- **Change in fare p_k of the other alternative k:** The fare of the second alternative flight k is changed, with a direct effect on its choice probability, which indirectly also affects the choice probability of alternative i as necessarily all probabilities sum to one (see, e.g., Train, 2003, p. 64):

$$E_{P_i,p_k} = \frac{\partial P_i}{\partial p_k}\frac{p_k}{P_i} = \frac{\partial\left(\frac{e^{V_i}}{\sum_j e^{V_j}}\right)}{\partial p_k}\frac{p_k}{P_i}$$

$$= \frac{\frac{\partial e^{V i}}{\partial p_k} \cdot \sum_j e^{V_j} - \frac{\partial \sum_j e^{V_j}}{\partial p_k} e^{V_i}}{(\sum_j e^{V_j})^2}\frac{p_k}{P_i}$$

$$= \left[0 - \frac{\partial V_k}{\partial p_k}\frac{e^{V_k} e^{V_i}}{(\sum_j e^{V_j})^2}\right]\frac{p_k}{P_i}$$

$$= -\frac{\partial V_k}{\partial p_k} P_k P_i \frac{p_k}{P_i}$$

$$= -\frac{\partial V_k}{\partial p_k} p_k P_k, \tag{12.4}$$

which based on the logarithmized fares $\ln(p_k)$ with coefficients $\beta^k_{\ln(p_k)}$ in V_k can be reduced to

$$= -\beta^k_{\ln(p_k)} \frac{1}{p_k} \cdot p_k P_k$$

$$= -\beta^k_{\ln(p_k)} P_k. \tag{12.5}$$

- **Change in fare $p_{c,t}$ of competitor c's flight in time segment t:** Competitor c adjusts the fare of its flight departure in time segment t, which, based on the universal formulation of the MNL, may affect the choice probabilities P_i and P_k of both flights i and k of the considered carrier (derived from Train, 2003, Sec. 3.6):

$$\begin{aligned}
E_{P_i,p_{c,t}} = \frac{\partial P_i}{\partial p_{c,t}} \frac{p_{c,t}}{P_i} &= \frac{\partial\left(\frac{e^{V_i}}{\sum_j e^{V_j}}\right)}{\partial p_{c,t}} \frac{p_{c,t}}{P_i} \\
&= \frac{\frac{\partial e^{V_i}}{\partial p_{c,t}} \cdot \sum_j e^{V_j} - \frac{\partial \sum_j e^{V_j}}{\partial p_{c,t}} e^{V_i}}{(\sum_j e^{V_j})^2} \frac{p_{c,t}}{P_i} \\
&= \left[\frac{\partial V_i}{\partial p_{c,t}} \frac{e^{V_i}}{\sum_j e^{V_j}} - \left(\frac{\partial V_i}{\partial p_{c,t}} e^{V_i} + \frac{\partial V_k}{\partial p_{c,t}} e^{V_k}\right) \frac{e^{V_i}}{(\sum_j e^{V_j})^2}\right] \frac{p_{c,t}}{P_i} \\
&= \left[\frac{\partial V_i}{\partial p_{c,t}} P_i - \frac{\partial V_i}{\partial p_{c,t}} P_i^2 - \frac{\partial V_k}{\partial p_{c,t}} P_k P_i\right] \frac{p_{c,t}}{P_i} \\
&= \left[\frac{\partial V_i}{\partial p_{c,t}} \big(1 - P_i\big) - \frac{\partial V_k}{\partial p_{c,t}} P_k\right] p_{c,t}, \qquad (12.6)
\end{aligned}$$

which based on the logarithmized competitor fares $\ln(p_{c,t})$ with coefficients $\beta^i_{\ln(p_{c,t})}$ in V_i for the first alternative i and $\beta^k_{\ln(p_{c,t})}$ in V_k for the second alternative k can be reduced to

$$\begin{aligned}
&= \left[\beta^i_{\ln(p_{c,t})} \frac{1}{p_{c,t}} \cdot \big(1 - P_i\big) - \beta^k_{\ln(p_{c,t})} \frac{1}{p_{c,t}} \cdot P_k\right] p_{c,t} \\
&= \beta^i_{\ln(p_{c,t})} \big(1 - P_i\big) - \beta^k_{\ln(p_{c,t})} P_k. \qquad (12.7)
\end{aligned}$$

If the competitor fare $p_{c,t}$ is not considered a significant driver for both own alternatives according to the estimation results, (12.6)/(12.7) reduce to the corresponding elasticities of the single case, depending on whether it is a driver of the same alternative (12.2)/(12.3) or the other own alternative (12.4)/(12.5).

Based on the derived formulas, Figures 12.8 and 12.9 below report the elasticities for the choice probabilities of the considered carrier's own flights (out- and inbound) on the various possible fare changes (internal and external). As can be taken directly from the above analysis, these elasticities depend on the prevalent choice probability when the change occurs and are

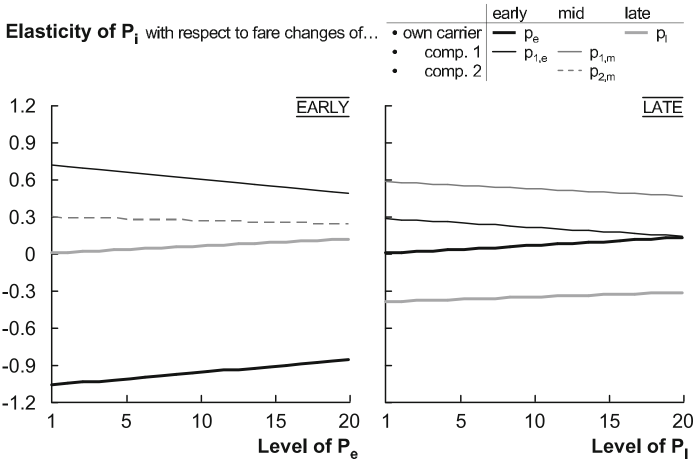


Figure 12.8: Elasticities of purchase probability – outbound
Source: Own design, based on results

therefore plotted along a reasonable window stretching up to 20% choice probability for the flight under examination.

For competitor fare changes, the elasticities depend on the prevalent choice probability regarding *both* alternatives – see (12.6)/(12.7). However, for dimensionality reasons the elasticities are illustrated along the unidimensional choice probability of the considered flight, with the remainder varied automatically, to keep the ratio of choice probabilities constant, depending on the averagely observed market shares ($\frac{P_e}{P_l} = 0.63$ for outbound and $\frac{P_m}{P_l} = 1.16$ for inbound flights). This approach allows a reasonable comparison, especially as the elasticities vary only slightly and linearly in the second choice probability.

Figure 12.8 reports the elasticities for the considered carrier's outbound flights. Naturally, the choice probability for the early departure is heavily (-1.06 – -0.86) elastic to its own underlying fare. Predictably, it also exhibits a strong elasticity with respect to the directly competing early departure of competitor 1, with a slightly smaller magnitude (0.49 – 0.72). The choice probability for the early departure is also slightly elastic (0.01 – 0.12) to the fare level of the own late flight. Similarly, choice is elastic to the fare of competitor 2's midday departure, which, according to Model 13, directly affects choice probability; however, the elasticity is rather constant (0.24 – 0.30). The fare level of competitor 1's midday departure is not a significant

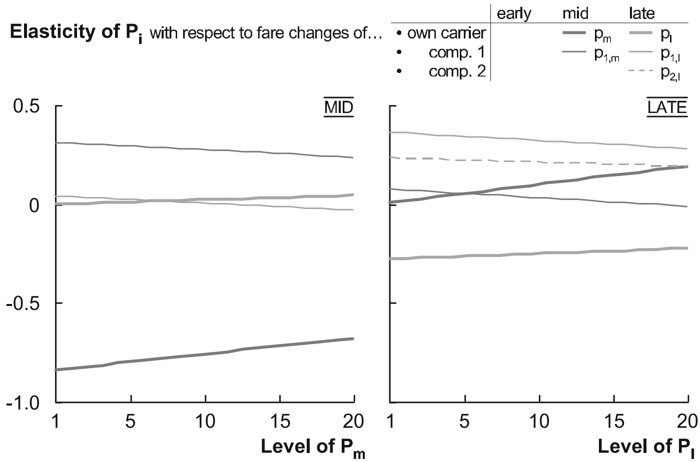


Figure 12.9: Elasticities of purchase probability – inbound
Source: Own design, based on results

driver behind the early departure of the considered carrier (see Table 11.23) and therefore the choice elasticity with respect to it is not shown.

The right side of Figure 12.8 illustrates that the choice probability for the late departure is genuinely less elastic to fare changes. Naturally, it is somewhat elastic to its own fare (-0.38 – -0.31), but even more elastic (0.47 – 0.58) with respect to the apparently most competitive midday departure of competitor 1. It also reacts in a slightly positive elastic manner to the own early departure (0.01 – 0.13) and that of competitor 1 (0.14 – 0.29). In case the choice probability is high ($\approx 20\%$), the elasticities are nearly equal, that is, reactions to fare changes on all early departures are similar. Here, the fare of the midday departure of competitor 2 has been found to not significantly affect the choice probability of the considered carrier's late departure above.

For the inbound direction, Figure 12.9 reports more dramatic results. Obviously, the midday departure is elastic solely to its own fare (-0.83 – -0.67) and somewhat to that of the directly competing midday departure of competitor 1 (0.24 – 0.32). Other fares – especially these of competitors – do not truly affect choice on the midday return.

For the late departure, the picture is more differentiated and somewhat similar to the outbound direction: Its choice probability is mostly elastic to its own fare (-0.27 – -0.22) and those of the late departing flights of competitor 1 (0.29 – 0.37) and competitor 2 (0.19 – 0.23). Fare changes of the own midday departure induce only a slight elasticity in the choice behavior for the own late flight (0.01 – 0.20).

This section has shown that the resulting choice models can be used not only to derive specific predictions for a particular flight's choice probability based on an environmental status quo or expectancy, but to precisely understand the importance of individual drivers in terms of how elastic demand will react to changes.

Both results can potentially be operationalized in airfare pricing. The next section takes a look at such possible utilizations and their limitations in real-life situations.

12.3 Applications to Dynamic Airfare Pricing

The ultimate evaluation of the developed multinomial choice models inevitably must consider whether the discussed results meet the initial claim of Section 3.4 in Part I. That is, does the model help in operationalizing dynamic pricing theory for actual airfare pricing? This final section of Chapter 12.3 starts a discussion on the extent to which this is possible with the results at hand, where potential limitations may exist and how the results can immediately help in conventional day-to-day airfare pricing.

Utilizing Results in Dynamic Pricing Models The discussed literature on dynamic pricing (DP) in Section 3.2 uniformly requires knowledge of a functional dependency between ticket price and purchase probability, which the here developed models are technically capable of delivering. However, two limitations to their operationalization exist.

First, the reported predictive performance of the models could hamper the overall performance of any DP model. As discussed, the accuracy could possibly be enhanced significantly through the inclusion of socio-demographic information about the decision makers, which is currently not included based on data restrictions. Conversely, for a possible application of the models, airlines would have to find a way to collect additional socio-demographic and personal information. Here, frequent flier programs and the corresponding CRM databases could help as long as potential customers are required or at least urged to identify themselves before searching for fares (e.g., by means of a mandatory login on the website). Nonetheless, national data protection laws may limit the amount of data that can be collected and used for the described purposes.

Besides that, it is already common practice at many airlines to manually enhance demand and booking forecasts through adjustments for calendar effects, special events or holidays. In addition to the described systematic model enhancements, such manual interventions could also improve the predictive performance significantly.

Secondly, in a competitive environment, purchase probabilities depend on multiple prices, i.e., fares of all relevant flight departures – including those of the competition. Thus, to calculate the optimal fare level for a carrier's own flights, any DP model will need a specific forecast or assumption about the expected fare changes of all competitors over the full booking period for each relevant flight. Naturally, this wealth of information cannot be known or reasonably forecasted in reality – in particular over the full forecast horizon.

In stable markets, it might be possible to derive the competitive behavior or pricing strategy based on recorded observations. For example, a smaller airline could tend to match the fares of the considered main carrier or a new entrant could try to universally undercut the incumbent carrier by a certain amount. Other carriers might work with somewhat fixed price curves.

If it is not possible to derive a competitor's "pricing schedule", a last resort could be to retreat toward scenario-based price optimization. According to multiple scenarios on how the competition might react to own fare setting, different DP model results could be calculated. A preferred model would then have to be picked based on some predefined criteria, e.g., minimizing risk in competitive reaction or preventing a price war.

Based on the described limitations in data availability, the usage of the models within a possibly automated dynamic pricing system would have to be closely monitored and manually fine-tuned by pricing analysts. By all means, the resulting process would still not result in the fully automated optimization system envisioned in many theoretical works (see Section 3.2).

Immediate Operationalization in Day-to-day Pricing The previous Section 12.2 on choice elasticities has already given an impression of how the results may support pricing analysts at low-cost carriers in their daily work. A thorough understanding of the prevalent price elasticities that affect purchase probability on individual flights can definitely help to improve both pricing efficiency and effectiveness.

First, the model coefficients and their estimates from Sections 11.2.3 and 11.2.4 reveal the truly relevant competitors and their corresponding flights. Analysis of the fare coefficients for individual competitor flights can tell the analyst where to look for relevant benchmarks when adjusting fares to competitive levels. Selected competitor flights may only partly affect demand on own departures, mostly independent of the fare level of the latter.

A thorough understanding of a market's fare dynamics may also help to discover possible market or price leaders, whose fare changes have to be monitored and possibly followed more closely than such of others.

Second, the derived price or choice elasticities from Section 12.2 may help the analyst to determine the right magnitude for decided fare changes. To optimize the overall yield, airlines should capture the available premiums on inelastic flights where demand is stable and sufficient (e.g., midday departures on the inbound direction) while selectively inducing additional and incremental sales through minor fare reductions or discounts on flights that do react elastic, i.e., where such pushes in fact induce higher sales and not pure windfalls for the customers (e.g., early departures on the outbound direction).

Besides the discussed purely quantitative steering measures, a *qualitative* understanding of the different effects of flight departure weekdays can help to set overall price levels correctly. Additionally, understanding customer behavior in terms of when bookings take place can help to manage expectations on booking run-ups and to deploy adequate countermeasures appropriately (i.e., the analyst should not panic based on low booking developments over the weekend).

Naturally, a constant – possibly manual – updating of the models and their estimates is required, but then the models may help analysts to keep track of changing market dynamics.

In sum, this chapter has taken a thorough look at the computational results and their possible application in real-world airfare pricing. The final Chapter 13 recapitulates the major results of Part III and provides an outlook on further research opportunities based on the findings.

Chapter 13

Summary and Outlook

The preceding Chapters 9 – 12 *developed*, *validated* and finally *evaluated* a *multinomial logit model* to understand the particular *customer choice probabilities* that convert latent demand in *low-cost air travel markets* to eventual *realized demand*, based on the prevalent price environment in the market and the decision makers' determining characteristics.

The current chapter concludes Part III of this work, in providing a final synthesis of the findings and results of these chapters as well as giving an outlook on possible further complementary research.

Model Description Chapter 11 derives a multinomial logit model that is capable of predicting choice behavior of customers who have expressed latent demand for low-cost air travel as analyzed in Part II of this work.

Based on this objective, a proprietary data basis has been compiled from three different sources, which are accessible from the perspective of a single airline. That is, besides the airline's own reservation and booking system, officially available schedule information and publicly collected fare data of competitors is used. Through the inclusion of competitor fares, the model explicitly acknowledges that purchase probability is a function of a full vector of fares including all available flights considered relevant substitutes by the decision maker, the latter often being omitted in theoretical works for means of simplicity.

While posted fares of all competing alternatives eventually become publicly available, the actual choice of prospective customers cannot be observed exhaustively because carriers that are not part of global distribution systems naturally can only observe their own bookings. Thus, these carriers can, at most, derive a pool of customers who may have bought at the competition or may not have bought a ticket at all.

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6_13,

However, the factually collected choice information in the available systems contains revealed preference data of the entire latent demand arriving at a carrier's reservations system. While it does not allow for modeling of possible hierarchical choice structures or processes, it may allow for a more realistic view than stated preference data collected in artificial field surveys.

Therefore, the conventional multinomial logit model is extended in the spirit of its universal formulation to include competitor fares in the systematic utility of an analyzed carrier's flights. That is, the model does not include multiple competitor flights by defining individual choice alternatives, but assumes the decision maker to know their fares and thus to carry that knowledge as part of her personal characteristics, eventually affecting travel choice at the considered airline (in a similar manner as, for example, her personal income level would).

If the decision maker is not observed to choose one of the available flight alternatives of the considered airline, its choice is attributed to a pooled "other/no-buy"-choice of fixed utility that serves as the reference point for the available and assessable options.

The fare levels of all flight alternatives enter log-linearly into the model, as a particular absolute fare increase is assumed to affect utility less if the base fare level is rather high compared to bargain entry level prices. Likewise, the model considers the remaining booking period to departure exponentially (here, quadratic) because – all else equal – analyses show a strongly increasing booking utility throughout the last days before departure.

Additionally, customer air travel choice seems to depend on both the particular weekday of the considered departure as well as the actual weekday when the booking is made. Finally, in addition to the competitor fares, a set of binary variables indicating the mere existence or availability of such alternative flights is included to account for cases where competing flights are sold out and thus do not exhibit a usable fare number.

Macro-seasonal demand variations are also found to affect the choice behavior as the share of serious demand within overall latent demand varies. That is, the fraction of customers who browse fares in search of a bargain flight without a true travel intention is higher during peak seasons. The model accounts for such effects through the inclusion of the inverted latent demand level.

Naturally, not all competitor fares are found to have a significant influence on the purchase probability of the considered carrier's flights. Predictably, customers seem to exhibit a preference for a particular departure time segment and therefore take as references only fares of those competitor flights that are relevant in relation to that specific time window.

Model Advantages The introduced discrete choice model developed in this part of the work exhibits a range of significant advantages compared to many other works, which foster its application and usability in dynamic price optimization models similar to the ones discussed in Chapter 3 as well as its usability in day-to-day pricing analyst tasks:

- **Consideration of competition:** The model includes competitive dynamics in an oligopolistic market where competition is heavily driven by price. Thus, it does not assume a monopolistic environment for simplification, but explicitly takes utility for a flight to be a functional composition of fare attributes from possibly all available offers that compete for a customer's choice.

- **Prevention of data bias:** Research on price sensitivity and customer reaction to pricing is traditionally prone to yielding biased results when based on stated perceptions of preferences. Specifically collected data from surveys may exhibit such a bias, e.g., based on respondents' divergent valuations of real or fictitious money. This work, therefore, is purposefully based on revealed preference data that report factual decisions of a representative sample spanning all customers who did purchase a flight at a specific airline.

- **High automation:** Data sources for the model have systematically been selected to allow for automated collection of relevant information from all underlying systems. That is, the model is technically transferable to other markets or points in time (i.e., for recurring parameter updates) without the need for additional manual data collection. It is generally transferable to truly realistic environments in which airlines typically serve hundreds of different markets.

- **Practicability:** The design is purposely not based on an exhaustive data basis that contains an artificially constructed full information view of the market, but rather resorts to sources that are realistically available to single market participants. That is, it answers the question to what extent the market behavior of customers can be comprehended based on only partial information about the process outcome.

- **Manifold usability:** The resulting parameterization of the model as well as the derived choice probabilities and elasticities can be used in two ways in lifelike situations. First, the resulting model provides specific choice probabilities for the observed latent demand based on the prevalent market environment, which can technically be employed as input for most conventional dynamic price optimization models throughout the literature as long as sufficient assumptions are provided about competitor pricing behavior. Second, the derived elasticities and identified particular utility drivers can serve as valuable information even for manual price intervention in current non-automated pricing environments.

- **Functionally broad:** The model design is based on conventional multinomial logit and is therefore linear in its variables, which permits easy enhancement through additional data when available (e.g., personal characteristics of decision makers or manual inputs as measures for calendar effects or special events). Design adjustments are intuitive and manageable as they are easy in execution – similar to conventional linear regressions.

Besides the listed advantages of the particular derived model, multinomial logit models in general provide an intuitive and mathematically traceable approach to uncovering the economic rationale behind customer behavior.

Performance and Accuracy The model's overall performance has been evaluated based on common ten-fold cross-validation, partitioning the data into ten equally sized subsets whereby the classification is based on a uniformly distributed index function. The model has then iteratively been estimated based on nine of the subsets, with the results being used to predict customer choice behavior in the remaining tenth dataset. The resulting forecasts adequately represent the model's predictive performance, mitigating the threat of over-fitting coefficients to a particular data sample.

The resulting total absolute percentage error, that is, the relative absolute deviation of the sum of predicted bookings for a particular departure over the considered booking period in comparison to the true cumulative value, is 27% for the early and 14% for the late departure on the outbound flights and 39% for the midday and 26% for the late departure on inbound flights.

These numbers document a satisfactory performance of the model, especially in light of the severe data limitations, namely the absence of any socio-demographic information on the decision makers as well as the lack of manual data adjustment for calendar effects and special events. Nevertheless, the results leave room for further improvements, e.g., based on the inclusion of socio-demographic customer characteristics.

Additionally, the research unveils the significant individual drivers behind customer choice for the analyzed carrier's own flights, which can help in manual steering and intervention in addition to plain forecasts of take-up rates. Naturally, the most important choice drivers are own fares in conjunction with specific fares of competing flight departures. That is, mainly flights that depart in the same, or at least an adjacent departure time segment, seem to affect mutual choice elasticities most heavily, which may be an indication that customers start the booking process with a somewhat clear preference for a particular departure time.

Understanding of such drivers and elasticities can drive actual steering performance simply by guiding the pricing analysts in terms of *where* to look for relevant competitor fares and *how* to (re-)act when aiming at a specific pricing result – higher booking rates induced through decided fare reductions or increased windfall profits through genuinely higher fares in inelastic market situations.

Finally, all results have been derived based on an informally and statistically sound model, where coefficients exhibit the expected signs and magnitudes and can be shown to significantly affect the choice probabilities. Based on the Akaike Information Criterion and the Likelihood Ratio test, the model has also been evaluated to be superior to more restrictive or general model formulations, which have been thoroughly tested in Chapter 11.

Ultimate Usage The models presented in this part of the work can be directly employed in lifelike applications in a two-fold manner. First, they provide a parameterized functional definition of choice probabilities depending on the prevalent price environment and therefore technically fit into most dynamic price optimization models discussed in Part I.

However, the meaningful application of these models would naturally require the predictive performance to be further improved to eventually yield the best possible overall pricing results. A potential lever to improve model performance would be to include truly customer-specific characteristics, such as socio-demographics or travel behavior. Most frequent flier programs should contain information providing a sufficient starting point for model extensions, although national data protection laws may limit their usability. Additionally, manual adjustments for calendar effects and special events may help to predict demand peculiarities around specific holidays or vacation seasons.

Nevertheless, final model performance will also depend on sufficient forecasts of how competitor prices may develop and react to a specific carrier's pricing action, as the model requires competitor fares as input variables.

The most realistic approach here would be to resort to scenario-based optimization depending on different behavioral assumptions or expected pricing strategies of the relevant competition. In markets where a particular competitor represents the unrivaled price leader, its fare could also be regarded simply as a given variable according to which the most appropriate own price point is sought.

Besides the integration into existing dynamic price optimization models with the aim of providing a constitutional part of the relevant input variables, derived results in the form of elasticities and model parametrization can also be used directly to support pricing analysts in their day-to-day work.

First, the functional composition of the models themselves reveals the relevant competition for each of the considered airline's own flights. That is, the sheer significance of coefficients and their estimated values relay insight on where to look in terms of relevant competitors. Specific choice elasticities in relation to competitor fares can amend the insight with information on how competitive fare changes may (or may not) affect realized demand on own departures.

Naturally, understanding of choice elasticities with respect to own fares may help in assessing the right magnitude for own proactive fare changes depending on the desired effect, which could be to induce additional demand based on reduced fares or simply to capture a fare premium in inelastic market situations.

Outlook and Further Research Despite being a younger academic field than revenue management, dynamic pricing is the subject of a growing number of academic studies. However, only a few of the works are explicitly based on lifelike assumptions or market settings. Most optimization models do not consider that realistic markets may exhibit partially irrational behavior based on market participants' evaluation and perception of choices.

Accordingly, there is still ample need for further research moving dynamic pricing and optimization models toward practical usability and acceptance. The development of specific models to understand and possibly predict choice behavior in highly dynamic and price-sensitive markets is a fundamental part in this direction.

In this regard, the described results still provide room for extensions and improvements of appropriate customer choice models, e.g., through the broadening of the underlying data basis. Also, the forecasting of purchasing behavior is only one portion of the necessary input for most price optimization models. Equally important is the forecast of latent demand or customer arrival rates for market participants. Part II provides a first look at the topic and possible solutions.

Regarding the specific multinomial choice model derived in Part III, the above chapters have traced some specific areas for possible improvement:

- **Broadening of external data basis:** The current model's predictive performance is naturally limited by the wealth of available external input variables that are considered. While possibly relevant choice attributes are considered in the formulation, the model does not include any personal characteristics of the decision maker at this stage due to data limitations. However, as such factors may influence individual choice, a possible further research enhancement could look at efficient options for extending the data basis in this direction.

- **Inclusion of competitor choice:** Similar to the above, currently no explicit information about the factual choice of competitor alternatives is included in the model, limiting the design options considerably and possibly also affecting predictive accuracy. Further research should therefore aim at assessing the counterfeiting effects. As lifelike application will always suffer from data limitations, for eventual practical usage, it would be valuable to understand the factual influence of such restrictions on model performance in terms of the additionally induced error.

- **Forecasting/simulation of competitor fares:** The usage of the presented results within dynamic pricing optimization requires the model to provide forecasts on the development of choice probabilities depending on a dynamic price environment. While the own fares are internal factors that can be adjusted to yield optimal results, competitor prices are considered external input variables. Henceforth, for true price optimization, an intelligent forecasting or simulation of competitor pricing behavior needs to exist to generate the necessary input variables.

  Enhancements should account for the dynamic and possibly stochastic price reactions of competitors to internal fare decisions of the considered individual market participant. Approaches in this regard could be learning of typical behavior based on recent observations or the development of true forecasting models, e.g., based on game theoretic approaches or agent-based simulations.

In general, further research in this area should continue to benefit from the growing understanding and tracking of customer needs and behavior through new developments on the Internet and other direct sales channels, which will develop further toward providing a truly personalized offer based on tailor-made product or service definitions with corresponding personally adjusted prices for each individual customer.

Appendix and Bibliography

Appendix

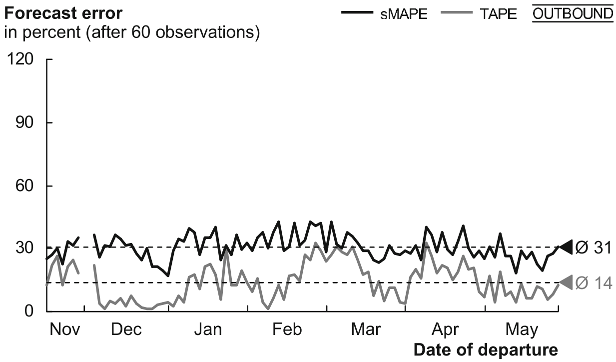


Figure A.1: Full-information forecast errors (sMAPE and TAPE) under noninformative learning – outbound
Source: own design based on collected data

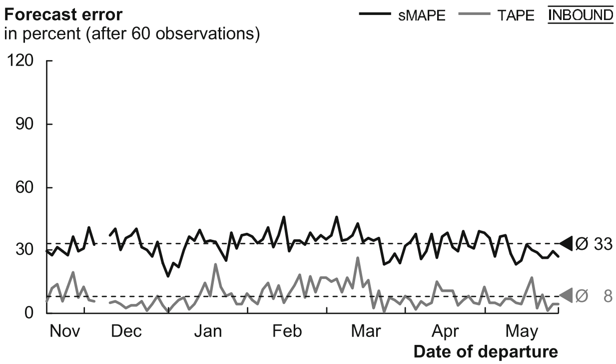


Figure A.2: Full-information forecast errors (sMAPE and TAPE) under noninformative learning – inbound
Source: own design based on collected data

For this book, **Tables A.1 – A.8** can be accessed online via OnlinePLUS on www.gabler.de.

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6,

Bibliography

Adida, E. and Perakis, G. (2004). A Nonlinear Fluid Model of Dynamic Pricing and Inventory Control with no Backorders. Working Paper, Massachusetts Institute of Technology, Cambridge, MA, U.S.A.

Adida, E. and Perakis, G. (2006). A Robust Optimization Approach to Dynamic Pricing and Inventory Control with no Backorders. *Mathematical Programming, Series B*, 107(1):97–129.

Akaike, H. (1973). Information Theory and an Extension of the Maximum Likelihood Principle. In B. N. Petrov and F. Csáki, editor, *Second International Symposium on Information Theory*, pages 267–281. Akadémiai Kiadó, Budapest.

Allenby, G. M. (1989). A Unified Approach to Identifying, Estimating and Testing Demand Structures with Aggregate Scanner Data. *Marketing Science*, 8(3):265–280.

Amman, H. M. and Kendrick, D. A. (1994). Active Learning: Monte Carlo Results. *Journal of Economic Dynamics and Control*, 18(1):119–124.

Amman, H. M. and Kendrick, D. A. (1997). Active Learning: A Correction. *Journal of Economic Dynamics and Control*, 21(10):1613–1614.

Anjos, M. F., Cheng, R. C. H., and Currie, C. S. M. (2004). Maximizing Revenue in the Airline Industry Under One-way Pricing. *Journal of the Operational Research Society*, 55(5):535–541.

Anjos, M. F., Cheng, R. C. H., and Currie, C. S. M. (2005). Optimal Pricing Policies for Perishable Products. *European Journal of Operational Research*, 166(1):246–254.

Anscombe, F. J. and Guttman, I. (1960). Rejection of Outliers. *Technometrics*, 2(2):123–147.

S. Christ, *Operationalizing Dynamic Pricing Models,* DOI 10.1007/978-3-8349-6184-6,

Araman, V. and Caldentey, R. (2005). Dynamic Pricing for Non-Perishable Products With Demand Learning. Working Paper, New York University, New York, NY, U.S.A.

Artin, M. (1991). *Algebra*. Prentice Hall.

Aviv, Y. and Pazgal, A. (2002). Pricing of Short Life-Cycle Products Through Active Learning. Working Paper, Washington University, St. Louis, MO, U.S.A.

Aviv, Y. and Pazgal, A. (2005). A Partially Observed Markov Decision Process for Dynamic Pricing. *Management Science*, 51(9):1400–1416.

Aviv, Y. and Pazgal, A. (2008). Optimal Pricing of Seasonal Products in the Presence of Forward-Looking Consumers. *Manufacturing & Service Operations Management*, 10(3):339–359.

Azoury, K. S. (1985). Bayes Solution to Dynamic Inventory Models Under Unknown Demand Distribution. *Management Science*, 31(9):1150–1160.

Backhaus, K., Erichson, B., Plinke, W., and Weiber, R. (2006). *Multivariate Analysemethoden*. Springer, 11th edition.

Baker, W. L., Lin, E., Marn, M. V., and Zawada, C. C. (2001). Getting Prices Right on the Web. *The McKinsey Quarterly*, (2):54–63.

Balvers, R. J. and Cosimano, T. F. (1990). Actively Learning About Demand and the Dynamics of Price Adjustments. *The Economic Journal*, 100(402):882–898.

Bamberg, G., Baur, F., and Krapp, M. (2009). *Statistik*. Oldenbourg, 15th edition.

Barlow, G. (2004). easyJet: An Airline that Changed our Flying Habits. In Ian Yeoman and Una McMahon-Beattie, editor, *Revenue Management and Pricing: Case Studies and Applications*, chapter 2. Thomson.

Bayes, T. (1763). An Essay towards Solving a Problem in the Doctrine of Chances. *Philosophical Transactions Series I*, 53:370–418. By the late reverend Bayes, communicated by Price, in a letter to John Canton.

Bell, P. C. and Zhang, M. (2006). Management Decision-making in the Single Period Optimum Pricing Problem. *Journal of the Operational Research Society*, 57(4):377–388.

Bellmann, R. E. (1957). *Dynamic Programming*. Princeton University Press.

Ben-Akiva, M. and Lerman, S. (1985). *Discrete Choice Analysis: Theory and Application to Travel Demand*. MIT Press.

Bertsimas, D. and Perakis, G. (2006). Dynamic Pricing: A Learning Approach. In Siriphong Lawphongpanich and Donald W. Hearn and Michael J. Smith, editor, *Mathematical and Regional Models for Congestion Charging*, volume 101 of *Applied Organization*, pages 45–79. Springer.

Besanko, D. and Winston, W. L. (1990). Optimal Price Skimming by a Monopolist Facing Rational Consumers. *Management Science*, 36(5):555–567.

Bhadra, D. (2003). Demand for Air Travel in the United States: Bottom-up Econometric Estimation and Implications for Forecasts by Origin and Destination Pairs. *Journal of Air Transportation*, 8(2):19–56.

Bhat, C. R. and Pulugurta, V. (1998). A Comparison of two Alternative Behavioral Choice Mechanisms for Household Auto Ownership Decisions. *Transportation Research, Part B*, 32(1):61–75.

Bierlaire, M. (1997). Discrete Choice Models. Working Paper, Massachusetts Institute of Technology, Cambridge, MA, U.S.A.

Bierlaire, M. (2003). BIOGEME: A Free Package for the Estimation of Discrete Choice Models. In *Proceedings of the 3rd Swiss Transportation Research Conference, Ascona, Switzerland*.

Bierlaire, M. (2009). Estimation of discrete choice models with BIOGEME Version 1.8. Technical report, URL: http://biogeme.epfl.ch.

Binggeli, U. and Pompeo, L. (2002). Hyped Hopes for Europe's Low-Cost Airlines. *The McKinsey Quarterly*, (4):87–97.

Binggeli, U. and Pompeo, L. (2005). The Battle for Europe's Low-Fare Flyers. *The McKinsey Quarterly*, (3):1–8.

Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer.

Bitran, G. R. and Caldentey, R. (2003). An Overview of Pricing Models for Revenue Management. *Manufacturing & Service Operations Management*, 5(3):203–229.

Bitran, G. R., Caldentey, R., and Mondschein, S. V. (1998). Coordinating Clearance Markdown Sales of Seasonal Products in Retail Chains. *Operations Research*, 46(5):609–624.

Bitran, G. R. and Mondschein, S. V. (1993). Pricing Perishable Products: An Application to the Retail Industry. Working Paper, Massachusetts Institute of Technology, Cambridge, MA, U.S.A.

Bitran, G. R. and Mondschein, S. V. (1997). Periodic Pricing of Seasonal Products in Retailing. *Management Science*, 43(1):64–79.

Bitran, G. R. and Wadhwa, H. K. (1996). A Methodology for Demand Learning with an Application to the Optimal Pricing of Seasonal Products. Working Paper, Massachusetts Institute of Technology, Cambridge, MA, U.S.A.

Blattberg, R. C. and Wisniewski, K. J. (1989). Price-Induced Patterns of Competition. *Marketing Science*, 8(4):291–309.

Bleymüller, J., Gehlert, G., and Gülicher, H. (2004). *Statistik für Wirtschaftswissenschaftler*. Vahlen, 14th edition.

Box, G. E. and Cox, D. R. (1964). An Analysis of Transformations. *Journal of the Royal Statistical Society*, 26(2):211–252.

Box, G. E. and Tiao, G. C. (1973). *Bayesian Inference in Statistical Analysis.* Wiley Classics.

Boyd, E. A. and Bilegan, I. C. (2003). Revenue Management and E-Commerce. *Management Science*, 49(10):1363–1386.

Boyd, E. A. and Kallesen, R. (2004). The Science of Revenue Management When Passengers Purchase the Lowest Available Fare. *Journal of Revenue and Pricing Management*, 3(2):171–177.

Boyd, J. H. and Mellman, R. E. (1980). The Effect of Fuel Economy Standards on the U.S. Automotive Market: An Hedonic Demand Analysis. *Transportation Research, Part A*, 14(5–6):367–378.

Bradley P. Carlin and Thomas A. Louis (2000). *Bayes and Empirical Bayes Methods for Data Analysis.* Chapman & Hall, 2nd edition.

Brynjolfsson, E. and Smith, M. D. (2000). Frictionless Commerce? A Comparison of Internet and Conventional Retailers. *Management Science*, 46(4):563–585.

Burger, B. and Fuchs, M. (2005). Dynamic Pricing – A Future Airline Business Model. *Journal of Revenue and Pricing Management*, 4(1):39–53.

Burnetas, A. N. and Smith, C. E. (2000). Adaptive Ordering and Pricing for Perishable Prioducts. *Operations Research*, 48(3):436–443.

Burnham, K. P. and Anderson, D. R. (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. *Sociological Methods & Research*, 33(2):261–304.

CAA (1998). The Single European Aviation Market: The First Five Years. CAP 685, Civil Aviation Authority.

Cardell, N. S. and Dunbar, F. C. (1980). Measuring the Societal Impacts of Automobile Downsizing. *Transportation Research, Part A*, 14(5–6):423–434.

Caro, F. and Gallien, J. (2007). Dynamic Assortment with Demand Learning for Seasonal Consumer Goods. *Management Science*, 53(2):276–292.

Carson, R. T., Louviere, J. J., Anderson, D. A., Arabie, P., Bunch, D. S., Hensher, D. A., Johnson, R. M., Kuhfeld, W. F., Steinberg, D., Swait, J., Timmermans, H., and Wiley, J. B. (1994). Experimental Analysis of Choice. *Marketing Letters*, 5(4):351–368.

Cascetta, E., Russo, F., Viola, F. A., and Vitetta, A. (2002). A Model of Route Perception in Urban Road Networks. *Transportation Research, Part B*, 36(7):577–592.

Chan, L. M., Shen, Z., Simchi-Levi, D., and Swann, J. L. (2004). Coordination of Pricing and Inventory Decisions: A Survey and Classification. In David Simchi-Levi and S. David Wu and Zuo-Jun Shen, editor, *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*, chapter 9. Kluwer Academic Publishers.

Chatwin, R. E. (2000). Optimal Dynamic Pricing of Perishable Products with Stochastic Demand and a Finite Set of Prices. *European Journal of Operational Research*, 125(1):149–174.

Chen, Y.-M. and Jain, D. C. (1992). Dynamic Monopoly Pricing under a Poisson-Type Uncertain Demand. *The Journal of Business*, 65(4):593–614.

Christopher Chatfield (1989). *The Analysis of Time Series – An Introduction.* Chapman & Hall, 4th edition.

Chun, Y. H. (2003). Optimal Pricing and Ordering policies for Perishable Commodities. *European Journal of Operational Research*, 144(1):68–82.

Clower, R. W. (1959). Some Theory of an Ignorant Monopolist. *The Economic Journal*, 69(276):705–716.

Coldren, G. M., Koppelman, F. S., Kasturirangan, K., and Mukherjee, A. (2003). Modeling Aggregate Air-travel itinerary Shares: Logit Model Development at a Major U.S. Airline. *Air Transport Management*, 9(6):361–369.

Cooper, W. L., de Mello, T. H., and Kleywegt, A. J. (2006). Models of the Spiral-Down Effect in Revenue Management. *Operations Research*, 54(5):968–987.

Cope, E. (2007). Bayesian Strategies for Dynamic Pricing in E-Commerce. *Naval Research Logistics*, 54(3):265–281.

Cortese, A. E. and Stepanek, M. (1998). E-Commerce: Good-bye to Fixed Pricing? *Business Week*, pages 70–84.

Cosslett, S. R. (1981). Efficient Estimation of Discrete Choice Models. In Charles F. Manski and Daniel L. McFadden, editor, *Structural Analysis of Discrete Data with Econometric Applications*, chapter 2. MIT Press.

Coy, P. (2000). The Power of Smart Pricing. *Business Week*.

Cross, R. G. (1997). *Revenue Management: Hardcore Tactics for Market Domination*. Broadway Books.

Currie, C. S. M., Cheng, R. C. H., and Smith, H. K. (2008). Dynamic Pricing of Airline Tickets with Competition. *Journal of the Operational Research Society*, 59(8):1026–1037.

Daganzo, C. (1979). *Multinomial Probit: The Theory and its Application to Demand Forecasting*. Academic Press.

Daly, A. and Zachary, S. (1979). Improved Multiple Choice Models. In David A. Hensher and M. Quasim Dalvi, editor, *Identifying and Measuring the Determinants of Mode Choice*. Teakfield.

Das Varma, G. and Vettas, N. (2001). Optimal Dynamic Pricing with Inventories. *Economics Letters*, 72(3):335–340.

Denison, D. G., Holmes, C. C., Mallick, B. K., and Smith, A. F. (2002). *Bayesian Methods for Nonlinear Classification and Regression*. John Wiley.

Desiraju, R. and Shugan, S. M. (1999). Strategic Service Pricing and Yield Management. *Journal of Marketing*, 63(1):44–56.

Dockner, E. and Jørgensen, S. (1988). Optimal Pricing Strategies for New Products in Dynamic Oligopolies. *Marketing Science*, 7(4):315–334.

Doganis, R. (2002). *Flying Off Course: The Economics of International Airlines*. Routledge.

Dolan, R. J. and Jeuland, A. P. (1981). Experience Curves and Dynamic Demand Models: Implications for optimal Pricing Strategies. *Journal of Marketing*, 45(1):52–62.

Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis*. John Wiley & Sons, Inc., 2nd edition.

Durbin, J. and Watson, G. S. (1950). Testing for Serial Correlation in Least Squares Regression: I. *Biometrika*, 37(3/4):409–428.

Durbin, J. and Watson, G. S. (1951). Testing for Serial Correlation in Least Squares Regression: II. *Biometrika*, 38(1/2):159–177.

Dutta, S., Bergen, M., Levy, D., Ritson, M., and Zbaracki, M. (2002). Pricing as a Strategic Capability. *MIT Sloan Management Review*, 43(3):61–66.

Easley, D. and Kiefer, N. M. (1988). Conterolling a Stochastic Process with Unknown Parameters. *Econometrica*, 56(5):1045–1064.

Efron, B. (1986). Why Isn't Everyone a Bayesian. *The American Statistician*, 40(1):1–11.

Eliashberg, J. and Steinberg, R. (1987). Marketing-Production Decisions in an Industrial Channel of Distribution. *Management Science*, 33(8):981–1000.

Eliashberg, J. and Steinberg, R. (1993). Marketing-Production Joint Decision-Making. In Jehoshua Eliashberg and Gary L. Lilien, editor, *Marketing*, volume 5 of *Handbooks in Operations Research and Management Science*, chapter 18, pages 827–880. Elsevier Science Publishers.

Elmaghraby, W., Gülcü, A., and Keskinocak, P. (2008). Designing Optimal Preannounced Markdowns in the Presence of Rational Customers with Multiunit Demands. *Manufacturing & Service Operations Management*, 10(1):126–148.

Elmaghraby, W. and Keskinocak, P. (2003). Dynamic Pricing in the Presence of Inventory Considerations: Research Overview, Current Practices, and Future Directions. *Management Science*, 49(10):1287–1309.

Erhardt, G. D., Koppelman, F. S., Freedman, J., Davidson, W. A., and Mullins, A. (2003). Modeling the Choice to Use Toll and High-Occupancy Vehicle Facilities. *Transportation Research Record*, 1854:135–143.

Farias, V. F. and van Roy, B. (2007). Dynamic Pricing with a Prior on Market Response. Working Paper, Stanford University, Stanford, CA, U.S.A.

Federgruen, A. and Heching, A. (1999). Combined Pricing and Inventory Control Under Uncertainty. *Operations Research*, 47(3):454–475.

Feng, Y. and Gallego, G. (1995). Optimal Starting Times for End-of-Season Sales and Optimal Stopping Times for Promotional Fares. *Management Science*, 41(8):1371–1391.

Feng, Y. and Gallego, G. (2000). Perishable Asset Revenue Management with Markovian Time Dependent Demand Intensities. *Management Science*, 46(7):941–956.

Feng, Y. and Xiao, B. (1999). Maximizing Revenues of Perishable Assets with a Risk Factor. *Operations Research*, 47(2):337–341.

Feng, Y. and Xiao, B. (2000a). A Continuous-Time Yield Management Model with Multiple Prices and Reversible Price Changes. *Management Science*, 46(5):644–657.

Feng, Y. and Xiao, B. (2000b). Optimal Policies of Yield Management with Multiple Predetermined Prices. *Operations Research*, 48(2):332–343.

Fisher, M. and Raman, A. (1996). Reducing the Cost of Demand Uncertainty through Accurate Response to Early Sales. *Operations Research*, 44(1):87–99.

Fisher, M. L. and Raman, A. (2000). Rocket Science Retailing is Almost Here – Are You Ready? *Harvard Business Review*, pages 115–124.

Fuller, W. A. (1996). *Introduction to Statistical Time Series*. Wiley Series in Probability and Statistics. John Wiley & Sons, 2nd edition.

Gallego, G. and Hu, M. (2006). Dynamic Pricing of Perishable Assets Under Competition. Working Paper, Columbia University, New York, NY, U.S.A.

Gallego, G., Iyengar, G., Phillips, R., and Dubey, A. (2004). Managing Flexible Products on a Network. CORC Technical Report TR-2004-01, Columbia University, New York, NY, U.S.A.

Gallego, G. and van Ryzin, G. (1994). Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons. *Management Science*, 40(8):999–1020.

Gallego, G. and van Ryzin, G. (1997). A Multiproduct Dynamic Pricing Problem and its Applications to Network Yield Management. *Operations Research*, 45(1):24–41.

Gallien, J. (2006). Dynamic Mechanism Design for Online Commerce. *Operations Research*, 54(2):291–310.

Galton, F. (1889). *Natural Inheritance*. Macmillan.

Garson, D. G. (2008). Regression Analysis from Statnotes: Topics in Multivariate Analysis. Online Textbook, North Carolina State University. URL: http://www2.chass.ncsu.edu/garson/pa765/statnote.htm.

Gelman, A. (2008). Objections to Bayesian Statistics. *Bayesian Analysis*, 3(3):445–450.

Gönsch, J., Klein, R., and Steinhardt, C. (2008a). Discrete Choice Modelling (Teil I) – Grundlagen. *WiSt - Wirtschaftswissenschaftliches Studium*, 37(7):356–362.

Gönsch, J., Klein, R., and Steinhardt, C. (2008b). Discrete Choice Modelling (Teil II) – Anwendungsbezogene Aspekte. *WiSt - Wirtschaftswissenschaftliches Studium*, 37(8):412–418.

Gönsch, J., Klein, R., and Steinhardt, C. (2009). Dynamic Pricing – State-of-the-Art. *Zeitschrift für Betriebswirtschaft*, Ergänzungsheft 3/2009:1–40.

Graybill, F. A. (2001). *Introduction to Matrices with Applications in Statistics*. Duxbury Press, 2nd edition.

Harvey, A. C. (1993). *Time Series Models*. Harvester Wheatsheaf, 2nd edition.

Hastie, T., Tibshirani, R., and Friedman, J. H. (2001). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.

Hausman, J. A. and McFadden, D. (1984). Specification Tests for the Multinomial Logit Model. *Econometrica*, 52(5):1219–1240.

Hausman, J. A. and Wise, D. A. (1978). A Conditional Probit Model for Qualitative Choice: Discrete Decisions Recognizing Interdependence and Heterogeneous Preferences. *Econometrica*, 46(2):403–426.

Heching, A., Gallego, G., and van Ryzin, G. (2002). Mark-down Pricing: An Empirical Analysis of Policies and Revenue Potential at One Apparel Retailer. *Journal of Revenue and Pricing Management*, 1(2):139–160.

Hensher, D. A. and Button, K. J. (2000). *Handbook of Transport Modelling*, volume 1. Pergamon.

Hensher, D. A., Smith, N. C., Milthorpe, F. W., and Barnard, P. O. (1992). *Dimensions of Automobile Demand: A Longitudinal Study of Household Automobile Ownership and Use*, volume 22 of *Studies in Regional Science and Urban Economics*. Elsevier.

Heun, C. T. (2001). Dynamic Pricing Boosts Bottom Line. *Information Week*. URL: http://www.informationweek.com/story/showArticle.jhtml-?articleID=6507202.

Hyndman, R. J. (2006). Another Look at Forecasting Accuracy Metrics For Intermittent Demand. *FORESIGHT: International Journal of Applied Forecasting*, (4):43–46.

IATA (2008a). Fact Sheet: Industry Statistics. Technical report, IATA.

IATA (2008b). Financial Forecast – September 2008. Technical report, IATA.

Infare (2009). Infare Solutions – Products. URL: http://www.infare.com/?-section=3 (July 14, 2009).

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

Joppien, M. G. (2006). *Strategisches Airline-Management*. Haupt Verlag.

Kachani, S., Perakis, G., and Simon, C. (2007). Modeling the Transient Nature of Dynamic Pricing with Demand Learning in a Competitive Environment. In Terry L. Friesz, editor, *Network Science, Nonlinear Science and Infrastructure Systems*, volume 102 of *International Series in Operations Research & Management Science*, chapter 11. Springer.

Kalish, S. (1983). Monopolist Pricing with Dynamic Demand and Production Cost. *Marketing Science*, 2(2):135–159.

Kalyanam, K. and Putler, D. S. (1997). Incorporating Demographic Variables in Brand Choice Models: An Indivisible Alternatives Framework. *Marketing Science*, 16(2):166–181.

Kannan, P. K. and Kopalle, P. K. (2001). Dynamic Pricing on the Internet: Importance and Implications for Consumer Behavior. *International Journal of Electronic Commerce*, 5(3):63–83.

Kephart, J. O., Hanson, J. E., and Greenwald, A. R. (2000). Dynamic Pricing by Software Agents. *Computer Networks*, 32(6):731–752.

Kiefer, N. M. and Nyarko, Y. (1989). Optimal Control of an Unknown Linear Process with Learning. *International Economic Review*, 30(3):571–586.

Kincaid, W. M. and Darling, D. A. (1963). An Inventory Pricing Problem. *Journal of Mathematical Analysis and Applications*, 7(2):183–208.

Klein, R. (2005). *Revenue Management: Grundlagen und Methoden der Kapazitätssteuerung*. TU Darmstadt. Professorial Dissertation.

Klein, R. and Steinhardt, C. (2008). *Revenue Management – Grundlagen und Mathematische Methoden*. Springer.

Kleywegt, A. J. (2001). An Optimal Control Problem of Dynamic Pricing. Working Paper, Georgia Institute of Technology, Atlanta, GA, U.S.A.

Klingenberg, C. (2005). The Future of Continental Traffic Program: How Lufthansa is Countering Competition from No-frills Airlines. In Werner Delfmann and Herbert Baum and Stefan Auerbach and Sascha Albers, editor, *Strategic Management in the Aviation Industry*, chapter 7. Ashgate.

Kmenta, J. (1997). *Elements of Econometrics*. University of Michigan Press, 2nd edition.

Knudsen, T. R., Randel, A., and Rugholm, J. (2005). The vanishing middle market. *The McKinsey Quarterly*, (4):6–9.

Kök, A. G. and Fisher, M. L. (2007). Demand Estimation and Assortment Optimization Under Substitution: Methodology and Application. *Operations Research*, 55(6):1001–1021.

Kolmogorov, A. N. (1933). On the Empirical Determination of a Distribution Function (Italian). *Giornale dell'Instituto Italiano degli Attuari*, 4:83–91.

Koppelman, F. S. and Bhat, C. (2006). A Self Instructing Course in Mode Choice Modeling: Multinomial and Nested Logit Models. Technical report, Prepared For U.S. Department of Transportation, Federal Transit Administration.

Koppelman, F. S. and Sethi, V. (2000). Closed-form Discrete-choice Models. In David A. Hensher and Kenneth John Button, editor, *Handbook of Transport Modelling*, volume 1 of *Handbooks in Transport*, chapter 13. Pergamon.

Krishnamurthi, L., Raj, S. P., and Sivakumar, K. (1995). Unique Inter-Brand Effects of Price on Brand Choice. *Journal of Business Research*, 34(1):47–56.

Lancaster, K. J. (1966). A New Approach to Consumer Theory. *The Journal of Political Economy*, 74(2):132–157.

Laplace, P.-S. (1774). Mémoire sur la probabilité des causes par les événements. *Mémoires de l'Académie Royale des sciences présentés par divers savans*, 6:621–656.

Lawton, T. C. (2002). *Cleared for Take-off: Structure and Strategy in the Low Fare Airline Business*. Ashgate Publishing.

Lazear, E. P. (1986). Retail Pricing and Clearance Sales. *American Economic Association*, 76(1):14–32.

Lee, P. M. (1989). *Bayesian Statistics: An Introduction*. Hodder Arnold, 3rd edition.

Lee, T. C. and Hersh, M. (1993). A model for Dynamic Airline Seat Inventory Control with Multiple Seat Bookings. *Transportation Science*, 27(3):252–265.

Lehn, J. and Wegmann, H. (2006). *Einfürung in die Statistik*. Teubner, 5th edition.

Levin, Y., McGill, J. I., and Nediak, M. (2006a). Dynamic Pricing in the Presence of Strategic Consumers and Oligopolistic Competition. Working Paper, Queen's University, Kingston, ON, U.S.A.

Levin, Y., McGill, J. I., and Nediak, M. (2006b). Optimal Dynamic Pricing of Perishable Items by a Monopolist Facing Strategic Consumers. Working Paper, Queen's University, Kingston, ON, U.S.A.

Levin, Y., McGill, J. I., and Nediak, M. (2007). Price Guarantees in Dynamic Pricing and Revenue Management. *Operations Research*, 55(1):75–97.

Levina, T., Levin, Y., McGill, J. I., and Nediak, M. (2006). Dynamic Pricing with Online Learning of General Reservation Price Distribution. In *Proceedings of INCOM 2006: 12th IFAC Symposium on Information Control Problems in Manufacturing (Etienne, France).*

Levina, T., Levin, Y., McGill, J. I., and Nediak, M. (2007). Linear Programming with Online Learning. *Operations Research Letters*, 35(5):612–618.

Lin, K. Y. (2004). A Sequential Dynamic Pricing Model and its Applications. *Naval Research Logistics*, 51(4):501–521.

Lin, K. Y. (2006). Dynamic Pricing with Real-time Demand Learning. *European Journal of Operational Research*, 174(1):522–538.

Lindley, D. V. (1983). Theory and Practice of Bayesian Statistics. *The Statistician*, 32(1/2):1–11. Proceedings of the 1982 I.O.S. Annual Conference on Practical Bayesian Statistics.

Liu, B. and Milner, J. (2006). Multiple-Item Dynamic Pricing Under a Common Pricing Constraint. Working Paper, University of Toronto, Toronto, ON, Canada.

Liu, Q. and van Ryzin, G. (2008a). On the Choice-Based Linear Programming Model for Network Revenue Management. *Manufacturing & Service Operations Management*, 10(2):288–310.

Liu, Q. and van Ryzin, G. (2008b). Strategic Capacity Rationing to Induce Early Purchases. *Management Science*, 54(6):1115–1131.

Lobo, M. S. and Boyd, S. (2003). Pricing and Learning with Uncertain Demand. Working Paper, Stanford University, Stanford, CA, U.S.A.

Luce, R. D. (1959). *Individual Choice Behavior: A Theoretical Analysis.* Wiley.

Makridakis, S. and Hibon, M. (2000). The M3-Competition: Results, Conclusions and Implications. *International Journal of Forecasting*, 16:451–476.

Manning, R. (1979). Market Research by a Monopolist: A Bayesian Analysis. *Economica*, 46(183):301–306.

Manski, C. F. (1977). The Structure of Random Utility Models. *Theory and Decision*, 8(3):229–254.

Manski, C. F. and McFadden, D. L. (1981). Alternative Estimators and Sample Designs for Discrete Choice Analysis. In Charles F. Manski and Daniel L. McFadden, editor, *Structural Analysis of Discrete Data with Econometric Applications*, chapter 1. MIT Press.

Marcus, B. and Anderson, C. K. (2008). Revenue Management for Low-cost Providers. *European Journal of Operational Research*, 188(1):258–272.

Marn, M. V. and Rosiello, R. L. (1992). Managing Price, Gaining Profit. *Harvard Business Review*, pages 84–93.

Mason, K. J. (2001). Marketing Low-Cost Airline Services to Business Travellers. *Journal of Air Transport Management*, 7(2):103–109.

McAfee, R. P. and te Velde, V. (2006). Dynamic Pricing in the Airline Industry. In Terrence Hendershott, editor, *Economics and Information Systems*, volume 1 of *Handbooks in Information Systems*, chapter 11. Elsevier.

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In P. Zarembka, editor, *Frontiers in Econometrics*, pages 105–142. Academic Press.

McFadden, D. (1975). On Independence, Structure, and Simultaneity in Transportation Demand Analysis. Technical Report Working Paper No. 7511, Institute of Transportation and Traffic Engineering, University of California, Berkeley, CA.

McFadden, D. (1978). Modeling the Choice of Residential Location. In A. Karlqvist and L. Lundqvist and F. Snickars and J. Wiebull, editor, *Spatial Interaction Theory and Planning Models*, pages 75–96. North Holland.

McFadden, D. (2000a). Disaggregate Behavioral Travel Demand's RUM Side – A 30-Year Retrospective. Technical Report, University of California, Berkeley, CA, U.S.A.

McFadden, D. (2000b). Economic Choices. Nobel Prize Lecture, University of California, Berkeley, CA, U.S.A.

McFadden, D., Train, K., and Tye, W. B. (1977). An Application of Diagnostic Tests for the Independence from Irrelevant Alternatives Property of Multinomial Logit Model. *Transportation Research*, (637):39–45.

McFadden, D. L. (1981). Econometric Models of Probabilistic Choice. In Charles F. Manski and Daniel L. McFadden, editor, *Structural Analysis of Discrete Data and Econometric Applications*, chapter 5. MIT Press.

McFadden, D. L. (1984). Econometric Analysis of Qualitative Response Models. In Zvi Griliches and Michael D. Intriligator, editor, *Handbook of Econometrics*, volume II of *Handbooks in Economics*, chapter 24. Elsevier.

McGill, J. I. and van Ryzin, G. J. (1999). Revenue Management: Research Overview and Prospects. *Transportation Science*, 33(2):233–256.

Mishra, S., Vinod, B., and Ratliff, R. (2005). New Generation in Demand Forecasting. In *Proceedings of AGIFORS Symposium (Guaruja, Brazil)*.

Monroe, K. B. and Della Bitta, A. J. (1978). Models for Pricing Decisions. *Journal of Marketing Research*, 15(3):413–428.

Mosteller, F. and Tukey, J. W. (1977). *Data Analysis and Regression: A Second Course in Statistics*. Addison-Wesley Reading.

Nagle, T. (1984). Economic Foundations for Pricing. *The Journal of Business*, 57(1):S3–S26.

Nason, S. D. (2007). Forecasting the Future of Airline Revenue Management. *Journal of Revenue and Pricing Management*, 6(1):64–66.

Neter, J., Wasserman, W., and Kutner, M. H. (1983). *Applied Linear Regression Models*. Richard D. Irwin, Inc.

Netessine, S. (2006). Dynamic Pricing of Inventory/Capacity with Frequent Price Changes. *European Journal of Operational Research*, 174(1):553–580.

Netessine, S. and Shumsky, R. A. (2005). Revenue Management Games: Horizontal and Vertical Competition. *Management Science*, 51(5):813–831.

Nguyen, D. (1984). The Monopolistic Firm, Random Demand, and Bayesian Learning. *Operations Research*, 32(5):1038–1051.

NIST/SEMATECH (2008). *e-Handbook of Statistical Methods*. URL: http://-www.itl.nist.gov/div898/handbook/.

OAG (2009). OAG Aviation Solutions – Schedules Data. URL: http://www.oagaviation.com/Solutions/DataProducts/schedulesdata/-schedulesdata.html (July 14, 2009).

O'Connell, J. F. and Williams, G. (2005). Passengers' Perceptions of Low Cost Airlines and Full Service Carriers: A Case Study Involving Ryanair, Aer Lingus, Air Asia and Malaysia Airlines. *Journal of Air Transport Management*, 11(4):259–272.

Pashigian, B. P. (1988). Demand Uncertainty and Sales: A Study of Fashion and Markdown Pricing. *The American Economic Review*, 78(5):936–953.

Pearson, K. (1896). Mathematical Contributions to the Theory of Evolution. III. Regression, Heredity, and Panmixia. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 187:253–318.

Petruzzi, N. C. and Dada, M. (2002). Dynamic Pricing and Inventory Control with Learning. *Naval Research Logistics*, 49(3):303–325.

Phillips, R. L. (2005). *Pricing and revenue optimization*. Stanford University Press.

Picard, R. R. and Cook, R. D. (1984). Cross-validation of Regression Models. *Journal of the American Statistical Association*, 79(387):575–583.

Pilz, J. (1991). *Bayesian Estimation and Experimental Design in Linear Regression Models*. Wiley.

Pole, A., West, M., and Harrison, J. (1994). *Applied Bayesian Forecasting and Time Series Analysis*. Chapman & Hall.

Pompl, W. (2006). *Luftverkehr: Eine ökonomische und politische Einführung*. Springer, 5th edition.

Press, S. J. (1989). *Bayesian Statistics: Principles, Models, and Applications*. Wiley.

Proussaloglou, K. and Koppelman, F. S. (1995). Air Carrier Demand. *Transportation*, 22(4):371–388.

Proussaloglou, K. and Koppelman, F. S. (1999). The Choice of Air Carrier, Flight, and Fare Class. *Journal of Air Transport Management*, 5(4):193–201.

Punt, A. E. and Hilborn, R. (1997). Fisheries Stock Assessment and Decision Analysis: the Bayesian Approach. *Reviews in Fish Biology and Fisheries*, 7(1):35–63.

Punt, A. E. and Hilborn, R. (2001). BAYES-SA - Bayesian Stock Assessment Methods in Fisheries - User's Manual. Computerized Information Series, Food and Organization of the United Nations. URL: http://www.fao.org/docrep/005/Y1958E/y1958e00.htm.

Qazaz, C. S., Williams, C. K. I., and Bishop, C. M. (1997). An upper bound on the Bayesian error bars for generalized linear regression. In Stephen W. Ellacott and John C. Mason and Iain J. Anderson, editor, *Mathematics of Neural Networks: Models, Algorithms and Applications*, pages 295–299. Kluwer.

Rajan, A., Steinberg, R., and Steinberg, R. (1992). Dynamic Pricing and Ordering Decisions by a Monopolist. *Management Science*, 38(2):240–262.

Raju, C. V. L., Narahari, Y., and Kumar, K. R. (2004). Learning Dynamic Prices in Multi-Seller Electronic Retail Markets with Price Sensitive Customers, Stochastic Demands, and Inventory Replenishment. Working Paper, Indian Institute of Science, Bangalore, India. Communicated to IEEE Transactions on Systems, Man, and Cybernetics.

Rao, V. R. (1984). Pricing Research in Marketing: The State of the Art. *The Journal of Business*, 57(1):S39–S60.

Rhoades, D. L. and Waguespack, B. (2005). Strategic Imperatives and the Pursuit of Quality in the US Airline Industry. *Managing Service Quality*, 15(4):344–356.

Riesenbeck, H. and Perrey, J. (2008). *Power Brands: Measuring, Making, and Managing Brand Success*. Wiley.

Rothkopf, M. (2009). *Innovation in Commoditized Service Industries: An Empirical Case Study Analysis in the Passenger Airline Industry* . Lit.

Roussas, G. G. (1973). *A First Course in Mathematical Statistics*. Addison-Wesley.

Sachs, L. (1997). *Angewandte Statistik: Anwendung statistischer Methoden*. Springer, 8th edition.

Schwarz, G. (1979). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464.

Sermons, M. W. and Koppelman, F. S. (1998). Factor Analytic Approach to Incorporating Systematic Taste Variation into Models of Residential Location Choice. *Transportation Research Record*, 1617:194–202.

Small, K. A. (1987). A Discrete Choice Model for Ordered Alternatives. *Econometrica*, 55(2):409–424.

Small, K. A. (2006). Fundamentals of Economic Demand Modeling: Lessons from Travel Demand Analysis. In Kemper E. Lewis and Wei Chen and Linda C. Schmidt, editor, *Decision Making in Engineering Design*, chapter 9. ASME Press.

Smirnov, N. V. (1939). On the Estimation of the Discrepancy Between Empirical Curves of Distribution for two Independent Samples (Russian). *Bulletin of Moscow University*, 2:3–16.

Smith, B. C., Leimkuhler, J. F., and Darrow, R. M. (1992). Yield Management at American Airlines. *Interfaces*, 22(1):8–31.

Smith, S. A. and Achabal, D. D. (1998). Clearance Pricing and Inventory Policies for Retail Chains. *Management Science*, 44(3):285–300.

Spedding, T. A. and Chan, K. K. (2000). Forecasting Demand and Inventory Management Using Bayesian Time Series. *Integrated Manufacturing Systems*, 11(5):331–339.

Stevens, S. S. (1946). On the Theory of Scales of Measurement. *Science*, 103(2684):677–680.

Strotz, R. H. (1957). The Empirical Implications of a Utility Tree. *Econometrica*, 25(2):269–280.

Strotz, R. H. (1959). The Utility Tree – A Correction and Further Appraisal. *Econometrica*, 27(3):482–488.

Su, X. (2007). Intertemporal Pricing with Strategic Customer Behavior. *Management Science*, 53(5):726–741.

Subrahmanyan, S. and Shoemaker, R. (1996). Developing Optimal Pricing and Inventory policies for Retailers Who Face Uncertain Demand. *Journal of Retailing*, 72(1):7–30.

Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press.

Talluri, K. T. and van Ryzin, G. J. (2004). Revenue Management Under a General Discrete Choice Model of Consumer Behavior. *Management Science*, 50(1):15–33.

Talluri, K. T. and van Ryzin, G. J. (2005). *Theory and Practice of Revenue Management*. Springer, Paperback edition.

Timmermans, H., Borgers, A., and van der Waerden, P. (1992). Mother Logit Analysis of Substitution Effects in Consumer Shopping Destination Choice. *Journal of Business Research*, 24(2):177–189.

Train, K. (1978). A Validation Test of a Disaggregate Mode Choice Model. *Transportation Research*, 12:167–174.

Train, K. E. (1998). Recreation Demand Models with Taste Differences over People. *Land Economics*, 74(2):230–239.

Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Train, K. E., Ben-Akiva, M., and Atherton, T. (1989). Consumption Patterns and Self-selecting Tariffs. *The Review of Economics and Statistics*, 71(1):62–73.

van Ryzin, G. and Vulcano, G. (2006). Simulation-based Optimization of Virtual Nesting Controls for Network Revenue Management. Working Paper, Columbia University, New York, NY, U.S.A. to appear in Operations Research.

Varian, H. R. (1980). A Model of Sales. *The American Economic Review*, 70(4):651–659.

Varian, H. R. (1992). *Microeconomic Analysis*. W.W. Norton & Co, 3rd edition.

Veblen, T. B. (1899). *The Theory of the Leisure Class. An Economic Study of Institutions*. The Macmillan Company.

Vovk, V. G. (1990). Aggregating Strategies. In *Proceedings of the Third Annual Workshop on Computational Learning Theory (San Francisco, CA, U.S.A.)*, pages 371–386. Morgan Kaufmann.

Vovsha, P. (1997). Application of Cross-Nested Logit Model to Mode Choice in Tel Aviv, Israel, Metropolitan Area. *Transportation Research Record*, 1607:6–15.

Waddell, P. (1993). Exogenous Workplace Choice in Residential Location Models: Is the Assumption Valid in a Multimodal Metropolis? *Geographical Analysis*, 25(1):65–82.

Walczak, D. and Brumelle, S. (2007). Semi-Markov Information Model for Revenue Management and Dynamic Pricing. *OR Spectrum*, 29(1):61–83.

Warburg, V., Bhat, C., and Adler, T. (2006). Modeling Demographic and Unobserved Heterogeneity in Air Passengers' Sensitivity to Service Attributes in Itinerary Choice. *Transportation Research Record*, 1951:7–16.

Weatherford, L. R. and Bodily, S. E. (1992). A Taxonomy and Research Overview of Perishable-asset Revenue Management: Yield Management, Overbooking, and Pricing. *Operations Research*, 40(5):831–844.

Wen, U.-P. and Chen, Y.-H. (2005). Dynamic Pricing on the Internet. *International Journal of Operations Research*, 2(2):72–80.

Wieland, V. (2000). Learning by Doing and the Value of Optimal Experimentation. *Journal of Economic Dynamics and Control*, 24(4):501–534.

Williams, H. (1977). On the Formation of Travel Demand Models and Economic Evaluation Measures of User Benefit. *Environment and Planning A*, 9(3):285–344.

Wilson, J. G., Anderson, C. K., and Kim, S.-W. (2006). Optimal Booking Limits in the Presence of Strategic Consumer Behavior. *International Transactions in Operational Research*, 13(2):99–110.

Xu, X. and Hopp, W. J. (2004). Dynamic Pricing and Inventory Control: The Value of Demand Learning. Working Paper, Northwestern University, Evanston, IL, U.S.A.

Yai, T., Iwakura, S., and Morichi, S. (1997). Multinomial Probit with Structured Covariance for Route Choice Behavior. *Transportation Research, Part B*, 31(3):195–207.

Zeni, R. H. and Lawrence, K. D. (2004). Unconstraining Demand Data at US Airways. In Ian Yeoman and Una McMahon-Beattie, editor, *Revenue Management and Pricing: Case Studies and Applications*, chapter 11. Thomson.

Zhang, D. and Cooper, W. L. (2005). Revenue Management for Parallel Flights with Customer-Choice Behavior. *Operations Research*, 53(3):415–431.

Zhang, D. and Cooper, W. L. (2009). Pricing Substitutable Flights in Airline Revenue Management. *European Journal of Operational Research*, 197(3):848–861.

Zhang, J.-L. and Chen, J. (2006). Bayesian Solution to Pricing and Inventory Control Under Unknown Demand Distribution. *Operations Research Letters*, 34(5):517–524.

Zhao, W. and Zheng, Y.-S. (2000). Optimal Dynamic Pricing for Perishable Assets with Nonhomogeneous Demand. *Management Science*, 46(3):375–388.