Choice-set Demand in Revenue Management: Unconstraining, Forecasting and Optimization

Haensel, A.

2012

Choice-set Demand in Revenue Management: Unconstraining, Forecasting and Optimization

Haensel, Alwin, 1982 –
Choice-set Demand in Revenue Management: Unconstraining, Forecasting and Optimization

ISBN: 978-90-819095-0-1

Publisher: Alwin Haensel, Amsterdam, the Netherlands.
Cover desinged by Jamie Lipps, New York, USA.
Printed by Ipskamp Drukkers, Enschede, the Netherlands.

VRIJE UNIVERSITEIT

Choice-set Demand in Revenue Management: Unconstraining, Forecasting and Optimization

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. L.M. Bouter,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de faculteit der Exacte Wetenschappen
op woensdag 27 juni 2012 om 13.45 uur
in de aula van de universiteit,
De Boelelaan 1105

door

Alwin Haensel

geboren te Osterburg, Duitsland.

promotor: prof.dr. G.M. Koole

leescommissie:	prof.dr.	M. Fleischmann
	dr.	K. Pak
	prof.dr.	R.D. van der Mei
	prof.dr.	J.I. van der Rest
	prof.dr.	G.J. van Ryzin

Preface

This book concludes three and a half years of research and work, both at the VU University Amsterdam and Bookit B.V. I very much enjoyed this exciting period, which comprises also the move to the Netherlands. Living in Amsterdam, with learning Dutch as well as the cultural habits was a fantastic experience. There are many people who helped and supported me on my path writing this book. I would like to take the opportunity to thank people who helped me, realizing that the list will by no means be complete.

First of all, I would like to mention my promotor Ger Koole and thank him for the freedom in my research and the great opportunity to do the PhD research at the VU University Amsterdam in combination with a part time assignment at the company Bookit B.V. Which leads me to Bookit itself, here I would like to thanks especially Karel Vos, the managing director, for his support, believe in me and most of all his challenging assignments. The opportunity to combine the theoretic research part of the PhD trajectory with the applied work at Bookit felt tailor-made for me, and a setup that I highly recommend to others. Further, I want to thank the reading committee for their interest in my research and their useful comments and suggestions.
I must say a huge thank you to my colleagues at Bookit and the OBP research group at the VU University for the nice and productive working environment. A special thank goes to Sandjai Bhulai, who always found time to discuss research questions and whose great positive attitude is contagious. My thanks goes also to the Business Optimization group at the IBM Research Lab in Zurich, which I had the pleasure to join for two months in spring 2011.
I may never forget to thank my dear friend Jamie Lipps for all his comments and corrections on my English writings over the years. I am fairly certain that the corrections became less by every year, but I am not sure if it's due improvement on my part, or just that my friend stopped trying to fill a bucket full of holes. Finally, I want to thank my lovely wife Sylvia for all her support during the last years, most of all for the necessary distractions and for the perfect mix of motivation and pressure, each given when needed most.

Alwin Haensel
April 2012

Die Mathematiker sind eine Art Franzosen: redet man zu ihnen, so übersetzen sie es in ihre Sprache, und dann ist es alsobald ganz etwas anderes.

- Johann Wolfgang von Goethe (1749 - 1832)

Table of Contents

1 Introduction **1**
1.1 Revenue Management . . . 1
1.2 Microeconomic Reflection . . . 3
1.3 Demand Models in Revenue Management . . . 9
1.4 Abstract of the Research . . . 12
1.5 Outline of the Book . . . 13

2 The Customer Choice-set Demand Model **15**
2.1 Customer Choice-set Model . . . 16
2.2 Demand Functions . . . 19
2.3 A Dynamic Programming Allocation Control . . . 21
2.4 A First Parameter Estimation Approach . . . 23
2.5 Numerical Tests . . . 28
2.6 Conclusion . . . 34

3 A Demand Unconstraining Case Study on Airline Data **37**
3.1 Available Dataset . . . 38
3.2 Customer Choice-set Demand . . . 42
3.3 Estimation Results . . . 46
3.4 Conclusion . . . 50

4 An Advanced Choice-set Demand Estimation Method **51**
4.1 Complete Data Knowledge . . . 51
4.2 Incomplete Data Knowledge . . . 60
4.3 Summary . . . 72
4.4 Algorithm . . . 73

5 Comparison of Discrete Choice Models on a Hotel Market **75**
5.1 Available Dataset . . . 76
5.2 Choice Models . . . 81
5.3 Numerical Results . . . 89
5.4 Conclusion . . . 98

6 A Forecasting with Updating Case Study on Hotel Data 99
6.1 Available Dataset . . . 101
6.2 Forecasting Method . . . 105
6.3 Forecast Updating . . . 112
6.4 Numerical Results . . . 115
6.5 Summary . . . 124

7 Time Dependent Bid Prices Models 125
7.1 Traditional Bid Price Approach . . . 127
7.2 Time Dependent Bid Price Model (TDB) . . . 129
7.3 Stochastic TDB Extensions . . . 131
7.4 Choice-set based TDB Model (CTDB) . . . 136
7.5 Summary . . . 141

8 A Combined Simulation Test 143
8.1 Simulation Setup . . . 143
8.2 Test Cases . . . 148
8.3 Numerical Results . . . 151
8.4 Conclusion . . . 159

9 Summary and Conclusion 161

Bibliography 165

Samenvatting 171

Curriculum Vitae 175

Chapter 1

Introduction

1.1 Revenue Management

Revenue Management (RM) is the art and science to maximize revenue or profit by selling products at the right time and at the best price to the right customer. Traditionally, RM is concerned with selling perishable products, e.g., airline tickets, by controlling price, availability and potential overbooking. The classic application areas, such as airlines, hotels and car rental companies, share four common characteristics: perishable products, time varying demand, relatively low variable costs and the possibility of market segmentation and price differentiation. Nowadays, the ideas, techniques and approaches of RM are also successfully applied to many other industries, which do not necessary share all of the above characteristics. The early RM work dates back to Littlewood (1972). He studied the airline problem on a single flight with two price classes and proposed the idea of focusing on profit maximization. The paper describes a basic passenger forecasting method and a revenue control, which is nowadays referred to as Littlewood's Rule. Low fare customers are assumed to arrive before high fare customers. Since the airplane has a finite capacity, the airline's revenue maximization problem is to determine the right passenger mix. In other words, the goal is to restrict the number of low fare customers to be accepted. Littlewood proposed a simple rule, stating that the airline should accept low fare customers as long as the lower fare exceeds the expected revenue from selling another seat at the higher fare. Belobaba (1989) extended this approach to multiple price classes with independent demand, called Expected Marginal Seat Revenue (EMSR) method.

Mostly when people talk about revenue management and certainly in scientific Operations Research papers, it is more about the technical side and problem specific, but not on the actual idea and attitude behind it. This book, as it claims a scientific standard and contribution, will be no exception and also focuses on technical aspects. Namely, we are proposing a practical approach to incorporate the customers' choice behavior in the demand modeling and further in the whole RM system. But behind revenue management stands an idea contrary to cost cutting, downsizing, concentration on the core business, etc. It is the idea of growth, of going into competition and fighting for every customer in order to increase revenue and profit. The core idea is to increase revenue from existing customers while simultaneously acquiring new ones. So the focus is always on growth and going into competition rather than reducing costs in order to boost the bottom line. A popular introduction to the ideas and basic approaches in revenue management is found in Cross (1997). He also illustrates the idea of 'what looks like the same product is not always the same product and therefore you can price it differently' on an example with airline tickets. Airlines charge different prices for technically the same seat on an airplane. So one might conclude that different prices are charged for the same product. But actually, it is not the same product if you buy the ticket 3 month before departure or on the day before departure or even at the same day. When reserving seats for customers requesting a ticket very close to departure, the airline takes the risk of not selling the seat if no customer arrives at a later time in the booking horizon. Therefore, the price of the ticket and also the willingness to pay of customers is generally increasing towards the departure of the airplane. This basic observation on the meaning of prices was already made by Silberston (1970), who explains the diversification of prices in the car market in essence by the variety of different products. A car product is more then the car type itself, it differs by configuration, payment form, etc. Since the 1980's, RM topics have been extensively investigated by the airline industry due to the deregulation of the US market in 1978. Over the last two decades there was also a huge growth in literature focusing on RM problems in other sectors such as hotels, car rental, cruise lines, advertising and many more. A popular track record of RM is found in Cross (1997). It is about the "battle" between American Airlines and People Express in 1985-1987. People Express was founded in 1981 and revolutioned the airline business by offering simple plain flight tickets with no extra services but at prices 50-70% less than the major airlines. By 1983, People Express became a serious competitor to the major airlines. In 1985, American Airlines introduced discount tickets, called Super Savers, directly targeting routes also served by People Express. American's newly formed yield management department was the key to success. It controlled the pricing and capacity availability in order to effectively compete on prices with airlines such as People Express and simultaneously preserving the full fare traffic when possible. Full and uncontrolled competition on price with the low-cost carrier would have killed American, since their fixed costs were considerably higher than the ones of People Express. This was the birth of customer segmentation and having variable prices on the same airplane. Two years later, People Express was forced out of business. Geraghty and Johnson (1997) report another success story, they describe the RM system implemented in 1993 and 1994 at National Car Rental. In 1993, National faced liquidation and their only hope was to generate substantial profit in the short term or 7500 jobs would have been lost. A RM program was introduced, focusing on pricing and capacity control at rental station level. The profit was immediately increased and in 1995 General Motors Corporation, as the parent company, sold National Car Rental for approximately 1.2 billion Dollar. Both applications opened the door for RM to be nowadays applied in many diverse companies, creating a whole industry of supporting software and consulting firms specialized on RM solutions. Also the academic research and education in the field increased with many courses on RM offered to Master or PhD students. There are also two research journal entirely devoted to RM, namely the Journal of Revenue and Pricing Management, first issue in 2002, and the International Journal of Revenue Management, first issue in 2007. Currently, RM is one of the fastest developing research areas in Operations Research.

We like to refer to two books for a broad overview of topics and a general introduction to revenue management: First, an extensive overview of RM aspects covering optimization, dynamic pricing, basic forecasting and choice models, industry profiles and implementation issues of RM systems is found in, what can be almost called the bible of RM, the book of Talluri and van Ryzin (2004b). Second, the PhD thesis of Pak (2005), who concentrates on an overview of RM techniques with a focus on airline, hotel and cargo. There are many good review papers, but we only like to mention here two of them. Starting with Chiang et al. (2007), who review 221 papers and five book and provide with it a very comprehensive overview of revenue management developments in research. Moreover, Bobb and Veral (2008) investigate the different components of a RM system and focus on the gaps between practice and research.

1.2 Microeconomic Reflection

Let us now access the idea of revenue management from a microeconomic perspective. This section follows the popular and standard textbook on microeconomics by Pindyck and Rubinfeld (2001). Generally speaking, microeconomics deals with economic units, individual buyers or sellers, and tries to explain the decisions made by these units as well as the trade-offs consumers and producers are facing. We initially concentrate on a perfect competitive market, which consists of many buyers and sellers and no single unit having a significant influence on the price and entering or exiting the market is fairly easy. The best example for perfect competition is the market of agriculture goods, as no single farmer on the supplier site nor a single buyer has a significant impact on the price. In such a market, supply and demand will come into equilibrium. In other words, there exists a general market price which determines the total quantity produced and ensures that the market is cleared, i.e., all produced quantity is sold at the market price. The relationship between producers' willingness to produce and the market price is illustrated by the supply curve, as shown in Figure 1.1 (a). The curve is increasing, because the higher the market price becomes, the more attractive it becomes for new producers to enter the market and producers already in the market are motivated to increase production. The complementary relationship between

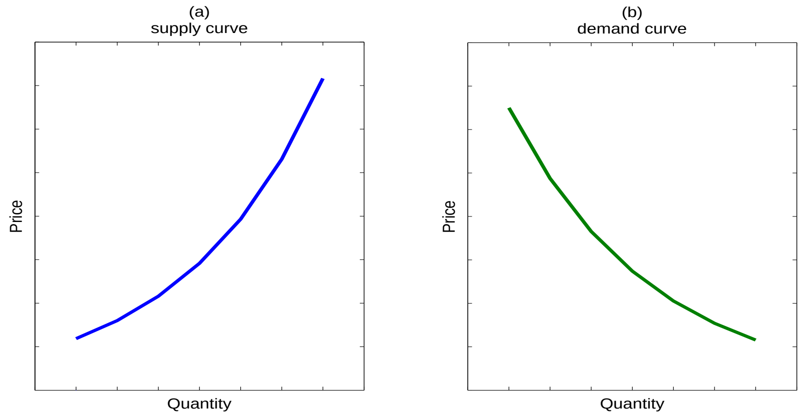

Figure 1.1. Example of supply and demand curves.

buyers and the quantity demanded related to the market price is given by the demand curve, shown in Figure 1.1 (b). In contrast, the demand curve is decreasing, because only few consumers are willing and are able to buy at very high prices, when the price lowers more consumers can afford to enter the market and buy the product. Free markets have the tendency for the price to change until the market clears, i.e., supply equals demand. This state is called equilibrium. Figure 1.2 illustrates the equilibrium situation. Suppose P_1 would be the market price. So producers would be willing to produce Q_1^S units, but only Q_1^D quantities are demanded due to the relatively high price. Hence, a surplus in produced products develops which will consequently lead

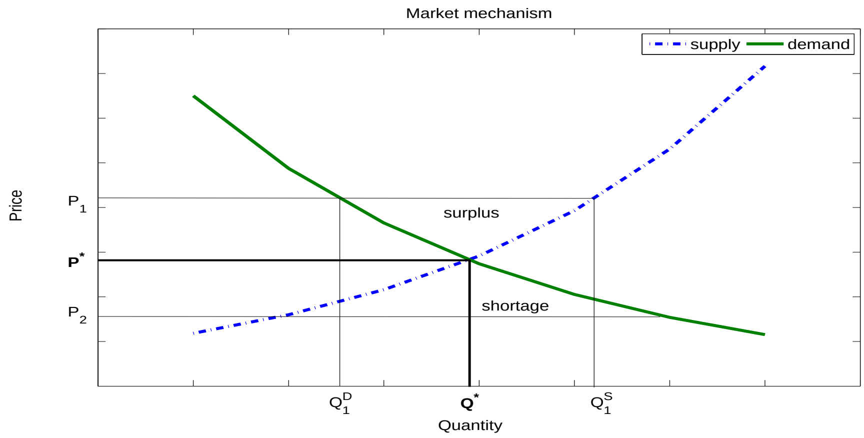

Figure 1.2. Market mechanism.

to price reductions. On the other hand, suppose P_2 is the current market price. In this situation, the quantity produced is much smaller than the quantity requested in the market at this price. Therefore, there is a shortage in the market, which results in a price rise. At a higher price, suppliers are willing to produce more and simultaneously the quantity requested will reduce. An equilibrium state will be a market price of P^*, where the quantity demanded and supplied equals at Q^*. Very interesting is of course the shape of the demand and supply functions. How do consumers or suppliers react to changes in the price? Are they price sensitive and react to relatively small changes or not? This can be answered by computing the elasticity of the demand or supply curve. The elasticity is defined as the % change of the variable of interest resulting from a 1% increase in the other, i.e., “elastic” means that a small change in price results in a large change in quantity. Figure 1.3 illustrates both extremes of the demand curve. Figure 1.3 (a) shows an infinitely elastic curve, where consumers buy as much as possible at price P, but stop if the price increases even a little. The opposite, a completely inelastic curve, is shown in Figure 1.3 (b). Here, consumers buy a fixed quantity with the price having zero influence. So far, we assumed that consumers will react to a price of a certain product with a yes or no decision, independent of other influences. This is a rather superficial assumption. What is in reality found is that consumers trade off between different products and their prices. Here, we distinguish between two groups: substitutes and complements. Two products are substitutes when an increase of the price of the one results in an increase of the quantity demanded of the other. An example of substitutes would be

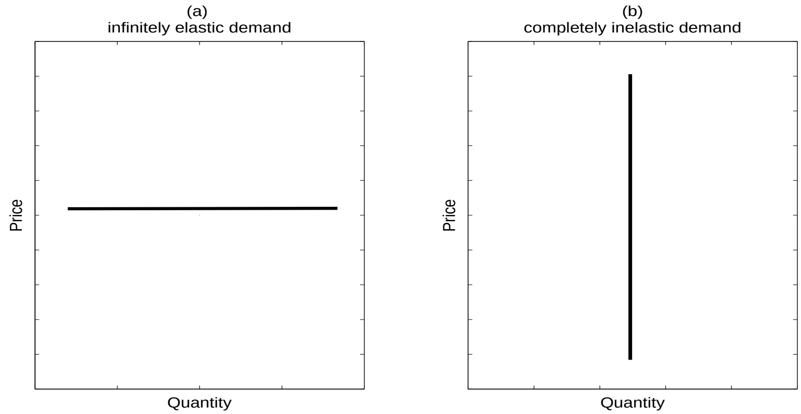

Figure 1.3. Extreme demand elasticity curves.

apples and oranges, as consumers can be assumed to substitute the one for the other. Complements are products, where a price rise for the one leads to a reduction in quantity demand for the other. An example would be electricity and air conditioners. Price elasticities can now also be computed in terms of price changes for different products; they are then called cross price elasticities. Consumers are supposed to make decisions in order to maximize their utility, which is a numerical measure of the satisfaction a consumer receives from buying a bundle of products. Producers on the other hand, are assumed to be profit maximizing. The profit maximizing point is shown in Figure 1.4.

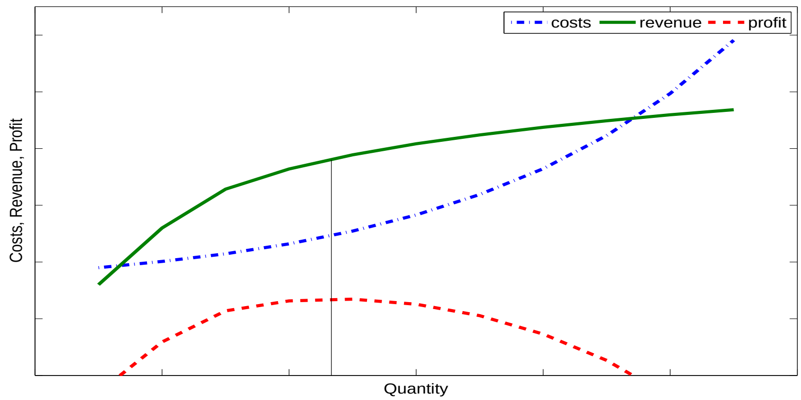

Figure 1.4. Profit maximization.

The optimal output is obtained when marginal revenue, change in revenue from one unit increase in output, equals marginal costs, similarly defined as the change in costs from one unit increase in output. There may be situations in which companies make reasonable decisions which are not directly focused on immediate profit targets. But generally and in the long run companies must focus on profitability to survive in a free market economy. This leads us to the definition of surpluses highlighted in Figure 1.5. The consumer surplus

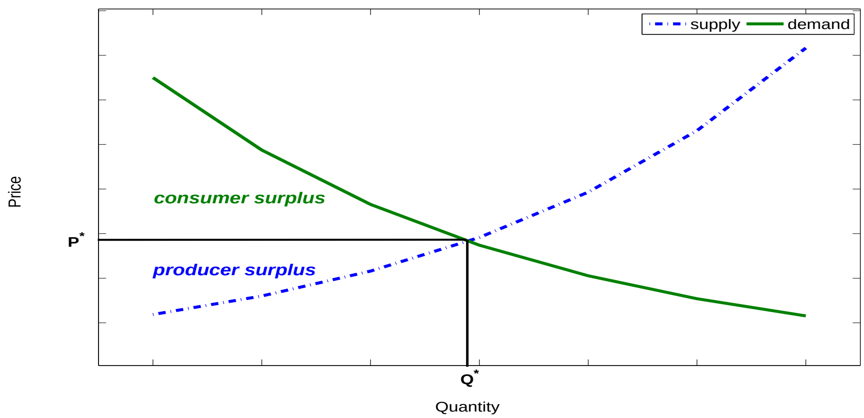

Figure 1.5. Consumer and producer surpluses.

is the difference between their highest acceptable price, i.e., maximum willingness to pay, and the actual paid price. The producer surplus on the other hand, is the difference between the market price and variable costs of production, whereas profit is defined as the difference between revenue and total costs.

Until now we assumed a perfect competitive market, which implied three assumptions on the producer side. First “price taking”: firms have no impact on market price and take it as given. Second “product homogeneity”: firms produce identical products. Third “free entry and exit”: firms can easily enter or exit an industry. The opposite to a perfect competitive market is a monopoly market, a single seller and many buyers. Normally, the production output will be lower and the resulting market price in a monopoly will be higher than in a competitive market and this is the reason why there are laws against monopolists. Still, in a monopoly the seller needs to determine the price to ask and the respective production output. Producing more units at a given price means extra revenue, but a production increase also means a lowering of the selling price of all products. In a competitive market, the production amount is chosen such that the marginal costs are equal to the price. The monopolist will choose a price higher than his marginal costs to increase his surplus, but it depends also inversely on the demand elasticity. Which means if the market demand is very elastic, the monopolist behaves as in competition. Most markets fall between both extremes. Usually, there are players in the market with market power, i.e., they have the ability to affect the market price. Or products are not complete substitutes, which allows firms to ask prices higher than the competitor without losing all its sales. So single firms often face a demand curve different from the market demand curve and contrary to a firm in a perfect competitive market, thus a price increase will not result in a loss of all customers nor will a price decrease result in the capturing of the complete market. Such companies are considered to have monopoly power. As Pindyck and Rubinfeld (2001) phrase it in the introduction of Chapter 11, pricing with market power, managers of firms with market power have a harder job than managers of firms in a perfect competitive market. In a perfect competitive market, prices are given and managers only have to focus on the cost side. But firms with market power must also think about competitor influences and the company's demand curve. In many cases, such companies can do much better by using more advanced pricing strategies than simply asking a single price. The idea is to turn consumer surplus into producer surplus by charging different prices to different customers. This requires very detailed information on demand and the customers' buying behavior. This process of charging different prices for technically the same product is called price discrimination. The essential problem is to identify different customers and to make them pay the different prices. The theory distinguishes three types of price discrimination. First degree price discrimination refers to the practice of charging each individual customers his or her maximum willing price to pay. This trick is almost impossible, since it is not in the consumer's interest to share his or her upper willingness to pay and it is therefore unknown to the seller. Second degree price discrimination corresponds to the practice of asking different unit prices depending on how much a customer demands, e.g., lower unit prices if one consumes more units in total. Third degree price discrimination tries to divide customers into different segments or groups and to ask different prices per group.

The third degree price discrimination refers to core revenue management. The problem is to identify or create groups. An example for identifying groups is the classical Saturday night stay in the airline business. A customer willing to stay on Saturday night away from his origin is very likely to be not on a business trip and therefore considered to be on a leisure trip. The creation of groups is possible by barriers such as checking IDs for age restrictions or University cards for student discounts. Some segments will be charged with a price lower then the optimal price in a single price setting, e.g., P^* in Figure 1.5. This means that price discrimination allows consumers to enter the market and buy products, who would have been excluded otherwise because of their relatively low upper willingness to pay. An example is the airline industry with discount fares for early bookers or opera halls and concert houses with student discounts. So revenue management practice can increase overall welfare such that producers and consumers are better off than in a single price environment.

The key challenge in revenue management is to get a correct understanding of the consumer demand in order to answer the following four questions:

- How big are the demand groups?
- What is the price elasticity of the demand groups?
- What are the substitution behaviors?
- What are the competitor influences?

With this input, the company is able to derive optimal prices and sales quantities per customer group in order to maximize the overall revenue/profit.

Also note that RM normally focuses on the short term, so output capacities are predefined and fixed. The question is which prices to ask or which products or price classes to offer to potential customers in order to maximize the revenue or profit.

1.3 Demand Models in Revenue Management

The focus in RM lies mostly on the optimization model, which is often based on numerous assumptions. The demand modeling and forecasting aspect, which provides the essential input for the optimization model, is considered as given and often neglected. Many RM setups work with the "independent demand model", i.e., assuming independent demand for different products or price classes. A detailed statistical analysis on different airline reservation datasets is found in Belobaba (1985). He focuses on identifying distributional patterns and finds a relationship between reservation distributions for flights and fare classes on certain weekdays and historical demand levels for those days. Lee

(1990) investigates in his PhD thesis probabilistic demand models in order to forecast airline reservations. The independent demand model is very convenient, because it assumes product demand to be equivalent with product sales and hence straightforward to forecast by simply applying some time series models to historic sales data. It is essential to account for non-observed demand, which did not turn into transactions because it was either turned down or the company's offer did not match the consumer's interests. Usually companies store only sales data at detailed level. Demand information not resulting in transactions is not available. Mostly there is no such data collection because it is either impossible, e.g., physical self-serve stores, or the data is simply not useful, e.g., consumers inquire airline's fare class availability multiple times. Ignoring this non-observed demand leads to an underestimation and results further in a spiral down effect, as investigated by Cooper et al. (2006). Therefore, stock out situations or product non-availabilities in the sales data must be so called unconstrained, to estimate the demand quantity one would have observed if the product had been available. Some research papers are devoted to the topic of unconstraining such censored data, see McGill (1995), Lui et al. (2002) and Weatherford and Pölt (2002). A comparison of all commonly known capacity unconstraining models is given in Queenan et al. (2007). The troubling part of the independent demand assumption is that it excludes any substitution effects by consumers. So, it is only valid in settings with very high and strict fences in the customer segmentation, e.g., student discount and regular price with the control of the student ID, or when products are very diverse such that a substitution effect is not existent, e.g., a car dealer with only compact cars and luxury cars but nothing in-between. But in most situations the substitution effect can not be neglected. For example, if the 8am flight costs EUR 800 and the 10am flight only EUR 200, many people may consider to book the later flight even though the earlier one is preferred. So, accounting for substitution effects is very important in order to get a clear understanding of the underlying demand and the customers' choice decisions. As van Ryzin (2005) formulates it: What is needed in revenue management research is a change from product demand models to models of customer behavior. The choice behavior of consumers was for a long time not considered in RM. In fact, a short self reflection shows that the availability of discount tickets in the airline case or the room prices of competing hotels in the same location have a huge influence in our buying decision. In recent years, the choice aspect has been an active field of research within the RM community. This starts with Andersson (1998) and Algers and Beser (2001), who are describing a pilot study at Scandinavian Airlines. They formulate a model for optimal seat allocation which takes into account that customers might buy up to higher fare classes or be recaptured at a different flight. The customer choice is modeled by the multinomial logit (MNL) model and its parameters are estimated from interviews and historical sales data. Talluri and van Ryzin (2004a) study customer behavior under a discrete choice model and state an optimal policy for the fare class availability control in a single leg setting. They suggest using maximum likelihood estimation for the attributes weight vector in an MNL to compute sale probabilities. The customer arrival rate is assumed to be constant in time. Ratliff et al. (2008) propose a multi-flight heuristic, generalizing the earlier approach. In their model they need, besides historical sales and fare class availability data, also information about the airline's market share and values for the relative customer attractiveness of flights and fare class alternatives. An empirical study on airline transaction data is found in Vulcano et al. (2010). They report a significant improvement by using choice behavior estimates in the booking control. Vulcano et al. (2011) focus on estimating the primary demand, i.e., customers' first choice demand if all alternatives are available, from historical sales and availability data. Demand is modeled by a Poisson process over multiple time periods and customers are assumed to choose among alternatives according to an MNL model. All of these articles, and to our best knowledge almost all other RM papers with choice considerations, work with the same choice model, namely the MNL model. The multinomial logit model has very appealing properties: parameters can be efficiently computed by maximum likelihood estimation, it has a clear and simple structure and choice probabilities of different alternatives can be easily computed. Its main shortcoming is the property of independence of irrelevant alternatives (IIA), which can result in abnormal choice probabilities when introducing or removing alternatives from the customer's consideration set. A small illustrative example of IIA: consider three hotels; a conference hotel with choice probability 0.6, a business hotel with choice probability 0.2 and a wellness hotel with probability 0.2 to be chosen by an arriving customer. The conference hotel is full and does not accept any more requests. The new choice probabilities under MNL will be 0.5 for both the business and the wellness hotel, because the MNL assumes that the ratio between choice probabilities of two alternatives is constant and independent of third alternatives. Whereas the substitution of business hotel instead of the non-available conference hotel will be significantly higher in reality than the substitution by the wellness hotel. Also, using a parametric model such as the MNL requires one to make strong assumptions on the choice process and there is the risk of over fitting when using more complex parametric choice models such as the latent class logit or mixed logit. Recently, Farias et al. (2012) and van Ryzin and Vulcano (2011) concentrate on a choice-model quite similar to our proposed choice-set model. Both use a non-parametric choice model originally introduced by Mahajan and van Ryzin (2001), which assumes preference lists or sets of alternatives with a strict rank ordering. Customers choose the available product with the highest preference, or make a non-purchase decision if none of the products of interest is available. Farias et al. (2012) concentrate on a robust approach finding the best demand distribution over preference lists, which best approximates the obtained revenue in the observed dataset. Closer to our work is van Ryzin and Vulcano (2011), who focus on the estimation of the distribution over preference lists representing the customer types. Additionally, they propose a market discovery mechanism to add relevant new customer types to an initial set of preferences lists. They assume a Bernoulli arrival process with a fixed rate, i.e., they assume at most one customer arrival per time period and the intensity over all periods is constant.

1.4 Abstract of the Research

In this thesis, we develop a procedure to estimate demand for different customer types, represented by what we call choice-sets. These choice-sets are sets of products or choice alternatives with a strict decreasing preference order and are very similar to the preference lists mentioned before. The estimation algorithm is developed for available transaction data as stored at most companies, i.e., sales data and product availability data at some aggregated level. So, we do not assume time periods with at most one customer arrival nor information on customers who do not result in a transaction. It is possible to extend the approach to incorporate competitor offers to account for their influences on sales. We also allow for more general demand rates, namely we are assuming an inhomogeneous Poisson process with an exponential rate function, motivated by statistical analysis on real sales data. The demand rate functions per choice-sets are used to estimate unobservable demand in periods with no overlap of choice-sets, representing the products of interest, and the set of offered products. The unconstraining method is tested on actual airline and hotel data and shows very promising results. We further develop a dynamic forecast updating procedure, which considers the correlations within the booking horizon as well as between successive horizons. The method is tested on hotel reservation data and shows a significant improvement in forecast accuracy. Finally, we propose a new optimization model for network revenue management problems, to compute time dependent bid prices. The model is extended to consider choice-set demand in order to include information on customer choices under offered alternatives. This book provides in essence a complete and fluent approach for incorporating the customers' choice behavior into a revenue management system.

1.5 Outline of the Book

The content of the book can be grouped into three main parts. The first and also major part concentrates on the choice-set model and the unconstraining problem, i.e., extraction of choice and demand information from given sales and offer data. The second part focuses on booking horizon forecasting with dynamic updating. And lastly, the third part proposes optimization models to compute optimal sales controls on the basis of forecasted demand with choice information, with the objective to maximize the overall revenue or profit.

Chapter 2 will introduce the choice-set demand model and propose a first unconstraining algorithm. Chapter 3 presents an unconstraining case study on real airline reservation data, followed by an advanced choice-set demand estimation algorithm developed in Chapter 4. The demand unconstraining part ends with Chapter 5 and a comparison study of different choice models on real hotel market data. The forecasting part is covered by Chapter 6, which presents a dynamic forecast updating method for booking horizon forecasting. The optimization part is comprised in Chapter 7, introducing a new method to compute time dependent bid prices in network revenue management problems. The general model is extended to work with choice-set demand in order to incorporate the customer's choice behavior under offered alternatives. Finally, Chapter 8 provides a small simulation study combining all three aspects of a RM process, namely unconstraining, forecasting and optimization with choice-set demand. We summarize the thesis in Chapter 9 and conclude with our findings.

Chapter 2

The Customer Choice-set Demand Model

This chapter is based on the paper Haensel and Koole (2011b).

The following sections introduce the idea of choice-sets to model the customer's buying behavior and choice decisions under offered alternatives. We are interested in establishing demand estimates for customer groups, reflecting different choice behaviors, based on historical sales data as input. Contrary to most of the previous research studies, see the introduction to demand models in RM in Chapter 1, we are not assuming a general customer arrival rate with sale probabilities resulting from offered alternatives. Rather, we assume the demand to be made up of different customer groups, representing different buying behaviors and preferences. Demand from each group can result in sales of different products, depending on the seller's and competitor's offer of alternatives. These groups or customer types are called customer choice-sets. We are presenting a demand estimation method for these choice-sets. The procedure is based on the maximum likelihood method and to over-come the problem of incomplete data or information, we additionally apply the Expectation Maximization (EM) algorithm. Using this demand information per choice-set, the revenue manager obtains a clear view of the underlying demand. In doing so, the sales consequences from different pricing and booking control actions can be compared and optimized in order to maximize the overall revenue.

The chapter continues with the introduction of our customer choice-set model, followed by some analysis on demand behavior in Section 2.2. In Section 2.3, an optimal allotment control based on dynamic programming is described. The parameter estimation method for the choice-set demand is explained in Section 2.4. The numerical results are given in Section 2.5 and the final Section 2.6 presents our conclusions.

2.1 Customer Choice-set Model

In this section we will explain the idea of a discrete choice-set, as earlier described in Ben-Akiva and Lerman (1985). The previous concept, where probabilities are attached to choice alternatives, is now changed into a preference ordering of choice alternatives. The customer's choice behavior is supposed to be represented by the notion of *choice-sets*, which we define as sets of substitutable products or choice alternatives with a strict preference order. We study the choice behavior on products, which are offered in a variety of classes or subclasses. A best example is to consider a seat on an airplane or a hotel room. This base product is offered to customers at different prices, denoted as price classes, and often combined with different conditions on the product itself, such as minimum length of stay, cancellation costs or membership credits. The conditions are introduced to create fences between customer segments, in order to enable us to ask different prices for technically the same product. Such segmentations are in practice rarely completely strict and exclusive. Therefore, substitution effects are generally existent, which means customers are considering and comparing different product offers and choose according their personal preferences and needs. Let us illustrate the choice-set concept on a small example, where we consider an airline which offers two fare classes A and B. Fare A is the discount ticket consisting of the seat and no extra services and fare class B is the full fare ticket, potentially including extras such as a meal to be served during the flight. The possible choice-sets are: $\{A\}$, $\{B\}$, $\{A, B\}$ and $\{B, A\}$; C will denote the set of all choice-sets. Choice-sets are written with a decreasing preference order from left to right. Therefore, the choice-set $\{A, B\}$ states that customers being represented by this choice-set are strictly preferring ticket A over ticket B. In contrast, customers with choice-set $\{B, A\}$ prefer B over A. See Table 2.1 for some example choice-set demand rates. The airline can control the booking availability of

choice-set c	$\{A\}$	$\{B\}$	$\{A, B\}$	$\{B, A\}$
Expected demand D_c	20	7	15	10

Table 2.1. Choice-set example.

all fare classes. O denotes the set of available products/open fare classes. The demand $D(f|O)$ of product/fare class f under offer set O is defined by

$$D(f|O) = \sum_{c \in C} D_c \cdot \mathbb{I}_{U(c,O)=f}, \tag{2.1}$$

with $\mathbb{I}$ denoting the indicator function and $U(c, O)$ returns the fare class contained in choice-set c with the highest preference or utility under the set of offered alternatives O, or zero if $c \cap O = \emptyset$. The amount of rejected customers, i.e., customers whose choice-set c is non-overlapping with O and therefore turned down, can be calculated by

$$D(0|O) = \sum_{c \in C} D_c \cdot \mathbb{I}_{U(c,O)=0}. \tag{2.2}$$

Hence, the sales probability of product $f \in O$ is given by

$$P(f|O) = \frac{D(f|O)}{D(0|O) + \sum_{h \in O} D(h|O)}. \tag{2.3}$$

The non-purchase probability is equivalently computed by

$$P(0|O) = \frac{D(0|O)}{D(0|O) + \sum_{h \in O} D(h|O)}. \tag{2.4}$$

By definition, we set $P(x|O) = 0$ if $x \notin O$. Returning to our small airline example, the probability of selling a certain ticket to an arriving customer for different O is straightforward computable

$$\begin{aligned}
P(A|O = \{A, B\}) &= \frac{20 + 15}{0 + 20 + 15 + 7 + 10} = 0.67, && (2.5) \\
P(B|O = \{A, B\}) &= \frac{7 + 10}{0 + 20 + 15 + 7 + 10} = 0.33, && (2.6) \\
P(A|O = \{A\}) &= \frac{20 + 15 + 10}{7 + 20 + 15 + 10} = 0.87, && (2.7) \\
P(B|O = \{B\}) &= \frac{7 + 15 + 10}{20 + 7 + 15 + 10} = 0.62. && (2.8)
\end{aligned}$$

A more general example can be given by assuming the market to be segmented in such a way that different conditions are attached to the product itself, for example cancellation possibilities, etc. See Table 2.2 for an example airline portfolio with three groups. Each of these groups is again differentiated into different price categories. Again, we assume that each customer has a set of classes which represent his willingness to buy, regarding price and conditions. These choice-set may consist of any coherent sequence of subclasses, e.g.,$\{L2, L1, M2\}$ or $\{M2, M1, H2\}$. The choice-sets have again a strict preference order from left to right, i.e., customers whose choice behavior can be represented by choice-set $\{M2, M1, H2\}$ want a possibility to change the ticket and their willingness to pay is larger or equal to 460 but strictly less to 520,

Booking Class	Miles earned	Changes	Cancellations	Price
H1	100 %	charge 50	charge 100	520
H2	100 %	charge 50	charge 100	460
M1	50 %	charge 50	No	370
M2	50 %	charge 50	No	320
L1	25 %	No	No	250
L2	25 %	No	No	230

Table 2.2. Choice-set example 2, with groups: high, medium and low.

since they are not willing to buy the H1 class. Of course the number of possible choice-sets can be very high, but with some reflection of marketing and sales ideas we can usually restrict ourselves to choice-sets which are coherent and are of limited length.
So far we presented the idea of choice-sets in the context of a single resource and a monopoly seller market. But it can be easily extended to multiple resources, e.g. multiple daily flights on one itinerary, as well as competitor offers can be easily embedded in the choice-sets structure. To illustrate the incorporation of competitor prices, let us return to our first example with two fare classes, A and B. Say, we have a competitor serving the same route and overing also two classes a and b, equivalently denoting the discount and full fare ticket. Our previous choice-set $\{A, B\}$ is now divided into three corresponding choice-sets which account for the different competitor offers

$$\{A, B\} \quad \Rightarrow \quad \left\{ \begin{array}{c} \{A[a, b, \emptyset], B[a, b, \emptyset]\} \\ \{A[a, b, \emptyset], B[b, \emptyset]\} \\ \{A[b, \emptyset], B[\emptyset]\} \end{array} \right\},$$

with $[\cdot]$ we denote the competitor's active classes under which we can observe a sale, $\emptyset$ represents the case that the competitor does not offer any class. For example, $\{A[b, \emptyset], B[\emptyset]\}$ means that customers represented by this choice-set prefer A over B for our company, but only buy class A if the competitor does not offer class a and they will only buy class B if we do not offer A and the competitor does not offer any class.

Consequently, a choice-set may consist of any sequence of substitutable products or subclasses from different comparable resources and sellers. We only need to have a strict preference ordering, so that the customer is never indecisive between two offers.

2.2 Demand Functions

For our analysis we were able to work with real sales data of a major airline, a low-cost airline and a European hotel reservation agency. Both problems are very similar, we have a perishable good, e.g., seats on a certain flight or hotel rooms for a certain date. For both datasets we observed that the demand is increasing as we are approaching the usage date of our product. Usually, we can not observe the true demand. The company may sell all its capacity or influence the booking process, when making certain price offers available or unavailable for booking. We say a price class is open when it is available for bookings and closed otherwise. Figures 2.1 shows the airline sales

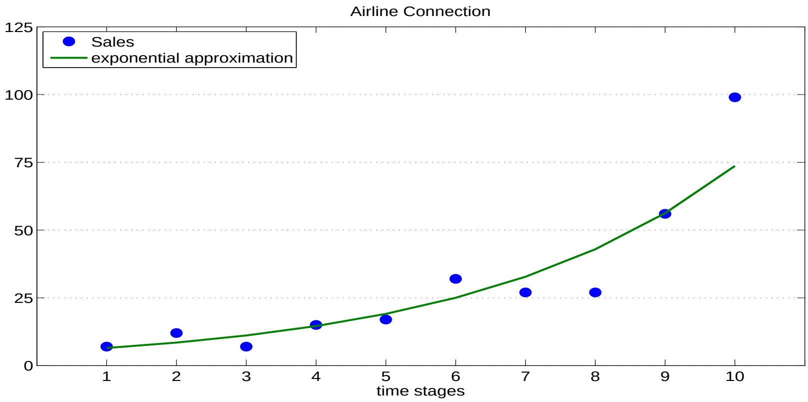

Figure 2.1. Airline sales behavior.

behavior for a fixed itinerary and departure day of week combination, with time stage 10 denoting the departure period. It contains all bookings with no differentiation of fare classes aggregated over some weeks. An example of the hotel sales behavior with aggregated sales data over multiple regions, as well as a single region is shown in Figure 2.2. The data corresponds to a fixed arrival date and time stages 11 and 13 denote the periods including the arrival at the hotel. Since the hotel dataset contains reservation data for multiple comparable hotels, we are very rarely observing total area sale outs, i.e. the case where a whole price class within an area is unavailable for bookings. Thus we are able to observe the untruncated booking behavior on some kind of bird's eye view. In addition to the monotonic increasing demand behavior, we observe that the quantiles on days in advance, i.e., distance between booking creation date and usage date of the product, in our hotel dataset are consistent in time. From a correlation analysis on the airline and

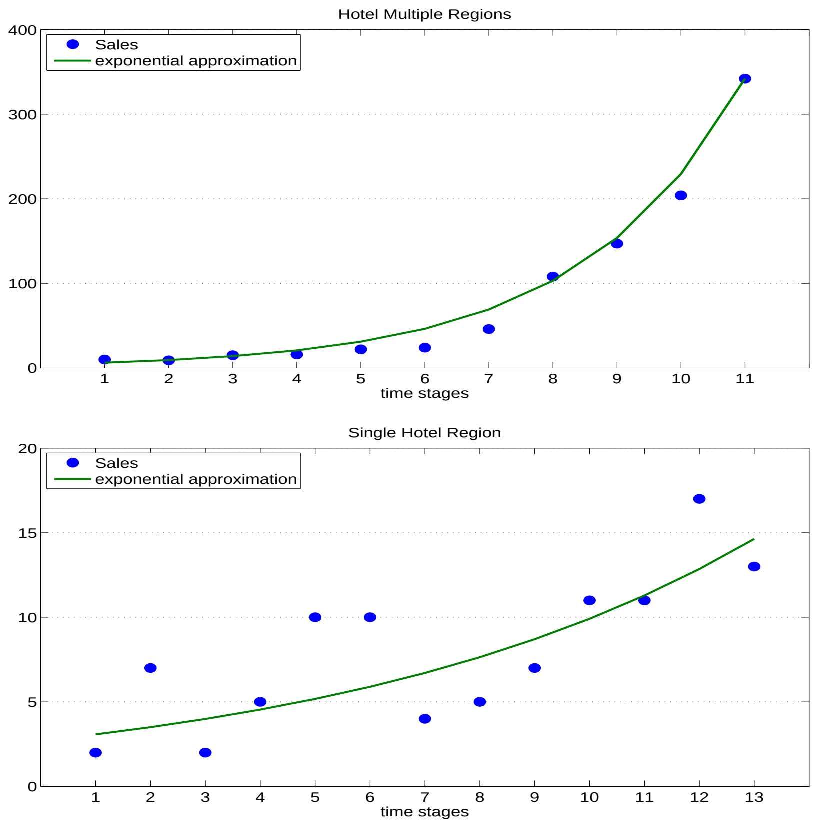

Figure 2.2. Airline and hotel sales behavior.

hotel datasets, we observe a high correlation between early and late bookings. This confirms our idea that demand rate is following a non-decreasing curve over time. From a least square analysis on an aggregated level, we learned that the demand rate per choice-set c follows approximately an exponential curve

$$\lambda_c(t) = \beta_c \cdot \mathrm{e}^{\alpha_c \cdot t}, \tag{2.9}$$

for time stage t and some parameters α_c and β_c, with β_c being non-negative. The sign of α_c depends on the definition of the booking horizon $1, \dots, T$. In our case, periods are defined as times towards usage of the product, e.g., departure of the airplane at T, the resulting curve is increasing in t and α_c is positive. But, if the time stages are defined as periods prior usage of the product, e.g., departure at $t = 1$, the resulting curve is decreasing in t and α_c is negative. Further, we tested the airline data if the demand per time stage is Poisson distributed. As statistical tests we used the likelihood ratio, the conditional chi-squared as well as the Neyman-Scott statistic. We find that the hypothesis can not be rejected. Consequently the demand of choice-set c is assumed to follow an inhomogeneous Poisson process with rate function $\lambda_c(t)$.

With the given demand rate functions for all choice-set, we are able to compute at time stage t the general arrival rate for a customer of any type by $\lambda(t) = \sum_{c \in C} \lambda_c(t)$, the sum of arrival rates over all choice-sets.

Also the primary demand (first choice demand), studied by Vulcano et al. (2011), can easily be computed. The estimated primary demand of product j at each time t is the sum of all demand rates $\lambda_c(t)$ over all choice-sets c, where j is the first choice class, i.e. the one with the highest preference.

2.3 A Dynamic Programming Allocation Control

An optimal control policy for a single resource problem including customer choice behavior is described in Talluri and van Ryzin (2004a) and Talluri and van Ryzin (2004b). In the following we will give a brief description of the dynamic programming model and how it is applied in our context. For detailed proofs the reader is referred to Talluri and van Ryzin (2004a). The original model considers a time horizon divided into intervals $i = 1, \ldots, I$ such that we have at most one customer arrival per interval. The probability of arrival is denoted λ and assumed to be constant in time.
In our case we are assuming an inhomogeneous Poisson process, where the arrival rate is supposed to be constant within each time stage $t = 1, \ldots, T$. Therefore each time stage will be divided into such intervals $\mathbb{I}_t$, such that we have at most one arrival per interval. The probability of any arrival at time stage t in all intervals i is given by $\Lambda(t) = \lambda(t)/\mathbb{I}_t$. Let us suppose we fixed a time stage t:
At each interval i we can make a decision on which classes to offer, i.e. choosing a set $O \subseteq N$, where N denotes the set of all classes. The probability that an arriving customer buys product j when O is offered is given by $P_j(O, t)$, where $P_j(O, t) = 0$ if $j \notin O$. We can compute this probability by

$$P_j(O, t) = \frac{\sum_{c \in C} \lambda_c(t) \cdot \mathbb{I}_{\{U(c,O)=j\}}}{\lambda(t)}. \tag{2.10}$$

So the probability for a sale of class j in interval i is given by $\Lambda(t) \cdot P_j(O, t)$, given that price classes in O are offered. Similar we can compute the no-sales probability $P_0(O,t)$ for each arrived customer by

$$P_0(O,t) = \frac{\sum_{c \in C} \lambda_c(t) \cdot \mathbb{I}_{\{U(c,O)=0\}}}{\lambda(t)}. \tag{2.11}$$

The no-sale probability at each interval i is then $\Lambda(t) \cdot P_0(O,t)$. The value function is defined as the maximal obtainable future revenue from period i in time stage t and inventory level x. The Bellman equation is then given by

$$V_i(x,t) = \Lambda(t) \cdot \max_{O \subseteq N} \{R(O,t) - Q(O,t)\Delta V_{i+1}(x,t)\} + V_{i+1}(x,t), \tag{2.12}$$

where $\Delta V_{i+1}(x,t) = V_{i+1}(x,t) - V_{i+1}(x-1,t)$. The purchase probability $Q(O,t)$ and the expected revenue $R(O,t)$ from offering set O are computed by

$$\begin{aligned} Q(O,t) &= \sum_{j \in O} P_j(O,t), \\ R(O,t) &= \sum_{j \in O} r_j \cdot P_j(O,t). \end{aligned} \tag{2.13}$$

On the boundaries the value function is defined as

$$\begin{aligned} V_i(0,t) &= 0 \qquad \forall i = 1, \dots, I, \\ V_{I+1}(x,t) &= \begin{cases} V_1(x,t+1) & \text{, if } t < T, \\ 0 & \text{, else} \end{cases} \qquad \forall x. \end{aligned}$$

Talluri and van Ryzin show in their paper that we only have to consider the subclass of efficient sets over all possible subsets of N, when making the decision which set of classes to offer.

Definition 2.1. *A set $S \subseteq N$ is* efficient *if there exist no non-negative weights $\alpha(M)$ on all $M \subseteq N \cup \{\emptyset\}$ with $\sum_{M \subseteq N \cup \{\emptyset\}} \alpha(M) = 1$ such that*

$$\frac{R(S,t)}{Q(S,t)} \geq \frac{\sum_{M \subseteq N \cup \{\emptyset\}} \alpha(M) R(M,t)}{\sum_{M \subseteq N \cup \{\emptyset\}} \alpha(M) Q(M,t)}.$$

We assume the collection of efficient sets E are indexed in increasing revenue and probability order. This is possible because for an indexing of E such that $Q(S_1,t) \leq Q(S_2,t) \leq \cdots \leq Q(S_m,t)$ follows that the expected revenues are also ordered in the same way. The problem of maximizing (2.12) is simplified by choosing among all sets in E

$$V_i(x,t) = \Lambda(t) \cdot \max_{k=1,\dots,m} \{R(S_k,t) - Q(S_k,t)\Delta V_{i+1}(x,t)\} + V_{i+1}(x,t). \tag{2.14}$$

Theorem 2.1. *Consider k^* such that S_{k^*} maximizes (2.14). For a fixed i, the largest optimal index k^* is increasing in the remaining capacity, and for any fixed x, k^* is decreasing in the remaining intervals.*

The problem of identifying the efficient sets can be solved in an inductive way. Initially we set $S_0 = \emptyset$. Having the r^{th} efficient set S_r we find the $(r+1)^{th}$ by

$$\text{argmax}_{S \subseteq N} \frac{R(S,t) - R(S_r,t)}{Q(S,t) - Q(S_r,t)}. \tag{2.15}$$

Definition 2.2. *A control policy is called nested if there exists an increasing family of subsets $S_1 \subseteq S_2 \subseteq \cdots \subseteq S_m$, with $S_j \subseteq N, j = 1, \ldots, m$. The set index $k_i(x)$, increasing in x, is chosen at interval i when the remaining capacity is x.*

Classes are seen as "higher" if they appear earlier in the sequence. And a policy is in fare order if the nesting order coincides with the fare order. In the case of a nested optimal policy we can compute the protection levels by

$$\begin{aligned} p_k^*(i) \quad = \quad & \max \quad x \\ & s.t. \quad \Delta V_{i+1}(x,t) > \frac{R(S_{k+1},t) - R(S_k,t)}{Q(S_{k+1},t) - Q(S_k,t)}. \end{aligned} \tag{2.16}$$

At interval i, only the classes contained in S_k will be open if the remaining capacity is less than $p_k^*(t)$. Nested booking limits are equivalently defined by $b_k^*(i) = capacity - p_{k-1}^*(i)$.
In most cases we have this nested policy, and usually the optimal policy is even nested by fare order. Nevertheless, the nested structure is not certain and the hypothesis has to be checked to apply (2.16) for the computation of optimal protection levels. In our test cases the policy will be in fare order.

2.4 A First Parameter Estimation Approach

The demand is assumed to follow an inhomogeneous Poisson process. As described in section 3 we assume the demand rate per choice-set to be approximately following

$$\lambda_c(t) = \beta_c \cdot \mathrm{e}^{\alpha_c \cdot t}, \tag{2.17}$$

for time stage t in the booking horizon $t = 1, \ldots, T$ and some parameter α and β depending on choice-set c, which are supposed to be positive. In our analysis we only work with observable sales data. Our goal is to unconstrain these data to get a good approximation of the real demand per choice-set. The given datasets contain the information if a class was open for bookings at a certain time and if yes the observed number of bookings or sales. For the parameter estimation we use the maximum likelihood method. The likelihood function for choice-set c is

$$L_c(\alpha_c, \beta_c) = \prod_{t=1}^{T} P(S_c(t)|\lambda_c(t)), \tag{2.18}$$

where $S_c(t)$ denotes the number of sales corresponding to choice-set c at time stage t. Remember that a sale corresponding to a certain choice-set will always materialize in its lowest available class. We will denote with "c open at t" the fact that at least one class in c was open for bookings at time t. So we obtain an expression for probabilities by

$$P(S_c(t)|\lambda_c(t)) = \begin{cases} \frac{\mathrm{e}^{-\lambda_c(t)} \cdot \lambda_c(t)^{S_c(t)}}{(S_c(t))!} & \text{, if } c \text{ open at } t \\ 1 & \text{, else.} \end{cases} \tag{2.19}$$

Since we are interested in the maximum of the likelihood function we will examine the log-likelihood function

$$\begin{aligned} \mathcal{L}_c(\alpha_c, \beta_c) &= \log\left(L_c(\alpha_c, \beta_c)\right) \\ &= \sum_{t=1}^{T} \log(P(S_c(t)|\lambda_c(t))) \\ &= \sum_{t=1}^{T} S_c(t)(\log(\beta_c) + \alpha_c t) - \beta_c \exp(\alpha_c t) - \log(S_c(t)!), \end{aligned} \tag{2.20}$$

for which the gradient and Hessian are given by

$$\nabla \mathcal{L}_c = \sum_{t=1}^{T} \begin{pmatrix} S_c(t)t - \beta_c t \exp(\alpha_c t) \\ \frac{S_c(t)}{\beta_c} - \exp(\alpha_c t) \end{pmatrix}, \tag{2.21}$$

$$H(\mathcal{L}_c) = \sum_{t=1}^{T} \begin{pmatrix} -\beta_c t^2 \exp(\alpha_c t) & -t \exp(\alpha_c t) \\ -t \exp(\alpha_c t) & -\frac{S_c(t)}{\beta_c^2} \end{pmatrix}. \tag{2.22}$$

The maximum likelihood estimates for parameter α and β and choice-set c are obtained by minimizing the negative log-likelihood function

$$(\hat{\alpha}_c, \hat{\beta}_c) = \arg\min_{\alpha,\beta>0} \quad -\mathcal{L}_c(\alpha, \beta). \tag{2.23}$$

We show in Chapter 4 that the negative log-likelihood function is unimodal and has a unique minimizer in $\mathbb{R} \times \mathbb{R}_+$. We omit the proof at this point and refer to the later chapter, which provides an in-depth analysis of the choice-set unconstraining problem. Having the optimal choice-set demand parameter, we can estimate the demand for unobservable time periods by a simple extrapolation of the demand rate function.

For now, we did not take the interaction of choice-sets into account and supposed that we know to which choice-set a sale corresponds. Let us illustrate this with our initial example with two fare classes A and B. Consider the two overlapping choice-sets $c_1 = \{A\}$ and $c_2 = \{A, B\}$. Say, we observe at a certain day a sale in class A, so we do not know if this is a realization corresponding to a customer of c_1 or c_2, since we only have information if a class was open, and if yes, the number of sales in it.

This leads to the following log-likelihood function

$$\mathcal{L} = \sum_{t=1}^{T} \sum_{f=1}^{F} \log P \left[X = S(t,f) \middle| X \sim \text{Poisson}\Big(\sum_{c \in C} \mathbb{I}_{\{U(c,t)=f\}} \cdot \lambda_c(t) \Big) \right], \tag{2.24}$$

where $\mathbb{I}$ denotes the indicator function, $U(c,t)$ returns the preferred available class in choice-set c at time t ($U(c,t) = 0$ means no available class in c at time t), F represents the number of classes and $S(t,f)$ denotes the observed sales in class f at time t. From simulation we learned that the negative log-likelihood function is in general not unimodal. Since we are usually confronted with multiple overlapping choice-sets we have to consider $2 \cdot |C|$ variables in the resulting non-convex optimization problem.

To overcome this problem we suggest the EM method, explained in Dempster et al. (1977). Initially we are estimating the parameter separately for all choice-sets by ignoring the interaction of choice-sets. Further, we use the fact that the random combination of Poisson processes gives again a Poisson process. Given the initial estimates of the corresponding rates $\lambda_c^0(t)$ for all times t and all choice-sets c, we start the iterative procedure at $i = 1$ by computing

$$P_c^i\big(S(t,f)\big) = \begin{cases} P\left[X = \left\lceil \frac{\lambda_c^{i-1}(t)}{\lambda_{overlap}^{i-1}(c,t)} \cdot S(t,f) \right\rceil \right] \cdot \mathbb{I}_{\{U(c,t)=f\}} & \text{, if } f > 0 \\ 1 & \text{, else} \end{cases}$$
$$\text{where } X \sim \text{Poisson}(\lambda_c^i(t)). \tag{2.25}$$

The $\lceil x \rceil$ operator returns the closest integer greater than or equal to x. The parameter $\lambda^j_{overlap}(c,t)$ denotes the sum of the estimated rates from iteration j over all choice-sets for which the preferred available classes coincide with choice-set c. In the maximization step the resulting negative log-likelihood function is then minimized separately for all choice-sets c to obtain new estimates of $\lambda^i_c(t) = \beta^i_c \cdot exp(\alpha^i_c \cdot t)$

$$(\alpha^i_c, \beta^i_c) = \arg\ \min_{\alpha,\beta>0} \quad -\mathcal{L}_c(i), \tag{2.26}$$

where the log-likelihood function at iteration i for choice-set c is given by

$$\mathcal{L}_c(i) = \sum_{t=1}^{T} \log P^i_c\left(S\Big(t, U(c,t)\Big)\right). \tag{2.27}$$

We are updating the previous estimates in equation (2.25) with the new rate estimates for choice-set c, which is the expectation step. For the maximization step we are again interested in the shape of the negative log-likelihood function and the formulation of the its gradient and Hessian. Since the log-likelihood function is computed separately for each choice-set, we can compute the gradient and Hessian at each iteration i as in the initial case for non-overlapping choice-sets

$$\nabla\mathcal{L}_c(i) = \sum_{t=1}^{T} \begin{pmatrix} S(c,t,i)t - \beta_c t\exp(\alpha_c t) \\ \frac{S(c,t,i)}{\beta_c} - \exp(\alpha_c t) \end{pmatrix} \cdot \mathbb{I}_{\{U(c,t)>0\}} \tag{2.28}$$

$$H(\mathcal{L}_c(i)) = \sum_{t=1}^{T} \begin{pmatrix} -\beta_c t^2 \exp(\alpha_c t) & -t\exp(\alpha_c t) \\ -t\exp(\alpha_c t) & -\frac{S(c,t,i)}{\beta_c^2} \end{pmatrix} \cdot \mathbb{I}_{\{U(c,t)>0\}}, \tag{2.29}$$

if c is open the number of corresponding sales are given by

$$S(c,t,i) = \left\lceil \frac{\lambda^{i-1}_c(t)}{\lambda^{i-1}_{overlap}(c,t)} \cdot S\big(c, U(c,t)\big) \right\rceil. \tag{2.30}$$

In Chapter 4, we show that the negative log-likelihood function (2.27) is again unimodal. This proves that the M-step is always well defined.
The procedure is repeated until it converges. Even though convergence is in general not certain, the EM method is known to be very robust and has been also satisfactory employed by McGill (1995), Weatherford and Pölt (2002), Talluri and van Ryzin (2004a) and Vulcano et al. (2011).

EM algorithm to unconstrain demand rate function per choice-set:
Initialization: compute separately for all choice-sets $c \in C$

$$(\alpha_c^0, \beta_c^0) = \arg\min_{\alpha,\beta>0} \quad -\sum_{t=1}^{T} \log \begin{cases} P\Big(X_t = S\big(t, U(c,t)\big)\Big) & \text{, if } U(c,t) > 0 \\ 1 & \text{, else} \end{cases},$$

where $X_t \sim \text{Poisson}\big(\lambda = \beta \exp(\alpha t)\big)$.

Iteration Loop $i = 1, ...$

E-step:
For all $c \in C$
For all $t = 1, ..., T$

$$P_c^i\Big(S\big(t, U(c,t)\big)\Big) = \begin{cases} P\left[X = \left\lceil \frac{\lambda_c^{i-1}(t)}{\lambda_{overlap}^{i-1}(c,t)} \cdot S(t, U(c,t)) \right\rceil\right] & \text{, if } U(c,t) > 0 \\ 1 & \text{, else} \end{cases}$$

where X $\sim$ Poisson$(\lambda_c^i(t) = \beta_c^i exp(\alpha_c^i t))$.
end for
$\mathcal{L}_c(i) = \sum_{t=1}^{T} \log P_c^i \left(S\big(t, U(c,t)\big)\right)$
end for

M-step:
For all $c \in C$
$(\alpha_c^i, \beta_c^i) = \arg\min_{\alpha,\beta>0} \quad -\mathcal{L}_c(i)$
end for

Until Stopping criteria reached.

The stopping criteria could be either a maximum number of iterations or some kind of numeric convergence bound on changes in α and β between iterations. The demand rate function for each choice-set c at time t is estimated as long as at least one class contained in c is offered at t. A no sale observation at t is regarded as a realization of the stochastic arrival process with no arrivals from all choice-sets which intersect with the set of offered price classes at time t. The estimation is extrapolated over periods when no price class contained in choice-set c is offered, i.e. no customer arrivals corresponding to c are observable. Notice that the EM method is independent of the demand rate function form. The exponential curve is a result from our data analysis of the airline and hotel datasets. Only the unimodality of the negative log-likelihood function needs to be checked for different demand functions, it can be easily shown that it also holds for constant or linear demand functions.

2.5 Numerical Tests

In this section we present the numerical results of our estimation method. We evaluate the estimation results on two criteria, first on how good the real rate function is approximated. And second on the expected revenue gain by using the estimated values compared to the real values in the optimal control policy, evaluated on 100 independent demand scenarios and the obtained revenues are averaged. The computation is done in MATLAB R2008a, for the optimization we used the function “fmincon” from the optimization toolbox. The starting parameters for fmincon are computed by a small Monte Carlo simulation. Our input sales simulations are generated by independent demand realization from the known demand functions per choice-sets. Here the booking process is controlled via given initial booking limits. Using this procedure we are generating 1000 sales observations from where we estimate the parameter α and β for each choice-set. The resulting estimates are averaged. In all examples the booking horizon is set to 10 time stages. The EM method is stopped if the maximal change of the minimized value of the negative log-likelihood function over all choice-sets is less than $10^{-6}\%$ compared to the optimum from the previous iterations. In all beneath cases this convergence bound is reached within less than 30 steps.

2.5.1 Two overlapping choice-sets

First we analyze the case of two overlapping choice-sets, considering two products A and B (prices are 50 and 100) with the corresponding choice-sets $Set_1 = \{A\}$ and $Set_2 = \{A, B\}$, parameters are given in Table 2.3. The initial booking limits are 10 and 10 for products A and B respectively, see Figure 2.3 to compare the estimated demand function with the original. The estimated

	α_{Set1}	β_{Set1}	α_{Set2}	β_{Set2}	Revenue
Original	0.1	1	0.2	0.3	1251
Estimates	0.1646	0.9249	0.0736	0.9124	1132

Table 2.3. Case 1 - parameter and generated revenue.

demand function for Set_1 is after time stage 6 a very poor estimate of the real underlying demand function. It is a fact that the booking limit for product A will be reached between time stage 5 and 7. This means that the parameters for Set_1 are only fitted to sales data up to this moment and extrapolated into the future. In contrast, we observe that the estimated demand function for Set_2 approximates the real function closer at the end of the booking horizon.

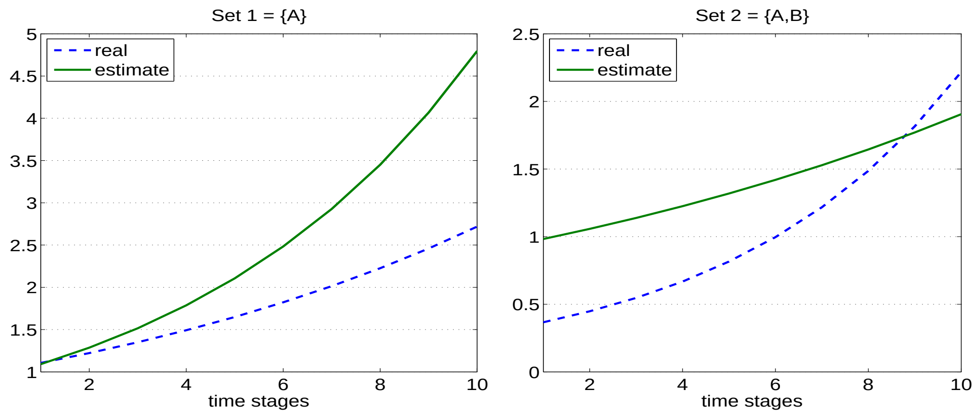

Figure 2.3. Case 1 - demand function of real parameters and estimates.

When A is closed we know for sure that we observe a realization of Set_2 and thus the estimate will be more precise. The estimated α and β parameter for both choice-sets, and the generated mean revenue using the estimates and real parameter in the optimal booking control, are shown in Table 2.3. Having perfect information about the demand functions for both choice-sets we generate in average 12.9 sales of product A and 6.06 of product B giving an average generated revenue of 1232. With our estimated parameters we generate in average 11.6 and 5.26 sales for products A and B respectively, which results in a mean revenue of 1159. This means from only given sales data of products A and B (generated using non-optimal booking limits), we were able to approximate the demand rate functions such that using these estimates we can gain 90.5% of the expected revenue from having perfect demand rate information($PI\ Rev$). For the moment we analyzed a case with small demand rate values. In case 2 we will increase the α and β parameter for Set_1 and Set_2, to see if the precision of our estimation performs better for higher rate values. The new parameter as well as the newly estimated are shown with their generated revenue in Table 2.4. The initial booking limits to generate the sales data are 50 and 100 for products A and B respectively, the prices are the same as above. See Figure 2.4 for a comparison of the real and estimated demand functions. The demand rate function for Set_2 is now much better approximated. As in the previous case we observe a better fit for time stages when A is closed. For Set_1 we observe now a underestimation of the demand rate, contrary to the previous case where the estimates where overestimating the demand rate. This underestimation seems to be very large, but analyzing the obtained revenue from the estimated parameter, we observe that this lack of accuracy does not imply a huge loss of revenue. In fact with our estimates we are very close to the revenue obtained by perfect demand rate information. Having this perfect information about the demand functions we can generate

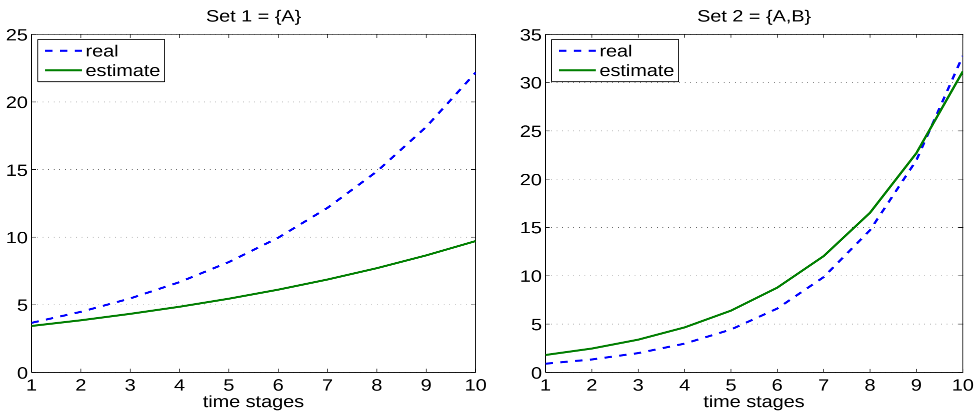

Figure 2.4. Case 2 - demand function of real parameters and estimates.

on average 63.52 sales of product A and 72.78 sales of product B. With the estimated parameter we generated on average 26.15 sales of product A and 88.26 sales of product B. This means by using estimated demand functions we can generate 97% of the *PI Rev*.

	α_{Set1}	β_{Set1}	α_{Set2}	β_{Set2}	Revenue
Original	0.2	3	0.4	0.6	10454
Estimates	0.1155	3.0593	0.3169	1.3101	10134

Table 2.4. Case 2 - parameter and generated revenue.

2.5.2 Using Rounding instead the Ceiling Operator

Another question is how the estimation would perform if we replace the upper integer operator ($\lceil\cdot\rceil$) in equations (2.25) and (2.30) with the rounding operator ($\lfloor\cdot\rceil$). To analyze this case we will use the same parameters as before, see Table 2.5 for a comparison of the estimates and original values. The resulting demand rate functions for both cases are shown in Figure 2.5. As we see from the graphs the change of operator does not have a big impact for problems on larger demand rates. But for smaller demand rates the rounding operator tends to over-estimate the demand much more than the ceiling operator. This is due to the generally ill conditioned problem of fitting an exponential

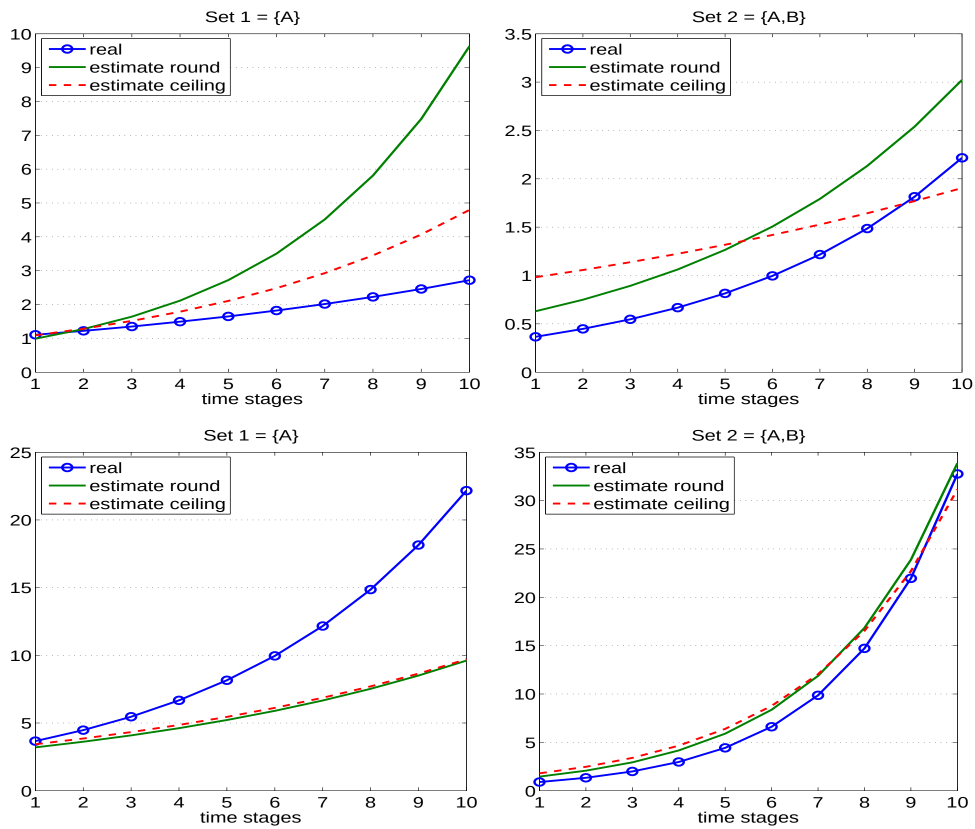

Figure 2.5. Demand function for smaller (top) and larger (bottom) rates.

	α_{Set1}	β_{Set1}	α_{Set2}	β_{Set2}	Revenue
Original	0.1	1	0.2	0.3	1251
Estimates $\lceil\cdot\rceil$	0.1646	0.9249	0.0736	0.9124	1132
Estimates $\lfloor\cdot\rceil$	0.2528	0.7687	0.1742	0.5293	1079
Original	0.2	3	0.4	0.6	10454
Estimates $\lceil\cdot\rceil$	0.1155	3.0593	0.3169	1.3101	10134
Estimates $\lfloor\cdot\rceil$	0.1221	2.8343	0.3495	1.0280	10124

Table 2.5. Parameter and generated revenue.

curve using maximum likelihood. The solver is in between multiple parameter combinations. With the ceiling operator small steps of rate changes result in larger steps of realization estimates and so the parameter estimation will be more conservative. Therefore we suggest to use the ceiling operator instead of the rounding operator as described in the previous section.

2.5.3 Stepwise Improvement

Further, we will test if the estimation can be improved stepwise. Therefore we compare the "first" estimates with the "second". The first estimates are generated as before. The second estimates are obtained by the estimation from sales realization generated using the first estimates instead of fixed booking limits. Results are shown in Table 2.6 and the demand rate functions are shown in Figure 2.6. Comparing the demand rate functions for Set_2 we ob-

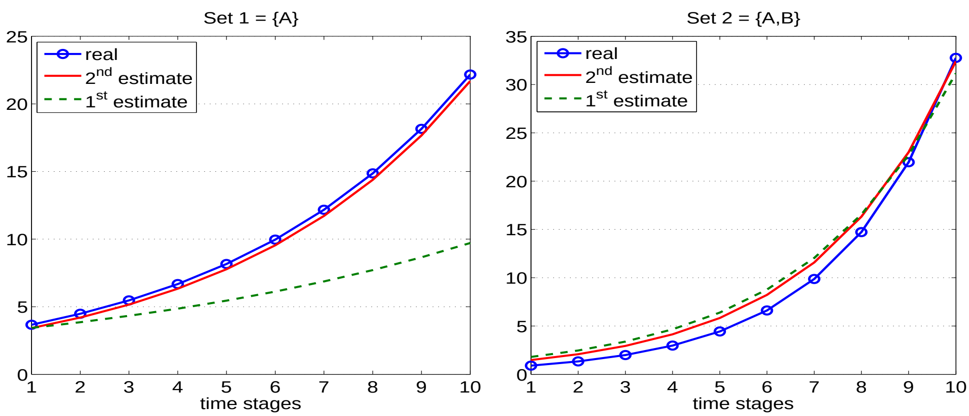

Figure 2.6. Demand function of real parameters and estimates.

serve only a small increase in fitting accuracy. But for Set_1 the results of the second estimates are almost a perfect fit, compared to the huge underestimation resulting from the first estimates. There is also no loss of accuracy for time stages when A is closed. Using the first estimates results already in 97% of the *PI Rev*, with the second estimates we even improve this value to 99%. Thus we conclude that the estimation method improves under more optimal booking controls, since the information contained in the sales data reflects the real demand better. And therefore it will not suffer from the spiral down effect as discussed in Cooper et al. (2006), when used in dynamic booking controls.

2.5.4 Multiple choice-sets

In reality we have of course more than two choice-sets, thus our final study case will consist of three price classes: Economy, Business and Premium with prices 100,150 and 200. The resulting choice-sets are: $\{E\}, \{E, B\}, \{B\}, \{B, P\}$ and

	α_{Set1}	β_{Set1}	α_{Set2}	β_{Set2}	Revenue	% PI Rev
Original	0.2	3	0.4	0.6	10454	100%
first Estimates	0.1155	3.0593	0.3169	1.3101	10134	97%
second Estimates	0.2053	2.7838	0.3429	1.0500	10442	99%

Table 2.6. Parameter and generated revenue.

choice-set	α	β	α'	β'	α''	β''
$\{E\}$	0.1	6	0.0989	4.6706	0.1191	3.8758
$\{E, B\}$	0.15	2	0.0334	5.1163	0.1611	2.4389
$\{B\}$	0.2	1.5	0.1801	1.8443	0.2089	1.6768
$\{B, P\}$	0.25	1	0.1506	2.0707	0.2098	1.6605
$\{P\}$	0.3	0.5	0.3784	0.3854	0.3064	0.5213

Table 2.7. Real parameters with first (') and second (") estimates.

$\{P\}$, in Table 2.7 the demand parameters are shown. Again we are considering a booking horizon of 10 time stages and the initial booking limits are set to 100,100 and 50 for fare classes E,B and P respectively. Figure 2.7 displays the resulting demand rate functions of the first and second estimates compared to the real values. Even the first estimates give a good approximation of the real demand functions. But there is still a big overestimation for set $\{P\}$ and the shape of set $\{E, B\}$ is very poorly approximated. With the second estimates the approximation becomes much tighter. In Table 2.8 the sales and revenue results from using the estimated values in the booking control are shown. With the first estimates we have already gained a *% PI Rev* of 99.4%, which is even increased to 99.6% by using the second estimates. The lower expected revenues are resulting from the underestimation of the $\{E\}$ demand and the overestimation of $\{B, P\}$ and $\{P\}$ demand. This explains also the lower capacity utilization (Cap. Util.) by using the estimates.

	Sales E	Sales B	Sales P	Cap. Util.	Revenue	% PI Rev
Original	92.3	119.5	36.7	248.5	34498	100%
1^{st} est	89.8	119.7	36.7	246.2	34286	99.4%
2^{nd} est	81.2	125	37.5	243.7	34360	99.6%

Table 2.8. Generated Revenue using real parameters and estimates (est).

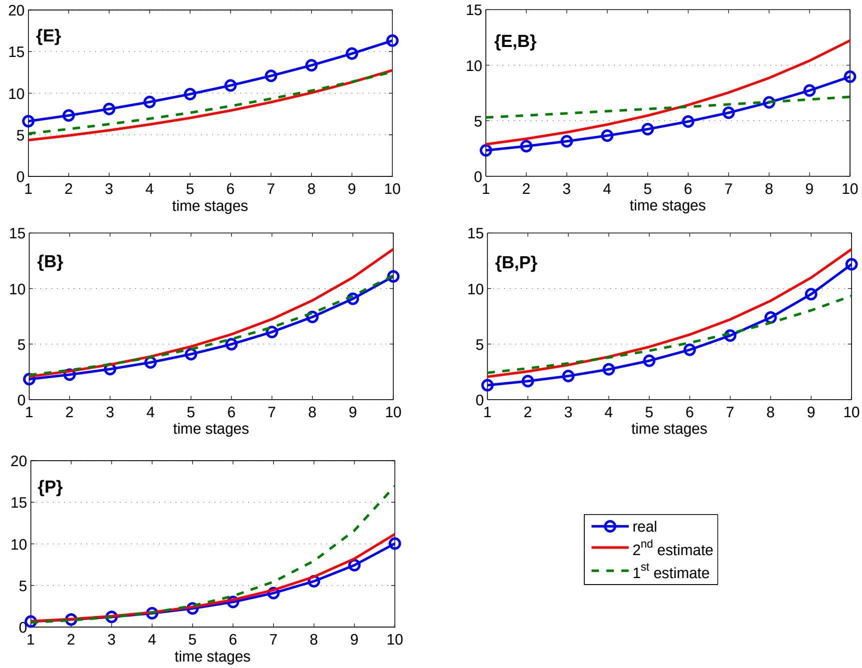

Figure 2.7. Demand function of real parameters and estimates.

2.6 Conclusion

In this chapter we have stated and analyzed a general method to estimate customer choice-set parameter from given sales observations. Previous research papers, such as van Ryzin (2005) and Bobb and Veral (2008), have shown that there is a huge need in revenue management for a move from product-based demand to customer choice demand. Based on the obtained results, we are convinced that our method will be helpful in achieving this goal. Even though using maximum likelihood to approximate an exponential curve from data points is in general ill conditioned, the resulting demand curves fit the original ones reasonably well and hence indicate a good approximation. Plus, using the excerpt information to control the future booking process will increase the resulting revenue, as well as the accuracy of future estimations. We even observe that using the estimates in the future booking process, we were able to generate revenue results close to optimality. The proposed unconstraining approach focuses on the two demand censoring dimensions. First, the unobservable demand due product unavailability, and second, the demand constraining into product sales dependent on the customer preferences and the company's offers. The proposed demand estimation method provides a revealing view on the actual demand with information on the choice behavior.

Chapter 3

A Demand Unconstraining Case Study on Airline Data

This chapter is build on the paper Haensel et al. (2011b).

Accurate demand information is essential for the success of all kinds of sophisticated booking or pricing controls in revenue management. Any successful RM systems needs customer information on the micro-market level. The information should not only contain the number of customers to expect, but also comprise information on customer behavior and preferences. What is actually available at companies, as input to the demand estimation, is only historical sales and availability data per price classes or products. But customers who book the same price classes or buy the same products are not necessarily equal; possibly some of them are also interested in other products or would also buy the product at a higher price. Therefore, the sales data needs to be unconstrained in order to obtain a clear understanding of the actual demand. The unconstraining is needed in two dimensions: First, we are interested in the number of unobserved demand, which was turned down by the company's offers and did not result in a transaction. Second, we want to know the preferences in the choice process of customers who made a transaction. The choice-set demand model and the corresponding demand rate functions are ideal to combine both dimensions, and we aim to estimate demand rates per choice-set, i.e., per choice behavior group. The objective of the chapter is to test the choice-set unconstraining method, as proposed in the previous chapter, on real airline transaction data.

The chapter continues in the following section with the explanation of the airline dataset, followed by the choice-set model in Section 3.2. The estimation results are presented in Section 3.3, and our general findings are concluded in the final section.

3.1 Available Dataset

In our case study, we are able to work on real airline booking data of two routes provided by Transavia Airlines, which we will call from this point simply Route 1 and Route 2. Both routes are connecting the Amsterdam airport Schiphol (AMS) with a Spanish airport and there is only one competitor airline serving the same direct connection. Unfortunately, we have no information of historic competitor prices available for our analysis. The datasets consist of the booking information for 11 consecutive departure day of weeks, i.e., we fixed a certain weekday for each route and work with the data of 11 weeks. The separation of different weekdays is very common in the airline business and based on statistical tests, which show a higher dependency and more common characteristics for consecutive weekdays than for consecutive days. The total bookings per departure day and route are shown in Figure 3.1. The usual possible booking horizon consist of several months and can span a

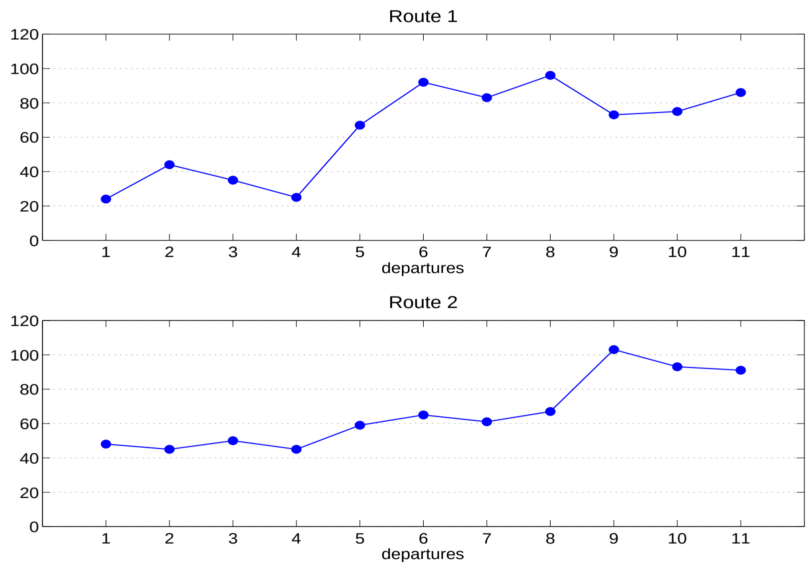

Figure 3.1. Total bookings for both routes and each departure.

period up to a whole year. Even so, we observe that most of the bookings are made in a much smaller time span, namely 12 weeks prior to departure. The average number of bookings per week are shown in Figure 3.2, where week 1 denotes the beginning of the booking horizon and week 12 the week including the departure day. On both routes we have $F = 12$ fare classes

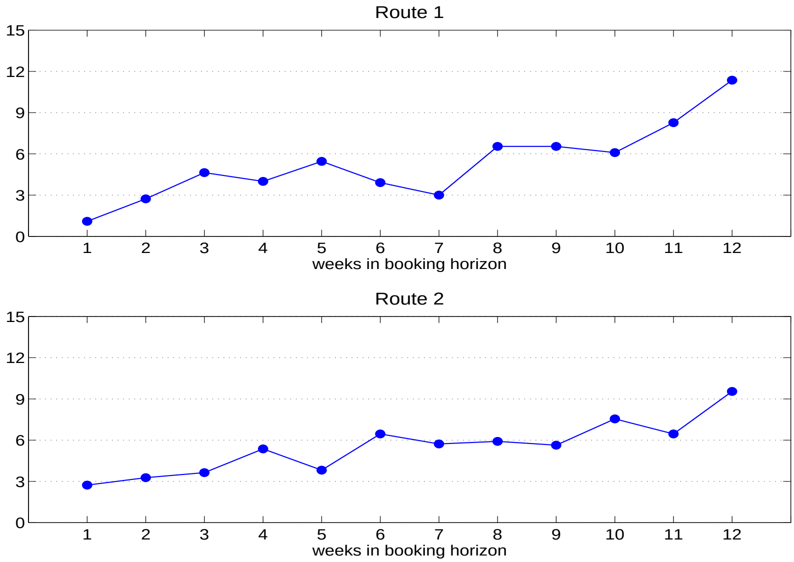

Figure 3.2. Average weekly bookings for both routes.

which only separate in price, as given in Table 3.1. There are no extra services

most expensive -	Y	Z	S	B	M	H	Q	V	K	L	T	N	- cheapest

Table 3.1. Fare classes.

or standards associated with different fare classes. Thus, the price is the only differentiator, so there is only one active fare class at a time. The fare class booking and availability data is given on daily level. This means we know for each day in the booking horizon which fare class is available for booking, which is open, and also how many bookings are made. A whole flight can be unavailable for bookings if all fare classes are not available/closed. Table 3.2 shows the summarized information per fare class for both routes. This information contains the number of departures when each fare class is open, the percentage on total booking days it is open ($11 \times 12 \times 7 = 924$ total booking days), the total number of sales/bookings and the averaged number of sales over all departures when the fare class was open. We find by adding the percentages of open days in Table 3.2 that Route 1 is 12.9% and Route 2 is 9.4 % of all considered booking days closed, i.e., no fare class is available.

Route 1

	Y	Z	S	B	M	H	Q	V	K	L	T	N
Dep. open	0	1	3	3	5	7	8	9	4	3	1	1
% open	-	0.6%	0.9%	3.4%	7.5%	20%	11%	29.5%	10.1%	3.7%	0.3%	0.1%
Total sales	0	2	18	24	48	149	105	160	45	45	7	7
Avg. sales	0	2	6	8	9.6	21.3	13.1	17.8	11.3	15	7	7

Route 2

	Y	Z	S	B	M	H	Q	V	K	L	T	N
Dep. open	2	2	3	3	4	6	8	8	6	3	1	0
% open	0.2%	2.4%	4.3%	6.3%	6.6%	10.2%	16.6%	26.5%	15%	2.2%	0.5%	-
Total sales	2	17	39	40	101	76	136	148	121	26	12	0
Avg. sales	1	8.5	13	13.3	25.3	12.7	17	18.5	20.2	13	12	0

Table 3.2. Performance data of Route 1 and 2.
(Dep. open - number of departures where this fare class is open, % open - fraction of possible booking days this class is open, Total sales - total sales per class over all departures, Avg. sales - averaged sale per class over open departures.)

The case that all fare classes are closed can have two causes: First, there were no available seats and no further seats could have been sold. And second, when data points are removed by our outlier detection. From the datasets we have only information about the total sales per day, but not how they are made up. For example, if we observe six sales for a given day, we don't know if they are six individual sales or two sales of size 3, etc. We observe in our datasets some days with very large daily bookings, see Figure 3.3. The average daily booking size of Route 1 is 2.8 with a standard deviation of 2.2, for Route 2 we observe small values with an average of 2.5 and a standard deviation of 1.8. The extreme booking sizes are likely generated by group bookings, which we choose to exclude from our computation. Group bookings are normally not made via the usual online sales channels, but by direct negotiation with airlines representatives. Therefore, we will exclude booking days with more than seven bookings and the availability for these days is set to zero. Thereby, we are not overwriting sales data, we only exclude outlier data points from our estimation analysis.

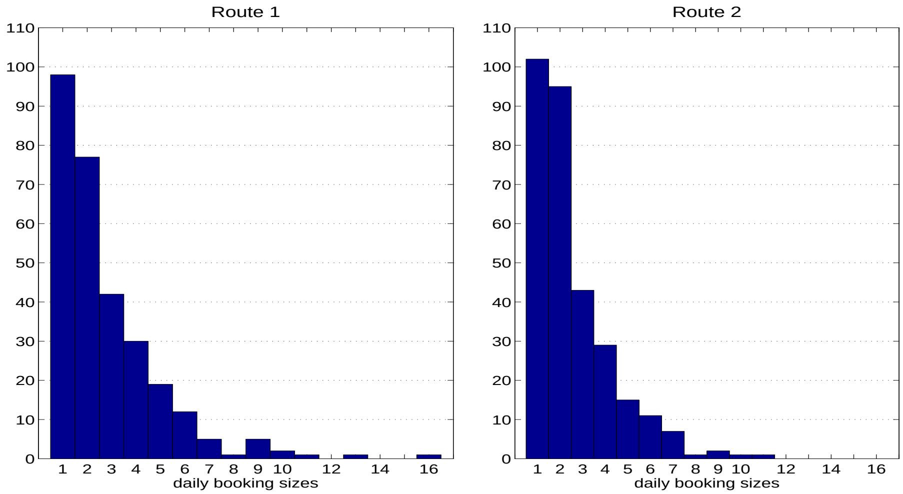

Figure 3.3. Histogram of daily booking sizes for both routes.

3.2 Customer Choice-set Demand

The concept of choice-sets is earlier introduced in Chapter 2. Customer choice-sets are sets of choice alternatives or products with a strict preference order and are assumed to model different choice behaviors of customers and their selections made. The usual approach in travel demand modelling is to divide customers into different non-overlapping groups or segments based on their characteristics, best example is a segmentation in business and leisure customers. Both segments are assumed to have a very different buying behavior. The first are seen to buy more on short notice and are considered to have a high willingness to pay, and the second are supposed to be more price sensitive but book long in advance. In fact, our airline observes many leisure customers, who have a relatively high willingness to pay and book close to departure, and on the other hand also observes early booking and price sensitive business customers. Therefore, our choice-set approach aims to distinguish customers only by choice behavior, independent of their individual characteristics. This can lead to choice-sets made up by a very homogeneous customer group, but also allows a mix of different types of customers if their observed buying behavior is similar. The proposed demand estimation method associates demand quantities with different choice-sets, representing different choice behaviors.
Usually, RM is concerned with the sale of a base product, e.g. in our case seats on an airplane. The product offer is divided into several price classes and different conditions may additionally be attached to the product itself, such as cancellation possibilities, etc. The different conditions on the product offers are introduced either as fences to separate customers into segments, or as optional extras to increase the customer evaluation of the product offer. In any case, the objective is to enable the seller to ask different prices for the same base product. All of the resulting product offer versions are called classes or subclasses. See Table 3.3, for an example airline portfolio with three main groups H, M and L. At each point in time all classes may or may not be available to buy. In our example, the conditions within the groups are equal and the price is the only differentiator. Customers are assumed to be rational and utility maximizers, compare to Ben-Akiva and Lerman (1985). Hence, there is only one active class per group, since the customer's utility of a product decreases with an increasing price. In our choice-set model we assume that each customer has a set of classes which represent his preferences, regarding price and conditions. In the airline example, the choice-sets may consist of any coherent sequence of classes, e.g., {L2,L1,M2} or {M2,M1,H2,H1}. In this context it makes no sense to consider choice-sets which are not coherent. From the customer point of view this would mean that a customer is interested in

Booking Class	Miles earned	Changes	Cancellations	Price
H1	100 %	charge 50	charge 100	520
H2	100 %	charge 50	charge 100	460
M1	50 %	charge 50	No	370
M2	50 %	charge 50	No	320
L1	25 %	No	No	250
L2	25 %	No	No	230

Table 3.3. Booking classes in airline example with groups high, medium and low.

buying low or high priced tickets, but no medium priced tickets. It is obvious that we can neglect these choice-sets. A customer will always choose the available class with the highest preference/ utility within his or her choice-set. The classes in the choice-set are displayed in hierarchic preference order, decreasing from left to right. If none of the classes, in which the customer is interested, is available, he or she will not buy at all.

The booking horizon is assumed to be divided into T time stages $t = t_1, \ldots, t_T$, where t_1 denotes the beginning of the booking horizon and t_T the last time stage before the departure of the airplane. In general, we observe strictly increasing demand curves over the booking horizon, i.e., the demand increases towards the time of departure in the airlines case. Even though one often observes a drop in demand very close to the airplane's time of departure, the width of time stages can be defined such that the assumption of increasing demand over time stages is justified. The assumption can be relaxed to allow more complex demand functions, at the costs of additional parameters and the loss of structural properties of the estimation problem. In our case of European flights, we only observe a very small demand drop just before departure.

The estimation of the choice-set model is divided into two steps: First, the identification of different choice-sets, and second, the demand estimation per choice-set. The identification of possible choice-sets is in our test case very straightforward. The airline offers no extras with the seat, such as extra services or cancellation possibilities, and the considered flights are at most operated once a day. Thus, the only product differentiation for a fixed itinerary is the price of the offered fare classes. Consequently the airline has only one available fare class at a time. All possible choice-sets can therefore be given by combinations of consecutive fare classes (in fare order) starting with the cheapest class. Since price is the only differentiator, the cheapest fare class has the highest preference for all customers and the choice-sets are only distinguishable by the upper willingness to pay. The choice-sets in our airline test case are shown in Table 3.4. For the second estimation step, the de-

highest acceptable fare class	choice-set
Y	N T L K V Q H M B S Z Y
Z	N T L K V Q H M B S Z
S	N T L K V Q H M B S
B	N T L K V Q H M B
M	N T L K V Q H M
H	N T L K V Q H
Q	N T L K V Q
V	N T L K V
K	N T L K
L	N T L
T	N T
N	N

Table 3.4. choice-sets in airline test case.

mand estimation per choice-set, we apply the algorithm proposed in Chapter 2. There, we describe a parameter estimation method for the case of Poisson distributed demand with exponential demand functions $\lambda_c(t) = \beta_c \cdot \exp(\alpha_c \cdot t)$, for choice-sets c and time stage t. The estimation method is based on maximum likelihood estimation (MLE) with an application of the EM-Algorithm. The open and closing decision of fare classes are on a daily level, but the time stages generally cover multiple days. Hence, we have to define $O_t = \bigcup_{d \in t} O_t(d)$ as the union of all sets of open classes $O_t(d)$ for each day d in time stage t. Further, we have to redefine the $U(c, O_t)$ for the input of sets of open classes by

$$U(c, O_t) = \bigcup_{d \in t} \begin{cases} \Big\{ U\big(c, O_t(d)\big) \Big\} & \text{, if } U\big(c, O_t(d)\big) > 0 \\ \emptyset & \text{, else.} \end{cases} \tag{3.1}$$

The general log-likelihood functions of the estimation problem is given by

$$\mathcal{L} = \sum_{c \in C} \sum_{t = t_1, \ldots, t_T} \log P \Big[X = \mathcal{S}(t, c) | X \sim \text{Poisson}\Big(\lambda_c(t) \cdot \mathbb{I}_{\{U(c, O_t) \neq \emptyset\}} \Big) \Big], \tag{3.2}$$

where $\mathcal{S}(t, c)$ denotes the number of sales/observed demand in time stage t corresponding to choice-set c. In our input data, we have only the information of sales per day and fare class $S(d,f)$ and not per choice-sets. As in Chapter 2, we propose to use the EM algorithm, introduced by Dempster et al. (1977), to overcome this information problem in the MLE. The EM algorithm is an iterative method, where parameters are computed under an expectation based on values from previous iterations. In our application, we compute the expected number of sales at time stage t corresponding to choice-set c in iteration i by

$$\mathcal{S}^i(t,c) = \left\lceil \sum_{d\in t} \frac{\lambda_c^{i-1}(t)}{\lambda_{overlap}^{i-1}(c,t,d)} \cdot S\Big(d, U\big(c, O_t(d)\big)\Big) \right\rceil, \tag{3.3}$$

where $\lambda_c^j(t)$ denotes the demand rate of choice-set c at time t in the j^{th} iteration of the EM algorithm, $\lceil\cdot\rceil$ denotes the ceiling operator and $\lambda_{overlap}^j(c,t,d)$ is defined by

$$\lambda_{overlap}^j(c,t,d) = \sum_{s\in C} \lambda_s^j(t) \cdot \mathbb{I}_{\{U(s,O_t(d))>0 \ and \ U(s,O_t(d))=U(c,O_t(d))\}}. \tag{3.4}$$

Another problem occurs when the demand of choice-set c is not observable at all days in times stage t, i.e., c is overlapping with $O_t(d)$ for some but not all $d \in t$. Days within a time stage do not have the same booking intensity, e.g., we observe different booking intensities for different weekdays. The estimated booking intensity $\pi_t(d)$ of day d in time stage t is computed over historic booking horizons and reflect the different weighting between days in the same time stage and $\sum_{d\in t} \pi_t(d) = 1$. The definition of $\mathcal{S}^i(t,c)$ is extended to incorporate the booking intensity with an application of the rule of proportion by

$$\mathcal{S}^i(t,c) = \left\lceil \frac{\sum_{d\in t} \frac{\lambda_c^{i-1}(t)}{\lambda_{overlap}^{i-1}(c,t,d)} \cdot S\Big(d, U\big(c, O_t(d)\big)\Big)}{\sum_{d\in t} \pi_t(d) \cdot \mathbb{I}_{\{U(c,O_t(d))>0\}}} \right\rceil. \tag{3.5}$$

The starting values of the EM algorithm are obtained by ignoring the intersection of choice-sets

$$\mathcal{S}^0(t,c) = \left\lceil \frac{\sum_{d\in t} S\Big(d, U\big(c, O_t(d)\big)\Big)}{\sum_{d\in t} \pi_t(d) \cdot \mathbb{I}_{\{U(c,O_t(d))>0\}}} \right\rceil. \tag{3.6}$$

λ_c^0, with the corresponding α_c^0 and β_c^0 parameters, is obtained by minimizing the negative log-likelihood function separately per choice-set

$$
\begin{aligned}
(\alpha_c^0, \beta_c^0) &= \arg\min_{\alpha,\beta>0} - \mathcal{L}_c^0 \\
&= \arg\min_{\alpha,\beta>0} - \sum_{t=1}^{T} \log \begin{cases} P\Big(X_t = S^0\big(t, c)\big)\Big) & \text{, if } U(c,t) > 0 \\ 1 & \text{, else,} \end{cases}
\end{aligned}
\tag{3.7}
$$

where $X_t \sim \text{Poisson}\big(\lambda = \beta_c \cdot \exp(\alpha_c \cdot t)\big)$. The EM algorithm can be separated in two steps: The expectation step (E-step), an application of equation 3.5 in the negative log-likelihood function for each choice-set. Second, the maximization step (M-step), which consists of minimizing $-\mathcal{L}_c$ separately for all choice-sets $c \in C$. The algorithm stops if one of the following criteria is satisfied:

- maximum number of iterations reached,
- no changes in α and β values between iterations.

In general, we observe $\mathcal{L}^{i+1} = \sum_{c \in C} \mathcal{L}_c^{i+1} \geq \sum_{c \in C} \mathcal{L}_c^i = \mathcal{L}^i$. This results in the optimal solution of the MLE. Occasionally, we observe that the EM algorithm reaches a likelihood maximum within the iteration cycle and that the final results hold a lower likelihood. In such a case we do not use the final EM output, but rather the intermediate results with the maximum likelihood.

3.3 Estimation Results

The proposed demand estimation method is tested on real airline reservation data to verify: First that the choice-set model approximates the underlying demand closely, and second that the estimation method is applicable for practitioners. In our estimation example we will consider the weeks in the booking horizon as time stages and the booking intensities π are obtained from the previous year's data. The estimation error is simply defined as

$$
error = actual - estimate.
$$

The demand estimate for any open fare class f at every booking day and departure combination is simply computed by

$$
D(f|O) = \sum_{c \in C} \lambda_c \cdot \mathbb{I}_{U(c,O)=f}, \tag{3.8}
$$

where O denotes the given set of open classes, in our case a singleton. The choice-set demand is estimated for all 11 departures in both datasets. In the following, we will examine the choice-set estimation errors from different perspectives, such as: total relative errors over all booking days and price classes, errors per time stages, and the error on fare class level. The total number of bookings over all fare classes and per departures is slightly overestimated by 1% for Route 1 and 2.6% for Route 2, i.e., the estimation error is negative if the estimate exceeds the actual. The total relative estimation errors for all departures are shown in Figure 3.4. We observe no pattern of constant over

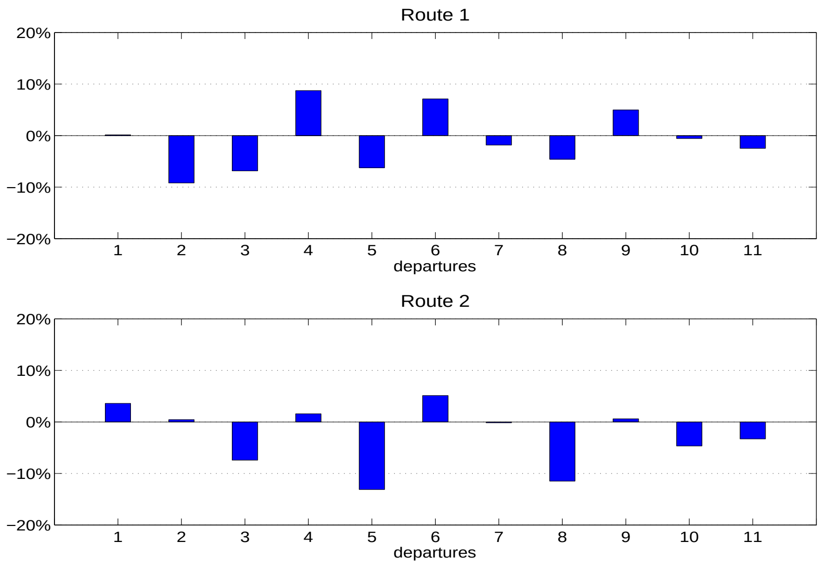

Figure 3.4. Total relative estimation errors per departure for both routes.

or under estimation and please note that we are considering total numbers of usually less than 100 bookings. Figure 3.5 shows the average weekly bookings and the corresponding estimation errors. We observe, especially for Route 1, a slight constant overestimation in booking weeks 7-11. The last time stage, week 12, is underestimated for both routes. But when comparing the estimation errors to the average bookings, we find the errors to be considerably small in proportion. Finally, we look into the estimation error on fare class level, with the results given in Table 3.5. As also shown in Table 3.2, we find that some fare classes are used much more often than other. Very high and very low classes are not often available for booking and thus we have limited data to estimate the corresponding choice-sets. But even with this limited data, the estimation errors are considerably small compared to the average booking number when the considered fare class was available. The consideration of

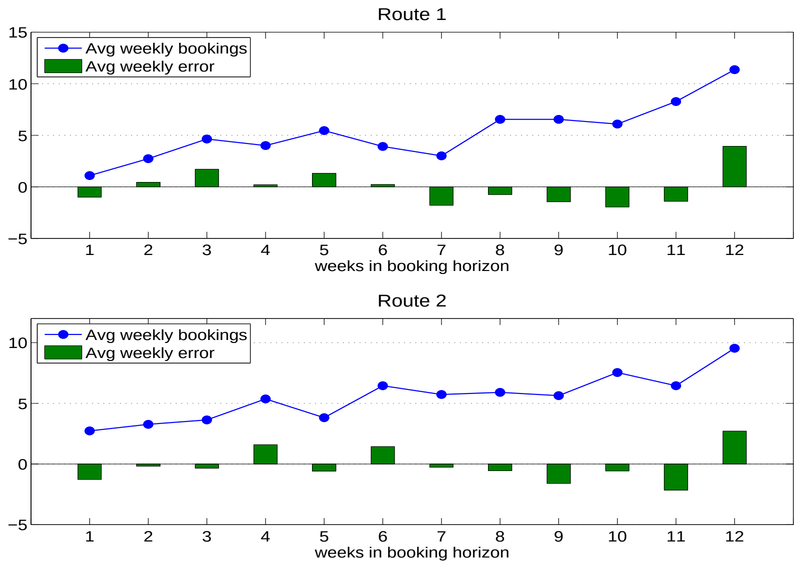

Figure 3.5. Average weekly bookings with the average approximation error for both routes.

error results at more frequently used classes shows very low average errors for both routes. These results reinforce our positive conclusion on the proposed choice-set based estimation method.

Route 1

	Y	Z	S	B	M	H	Q	V	K	L	T	N
% open	-	0.6%	0.9%	3.4%	7.5%	20%	11%	29.5%	10.1%	3.7%	0.3%	0.1%
Avg. sales	0	2	6	8	9.6	21.3	13.1	17.8	11.3	15	7	7
Est. error	-	-0.5	2.3	-0.3	0.8	1.3	-0.5	-1.3	-1.0	-2.6	-0.1	-0.1

Route 2

	Y	Z	S	B	M	H	Q	V	K	L	T	N
% open	0.2%	2.4%	4.3%	6.3%	6.6%	10.2%	16.6%	26.5%	15%	2.2%	0.5%	-
Avg. sales	1	8.5	13	13.3	25.3	12.7	17	18.5	20.2	13	12	0
Est. error	0.5	-1.8	-1.7	0	-1.5	1.4	-0.6	0	-2.3	1	0.6	-

Table 3.5. Error Results per fare class for both routes.
(% open - fraction of possible booking days this class is open, Avg. sales - averaged sale per class over open departures, Est. error - averaged estimation error per class.)

3.4 Conclusion

In this chapter, we studied the problem of unconstraining sales data per price classes into demand estimates per customer choice behavior. Our proposed estimation method is tested on real airline data. The results show a slight overestimation on the total demand over all fare classes, but it should be noted that the total considered sales figures are usually smaller than 100. Hence, estimation errors of 5% are equivalent to an actual error of at most 5 bookings. Much more interesting than the results on the total bookings are the estimation errors on the fare class level. Here we observe small errors for all fare classes and especially very low values for frequently used classes. Further, the estimation method shows a very good computational behavior; the EM method converges in general within 10-15 iterations. This makes the algorithm feasible for practitioners. Overall, we find that the choice-set model gives a very close approximation of the real underlying customer choice behavior, and that the estimation method can be successfully implemented in real-world applications. Demand information on choice-set level provides the revenue manager with detailed information on the price elasticity and the choice preferences of his customers. This information is crucial for any form of pricing or booking control.

Chapter 4

An Advanced Choice-set Demand Estimation Method

This chapter deals with an in-depth analysis of the choice-set demand unconstraining problem, as described in Chapter 2. A special emphasis is put on the twofold incomplete data knowledge. Namely first, the incomplete knowledge occurring when demand is not observable due to product unavailabilities. And second, in case of sales realization the actual contributing choice-set/ customer group which created the sale is not exactly known when multiple choice-sets overlap. We start the analysis in Section 4.1 by concentrating on the complete knowledge case and derive some helpful properties. The results are then extended to the incomplete knowledge problem in Section 4.2. Section 4.3 summarizes our findings. The chapter concludes with an advanced demand unconstraining algorithm, stated in pseudo code in Section 4.4. The new algorithm is used in the choice-set demand estimation problems in the remainder of the book.

Remember that the demand is assumed to follow an inhomogeneous Poisson process and the demand rate per choice-set c is given by

$$\lambda_c(t) = \beta_c \mathrm{e}^{\alpha_c t}, \tag{4.1}$$

for time stage t in the booking horizon $t = 1, \ldots, T$ and some parameter α and β depending on choice-set c. The Poisson distribution is only defined for positive realizations, so λ_c is not defined for $\beta_c < 0$.

4.1 Complete Data Knowledge

For the parameter estimation we use the maximum likelihood method. The likelihood function for choice-set c is

$$L_c(\alpha_c, \beta_c) = \prod_{t=1}^{T} P(S_c(t)|\lambda_c(t)), \tag{4.2}$$

where $S_c(t)$ denotes the number of sales corresponding to choice-set c at time stage t. Remember that a sale corresponding to a certain choice-set will always materialize in the most preferred available class/ product. We will denote with "c open at t" the fact that at least one product in c was open for bookings at time t. So we obtain an expression for probabilities by

$$P(S_c(t)|\lambda_c(t)) = \begin{cases} \frac{\exp(-\lambda_c(t))\lambda_c(t)^{S_c(t)}}{(S_c(t))!} & \text{, if } c \text{ open at } t \\ 1 & \text{, else.} \end{cases} \tag{4.3}$$

Since we are interested in the maximum of the likelihood function we will examine the log-likelihood function

$$\begin{aligned} \mathcal{L}_c(\alpha_c, \beta_c) &= \log\left(L_c(\alpha_c, \beta_c)\right) \\ &= \sum_{t=1}^{T} \log(P(S_c(t)|\lambda_c(t))) \\ &= \sum_{t=1}^{T} \Big(S_c(t)(\log(\beta_c) + \alpha_c t) - \beta_c \mathrm{e}^{\alpha_c t} - \log(S_c(t)!)\Big)\mathbb{I}_{\{U(c,t)>0\}}, \end{aligned} \tag{4.4}$$

where $\mathbb{I}$ denotes the indicator function, $U(c,t)$ returns the most preferred class in choice-set c at time t ($U(c,t) = 0$ means no available class in c at time t). The gradient and Hessian are given by

$$\nabla \mathcal{L}_c = \sum_{t=1}^{T} \begin{pmatrix} S_c(t)t - \beta_c t \mathrm{e}^{\alpha_c t} \\ \frac{S_c(t)}{\beta_c} - \mathrm{e}^{\alpha_c t} \end{pmatrix} \mathbb{I}_{\{U(c,t)>0\}} \tag{4.5}$$

$$H(\mathcal{L}_c) = \sum_{t=1}^{T} \begin{pmatrix} -\beta_c t^2 \mathrm{e}^{\alpha_c t} & -t\mathrm{e}^{\alpha_c t} \\ -t\mathrm{e}^{\alpha_c t} & -\frac{S_c(t)}{\beta_c^2} \end{pmatrix} \mathbb{I}_{\{U(c,t)>0\}}. \tag{4.6}$$

Let us denote with $\mathcal{T}_c$ the set of all time stages t with $U_{c,t} > 0$. If there exist no time stages t with $U_{c,t} > 0$, i.e., the choice-set is not observable, the choice-set will not be considered in the estimation. In the case that there exists only one observable time stage t, i.e., $|\mathcal{T}_c| = 1$, we simply set $\hat{\alpha}_c = 0$ and $\hat{\beta}_c = S_c(t)$. Let us compute the gradient and the Hessian at the corresponding point

$$\nabla \mathcal{L}_c(\hat{\alpha}_c, \hat{\beta}_c) = \begin{pmatrix} S_c(t)t - \hat{\beta}_c t \\ \frac{S_c(t)}{\hat{\beta}_c} - 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{4.7}$$

$$H\big(\mathcal{L}_c(\hat{\alpha}_c, \hat{\beta}_c)\big) = \begin{pmatrix} -\hat{\beta}_c t^2 & -t \\ -t & -\frac{1}{\hat{\beta}_c} \end{pmatrix}. \tag{4.8}$$

We further check for the negative definiteness of the Hessian at our stationary point, by computing

$$(x\ y)\begin{pmatrix} -\hat{\beta}_c t^2 & -t \\ -t & -\frac{1}{\hat{\beta}_c} \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix} = -\left(xt\sqrt{\hat{\beta}_c} + \frac{y}{\sqrt{\hat{\beta}_c}}\right)^2, \tag{4.9}$$

for some point $(x \quad y)$ with $y > 0$. Hence, for $\hat{\beta}_c > 0$ we have negative definiteness of the Hessian, in other words, $(\hat{\alpha}_c, \hat{\beta}_c)$ is a maximizer of the log-likelihood function. In case of $|\mathcal{T}_c| = 1$ and $S_c(t) = \hat{\beta}_c = 0$, we assume by common sense that the probability of a zero outcome is maximized by setting the Poisson rate to zero. In the remainder, we will consider the general situation with $|\mathcal{T}_c| \geq 2$. If $S_c(t) = 0$ for all observed time stages t, we simply set α_c and β_c to zero, by a similar reasoning that the probability of only zero outcomes is maximized by the zero demand rate. We like to exclude such trivial and not interesting process with $\lambda_c = 0$ from the investigation and assume in the remainder that at least one sales observation is strictly greater than zero and this implies $\beta_c > 0$.
The following proposition will help us in the maximization of our log-likelihood function.

Proposition 4.1. *The negative log-likelihood function is unimodal in* $\mathbb{R} \times \mathbb{R}_+$.

Proof. We will use Theorem 50 from Demidenko (2004) *Criterion of Unimodality: Let $F(u)$ be a twice differentiable function of $u \in \mathbb{R}^n$ such that $\|u\| \to \infty$ implies that $F(u) \to \infty$. If at each point where the gradient is zero, the Hessian is positive definite, then F has a unique minimum on $\mathbb{R}^n$, i.e. the function is unimodal.*
We observe that $\|(\alpha_c, \beta_c)\| \to \infty$ implies that the Poisson rate $\lambda_c \to \infty$ or 0, this implies for the likelihood function (4.2) $L_c(\alpha_c, \beta_c) \to 0$, since any sales realization $S_c(t) < \infty$ and at least one is greater than zero. Hence, this implies that $-\mathcal{L}_c(\alpha_c, \beta_c) \to \infty$. It remains to show that the Hessian of the negative log-likelihood function is positive definite at all stationary points, i.e. where $\nabla(-\mathcal{L}_c) = 0$. To check positive definiteness of a matrix, we have to verify that all determinants of its leading principal minors are positive. By differentiating $-\mathcal{L}_c(\alpha, \beta)$, see (4.4), we easily conclude that the gradient is given by $\nabla(-\mathcal{L}_c) = -\nabla\mathcal{L}_c$ and from that we can compute the Hessian by $H(-\mathcal{L}_c) = -H(\mathcal{L}_c)$ and obtain

$$H(-\mathcal{L}_c) = \sum_{t \in \mathcal{T}_c} \begin{pmatrix} \beta_c t^2 \mathrm{e}^{\alpha_c t} & t\mathrm{e}^{\alpha_c t} \\ t\mathrm{e}^{\alpha_c t} & \frac{S_c(t)}{\beta_c^2} \end{pmatrix}.$$

The determinant of the first leading principal minor is $\sum_{t\in\mathcal{T}_c} \beta_c t^2 \mathrm{e}^{\alpha_c t} > 0$, since $\beta_c > 0$. For the second leading principal minor, which coincides with the Hessian, we need the condition that the gradient is zero.

$$\sum_{t\in\mathcal{T}_c} \left(S_c(t)t - \beta_c t\mathrm{e}^{\alpha_c t}\right) = 0 \tag{4.10}$$

$$\sum_{t\in\mathcal{T}_c} \left(\frac{S_c(t)}{\beta_c} - \mathrm{e}^{\alpha_c t}\right) = 0. \tag{4.11}$$

The determinant of the second leading principal minor is given by

$$\left(\sum_{t\in\mathcal{T}_c} t^2\mathrm{e}^{\alpha_c t}\right)\left(\sum_{t\in\mathcal{T}_c} \frac{S_c(t)}{\beta_c}\right) - \left(\sum_{t\in\mathcal{T}_c} t\mathrm{e}^{\alpha_c t}\right)^2. \tag{4.12}$$

Using (4.11) we can eliminate $S_c(t)$ in (4.12)

$$\left(\sum_{t\in\mathcal{T}_c} t^2\mathrm{e}^{\alpha_c t}\right)\left(\sum_{t\in\mathcal{T}_c} \mathrm{e}^{\alpha_c t}\right) - \left(\sum_{t\in\mathcal{T}_c} t\mathrm{e}^{\alpha_c t}\right)^2, \tag{4.13}$$

which can further be simplified to

$$\sum_{t_1,t_2\in\mathcal{T}_c} (t_1^2 - t_1 t_2)\exp\big(\alpha_c(t_1+t_2)\big). \tag{4.14}$$

Let us for a moment concentrate an a slightly different expression

$$\sum_{t_1,t_2\in\mathcal{T}_c} \underbrace{\exp\big(\alpha_c(t_1+t_2)\big)}_{>0}\underbrace{(t_1^2 - t_1t_2 + t_2^2 - t_1t_2)}_{=(t_1-t_2)^2\geq 0}, \tag{4.15}$$

which is clearly positive for $|\mathcal{T}_c| > 1$. From symmetry we have that

$$\sum_{t_1,t_2\in\mathcal{T}_c} (t_1^2 - t_1t_2) = \sum_{t_1,t_2\in\mathcal{T}_c} (t_2^2 - t_1t_2)$$

and so we can rewrite (4.15) into

$$\sum_{t_1,t_2\in\mathcal{T}_c} \exp\big(\alpha_c(t_1+t_2)\big)2(t_1^2 - t_1t_2) > 0 \tag{4.16}$$

Dividing (4.16) by two gives that (4.14) is positive and hence the second leading principal minor is positive, by the fact that

$$2x > 0 \qquad \Rightarrow \qquad x > 0.$$

Thus we have shown unimodality of the negative log-likelihood function. $\square$

We obtain the maximum likelihood estimates for the parameter α and β for choice-set c by minimizing the negative log-likelihood function

$$(\hat{\alpha}_c, \hat{\beta}_c) = \arg\,\min_{\alpha,\beta>0} \quad -\mathcal{L}_c(\alpha, \beta) \tag{4.17}$$

The next proposition shows that problem (4.17) has a unique solution.

Proposition 4.2. *Suppose that the choice-set c is observable at all time stages, i.e.,* $|\mathcal{T}_c| = T$*. Then the following holds: The optimization problem (4.17) has a unique minimizer* $(\hat{\alpha}_c, \hat{\beta}_c)$*. Parameter* $\hat{\alpha_c}$ *corresponds to the unique root of the function*

$$f(\alpha_c) = \sum_{t=0}^{T-1} \left(\frac{\mathrm{e}^{\alpha_c t} - \mathrm{e}^{\alpha_c T}}{1 - \mathrm{e}^{\alpha_c T}} \right) - \frac{\Sigma_2}{\Sigma_1}. \tag{4.18}$$

$\hat{\beta}_c$ *can be directly computed by*

$$\hat{\beta}_c = \frac{\Sigma_1}{\sum_{t=1}^{T} \mathrm{e}^{\hat{\alpha}_c t}} = \frac{\Sigma_1 (1 - \mathrm{e}^{\hat{\alpha}_c})}{(\mathrm{e}^{\hat{\alpha}_c} - \mathrm{e}^{\hat{\alpha}_c (T+1)})}. \tag{4.19}$$

Σ_1 *and* Σ_2 *are aggregated sales numbers and defined as*

$$\Sigma_1 = \sum_{t=1}^{T} S_c(t) \qquad \text{and} \qquad \Sigma_2 = \sum_{t=1}^{T} t S_c(t). \tag{4.20}$$

Proof. At the beginning we need to derive some results for the geometric series of e^{α_c} and its derivative. The geometric series is

$$\sum_{i=1}^{n} \mathrm{e}^{\alpha_c i} = \frac{1 - \mathrm{e}^{\alpha_c (n+1)}}{1 - \mathrm{e}^{\alpha_c}} - 1 = \frac{\mathrm{e}^{\alpha_c} - \mathrm{e}^{\alpha_c (n+1)}}{1 - \mathrm{e}^{\alpha_c}}, \tag{4.21}$$

and its derivative is

$$\begin{aligned} \sum_{i=1}^{n} i \mathrm{e}^{\alpha_c i} &= \frac{-(n+1)\mathrm{e}^{\alpha_c (n+1)}(1 - \mathrm{e}^{\alpha_c}) - (1 - \mathrm{e}^{\alpha_c (n+1)})(-\mathrm{e}^{\alpha_c})}{(1 - \mathrm{e}^{\alpha_c})^2} \\ &= \mathrm{e}^{\alpha_c} \frac{1 - \mathrm{e}^{\alpha_c (n+1)} - (n+1)\mathrm{e}^{\alpha_c n}(1 - \mathrm{e}^{\alpha_c})}{(1 - \mathrm{e}^{\alpha_c})^2}. \end{aligned} \tag{4.22}$$

Next, we will show that there exists a unique stationary point of the gradient (4.5), that is, $\nabla{-\mathcal{L}_c}(\hat{\alpha}_c, \hat{\beta}_c) = 0$ and that the point $(\hat{\alpha}_c, \hat{\beta}_c)$ is uniquely defined.

Therefore, we set the gradient equal to zero, split it into the two dimensions and observe

$$0 = \sum_{t=1}^{T} \left(S_c(t)t - \beta_c t \mathrm{e}^{\alpha_c t} \right) \Leftrightarrow 0 = \Sigma_2 - \sum_{t=1}^{T} \beta_c t \mathrm{e}^{\alpha_c t} \tag{4.23}$$

$$0 = \sum_{t=1}^{T} \left(\frac{S_c(t)}{\beta_c} - \mathrm{e}^{\alpha_c t} \right) \Leftrightarrow 0 = \Sigma_1 - \sum_{t=1}^{T} \beta_c \mathrm{e}^{\alpha_c t}. \tag{4.24}$$

Reorganizing (4.24) and applying (4.21), we obtain an expression for β

$$\beta_c = \frac{\Sigma_1}{\sum_{t=1}^{T} \mathrm{e}^{\alpha_c t}} = \frac{\Sigma_1(1 - \mathrm{e}^{\alpha})}{(\mathrm{e}^{\alpha} - \mathrm{e}^{\alpha(T+1)})}. \tag{4.25}$$

From (4.23) we have

$$\Sigma_2 = \beta_c \sum_{t=1}^{T} t\mathrm{e}^{\alpha_c t}. \tag{4.26}$$

For a better visualization, we will substitute $x = \mathrm{e}^{\alpha_c}$. Combining the last equation with (4.22) and (4.25), we obtain

$$\begin{aligned} \Sigma_2 &= \frac{\Sigma_1(1-x)}{(x - x^{T+1})} \cdot x \cdot \frac{1 - x^{T+1} - (T+1)x^T(1-x)}{(1-x)^2} \\ \frac{\Sigma_2}{\Sigma_1} &= \frac{1 - x^{T+1}}{(1-x^T)(1-x)} - \frac{(T+1)x^T}{(1-x^T)} \\ &= \frac{Tx^{T+1} - (T+1)x^T + 1}{(1-x^T)(1-x)}. \end{aligned} \tag{4.27}$$

We use the expanded formulation of the divider

$$(1-x^T)(1-x) = 1 - x + x^{T+1} - x^T,$$

in order to cancel and simplify the equation

$$\begin{aligned} \frac{\Sigma_2}{\Sigma_1} &= 1 + \frac{(T-1)x^{T+1} - Tx^T + x}{(1-x^T)(1-x)} \\ &= 1 + x\frac{(T-1)x^T - Tx^{T-1} + 1}{(1-x^T)(1-x)} \\ &= 1 + x\frac{(T-1)x^T - Tx^{T-1} + 1}{(1-x^T)(1-x)} \frac{(1-x^{T-1})}{(1-x^{T-1})}. \end{aligned}$$

The same trick can be iteratively applied

$$
\begin{aligned}
\frac{\Sigma_2}{\Sigma_1} &= 1 + x\frac{(1-x^{T-1})}{(1-x^T)}\left(1 + x\frac{(T-2)x^{T-1}-(T-1)x^{T-2}+1}{(1-x^{T-1})(1-x)}\frac{(1-x^{T-2})}{(1-x^{T-2})}\right) \\
&= 1 + \frac{1}{1-x^T}\sum_{t=1}^{T-1} x^t(1-x^{T-t}) \\
&= 1 + \sum_{t=1}^{T-1}\frac{x^t - x^T}{1-x^T} \qquad \left(= 1 + \frac{1}{1-x} - \frac{1}{1-x^T} - \frac{(T-1)x^T}{1-x^T}\right) \\
&= \sum_{t=0}^{T-1}\frac{x^t - x^T}{1-x^T}.
\end{aligned} \tag{4.28}
$$

Hence, we define a function $f(x)$

$$
f(x) = \sum_{t=0}^{T-1}\frac{x^t - x^T}{1-x^T} - \frac{\Sigma_2}{\Sigma_1}. \tag{4.29}
$$

The argument $x = 1$ would result in a term $\frac{0}{0}$, which is not defined, so $x = 1$ is not contained in the domain of f. Let us inspect the limits of $f(x)$, remember that Σ_1 and Σ_2 are independent of x. The results are

$$
\begin{aligned}
\lim_{x\to 0} \quad \sum_{t=0}^{T-1}\frac{x^t - x^T}{1-x^T} - \frac{\Sigma_2}{\Sigma_1} &= \lim_{x\to 0} \quad 1 + \sum_{t=1}^{T-1}\underbrace{\frac{x^t - x^T}{1-x^T}}_{\to 0} - \frac{\Sigma_2}{\Sigma_1} \\
&= 1 - \underbrace{\frac{\Sigma_2}{\Sigma_1}}_{\geq 1} \leq 0 \\
\lim_{x\to \infty} \quad \sum_{t=0}^{T-1}\frac{x^t - x^T}{1-x^T} - \frac{\Sigma_2}{\Sigma_1} &= \lim_{x\to\infty} 1 + \underbrace{\frac{1}{1-x}}_{\to 0} - \underbrace{\frac{1}{1-x^T}}_{\to 0} - \underbrace{\frac{(T-1)x^T}{1-x^T}}_{\to -T-1} - \frac{\Sigma_2}{\Sigma_1} \\
&= T - \frac{\Sigma_2}{\Sigma_1} = T - \underbrace{\sum_{t=1}^{T} t\frac{S_c(t)}{\Sigma_1}}_{\leq T} \geq 0.
\end{aligned}
$$

So the function goes from negative values to positive values. The roots x_0 of the function $f(x)$ are corresponding to our stationary points of the gradient (4.5), by simply re-substituting

$$
\exp(\hat{\alpha_c}) = x_0 \qquad \Rightarrow \qquad \hat{\alpha_c} = \log(x_0). \tag{4.30}
$$

We will further show that there exists exactly one x_0 such that $f(x_0) = 0$. The first derivative is given by

$$\begin{aligned} f'(x) &= \sum_{t=1}^{T-1} \frac{(tx^{t-1} - Tx^{T-1})(1-x^T) + Tx^{T-1}(x^t - x^T)}{(1-x^T)^2} \\ &= \sum_{t=1}^{T-1} \frac{tx^{t-1} + (T-t)x^{t+T-1} - Tx^{T-1}}{(1-x^T)^2}. \end{aligned} \tag{4.31}$$

With the help of Lemma 4.1, we observe that all terms in the sum of $f'(x)$ are non-negative and so the function is monotone increasing.
Let us now investigate the undefined point $x = 1$, by computing the limits from the left and right hand side. First, we rewrite function f into a composition of two differentiable functions g_1 and g_2

$$f(x) = 1 + \frac{g_1(x)}{g_2(x)} - \frac{\Sigma_2}{\Sigma_1} \tag{4.32}$$

$$g_1(x) = \sum_{t=1}^{T-1} \left(x^t - x^T\right)$$

$$g_2(x) = 1 - x^T.$$

The limits of both g functions are

$$\lim_{x\to 1} g_1(x) = \lim_{x\to 1} \sum_{t=1}^{T-1} x^t - x^T = 0 \tag{4.33}$$

$$\lim_{x\to 1} g_2(x) = \lim_{x\to 1} 1 - x^T = 0. \tag{4.34}$$

Thus, we can apply l'Hospital's rule

$$\begin{aligned} \lim_{x\to 1} \frac{g_1(x)}{g_2(x)} &= \lim_{x\to 1} \frac{g_1'(x)}{g_2'(x)} \\ &= \lim_{x\to 1} \sum_{t=1}^{T-1} \frac{tx^{t-1} - Tx^{T-1}}{1 - Tx^{T-1}} \\ &= 1 + \lim_{x\to 1} \sum_{t=2}^{T-1} \frac{tx^{t-1} - Tx^{T-1}}{1 - Tx^{T-1}} \\ &= 1 + \sum_{t=2}^{T-1} \frac{t-T}{1-T}. \end{aligned} \tag{4.35}$$

Using this in the limit investigation for f, we obtain

$$\lim_{x\nearrow 1} f(x) = 1 - \frac{\Sigma_2}{\Sigma_1} + \lim_{x\nearrow 1} \frac{g_1(x)}{g_2(x)} = 2 + \sum_{t=2}^{T-1} \frac{t-T}{1-T} - \frac{\Sigma_2}{\Sigma_1} \tag{4.36}$$

$$\lim_{x\searrow 1} f(x) = 1 - \frac{\Sigma_2}{\Sigma_1} + \lim_{x\searrow 1} \frac{g_1(x)}{g_2(x)} = 2 + \sum_{t=2}^{T-1} \frac{t-T}{1-T} - \frac{\Sigma_2}{\Sigma_1}. \tag{4.37}$$

We observe that $\lim_{x\nearrow 1} f(x) = \lim_{x\searrow 1} f(x) < \infty$ and this implies that $f(x)$ is continuously extendable at $x = 1$. For all $x > 0$: $\lim_{x\to c} f(x) = f(c)$ and so f is continuous. Thus, we can apply the *intermediate value theorem*, which states that if f is a real valued continuous function on $[a, b]$ and u is a number between $f(a)$ and $f(b)$, then there exists a $c \in [a, b]$ such that $f(c) = u$. This implies that there exists x_0 on the domain of f, such that $f(x_0) = 0$. Further, since $f'(x)$ is monotone increasing, x_0 is unique.
The exponential and the logarithmic function are one-to-one mappings and therefore the expressions for α_c and β_c are uniquely defined by (4.25) and (4.30). It remains to show that the stationary point $(\hat{\alpha}_c, \hat{\beta}_c)$ of the gradient is indeed a minimum and no saddle point of the negative log-likelihood function $\mathcal{L}_c$. In the proof of Proposition 4.1, we have already shown that the determinate of both leading principle minors of $-\mathcal{L}_c$ are strictly positive. Thus, the point $(\hat{\alpha}_c, \hat{\beta}_c)$ is a minimizer of (4.17). □

Lemma 4.1. *For all $t, T \in \mathbb{N}$ such that $t \leq T$ and for all $x > 0$ holds:*

$$(T-t)x^{T+t-1} + tx^{t-1} \geq Tx^{T-1}. \tag{4.38}$$

Proof. The proof goes in two parts, first we proof the lemma for $x \in (0, 1]$ and afterwards for $x > 1$. Let us suppose the contrary holds for $x \in (0, 1]$, such that

$$\begin{aligned}(T-t)x^{T+t-1} + tx^{t-1} &< Tx^{T-1}\\ (T-t)x^{T} + t &< Tx^{T-t}.\end{aligned}$$

Moving all terms on one side, we obtain

$$0 < \underbrace{Tx^{T-t} - (T-t)x^{T}}_{=f(x)} - t.$$

Let us investigate $f(x)$ at the boundaries

$$\lim_{x\to 0} f(x) = 0 \qquad \text{and} \qquad f(1) = t.$$

We would like to show that $f(x) \leq t$ for all x in $(0, 1]$, therefore we compute the first derivative

$$f'(x) = T(T-t)x^{T-t-1} - T(T-t)x^{T-1} = T(T-t)\underbrace{\left(x^{T-t-1} - x^{T-1}\right)}_{\geq 0} \geq 0.$$

So, we can conclude

$$\begin{aligned} 0 < Tx^{T-t} - (T-t)x^T - t &= \underbrace{x^{T-t}\big(T - (T-t)x^t\big)}_{\leq t} - t \\ &\leq t - t = 0. \end{aligned}$$

That is a contradiction, so the lemma is true for $x \in (0, 1]$. Let us split both sides of (4.38) into two functions f and g

$$\begin{aligned} f(x) &= (T-t)x^{T+t-1} + tx^{t-1} \\ g(x) &= Tx^{T-1}. \end{aligned}$$

We know that $f \geq g$ for $x \in (0, 1]$ and in fact for $x = 1$ we observe $f = g$. Let us investigate both functions according to (4.38) for $x > 1$

$$\begin{aligned} f(x) &> g(x) \\ f(x) - g(x) &> 0 \\ (T-t)x^{T+t-1} + tx^{t-1} - Tx^{T-1} &> 0 \\ (T-t)x^t + \frac{t}{x^{T-t}} - T &> 0 \\ \underbrace{(T-t)\underbrace{x^t}_{>1}}_{>T-t} - \underbrace{\Big(T - \underbrace{\frac{t}{x^{T-t}}}_{<t}\Big)}_{<T-t} &> 0. \end{aligned} \tag{4.39}$$

This proofs the lemma for $x > 1$. □

4.2 Incomplete Data Knowledge

So far, we disregarded the interaction of choice-sets and supposed that we know to which choice-set a sale corresponds. Let us consider three products $p1, p2, p3$ and the two choice-sets, which are given by $c_1 = \{p3, p2\}$ and $c_2 = \{p3, p2, p1\}$. Say, we observe a sale of product $p3$. Then, we do not know if this is a realization corresponding to a customer of c_1 or c_2, since we only have information if a product was available and the corresponding number of sales.

This leads us to the following incomplete data log-likelihood function

$$\mathcal{L} = \sum_{t=1}^{T} \sum_{f=1}^{F} \log P \left[X = S(t,f) \middle| X \sim \text{Poisson}\Big(\sum_{c \in C} \mathbb{I}_{\{U(c,t)=f\}} \lambda_c(t) \Big) \right] \quad (4.40)$$

with $\mathbb{I}$ denoting the indicator function, $U(c,t)$ returns the most preferred product or class in choice-set c at time t ($U(c,t) = 0$ means that c is non-observable at time t), F represents the number of classes and $S(t,f)$ denotes the observed sales in class f at time t. From simulation, we learned that the negative log-likelihood function is in general not unimodal. Since we are usually confronted with multiple overlapping choice-sets we have a to consider $2|C|$ variables in the resulting non-convex optimization problem.

To overcome this incomplete data problem we suggest an application of the Expectation-Maximization (EM) algorithm, first published in Dempster et al. (1977) and detailed described in McLachlan and Krishnan (1997). The EM algorithm is an iterative procedure to compute maximum likelihood estimates (MLE) in situations with missing data or incomplete data knowledge. To illustrate the algorithm, let us denote the observed data vector by y, the complete data by x and unknown parameter by θ. Further, the p.d.f. of the complete data random variable is denoted by $f(x, \theta)$. Thus, the complete data log-likelihood function is given by

$$\mathcal{L}(\theta) = \log f(x, \theta).$$

The complete data MLE is obtained by

$$\theta^* = \arg\max_{\theta} \mathcal{L}(\theta).$$

The idea of the EM approach is to replace the complete data log-likelihood with the conditional expectation given the observed data y and a current estimate of θ. The procedure is therefore split into the E-step, expectation step, and the M-step, maximization step. In addition, we need some initial estimate of the unknown parameter, denoted by $\theta^{(0)}$. The algorithm is then defined at each iteration $i = 1, 2, \ldots$ by:
The E-step calculating the conditional expectation

$$Q(\theta, \theta^{(i-1)}) = \mathbb{E}[\mathcal{L}^{(i)}(\theta) | y, \theta^{(i-1)}]. \quad (4.41)$$

The M-step computing a new estimate of θ by

$$\theta^{(i)} = \arg\max_{\theta} Q(\theta, \theta^{(i-1)}). \tag{4.42}$$

The EM method satisfies that the log-likelihood function is non-decreasing during the iterations

$$\mathcal{L}(\theta^{(i+1)}) \geq \mathcal{L}(\theta^{(i)}).$$

Returning to our problem, at iteration $i = 0$, we are initially estimating the parameter separately for all choice-sets by ignoring the interaction between them. Further we use the fact that the random combination of Poisson processes gives again a Poisson process. Being in iteration $i \geq 1$ and given the previous estimates of rates $\lambda_c^{i-1}(t)$ for all times t and all choice-sets c, we compute the estimated sales corresponding to choice-set c and time stage t by

$$S_c^i(t) = \frac{\lambda_c^{i-1}(t)}{\lambda_{overlap}^{i-1}(c,t)} S(t, U(c,t)), \tag{4.43}$$

where $\lambda_{overlap}^{j}(c,t)$ denotes the sum of the estimated rates from iteration j over all choice-sets for which the most preferred available classes coincide with the one of choice-set c. Remember that $U(c,t)$ returns the most preferred class of choice-set c at time t, and so $S(t, U(c,t))$ represents the sales made in the chosen class. Since $\lambda_c > 0$ for all choice-sets c, $S_c^i(t)$ is not restricted to integer values, but in the likelihood function we work with the Poisson distribution, which only assumes integer values of S_c. We will denote the decimal place of $S_c^i(t)$ with

$$\gamma_{c,t,i} = S_c^i(t) - \left\lfloor S_c^i(t) \right\rfloor, \tag{4.44}$$

with $\lfloor x \rfloor$ denotes the *floor* operator, returning the closest integer less or equal to x. If γ takes a positive value, the algorithm is actually in between two values for $S_c(t)$, the *floor* $\lfloor S_c(t) \rfloor$ and the *ceiling* $\left\lceil S_c^i(t) \right\rceil$. We consider γ as a probability measure for both values. As mentioned before, the Poisson distribution only assumes integer values. From Johnson (2007), we know that a Poisson distributed random variable is logarithmically concave (log-concave). An example of the log-concave combination of the Poisson distribution and the equivalent convex combination of the negative logarithmic Poisson distribution are shown in Figure 4.1. A function is called log-concave if the logarithm of the function is concave. Since our approach is to minimize the negative log-likelihood function, we aim to have a convex approximation of the probability

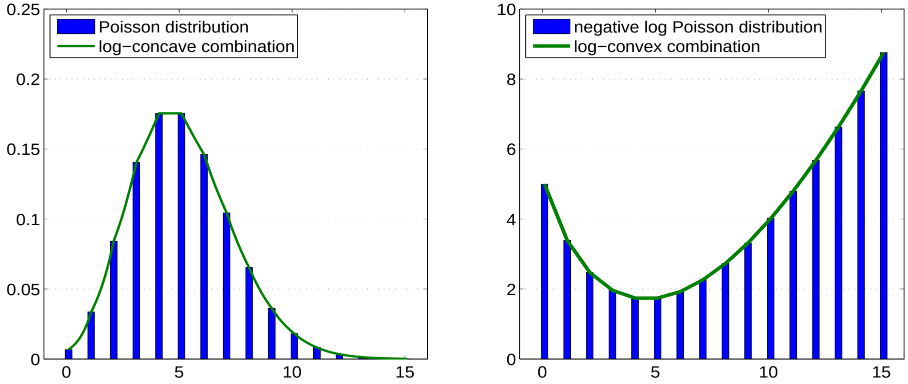

Figure 4.1. Poisson distribution with its log-concave combination and the negative log Poisson distribution and the convex combination, an example for $\lambda = 5$.

of rational sales estimates in the negative log-likelihood function. Thus, we need a concave approximation in the log-likelihood function and a log-concave approximation in the probability distribution function. Therefore, we compute the log-concave γ-combination of the probabilities of both possible integer values

$$P\big(\lceil S_c^i(t)\rceil\,|\lambda_c(t)\big)^{\gamma_{c,t,i}} P\big(\lfloor S_c^i(t)\rfloor\,|\lambda_c(t)\big)^{1-\gamma_{c,t,i}} = \frac{\mathrm{e}^{-\lambda_c(t)}\lambda_c(t)^{S_c^i(t)}}{(\lceil S_c(t)\rceil!)^{\gamma_{c,t,i}}(\lfloor S_c(t)\rfloor!)^{1-\gamma_{c,t,i}}}.$$

The probability for the estimated rational sales value is consequently given by

$$P\big(S_c^i(t)|\lambda_c(t)\big) = \begin{cases} \frac{\mathrm{e}^{-\lambda_c(t)}\lambda_c(t)^{S_c^i(t)}}{(\lceil S_c(t)\rceil!)^{\gamma_{c,t,i}}(\lfloor S_c(t)\rfloor!)^{1-\gamma_{c,t,i}}} & \text{, if } U(c,t) > 0 \\ 1 & \text{, else.} \end{cases} \tag{4.45}$$

The expected likelihood function for choice-set c at the i^{th} iteration of the EM algorithm is then

$$L_c^i(\alpha_c, \beta_c) = \prod_{t=1}^{T} P\big(S_c^i(t)|\lambda_c(t)\big). \tag{4.46}$$

Thus, the corresponding expected log-likelihood function is

$$\begin{aligned}\mathcal{L}_c^i(\alpha_c,\beta_c) &= \log\left(L_c^i(\alpha_c,\beta_c)\right)\\ &= \sum_{t=1}^{T}\log\Big(P\big(S_c^i(t)|\lambda_c(t)\big)\Big)\\ &= \sum_{t=1}^{T}\Big(S_c^i(t)(\log(\beta_c)+\alpha_c t)-\beta_c \mathrm{e}^{\alpha_c t}\\ &\quad -\log\left((\lceil S_c(t)\rceil !)^{\gamma_{c,t,i}}(\lfloor S_c(t)\rfloor !)^{1-\gamma_{c,t,i}}\right)\Big)\mathbb{I}_{\{U(c,t)>0\}} \qquad (4.47)\end{aligned}$$

for which the gradient and Hessian are given by

$$\nabla\mathcal{L}_c^i = \sum_{t=1}^{T}\begin{pmatrix} S_c^i(t)t-\beta_c t\mathrm{e}^{\alpha_c t}\\ \frac{S_c^i(t)}{\beta_c}-\mathrm{e}^{\alpha_c t}\end{pmatrix}\mathbb{I}_{\{U(c,t)>0\}} \qquad (4.48)$$

$$H(\mathcal{L}_c^i) = \sum_{t=1}^{T}\begin{pmatrix} -\beta_c t^2\mathrm{e}^{\alpha_c t} & -t\mathrm{e}^{\alpha_c t}\\ -t\mathrm{e}^{\alpha_c t} & -\frac{S_c^i(t)}{\beta_c^2}\end{pmatrix}\mathbb{I}_{\{U(c,t)>0\}}. \qquad (4.49)$$

Remember from the previous section that $\mathcal{T}_c$ denotes the set of observable time stages t of choice-set c. Choice-sets with $|\mathcal{T}_c| = 0$ are disregarded in the estimation. If $|\mathcal{T}_c| = 1$, we set $\alpha_c = 0$ and $\beta_c = S_c(t)$ for $t \in \mathcal{T}_c$. In the remainder we assume the general situation that $|\mathcal{T}_c| \geq 2$.

Next, we need to validate that Proposition 4.1 and Proposition 4.2 also hold for the incomplete-data formulation.

Proposition 4.3. *The negative of the log-likelihood function (4.47) is unimodal in* $\mathbb{R}\times\mathbb{R}_+$.

Proof. The proof of Proposition 4.1 will be used and we apply again Theorem 50 from Demidenko (2004). We have that if $\|(\alpha_c,\beta_c)\| \to \infty$ the likelihood (4.46) $L_c(\alpha_c^i,\beta_c) \to 0$ and this implies that $-\mathcal{L}_c^i(\alpha_c,\beta_c) \to \infty$. It remains to show that the Hessian of the negative log-likelihood function is positive definite at all stationary points, i.e. where $\nabla(-\mathcal{L}_c^i) = 0$. To check positive definiteness of a matrix, we have to verify that all determinants of its leading principal minors are positive.

In the remainder of the proof, we only need the gradient and the Hessian of the negative log-likelihood function. By comparing the gradient and the Hessian of the incomplete-data knowledge case ((4.48) and (4.49)) with the ones for the the complete-data knowledge case ((4.5) and (4.6)), we find that they are identical. Only the S values were in the complete-data knowledge case integer and will now also take rational values, but they are still non-negative and that is all we need. So the proof is simply a repetition of the steps in the proof of Proposition 4.1 and we omit them here. □

Proposition 4.4. *Suppose that the choice-set c is observable at all time stages, i.e., $|\mathcal{T}_c| = T$. Then the following holds: The minimization of the negative of the log-likelihood function (4.47) has a unique minimizer $(\hat{\alpha}_c, \hat{\beta}_c)$. Parameter $\hat{\alpha_c}$ corresponds to the unique root of the function*

$$f(\alpha_c) = \sum_{t=0}^{T-1} \left(\frac{\mathrm{e}^{\alpha_c t} - \mathrm{e}^{\alpha_c T}}{1 - \mathrm{e}^{\alpha_c T}} \right) - \frac{\Sigma_2}{\Sigma_1}, \tag{4.50}$$

and $\hat{\beta}_c$ can be directly computed by

$$\hat{\beta}_c = \frac{\Sigma_1}{\sum_{t=1}^{T} \mathrm{e}^{\hat{\alpha}_c t}} = \frac{\Sigma_1 (1 - \mathrm{e}^{\hat{\alpha}_c})}{(\mathrm{e}^{\hat{\alpha}_c} - \mathrm{e}^{\hat{\alpha}_c (T+1)})}. \tag{4.51}$$

Σ_1 and Σ_2 are aggregated sales numbers and defined as

$$\Sigma_1 = \sum_{t=1}^{T} S_c(t) \qquad \text{and} \qquad \Sigma_2 = \sum_{t=1}^{T} t S_c(t). \tag{4.52}$$

Proof. The proof is a full repetition of the proof for Proposition 4.2. It first shows that there exists a unique stationary point of the gradient (4.48). And afterwards shows that this point corresponds to the unique root of function f, defined by (4.50). The gradient and the Hessian are identical for both cases, except that S is now rational, but this fact is not used in the proof. All steps are omitted and we refer to the proof of Proposition 4.2. □

Algorithm 4.1. *If $|\mathcal{T}_c| < T$, we have initial estimates (α_0, β_0).*
Do Loop *for iterations $i = 1, \ldots$ over the following steps:*

1. *step: set* $\hat{S}_c^{(i)}(t) = \begin{cases} S_c(t) & \text{, if } U_c(t) > 0 \\ \beta_{i-1} \mathrm{e}^{\alpha_{i-1} t} & \text{, else} \end{cases}$

2. *step:* $\Sigma_1 = \sum_{t=1}^{T} \hat{S}_c^{(i)}(t)$ *and* $\Sigma_2 = \sum_{t=1}^{T} t \hat{S}_c^{(i)}(t)$

3. *step: find the unique root α_i of* $f(\alpha) = \sum_{t=0}^{T-1} \left(\frac{\mathrm{e}^{\alpha t} - \mathrm{e}^{\alpha T}}{1 - \mathrm{e}^{\alpha T}} \right) - \frac{\Sigma_2}{\Sigma_1}$

4. step: compute $\beta_i = \frac{\Sigma_1}{\sum_{t=1}^{T} \mathrm{e}^{\hat{\alpha}_i t}} = \frac{\Sigma_1 (1-\mathrm{e}^{\hat{\alpha}_i})}{(\mathrm{e}^{\hat{\alpha}_i} - \mathrm{e}^{\hat{\alpha}_i (T+1)})}$

Until (α_i, β_i) *converged*

Note that in case of $|\mathcal{T}_c| = T$, we have convergence of the algorithm after the first iteration. This results from the uniqueness of the solution (α^*, β^*) in Proposition 4.2. If $|\mathcal{T}_c| < T$, we assume to have some rough estimates (α'_c, β'_c) such that we can compute

$$\hat{S}'_c(t) = \begin{cases} S_c(t) & \text{, if } U_c(t) > 0 \\ \beta'_c \mathrm{e}^{\alpha'_c t} & \text{, if } U_c(t) = 0 \end{cases}, \tag{4.53}$$

and apply Algorithm 4.1 to compute $(\hat{\alpha}_c, \hat{\beta}_c)$.

Proposition 4.5. *Suppose that the choice-set c is not observable at all time stages, i.e.,* $|\mathcal{T}_c| < T$*. Then the following holds: For some starting values* (α_0, β_0) *the previous Algorithm 4.1 converges to some values* (α^*, β^*)*. And* (α^*, β^*) *is a minimizer of the negative of the log-likelihood function (4.47).*

Proof. Algorithm 4.1 describes a separate EM algorithm for the incomplete data problem with $2 \leq |\mathcal{T}_c| < T$. Sales values $S_c(t)$ are given for $t \in \mathcal{T}_c$ and unknown for $t \in \bar{\mathcal{T}}_c$, with $\bar{\mathcal{T}}_c$ denoting the complement of $\mathcal{T}_c$, i.e., the set of non-observed time stages. We are aiming to compute parameters (α^*, β^*), which maximize the complete data log-likelihood function

$$\begin{aligned} (\alpha^*, \beta^*) &= \arg\max_{\alpha,\beta>0} \quad \tilde{\mathcal{L}}_c(\alpha, \beta) \\ &= \arg\max_{\alpha,\beta>0} \quad \sum_{t=1}^{T} \log\Big(P\big(X = S_c(t) | X \sim \text{Poisson}(\lambda(t) = \beta \mathrm{e}^{\alpha t})\big)\Big). \end{aligned} \tag{4.54}$$

Having some parameter estimates (α', β'), we are able to compute estimated sales values for the unknown time stages, as in step 1 of the algorithm, by

$$\hat{S}_c(t, \alpha', \beta') = \begin{cases} S_c(t) & \text{, if } U_c(t) > 0 \\ \beta' \mathrm{e}^{\alpha' t} & \text{, else} \end{cases}. \tag{4.55}$$

Hence, we can compute the conditional expectation of the complete data log-likelihood function

$$\mathbb{E}[\tilde{\mathcal{L}}(\alpha, \beta) | \hat{S}_c] = \sum_{t=1}^{T} \log\Big(P\big(X = \hat{S}_c(t) | X \sim \text{Poisson}(\lambda(t) = \beta \mathrm{e}^{\alpha t})\big)\Big). \tag{4.56}$$

$\hat{S}_c$ will generally take rational values and is not restricted to integer numbers. This is similar to our problem with the estimated sales corresponding to a certain choice-set, equation (4.43). As in the previous case (4.45), we will work with the log-concave combination of the Poisson probabilities of the closest lower and upper integer realization for all $t = 1, \ldots, T$ by

$$P(\hat{S}_c(t)|\lambda(t)) = P\left(\left\lceil \hat{S}_c(t) \right\rceil |\lambda_c(t)\right)^{\gamma_{c,t}} P\left(\left\lfloor \hat{S}_c(t) \right\rfloor |\lambda_c(t)\right)^{1-\gamma_{c,t}}, \tag{4.57}$$

with $\gamma_{c,t}$ denoting the decimal place of $\hat{S}_c(t)$. Our conditional log-likelihood function (4.56) is equivalent to the log-likelihood function (4.47) and by Proposition 4.3 also unimodal. Since $\hat{S}_c(t)$ is defined for all $t = 1, \ldots, T$, we can straight forward compute the unique minimizer of the negative log-likelihood function, i.e., equivalent to the maximizer of the log-likelihood function. Further, since the conditional expectation of the log-likelihood function is clearly continuous in (α, β) and (α', β'), the EM algorithm will converge monotonically to some value $\tilde{\mathcal{L}}(\alpha^*, \beta^*)$ for some stationary point (α^*, β^*). This follows directly from Theorem 3.2 in McLachlan and Krishnan (1997), and is based on the main convergence theorem of the generalized EM (GEM) algorithm given by Wu (1983).

It remains to show that the point (α^*, β^*) is indeed a minimizer of the negative of the log-likelihood function (4.47). Let us define two functions

$$A(\mathcal{V}, S, \alpha, \beta) = \sum_{t \in \mathcal{V}} \left(S(t) - \beta e^{\alpha t}\right) \tag{4.58}$$

$$B(\mathcal{V}, S, \alpha, \beta) = \sum_{t \in \mathcal{V}} \left(tS(t) - t\beta e^{\alpha t}\right). \tag{4.59}$$

Per definition of $\hat{S}_c$, we have

$$A(\bar{\mathcal{T}}_c, \hat{S}_c^{(i-1)}, \alpha_{i-1}, \beta_{i-1}) = 0 \qquad \text{and} \qquad B(\bar{\mathcal{T}}_c, \hat{S}_c^{(i-1)}, \alpha_{i-1}, \beta_{i-1}) = 0. \tag{4.60}$$

From the proof of Proposition 4.2 we obtain for (α_i, β_i) that

$$A(\mathcal{T}_c \cup \bar{\mathcal{T}}_c, \hat{S}_c^{(i-1)}, \alpha_i, \beta_i) = 0 \qquad \text{and} \qquad B(\mathcal{T}_c \cup \bar{\mathcal{T}}_c, \hat{S}_c^{(i-1)}, \alpha_i, \beta_i) = 0. \tag{4.61}$$

This implies in case of convergence, i.e. $(\alpha_i, \beta_i) = (\alpha_{i-1}, \beta_{i-1}) = (\alpha^*, \beta^*)$, that

$$A(\mathcal{T}, \hat{S}_c^{(i-1)}, \alpha^*, \beta^*) = 0 \qquad \text{and} \qquad B(\mathcal{T}, \hat{S}_c^{(i-1)}, \alpha^*, \beta^*) = 0. \tag{4.62}$$

Thus, $\nabla - \mathcal{L}_c(\alpha^*, \beta^*) = 0$ and we know that the negative log-likelihood function is unimodal. As shown in the proof of Proposition 4.3, the Hessian of the negative log-likelihood function is positive definite at all roots of the gradient. This implies that (α^*, β^*) can not be a saddle point of the negative of the log-likelihood function (4.47) and is indeed a minimizer.

□

The pseudo code of the advanced EM algorithm, for the choice-set demand rate estimation, is stated in Section 4.4.

Theorem 4.1. *All limit points of any instance ($\alpha_c^{(i)}, \beta_c^{(i)}$, $i = 1, 2, \ldots$) of the choice-set EM-Algorithm are stationary points of the corresponding incomplete date log-likelihood and converge monotonically to some value $L(\alpha_c^*, \beta_c^*)$ for some stationary point (α_c^*, β_c^*).*

Proof. The results follow directly from Theorem 3.2 in McLachlan and Krishnan (1997), which itself is based on the main convergence theorem of the generalized EM (GEM) algorithm given by Wu (1983). Theorem 3.2 only asks to ensure that the conditional expectation $Q(\alpha_c, \beta_c|\alpha_c^{(i-1)}, \beta_c^{(i-1)})$ in the E-step (4.41) is continuous in both (α_c, β_c) and $(\alpha_c^{(i-1)}, \beta_c^{(i-1)})$, for all iterations i. S_c^i in (4.43) is clearly continuous in $(\alpha_c^{(i-1)}, \beta_c^{(i-1)})$. Also the log-concave combination of both integer realizations of S_c^i given by $P(\lceil S_c^i(t) \rceil)^{\gamma_{c,t,i}} \cdot P(\lfloor S_c^i(t) \rfloor)^{1-\gamma_{c,t,i}}$ in (4.45) is obviously continuous in λ^{i-1} and so in $(\alpha_c^{(i-1)}, \beta_c^{(i-1)})$. Further, it is straightforward to see that the expected log-likelihood (4.47) is continuous in (α_c, β_c). Also the conditional expected log-likelihood function (4.56) within the inner EM algorithm (Algorithm 4.1) is continuous in the old and new parameter. Hence, the conditions of Theorem 3.2 in McLachlan and Krishnan (1997) are satisfied and that proofs the theorem. □

The stationary point (α_c^*, β_c^*), is not guaranteed to be a global or even local maximum. McLachlan and Krishnan (1997) show in section 3.6 examples where the algorithm converges to saddle points or even to a local minimum. However, the EM algorithm is known to be very robust and was successfully applied in many incomplete data problems. Besides, as pointed out by Vulcano et al. (2011), the drawback that the EM algorithm does not guarantee a convergence to a global maximum is shared with any other standard non-linear optimization method for MLE on incomplete data. Wu (1983) mentions that the convergence of the log-likelihood values $L^{(i)}(\alpha_c, \beta_c)$ does not automatically also imply a convergence of $(\alpha_c^{(i)}, \beta_c^{(i)})$. We check the numerical convergence of the parameter points within the M-step and observed in our experiments that the sequence of parameters $(\alpha_c^{(i)}, \beta_c^{(i)})$ converged in almost all cases.

The maximization, i.e., minimization of the negative log-likelihood function, in the M-step of the EM algorithm reduces now to finding the root of function f, as defined in (4.50). An example of function f for sales values $S = (4, 5, 1, 2, 1)$, for $t = 1, \ldots, 5$, is shown in Figure 4.2. We suggest an application of New-

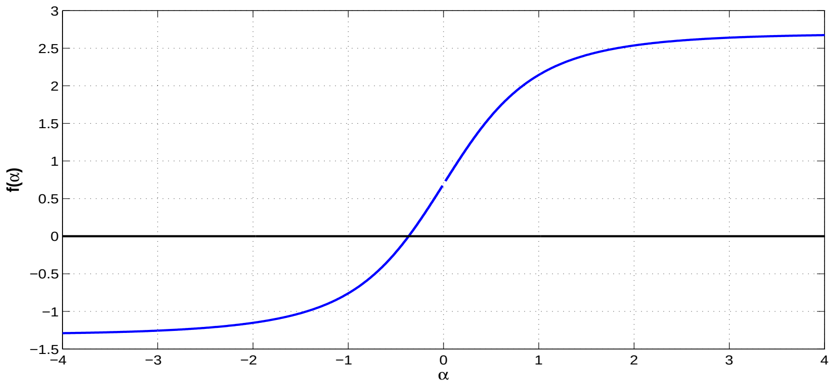

Figure 4.2. Example plot of function (4.50).

ton's method, also called Newton-Raphson method, for finding the root of the function. The α formulation of f and its derivatives are given by

$$f(\alpha) = \sum_{t=0}^{T-1} \left(\frac{\mathrm{e}^{\alpha t} - \mathrm{e}^{\alpha T}}{1 - \mathrm{e}^{\alpha T}} \right) - \frac{\Sigma_2}{\Sigma_1} \tag{4.63}$$

$$f'(\alpha) = \sum_{t=0}^{T-1} \frac{t\mathrm{e}^{\alpha t} + (T-t)\mathrm{e}^{\alpha(t+T)} - T\mathrm{e}^{\alpha T}}{(1 - \mathrm{e}^{\alpha T})^2} \tag{4.64}$$

$$\begin{aligned} f''(\alpha) = \sum_{t=0}^{T-1} \Bigg(& \frac{\big(t^2\mathrm{e}^{\alpha(t-1)} + (T^2 - t^2)\mathrm{e}^{\alpha(t+T-1)} - T^2\mathrm{e}^{\alpha(T-1)}\big)(1 - \mathrm{e}^{\alpha T})^2}{(1 - \mathrm{e}^{\alpha T})^4} \\ & - \frac{\big(2T\mathrm{e}^{\alpha(2T-1)} - 2T\mathrm{e}^{\alpha(T-1)}\big)\big(t\mathrm{e}^{\alpha t} + (T-t)\mathrm{e}^{\alpha(t+T)} - T\mathrm{e}^{\alpha T}\big)}{(1 - \mathrm{e}^{\alpha T})^4} \Bigg) \end{aligned} \tag{4.65}$$

The Newton-step, calculating the $(n+1)^{th}$ approximation for the root of f is given by

$$\alpha_{n+1} = \alpha_n - \frac{f(\alpha_n)}{f'(\alpha_n)}. \tag{4.66}$$

The iteration stops when $f(\alpha_n)$ is sufficient close to zero or when we reach a maximum iteration limit.

As outlined in Press et al. (2007) and Deuflhard and Hohmann (2002), the big advantage of Newton's method is the rate of convergence in a neighborhood of the root. Suppose $\hat{\alpha}$ is the root of f, then the Taylor expansion of $f(\hat{\alpha})$ at $\alpha_n = \hat{\alpha} - \epsilon_n$ for some distance ϵ_n is

$$0 = f(\hat{\alpha}) = f(\alpha_n) + \epsilon_n f'(\alpha_n) + \underbrace{\epsilon_n^2 \frac{f''(\bar{\alpha})}{2!}}_{=R_1}, \tag{4.67}$$

where R_1 represents the Lagrange form of the expansion remainder with $\bar{\alpha}$ being between $\hat{\alpha}$ and α_n. We divide the expression by $f'(\alpha_n)$ and obtain

$$\frac{f(\alpha_n)}{f'(\alpha_n)} + \underbrace{\epsilon_n}_{=\hat{\alpha}-\alpha_n} = -\epsilon_n^2 \frac{f''(\bar{\alpha})}{2f'(\alpha_n}. \tag{4.68}$$

Remembering the Newton step (4.66), we can replace α_n on the left hand side of the equation

$$\underbrace{\alpha - \alpha_{n+1}}_{=\epsilon_{n+1}} = -\epsilon_n^2 \frac{f''(\bar{\alpha})}{2f'(\alpha_n}. \tag{4.69}$$

Taking the absolute values, we obtain

$$|\epsilon_{n+1}| = \frac{|f''(\bar{\alpha}))|}{2|f'(\hat{\alpha_n})|}\epsilon_n^2. \tag{4.70}$$

So, we have a quadratic convergence of the Newton method within an interval $\mathcal{U} = [\hat{\alpha} - \epsilon_0, \hat{\alpha} + \epsilon_0]$ for which holds

1. $f'(\alpha) \neq 0$ for all $\alpha \in \mathcal{U}$
2. $f''(\alpha) < \infty$ for all $\alpha \in \mathcal{U}$
3. the start value α_0 is sufficiently close to $\hat{\alpha}$.

We can rewrite (4.70) into

$$|\epsilon_{n+1}| \leq \underbrace{\left(\frac{\max_{\alpha\in\mathcal{U}} f''(\alpha)}{\min_{\alpha\in\mathcal{U}} 2f'(\alpha)}\right)}_{=K} \epsilon_n^2. \tag{4.71}$$

Defining $\delta_n = K\epsilon_n$ and by the last inequality we observe

$$\delta_{n+1} \leq KK\epsilon_n^2 = \delta_n^2 \qquad \Rightarrow \qquad \delta_n \leq \delta_0^{2n} = (K\epsilon_0)^{2n}. \tag{4.72}$$

This leads us to a reformulation of condition 3: to require $K|\epsilon_0| < 1$. Consequently, our sequence $\{\delta_n\}$ is a zero-sequence fulfilling all inequalities above. Let us visualize this convergence interval for our example, shown in Figure 4.2. The root of f is attained at $\hat{\alpha} = -0.3658$ and let us define the following function

$$\delta(\alpha) = \frac{\max_{\xi \in [\hat{\alpha}-|\hat{\alpha}-\alpha|, \hat{\alpha}+|\hat{\alpha}-\alpha|]} f''(\xi)}{\min_{\xi \in [\hat{\alpha}-|\hat{\alpha}-\alpha|, \hat{\alpha}+|\hat{\alpha}-\alpha|]} 2f'(\xi)} |\hat{\alpha} - \alpha|. \tag{4.73}$$

The δ function for our example is shown in Figure 4.3. The convergence interval for the example is $\mathcal{U} = [-0.9, 0.16]$, which is relatively large.

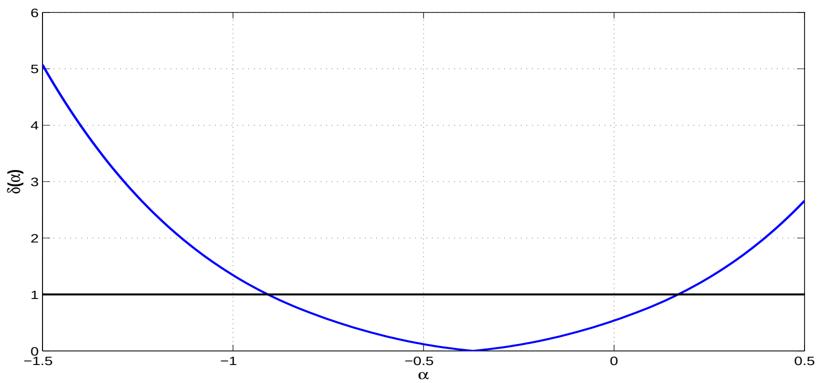

Figure 4.3. δ function corresponding to the function shown in Figure 4.2.

For starting values outside the interval $\mathcal{U}$, the convergence rate can be considerably lower, and the convergence itself is not guaranteed. Therefore, the Newton method is only recommended if the neighborhood of the root is known with some certainty. For other cases, when the approximated position of the root is completely unknown, we recommend the bisection method. The method converges only linearly, but is guaranteed to converge to the root of the underlying function f, if f is continuous on an interval $[a, b]$ and $f(a), f(b)$ have opposite signs. In the proof of Proposition 4.2, we have already shown that our function f (4.50) fulfills $\lim_{\alpha \to -\infty} f(\alpha) < 0$ and $\lim_{\alpha \to \infty} f(\alpha) > 0$. Further, f is continuous on it's domain. It remains to choose the initial values a and b, such that the respective function values are negative and positive. The algorithm is very intuitive and takes the following steps:

1. set $a = a_0$ and $b = b_0$,
2. test if interval $[a, b]$ contains a root $\big(f(a) < 0$ and $f(b) > 0\big)$, if yes continue,
3. test if $b - a < \epsilon$, if yes stop and return as approximation of the root either a or b,
4. divide $[a, b]$ into two equal sized intervals $[a, a + \frac{b-a}{2}]$ and $[a + \frac{b-a}{2}, b]$ and go to step 2 with each interval,

where ϵ defines the stopping criteria of the algorithms.

4.3 Summary

In this chapter, we analyzed the choice-set unconstraining problem. We started with the complete data knowledge problem and later extending the obtained property results to the incomplete data knowledge situation. We find that the minimization of the negative log-likelihood function can be replaced by the problem of finding the unique root of a certain function. The resulting EM algorithm is displayed in pseudo code in the following section. The new proposed advanced choice-set unconstraining algorithm overcomes the rounding problem of the algorithm proposed in Chapter 2. Further, the new algorithm does not require any advanced mathematical programming software, since it is only based on Newton's method or the Bisection method. Hence, the new algorithm can be straightforward coded in any common programming language and is therefore easily implementable into any RM system.

4.4 Algorithm

Pseudo code of the Advanced Choice-set Estimation Algorithm:
Initialization: with estimates $(\hat{\alpha}_c, \hat{\beta}_c)$ for all choice-sets c with $|\mathcal{T}_c| < T$

For all $c \in C$
  For all $t = 1, \ldots, T$

$$S_c^{(0)}(t) = \begin{cases} S(t, U(c,t)) & \text{, if } U(c,t) > 0 \\ \hat{\beta}_c \exp(\hat{\alpha}_c t) & \text{, if } U(c,t) = 0 \end{cases}$$

  end for
  If $|\mathcal{T}_c| \geq 2$: Run Algorithm 4.1 to compute $\alpha_c^{(i)}$ and $\beta_c^{(i)}$
  If $|\mathcal{T}_c| = 1$: $\alpha_c^{(i)} = 0$ and $\beta_c^{(i)} = S_c^0$
  $\lambda_c^{(0)}(t) = \beta_c^{(0)} \exp(\alpha_c^{(0)} t)$
end for

Iteration Loop $i = 1, ...$

Expectation-step:
For all $c \in C$
  For all $t = 1, ..., T$

$$S_c^i(t) = \begin{cases} \frac{\lambda_c^{(i-1)}(t)}{\lambda_{overlap}^{(i-1)}(c,t)} S(t, U(c,t)) & \text{, if } U(c,t) > 0 \\ \beta_c^{(i-1)} \exp(\alpha_c^{(i-1)} t) & \text{, if } U(c,t) = 0 \end{cases}$$

$$\gamma_{c,t,i} = S_c^i(t) - \lfloor S_c^i(t) \rfloor$$

$$P\big(S_c^i(t) | \lambda_c(t)\big) = \begin{cases} \frac{e^{-\lambda_c(t)} \lambda_c(t)^{S_c^i(t)}}{(\lceil S_c(t) \rceil !)^{\gamma_{c,t,i}} (\lfloor S_c(t) \rfloor !)^{1-\gamma_{c,t,i}}} & \text{, if } U(c,t) > 0 \\ 1 & \text{, if } U(c,t) = 0 \end{cases}$$

  end for
  $\Sigma_1 = \sum_{t=1}^T S_c^i(t)$ and $\Sigma_2 = \sum_{t=1}^T t S_c^i(t)$
  $\mathcal{L}_c^{(i)}(\alpha_c, \beta_c) = \sum_{t=1}^T \log P\big(S_c^i(t) | \lambda_c(t)\big)$
end for

Maximization-step:
For all $c \in C$
  If $|\mathcal{T}_c| \geq 2$: Run Algorithm 4.1 to compute $\alpha_c^{(i)}$ and $\beta_c^{(i)}$
  If $|\mathcal{T}_c| = 1$: $\alpha_c^{(i)} = 0$ and $\beta_c^{(i)} = S_c^i$
  $\lambda_c^{(i)}(t) = \beta_c^{(i)} \exp(\alpha_c^{(i)} t)$
end for

Until Stopping criteria reached.

The stopping criteria could be either a maximum number of iterations or some kind of numerical convergence bound on changes in $\mathcal{L}_c^{(i)}$ between iterations, as well as on changes in $\alpha_c^{(i)}$ and $\beta_c^{(i)}$. The demand rate function for each choice-set c at time t is estimated as long as at least one class contained in c is offered at t. A no sale observation at t is regarded as a realization of the stochastic arrival process with no arrivals from all choice-sets which intersect with the set of offered classes at time t. The estimation is extrapolated over periods when no class contained in choice-set c is offered, i.e. no customer arrivals corresponding to c are observable. Notice that the EM method is independent of the demand rate function form. The exponential curve is a result from our data analysis of the airline and hotel data sets. Only the unimodality of the negative log-likelihood function needs to be checked for different demand functions, it can be easily shown that it also holds for constant or linear demand functions. Also the straight forward computation of the minimizers, as given in Proposition 4.2, needs to be adjusted for other demand functions.

Chapter 5

Comparison of Discrete Choice Models on a Hotel Market

In this chapter, we present a comparison study of different choice models from the literature and an extension of our choice-set model.

With the exception of Farias et al. (2012) and van Ryzin and Vulcano (2011), who propose a non-parametric choice model with preference lists similar to our choice-set model, most choice based models in RM use the multinomial logit model (MNL). The MNL model is a well known and widely used choice model in practice. But when we take a look into the choice modeling research, we find that the MNL is in general only used as a reference model for more advanced choice models. So, clearly there is a need for a comparison study of different choice models in a RM context. The focus of this paper is to provide such a comparison of different choice models from the literature, plus an extension of the choice-set demand model. The choice-set model is extended with the MNL to model a two level choice process, which allows to incorporate multiple alternative characteristics as well as a consideration of many choice alternatives. The first level of the choice process is modeled by the choice-set idea on groups / classes of alternatives. The choice among alternatives within the chosen group is modeled by the MNL model.

The comparison is performed on a real hotel dataset consisting of reservations of multiple hotels located in a larger city in the nations of the Benelux (Belgium, Netherlands and Luxembourg). For a specific study on hotel RM, we refer to Vinod (2004), who provides a description of steps and challenges of a hotel RM system. The choice models are estimated and compared on real reservations made in the hotel market. We only work with reservation transaction data and not individual customer information for two reasons: first, the data availability, and second, the usability of information. The information of historical product availability and the amount of sales per time periods is already available and easy to extract from corporate databases. Companies do not store many individual characteristics of customers, and usually not more than name, gender, address and payment details. Further, even with potential detailed information on the choice behavior based on individual characteristics, the question of the usability remains. In the RM setting and even more in the time of e-commerce, the seller does not know who is requesting a product. For example, from a flight request on an airline website for a certain flight from Amsterdam to New York we learn nothing about the flight purpose, e.g., whether it's business or leisure, nor anything else about the customer.

The chapter continues with an introduction of the dataset, followed by the description of the different choice models in Section 5.2. The results are reported in Section 5.3 and the final Section 5.4 presents our conclusions.

5.1 Available Dataset

We are able to perform our analysis on real market data, provided by Bookit B.V., a European short break specialist for hotels and holiday parks with a market leading position in the Netherlands. The chosen location for our test case is an economical, cultural and historical important city in the nations of the Benelux (Belgium, Netherlands and Luxembourg). The town has a population of approximately 500,000 people. Bookit holds data of 31 hotels for the given location on its website, but 98% of all reservations are made from our chosen subset of 22 hotels. The considered hotel product is a two-night hotel stay with arrival on Friday afternoon, a pure leisure product. The dataset contains all hotel offers available to the online visitors who made a reservation at one of our selected hotels. The dataset is incomplete in the sense that we have no data on offered alternatives for visitors who make no transaction. This data incompleteness is very common in company databases, since it is often impossible to collect and store data for all visiting customers. Nevertheless, having data of all alternatives offered to a reserving customer enables us to estimate the underlying decision process. The dataset covers a period of 2 years, resulting in 104 different instances of our hotel product. We considered a booking horizon of 56 days/ 8 weeks. The individual hotel offers are distinguished by: the hotel, the distance between booking and arrival day, the hotel standard measured by the five star rating system, the hotel distance to the city center and the price. The set of hotels consists of 13 3-star hotels, 8 4-star hotels and 1 5-star hotel.

Figure 5.1 shows the aggregated reservations per month and separated per hotel stars. We observe three large peaks in the first year for February, August and November. The second year is less volatile with three peaks in Jan-Feb, July and October. These peaks correspond to the main holiday seasons: winter, summer and autumn. The average offer prices per year, month and

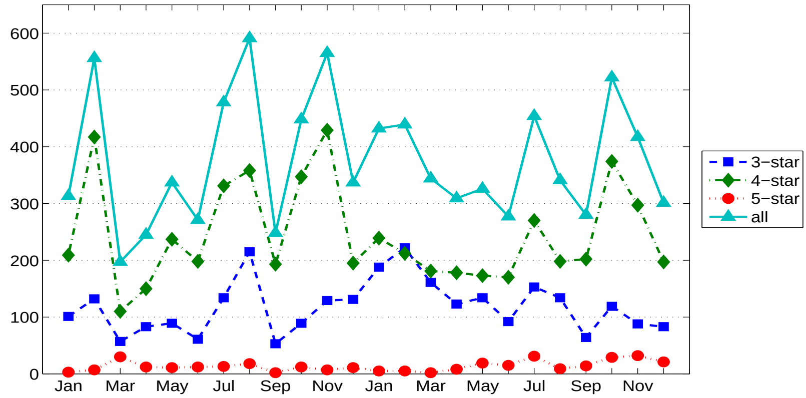

Figure 5.1. Total reservations per month and hotel stars.

star category are shown in Figure 5.2. The jump for the 5-star offers in April year 2 is very obvious. In the first 15 months the average 5-star price is 170, which is increased to 280 for the last 9 month (Apr-Dec year 2). It might be

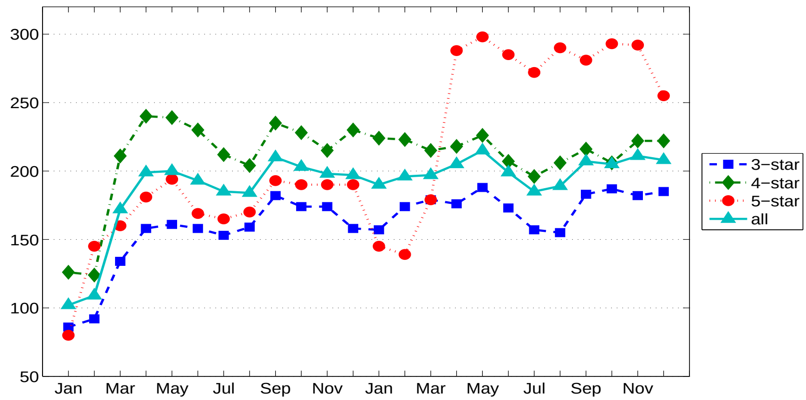

Figure 5.2. Averaged offer prices per month and hotel stars.

surprising that the 5-star hotels offers rooms for prices less than the averaged 4-star price. But since Bookit is only one sales channel, the prices are only representative for the offers at Bookit's websites and not for all reservations made at a specific hotel. Despite this fact, the Bookit data provides a clear view of the online hotel market with numerous competitive hotel offers and very high sales volumes. The 5-star sales volume in Figure 5.1 indicate that the 5-star hotel uses Bookit not as the major sales channel. We further observe that the 3-star hotels slightly increase and the 4-star hotels slightly decrease their prices between both years. It is interesting to observe also that the first year prices for January and February are very low for all categories. For the remainder we will divide our 2 year spanning dataset in four smaller sets, namely year 1 and 2 separately, and also the summer period July and August separately, for both years. Figure 5.3 shows the booking curves for all four datasets. The 56^{th} day denotes the beginning of the booking horizon and booking day 1 the day of arrival at the hotel. For comparability of the booking speed and buildup we show the booking curves in percentages of the total received reservations. We observe that most of the bookings are made close to the hotel arrival day of the product. The averaged price developments

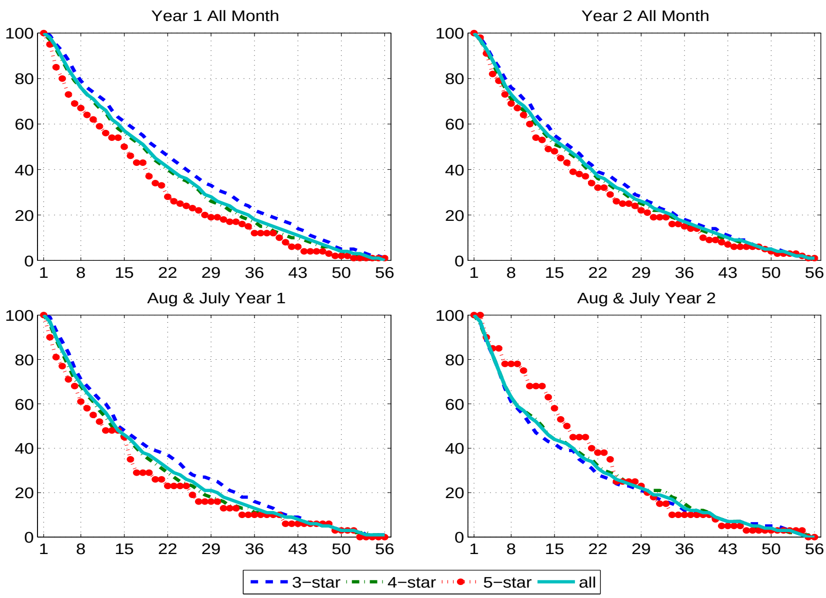

Figure 5.3. Booking curves in percent of total reservations per hotel stars.

of the offers during the booking horizon are displayed in Figure 5.4. Price changes are shown in percentages to the averaged offered price (level = 100). Interesting to observe is to compare the pricing policy of the 5-star hotel for both years. In year one, the 5-star hotel decreased the price to the end of the booking horizon, contrary to the behavior of the 3 or 4-star hotels. In the second year, we observe an adaptation of the general market practice to increase the prices towards the arrival day at the hotel.

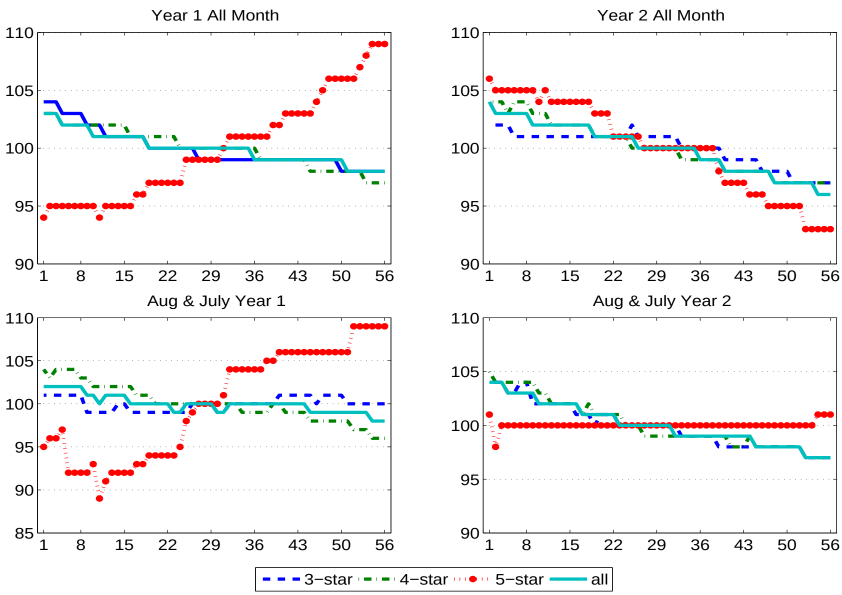

Figure 5.4. Price development within the booking horizon in percentages of the average price per hotel stars.

In the last part of our data analysis, we investigate the frequency of prices per star category over all offers and made reservations. The corresponding histograms for year 1 and 2 are displayed in Figures 5.5 and 5.6. We observe that the distributions of offer and reservations prices are not equal; we find proportionally more lower priced reservations. This higher attractiveness of low price offers for the same hotel stars clearly shows the price elasticity. We also observe that the disproportion is stronger for the 3-star products than for the 5-star products, which is in line with marketing knowledge that one observes a smaller price sensitivity for higher priced products.

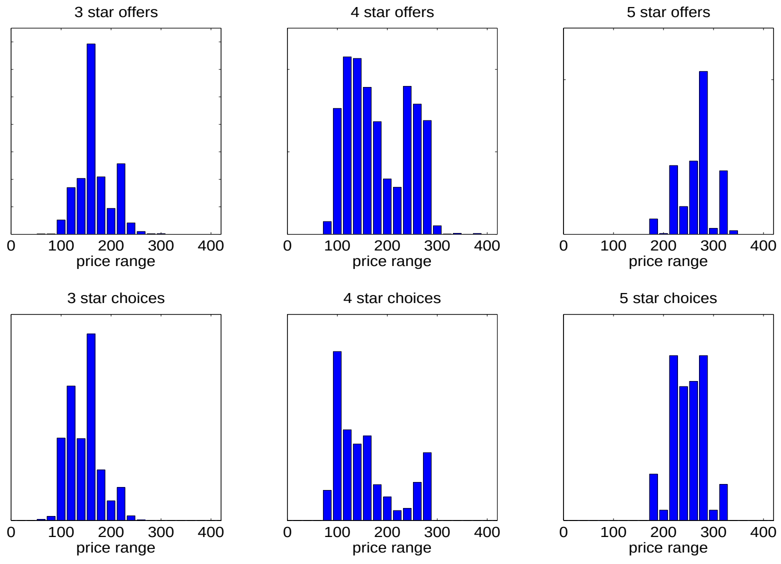

Figure 5.5. Histograms of price offers and choices for year 1.

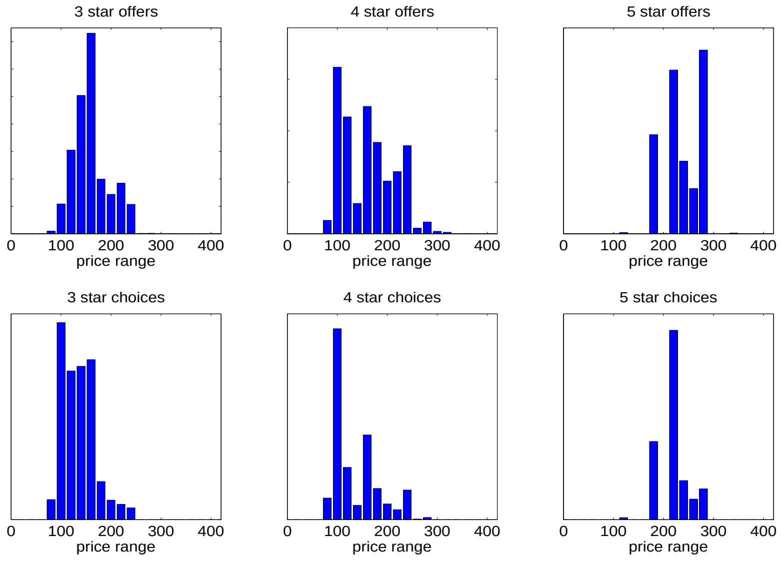

Figure 5.6. Histograms of price offers and choices for year 2.

5.2 Choice Models

In the current section we list and describe the most common discrete choice models. In every purchase decision, the individual is faced with multiple alternative offers, each having different properties. The individual has to chose one of the offered alternatives or make a non-purchase decision. Let us introduce the following notation: $\mathcal{J}$ denotes the set of alternatives ($J = |\mathcal{J}|$), m denotes number of explanatory variables, $x_j \in \mathbb{R}^m$ represents the explanatory variables of alternative $j \in \mathcal{J}$ and $\beta \in \mathbb{R}^m$ represent the choice model parameter.

5.2.1 Multinomial Logit (MNL) Model

We start with the best known and most studied choice model, the multinomial logit model. The choice process is modeled as follows: n individuals, each confronted with a set of alternatives (not necessarily the same for all individuals), are supposed to choose one of these alternatives. The choice of individual i is denoted by y_i. Each individual is supposed to make a choice purely based on utility maximization. In other words, each offer characterized by its explanatory variables comprises a certain utility for the individual. The utility of alternative j for individual i is computed by a linear function

$$U_{i,j} = \beta^\top x_j + \epsilon_{i,j}, \tag{5.1}$$

where $\epsilon_{i,j}$ is an i.i.d. extreme value random variable with zero mean, representing the unobservable decision factors and noise. Since the disturbance term is i.i.d and has mean zero, we will further suppose that the general utility of alternative j for any individual can be calculated by

$$U_j = \beta^\top x_j. \tag{5.2}$$

Choosing none out of the offered alternatives is equalized with a utility of zero. The probability that individual i chooses alternative j is computed by

$$P(y_i = j) = \frac{\exp(U_j)}{1 + \sum_{k \in \mathcal{J}_i} \exp(U_k)}, \tag{5.3}$$

where $\mathcal{J}_i$ denotes the set of alternatives which are presented to individual i and $P(y_i = j) = 0$ if alternative j is not contained in $\mathcal{J}_i$. In our dataset we can not observe non-purchase decisions, therefore equation 5.3 can be simplified into

$$P(y_i = j) = \frac{\exp(U_j)}{\sum_{k \in J_i} \exp(U_k)}. \tag{5.4}$$

To avoid numerical computation problems with very large or very small $\exp(U)$ values, we reformulate (5.4) to work with the utility differences from the chosen alternative

$$P(y_i = j) = \frac{1}{\sum_{k\in J_i} \exp(U_k - U_j)} = \frac{1}{1 + \sum_{k\in J_i, k\neq j} \exp(U_k - U_j)}. \tag{5.5}$$

The parameter estimations is done by maximum likelihood estimation (MLE) on the log-likelihood function

$$\log(L) = \sum_{i=1}^{n} \log\left(P(y_i)\right). \tag{5.6}$$

5.2.2 Latent Class MNL Model (LCM)

A detailed introduction on general Latent Class (LC) models is given by Vermunt (2010). The LC model assumes that the individuals in the choice experiment are not identical and that they can be grouped into C classes. The membership of the individual to a certain class is unknown. Thus, we define the latent variable v_i of individual i as the possible class membership. $P(v_i = c)$ denotes the probability that individual i belongs to class c. Contrary to the MNL model the parameter vector β on the explanatory variables in the utility function is not unique for all individuals in the population, but for all individuals in a certain group. Therefore, the utility of alternative j for all individuals i in class c is given by

$$U_{c,j} = \beta_c^\top x_j. \tag{5.7}$$

Hence, the choice probability for individual i is expressed by

$$\begin{aligned} P(y_i = j | v_i = c) &= \frac{\exp(U_{c,j})}{\sum_{k\in J_i} \exp(U_{c,k})}, \\ &= \frac{1}{\sum_{k\in J_i} \exp(U_{c,k} - U_{c,j})}. \end{aligned} \tag{5.8}$$

The general choice probability, without the exact class information, is defined by the a priori membership probabilities

$$P(y_i = j) = \sum_{c=1}^{C} P(v_i = c) P(y_i = j | v_i = c). \tag{5.9}$$

The parameter estimations is, similar to the MNL model, done by MLE. The log-likelihood function is

$$\log(L) = \sum_{i=1}^{n} \log\left(P(y_i)\right) = \sum_{i=1}^{n} \log\left(\sum_{c=1}^{C} P(v_i = c)P(y_i|v_i = c)\right), \tag{5.10}$$

where y_i denotes the observed choice of individual i. The estimation problem contains C parameter vectors β_c and C latent class parameter $p_c = p_{i,c} = P(v_i = c)$ with $\sum_{c=1}^{C} p_c = 1$. Since the analyst is assumed to be unable to distinguish between individuals, the class membership probabilities are equal over the whole population. An important prior problem to the estimation itself is the choice of C, the number of classes. As explained in Greene and Hensher (2003), the choice of C is very important. If the chosen C is larger than the "true" C^*, it is possible to test down to C^*. The reverse is not possible since estimates for $C < C^*$ are inconsistent.

5.2.3 Mix Logit Model (MLM)

For a fundamental discussion of the MLM, we refer to McFadden and Train (2000). They show under the assumption of mild regularity conditions that the choice probabilities of any discrete choice model, which is derived from a random utility maximization model, can be arbitrarily close approximated by a MLM model. Further, they describe a Maximum Simulated Likelihood and a Simulated Moments method to estimate MLM models. A brief discussion of the MLM is also found in Greene and Hensher (2003).
As in the LCM methodology, the MLM models assumes that the individual are not identical. The parameter vector β is assumed to be continuously distributed on the population of all individuals i

$$\beta_i = \beta + \Lambda\zeta. \tag{5.11}$$

with $\beta \in \mathbb{R}^m$ denoting the mean parameter coefficients, $\Lambda \in \mathbb{R}^{m\times r}$ factor loadings and $\zeta \in \mathbb{R}^r$ the random vector of independently but not necessarily identical distributed factor levels with density $f(\zeta)$. The choice probability of alternative j for individual i is given by

$$P(y_i = j) = \int g_i(j|\beta + \Lambda\zeta)f(\zeta|\Lambda)\ \mathrm{d}\zeta, \tag{5.12}$$

where $g_i(j|\beta)$ represents the logit function

$$g_i(j|\beta) = \frac{\exp(\beta^\top x_j)}{\sum_{k\in J_i} \exp(\beta^\top x_k)}, \\ = \frac{1}{\sum_{k\in J_i} \exp\left(\beta^\top (x_k - x_j)\right)}. \tag{5.13}$$

The integral in (5.12) is generally intractable and so the usual parameter estimation approach based on maximum likelihood is not feasible. But, by means of simulation it is possible to compute a simulated log-likelihood function L_s

$$\log(L_s) = \sum_{i=1}^{n} \log\left(P_s(y_i)\right). \tag{5.14}$$

The simulated choice probabilities are computed by S independent draws of ζ_s from $f(\zeta)$ by

$$P_s(y_i = j) = \frac{1}{S}\sum_{s=1}^{S} g_i(j|\beta + \Lambda\zeta_s). \tag{5.15}$$

Train (2003), chapters 6 and 10, describes in brief the Maximum Simulated Likelihood Estimation (MSLE) method and its properties. MSLE is consistent, efficient and asymptotically equivalent to MLE. We will use the Matlab package Train (2006) for the MSLE parameter estimation.

5.2.4 Nested Logit Model (NLM)

For a detailed introduction to the NLM, we refer to Train (2003) chapter 4. An interesting and widely discussed property of the MNL model is the independence from irrelevant alternative (IIA), which states that the ratio between the choice probabilities of two alternatives is independent of other alternatives. This property is often seen as a drawback of the MNL model. The most popular example is the "red bus / blue bus paradoxon": Choice of transport mode, the alternatives are car or bus-both with the same travel time. The utility is supposed to be a function of the travel time. Since the travel time for both modes are equal, the choice probabilities for car and bus are both $\frac{1}{2}$. The introduction of a second bus with the same travel time only with a different color, will result in a choice probability of $\frac{1}{3}$ for each alternative. One would expect that the introduction of a second bus will not change the probability of choosing the car as the transport mode and that the choice probability between both bus alternatives are equal.
This drawback is overcome by the nested logit model (NLM), which is used

when the set of alternatives can be divided into non-overlapping subsets, the nests, such that the following two properties are satisfied:

1. The IIA holds for all alternatives within a nest.

2. The IIA does generally not hold for alternatives between different nests.

Let us further suppose that the set of alternatives is divided into K disjoint nests $B_1, \ldots, B_K$. The utility of alternative j for individual i is given by

$$U_{i,j} = \beta^\top x_j + \epsilon_{i,j}, \tag{5.16}$$

where $\epsilon_i = \{\epsilon_{i,1}, \ldots, \epsilon_{i,J}\}$ is a random variable, following a multivariate generalized extreme value (GEV) distribution with cdf

$$f(\epsilon_i) = \exp\left(-\sum_{k=1}^{K}\Big(\sum_{j\in B_k}\exp(\frac{-\epsilon_{i,j}}{\lambda_k})\Big)^{\lambda_k}\right). \tag{5.17}$$

λ_k represents the measure for the independence among alternatives in nest k, i.e., high values indicate a high independence and thus smaller correlation among the alternatives in the nest. Therefore $1-\lambda_k$ can be used as an indicator for correlation. Any two alternatives from different nests are supposed to be uncorrelated, $Cov(\epsilon_{i,j}, \epsilon_{i,l}) = 0$ for all $j \in B_r$ and $l \in B_s$ with $r \neq s$. In the case of $\lambda_k = 1$ for all nests k, ϵ_i is the product of J independent univariate extreme value distributed random variables and thus reduces the NLM to the MNL model. (See Figure 5.7 for an illustration of the MNL versus the NLM.) The probability that individual i choses alternative $j \in B_k$ is given by

$$P(y_i = j) = \frac{\exp(\frac{\beta^\top x_j}{\lambda_k})\Big(\sum_{l\in B_k}\exp(\frac{\beta^\top x_l}{\lambda_k})\Big)^{\lambda_k - 1}}{\sum_{r=1}^{K}\Big(\sum_{m\in B_r}\exp(\frac{\beta^\top x_m}{\lambda_r})\Big)^{\lambda_r}}. \tag{5.18}$$

An alternative way of expressing the choice probability is to write it as a product of a conditional and a marginal probability

$$P(y_i = j) = P(y_i = j | y_i \in B_k) \cdot P(y_i \in B_k), \tag{5.19}$$

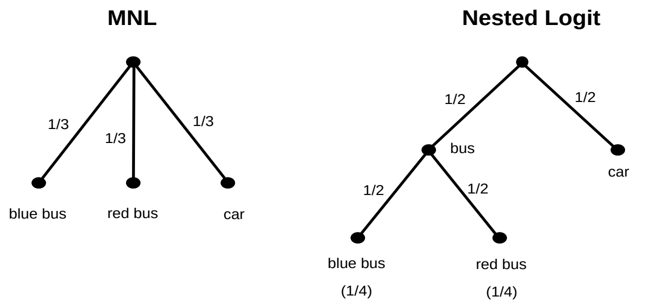

Figure 5.7. Red bus/ blue bus paradoxon, without captured correlations (MNL) and with captured correlation (Nested Logit).

with

$$\begin{aligned} P(y_i \in B_k) &= \frac{\exp\left(\beta_1^\top z_k + \lambda_k I_{i,k}\right)}{\sum_{l=1}^K \exp\left(\beta_1^\top z_l + \lambda_l I_{i,l}\right)}, \\ &= \frac{1}{\sum_{l=1}^K \exp\left(\beta_1^\top (z_l - z_k) + \lambda_l I_{i,l} - \lambda_l I_{i,l}\right)}, \end{aligned} \tag{5.20}$$

$$\begin{aligned} P(y_i = j | y_i \in B_k) &= \frac{\exp\left(\frac{\beta_2^\top x_j}{\lambda_k}\right)}{\sum_{l \in B_k} \exp\left(\frac{\beta_2^\top x_l}{\lambda_k}\right)}, \\ &= \frac{1}{\sum_{l \in B_k} \exp\left(\frac{\beta_2^\top (x_l - x_j)}{\lambda_k}\right)}, \end{aligned} \tag{5.21}$$

$$I_{i,k} = \log \sum_{l \in B_k} \exp\left(\frac{\beta_2^\top x_l}{\lambda_k}\right), \tag{5.22}$$

where z_k denotes the explanatory variables of attributes of nest k, β_1 its respective parameter vector, x_j represents the explanatory variable of attributes of alternative j and β_2 the corresponding parameter vector. Thus we are now faced with a two level logit model: The first level models the choice of the nest and the second level models the choice among alternatives in the chosen nest. $I_{i,k}$ links both choice levels and the product $\lambda_k I_{i,k}$ can be interpreted as the expected utility of individual i from choosing among the alternatives in nest k.

The parameter estimation is commonly done by MLE, with the choice probabilities (5.18) we obtain the log-likelihood

$$\log(L) = \sum_{i=1}^{n} \log\Big(P(y_i)\Big), \tag{5.23}$$

which is in general not concave. Because of the leveled choice structure it is possible to sequentially estimate the model parameter; first the choice of nests and afterwards the choices within the nests. Train (2003) argues against this sequential estimation and recommends only to use the sequential technique to generate starting values as input for the simultaneous parameter estimation.

5.2.5 Choice-set & MNL Combination Model (CSM)

The CSM model divides the choice process into two levels:

1. a choice-set based process which focuses on the choice between groups/classes of alternatives distinguished by some main characteristics

2. an MNL model which describes the exact choice among alternatives in the chosen group based on minor and possibly group specific characteristics

The concept of discrete choice-sets dates back to Ben-Akiva and Lerman (1985). As Train (2003) states, choice-sets are sets of alternatives, similar to $\mathcal{J}$, with three characteristics. First, the alternatives are mutually exclusive, i.e., choosing one strictly implies not choosing any other. Second, the choice-set is exhaustive, i.e., all possible choice decisions are included. Third, the choice-set has finite cardinality. In Chapter 2 we extend the above choice-set definition by forcing a strict preference order among alternatives within the choice-set and we defined choice-sets per customer types. Further, individuals are only distinguished by their choice-set. So if the choice-sets of two persons coincide, i.e., both are interested in the same set of products and have the same preference order, both person are not distinguished. This is in line with the situation at most companies. Usually companies do not have many detailed information of customers prior to their transaction, some data can be collected after the sale is made. But when deciding on the offer-set, i.e., which products to make available for sales, or which price to ask, normally no detailed information on individual customers is known. This model of an individual's choice process based on a set of alternatives of interest with a strict preference order, can be assumed to approximate the real choice process very well. In Chapter 3, we tested the choice-set demand unconstraining algorithm on a real airline reservation dataset. The results fit the actual sales data very well with only a slight overestimation and are therefore absolutely promising. Problem arise when the choice-set definition, i.e., definition of possible customer types, is not as straight forward as in the airline case with its 12 fare classes. Namely, when we have many alternatives with multiple different attributes. One clearly observes that the number of possible choice-sets grows exponentially in the number of possible choice alternatives. Our extended choice-set approach is to group alternatives with similar major choice attributes into groups or classes and model the choice among this classes with the choice-set model. The choice within the chosen class is further modeled with a different choice model, in our case the MNL, with a possible consideration of additional alternative attributes. The idea is to use choice-sets to divide customers into major behavior groups and estimate or forecast demand rates at this upper group level. The choice process within product classes may involve many attributes with each having similar impact in the decision. The MNL is, as described before, a standard model for this kind of choice processes with many explanatory variables.

The estimation of the choice-set model is divided into two steps: First, the identification of different choice-sets, and second, the demand estimation per choice-set. The identification of possible choice-sets is performed with the expert knowledge of marketing and sales persons in the specific business area. The demand rates per choice-set are estimated by the unconstraining EM algorithm developed in Chapter 4. The demand is assumed to follow an inhomogeneous Poisson process with an exponential rate function $\lambda_c(t) = \beta_c \cdot \exp(\alpha_c \cdot t)$, for choice-sets c and time stage t. The estimation method is applied separately for each product instance $n = 1, \ldots, N$ and we obtain choice-set estimates $\lambda_{c,n}(t) = \beta_{c,n} \cdot \exp(\alpha_{c,n} \cdot t)$. Remember in our case that product instances are the different arrival days of our hotel product, each having a booking horizon of 8 weeks. For a fair comparison with the other choice models, we need to find general parameter $\hat{\alpha_c}$ and $\hat{\beta_c}$ over the set of all instances. This is done by a separate MLE estimation

$$(\hat{\alpha_c}, \hat{\beta_c}) = \arg\min_{\alpha,\beta>0} - \mathcal{L}_c, \tag{5.24}$$

$$\mathcal{L}_c = \sum_{n=1}^{N} \sum_{t=t_1}^{t_T} \log P\left[X = \lambda_{c,n}(t) | X \sim \text{Poisson}\Big(\lambda_c(t)\Big)\right]. \tag{5.25}$$

The probability that individual i chooses class or group f at day d in time stage t is computed by

$$P(y_i \in f) = \frac{\sum_{c \in \mathcal{C}} \lambda_c(t) \cdot \mathbb{I}_{U(c,\mathcal{J}_t(d))=f}}{\sum_{c \in \mathcal{C}} \lambda_c(t) \cdot \mathbb{I}_{U(c,\mathcal{J}_t(d))>0}}. \tag{5.26}$$

Remember that $\mathcal{C}$ denotes the set of all choice-sets, $\mathbb{I}$ the indicator function and function U returns the chosen product of choice-set c under offer-set $\mathcal{J}$ or zero. To determine the probability of choosing hotel j in chosen class f, we apply the MNL model restricted to the available alternatives belonging to class f. We use the formula (5.1) with the optimal MNL choice parameter to compute the utility U_j of alternative j. The probability that individual i chooses hotel j is then given by

$$\begin{aligned} P(y_i = j) &= \sum_{f \in F} P(y_i \in f) \cdot \mathbb{I}_{j \in f} \cdot \frac{\exp(U_j)}{\sum_{k \in f, k \in J_i} \exp(U_k)}, \\ &= \sum_{f \in F} P(y_i \in f) \cdot \mathbb{I}_{j \in f} \cdot \frac{1}{\sum_{k \in f, k \in J_i} \exp(U_k - U_j)}, \end{aligned} \tag{5.27}$$

where F denotes the set of classes and J_i the set of alternatives offered to individual i. The parameter of the MNL model used in the second choice phase are estimated according to Section 5.2.1 over all alternatives j and their set of explanatory variables. The MNL choice parameter can be specific estimated per product class or general used for all groups. Class specific MNL parameter have the advantage to take account for different preference factors between different groups, e.g., the presence of a Champagne bar may be differently evaluated for 3-star or 5-star hotels. On the other hand, more parameters increase the estimation problem and its complexity.

5.3 Numerical Results

In the numerical results section, we estimate the choice models described in the preceding section. As input data, we use the reservation dataset provided by Bookit B.V., as introduced in Section 5.1. The entire dataset spans over a period of two years and is split into four smaller datasets to compare the estimation results of the choice models at different time periods. The datasets are the complete first and second year, as well as the two summer periods respectively for both years. The estimation software is coded and solved in MATLAB R2011b.

We divided the prices in the dataset by 50 to scale them to a similar level as the hotel stars or the geo-distance to the city center in kilometer. The maximum likelihood estimation for the MNL, LCM and NLM choice models is implemented by applying the MATLAB *fmincon* function on the negative log-likelihood function. Initial starting values are obtained by an upfront solved genetic algorithm, i.e., MATLAB *ga* function. We estimated the LCM with two numbers of latent classes $C = 2$ and 4. For the NLM, we also test two versions, the first, denoted *NLM 1*, with the nests representing distance groups $([0km, 1.5km), [1.5km, 2.5km)$ and $[2.5km, 10km])$ and the second, denoted *NLM 2*, representing star groups (3 stars, 4 stars and 5 stars). The MLM model is estimated by the MATLAB code provided by Train (2006). The distribution for the coefficients of the negative scaled price, the hotel stars and the negative geo-distance to the city center are set to log-normal and the coefficients for the booking day and the hotel alternatives are supposed to be normal distributed. Low prices, high stars and small geo-distances are considered to have a positive impact in the choice utility and therefore the coefficients are restricted to being positive by assuming a log-normal distribution. We estimate two version of the MLM, in the first the hotel-alternative coefficients are constant and only the coefficients for price, stars, geo-distance and booking day are random, the model is abbreviated with *MLM F4Rnd*. The second MLM version is denoted with *MLM AllRnd* and here we set all coefficients, including the hotel alternatives, to be random. Turning now to the CSM model, the first question is how to define the product classes for the choice-set model. From discussions with sales and marketing persons at Bookit, we learned that the price is the strongest separator for customer groups. Therefore, the product classes in our CSM model are defined by the hotel prices. We separate the offers into 4 price classes, which are chosen such that the price class (pc) cardinalities, number of corresponding sales realizations, are as equal as possible subject to some minimum price spread constraint of EUR 25. The price classes per dataset are shown in Table 5.1. The price class results show that customers become more price sensitive over both years. As shown in Figure

	pc1	pc2	pc3	pc4
Year 1	[70,118]	(118,149]	(149,193]	(193,300]
Aug & July Year 1	[70,101]	(101,138]	(138,169]	(169,300]
Year 2	[70,103]	(103,130]	(130,168]	(168,300]
Aug & July Year 2	[70,103]	(103,128]	(128,158]	(158,300]

Table 5.1. Price class definitions per dataset.

5.2, the price offers are not greatly reduced in year two but the reservations shift towards the lower prices. As choice-sets, we consider all combinations of coherent price classes

$$\begin{aligned}\mathcal{C} = \Big\{&\{pc1\}, \{pc2\}, \{pc3\}, \{pc4\},\\ &\{pc1, pc2\}, \{pc2, pc3\}, \{pc3, pc4\},\\ &\{pc1, pc2, pc3\}, \{pc2, pc3, pc4\},\\ &\{pc1, pc2, pc3, pc4\}\Big\}.\end{aligned}$$

Further, two versions of the CSM model are tested: *CSM 1* assumes constant MNL parameter for all price classes and *CSM 2* works with class specific MNL parameter. The choice-set demand parameter are estimated by the Algorithm proposed in Chapter 4 and the stopping criteria is set to maximum 100 iterations or a simultaneous satisfaction of the following three bounds on value changes: $\Delta\alpha$-tolerance 0.001, $\Delta\beta$-tolerance 0.01 and a Δ log-likelihood-tolerance of 0.1. Figure 5.8 shows an example trajectory of the log-likelihood development in the choice-set EM algorithm. The algorithm stopped in average after 30.9 seconds and took 31.2 iterations.

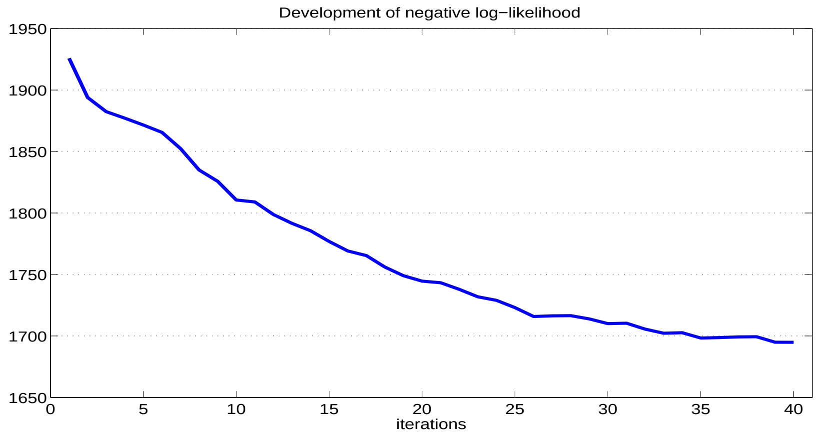

Figure 5.8. Example of the negative log-likelihood trajectory in the choice-set EM algorithm.

Let us now turn to the results of our comparison study. All models are fitted to the market data and we compare the results in order to evaluate the goodness of fits. As usual in model fitting, there is no one measure or criterion to base a strict decision on. Therefore, we focus on different measures and report the results of the different models. Table 5.2 shows the results of all models with their variations estimated on the complete year datasets, i.e., the complete first and second year. The corresponding results for the summer periods for both years are displayed in Table 5.4. As an initial goodness-of-fit measures we present the negative log-likelihood (NegLLH), the Akaike information criterion (AIC) and the prediction ration (PR) Top1, Top3 and Top5. The AIC, takes the negative log-likelihood value and punishes for two much parameter in the model, since a sparser model is due estimation reasons always preferred over a more complex model with the same fitting quality. We concentrate in the report on the NegLLH, as a statistical measure of the fit, and on the PR Top1, for a more practical view to answer 'how often is the choice model right'.

Dataset	Choice model	NegLLH	AIC	PR Top1	PR Top3	PR Top5
Year 1	MNL	10097	20246	0.2469	0.5372	0.7115
	LCM C=2	10009	20126	0.2555	0.5367	0.7165
	LCM C=4	9983	20182	0.2578	0.5488	0.7208
	MLM F4Rnd	10046	20152	0.2334	0.4446	0.5984
	MLM AllRnd	10022	20148	0.2596	0.5247	0.7104
	NLM 1	10083	20224	0.2580	0.5383	0.7156
	NLM 2	10114	20286	0.2469	0.5317	0.7115
	CSM 1	10488	21068	0.2203	0.4673	0.6874
	CSM 2	10336	20920	0.2291	0.4721	0.6974
Year 2	MNL	10434	20920	0.2175	0.5123	0.6801
	LCM C=2	10276	20660	0.2514	0.5239	0.6826
	LCM C=4	10349	20914	0.2467	0.5283	0.6889
	MLM F4Rnd	10372	20804	0.2309	0.5156	0.6931
	MLM AllRnd	10273	20650	0.2393	0.5276	0.6898
	NLM 1	10446	20950	0.2249	0.5016	0.6777
	NLM 2	10404	20866	0.2333	0.5067	0.6766
	CSM 1	10753	21598	0.2147	0.4742	0.6410
	CSM 2	10560	21368	0.2177	0.4961	0.6608

Table 5.2. Likelihood and prediction ratio results after estimation on the full year datasets.

The *PR TopX* represents the success probability over all choice situations

$$PR = \frac{right\ predictions}{all\ choice\ situations}, \tag{5.28}$$

where a success is given if the chosen hotel is under the hotels with the highest X choice probabilities computed by the choice model. Thus, PR Top1 is the success ratio that the choice model predicts the right outcome over the 22 possible hotel alternatives. We find that the PR Top1 ratio is always above 20%; a random chance model would have an expected prediction ratio of only

4.55%. Let us first concentrate on the full year results in Table 5.2. In both cases, the more complex models from the literature, the LCM, MLM and NLM, are best performing. The CSM 2, with specific MNL parameter per price class, is as expected outperforming the CSM 1, with only one MNL parameter set for all classes. Also, the difference between the CSM and the other models is larger for the first year. So the CSM seems to fit the second year data better, in comparison with the other models. When we compare the demand fluctuation for both years, see Figure 5.1, we clearly see a higher variation in year 1. The demand fit in the CS-part is better for the second year dataset, because the demand is more smooth. The CSM is essentially a demand model with choice information, whereas the other choice models are pure choice models with no need of information on the demand. In other words, the CSM needs to know how much customers of a given type are expected to arrive at a certain period to generate good prediction results. The other models, do not need such information and if a customer arrives they compute choice probabilities of alternatives. So the CSM is more complex, because it combines choice and demand. But this means, it also provides information on both, the choice behavior of different customer types and the expected demand rates per types. The choice-set demand rates are unconstrained per instance, and to ensure a fair comparison we use general choice-set parameter, see equation (5.24) and (5.25). Only as a reference, we show the corresponding CSM results obtained by the individual demand rates per instance, see Table 5.3. Note that we can

Dataset	Choice model	NegLLH	AIC	PR Top1	PR Top3	PR Top5
Year 1	CSM 1	9398	18888	0.3210	0.5863	0.7440
	CSM 2	9247	18742	0.3322	0.6000	0.7579
Year 2	CSM 1	9953	19998	0.2944	0.5627	0.6919
	CSM 2	9759	19766	0.2965	0.5809	0.7145

Table 5.3. Test with individual choice-set demand rates, likelihood and prediction ratios on the full year datasets.

not really compare these results with the other choice models, because this would mean comparing individual with general estimates and of course the individual ones perform better. Only the choice-set demand rates are now on individual level and the MNL parameter are still the same general ones as before. We observe a high increase in prediction quality and also very low NegLLH.

Next, we like to ask how the CSM performs in the summer periods, which are much more homogeneous in demand. The results are shown in Table 5.4. As expected we find that the CSM performs much better, CSM 2 even outperforms all other models by the PR Top1 ratio. Especially on the second year data, we find that the both CSM models outperform the other quite significantly. Out of the other models, which are closer together than in the full year periods, the MLM with all random parameter scores the best results.

Dataset	Choice model	NegLLH	AIC	PR Top1	PR Top3	PR Top5
Year 1	MNL	2504	5060	0.2050	0.4881	0.6826
	LCM C=2	2491	5090	0.2031	0.5167	0.6911
	LCM C=4	2495	5206	0.2126	0.4805	0.7073
	MLM F4Rnd	2500	5060	0.2040	0.5005	0.6826
	MLM AllRnd	2479	5062	0.2221	0.5081	0.6959
	NLM 1	2499	5056	0.2088	0.4919	0.6854
	NLM 2	2508	5074	0.2040	0.4890	0.6883
	CSM 1	2545	5182	0.2173	0.4643	0.6654
	CSM 2	2508	5263	0.2316	0.5033	0.6892
Year 2	MNL	1932	3916	0.2153	0.4815	0.6280
	LCM C=2	1919	3946	0.2140	0.4930	0.6548
	LCM C=4	1897	4010	0.2153	0.4790	0.6420
	MLM F4Rnd	1928	3916	0.2178	0.4777	0.6369
	MLM AllRnd	1898	3900	0.2204	0.4815	0.6599
	NLM 1	1936	3930	0.2178	0.4815	0.6318
	NLM 2	1933	3924	0.2153	0.4764	0.6268
	CSM 1	1937	3966	0.2471	0.4471	0.6293
	CSM 2	1891	4030	0.2611	0.4662	0.6701

Table 5.4. Likelihood and prediction ratio results after estimation on the summer (July & Aug.) datasets.

In the remainder, we will concentrate on the second year summer dataset and have a deeper look into the model performance. For a better visualization we will only report the results for the MNL, MLM AllRnd and the CSM choice models. The MNL represents the best known and standard choice model, the MLM is the most sophisticated and, as seen in Tab 5.4, the best performing model from the literature. Table 5.5 provides the choice predictions at hotel level on the second year summer dataset. The first 13 hotels are the 3-star hotels, followed by the 4-star hotels and hotel 22 is the single 5-star hotel in our dataset.

		Top1 Predictions				Top3 Predictions				Top5 Predictions			
Hotel	Sales	MNL	MLM	CSM1	CSM2	MNL	MLM	CSM1	CSM2	MNL	MLM	CSM1	CSM2
hotel 1	9	0	0	0	0	0	0	0	0	0	1	0	0
hotel 2	40	0	0	0	0	0	6	0	5	3	22	10	17
hotel 3	9	0	0	0	0	0	0	0	0	0	0	0	0
hotel 4	1	0	0	0	0	0	0	0	0	0	0	0	0
hotel 5	36	0	1	3	0	1	10	6	10	15	16	13	20
hotel 6	13	1	0	1	0	2	1	1	1	2	2	1	2
hotel 7	46	0	1	0	10	17	34	14	31	43	39	35	36
hotel 8	26	0	0	0	7	0	0	7	12	0	9	8	17
hotel 9	2	0	0	0	0	0	0	0	0	0	0	0	0
hotel 10	6	0	1	0	0	0	1	1	1	0	1	1	1
hotel 11	13	0	0	0	0	0	1	0	0	2	3	0	1
hotel 12	3	0	0	0	0	0	0	0	0	0	0	0	0
hotel 13	77	32	32	11	22	56	44	53	52	74	65	65	61
hotel 14	36	0	0	0	0	3	1	14	5	14	12	23	22
hotel 15	35	6	0	0	0	6	6	0	0	8	10	6	12
hotel 16	47	0	0	0	2	36	28	19	18	36	37	36	36
hotel 17	13	0	0	0	0	0	0	0	0	0	0	0	0
hotel 18	29	0	0	0	0	0	0	0	0	1	3	5	5
hotel 19	93	0	18	77	64	93	81	91	73	93	92	91	82
hotel 20	148	130	114	102	100	148	148	138	140	148	148	147	147
hotel 21	63	0	6	0	0	15	17	6	17	45	47	36	47
hotel 22	40	0	0	0	0	1	0	1	1	9	11	17	20

Table 5.5. Predicted sales per hotel on the second year summer dataset.

We observe that all models generate quite good results for the best selling hotel, number 20. But the CSM models focus less on the best selling one and score instead better on the other hotels. As we go from Top1 to Top5 results, we see that more and more hotels get accounted.
Let us define three additional very informative model fit measures, with the results given in Table 5.6. First, the adjusted prediction ratio, measuring the right prediction proportion and correcting against a null model which would always choose the most frequent outcome

$$Adj\ PR = \frac{right\ predictions - n}{all\ choice\ situations - n}, \tag{5.29}$$

with n denoting the most frequent outcome, in our case hotel 20 as the best selling hotel. Next, we define two pseudo R^2, one measuring the predictions per hotel and the other measures the predicted choice probabilities of the hotels. We denote the actual sales of hotel h with s_h and with e_h the estimated or predicted sales. Now we can define both pseudo R^2 values by

$$\begin{aligned} pseudo\ R^2 I &= 1 - \frac{\sum_{h \in hotels}(s_h - e_h)^2}{\sum_{h \in hotels}(s_h - \bar{s})^2}, \qquad (5.30)\\ with\ \bar{s} &= \frac{1}{|hotels|} \sum_{h \in hotels} s_h, \\ pseudo\ R^2 II &= 1 - \frac{\sum_{h \in hotels}(ps_h - pe_h)^2}{\sum_{h \in hotels}(ps_h - \bar{ps})^2}, \qquad (5.31)\\ with\ ps_h &= \frac{s_h}{\sum_{k \in hotels} s_k}, \\ pe_h &= \frac{e_h}{\sum_{k \in hotels} e_k}, \\ \bar{ps} &= \frac{\bar{s}}{\sum_{k \in hotels} s_k}. \end{aligned}$$

The resulting pseudo R^2s are not necessary between 0 and 1, since they are only pseudo values. This is especially true because they are not bounded from below by zero and can be negative. Nevertheless, higher values indicate a better model fit. The Adj PR gives a good indication on the added value of the choice model. In line with the predictions on hotel level, we see that the MNL and the MLM focus very much on the best selling hotel. The CSM models maintain also with the adjusted PR measure a significant large score.

	Top1 Predictions				Top3 Predictions				Top5 Predictions			
	MNL	MLM	CSM1	CSM2	MNL	MLM	CSM1	CSM2	MNL	MLM	CSM1	CSM2
PR	0.2153	0.2204	0.2471	0.2611	0.4815	0.4815	0.4471	0.4662	0.6280	0.6599	0.6293	0.6701
Adj PR	0.0612	0.0926	0.1444	0.1648	0.3611	0.3611	0.3344	0.3548	0.5416	0.5808	0.5447	0.5950
pseudo R^2 I	-0.081	0.021	0.076	0.139	0.532	0.557	0.487	0.554	0.749	0.813	0.782	0.844
pseudo R^2 II	-8.14	-4.97	-4.21	-2.56	-0.79	-0.50	-0.88	-0.36	0.32	0.56	0.47	0.69

Table 5.6. Model fit measures for MNL, MLM and CSM; on the second year summer dataset.

		Top1 Predictions				Top3 Predictions				Top5 Predictions			
Star-group	Sales	MNL	MLM	CSM1	CSM2	MNL	MLM	CSM1	CSM2	MNL	MLM	CSM1	CSM2
3 star	281	33	35	15	39	76	97	82	112	139	158	133	155
	0.36	0.12	0.12	0.05	0.14	0.27	0.35	0.29	0.4	0.49	0.56	0.47	0.55
4 star	464	136	138	179	166	301	281	268	253	345	349	344	351
	0.59	0.29	0.3	0.39	0.36	0.65	0.61	0.58	0.55	0.74	0.75	0.74	0.76
5 star	40	0	0	0	0	1	0	1	1	9	11	17	20
	0.05	0	0	0	0	0.03	0	0.03	0.03	0.23	0.28	0.43	0.50

Table 5.7. Predicted sales and percentages per hotel-star group on the second year summer dataset.

The observations is the same for the pseudo R^2s, the CSM models outperform the other two models. When considering the Top3 and Top5 predictions, we see that the results become naturally higher and that differences between all models decrease. Finally, we present the predicted sales and the corresponding percentages per hotel-star groups in Table 5.7. The 4-star hotels have the largest share with 0.6% and the 5-star hotel the lowest with only 5%. The 3-star and 4-star groups are fairly well predicted. The 5-star hotel is very poorly predicted by all four models. Only at the Top5 level, we obtain significant predictions.

5.4 Conclusion

In this chapter, we studied the performance of different choice models estimated on real reservation data of an entire hotel market. We tested the well known discrete choice models, starting with the standard choice model, the multinomial logit model. Followed by its extensions such as the latent class logit, the nested logit and the mixed logit model. These models are compared with a newly proposed choice-set based model. Overall, no model is tremendously outperforming the other. We observe that the pure choice models are superior to the choice-based approach, with constant rates, on datasets with high variability. The results can be greatly improved by using individual choice-set demand estimates instead of general demand rates. On datasets with fairly homogeneous demand pattern, we find that the choice-set based model outperforms the other significantly. The results show a strong increase in added information and prediction quality by the choice-set model. A big advantage of the choice-set based model is the merger of demand and choice. It models the choice process and simultaneously computes demand rates per customer types, which are also very suitable for forecasting. The other models are purely choice models, evaluating the probability for each alternative to be chosen by a customer, with no information on the number or types of customers to expect. The demand process is completely separated from these models. As we have seen in this study, the estimation of choice models on reservation data is practical, feasible and results in relatively good estimates. The question remaining, is how to forecast detailed future demand with choice behavioral information in such competitive and changing environments as the hotel or airline business. The choice-set approach already divides the demand into customer behavior groups. The corresponding demand functions are suitable for booking horizon forecasting techniques, which we will study in the next chapter.

Chapter 6

A Forecasting with Updating Case Study on Hotel Data

This chapter is based on the paper Haensel and Koole (2011a).

At the heart of every revenue management model always lies a demand forecast, whose accuracy is crucial for the success of the model. Pölt (1998) estimates for the airline industry that a 20% increase in forecast accuracy can be translated in a 1% increase in revenue generated by the underlying RM system. van Ryzin (2005) and Zeni (2007) argue that new models of demand forecasts are needed to adjust to the new market situation with more competition and less restricted products.
Besides the choice of the forecasting model and its adjustment to the demand time series, there are three important steps to include into the forecasting process. The first step is data unconstraining. It is important to note that sales figures are usually not equal to the real demand. This follows from capacity restrictions, booking control actions and the existence of competitors. Second, the customers' choice behavior has to be considered. A variety of product offers from a company or its competitors influence the customers' purchasing decision and thus the demand. The third point is the dynamic updating of forecasts when new information becomes available. As shown in O'Connor et al. (2000) the forecast accuracy can be improved by updating, especially when the time series is trended. In case of travel, accommodation and holiday products, the usual long booking horizons (plus the dependency of customer decisions due to holidays, special events or super offers) give additional hope for benefits from forecast updating. The first two points are covered by our choice-set demand model and the proposed unconstraining algorithm. The choice-set approach divides the demand into customer groups or types with different choice behavior. The unconstraining algorithm estimates demand rates per choice-set. Since the demand rates are independent per choice-set,

we can separately apply forecasting methods to the different choice-set demands. The focus of this chapter is the third point, the problem of forecast updating. Intensive research on forecast updating is done in the context of call centers. A significant correlation between within-day (morning, midday, evening) call arrivals is found. Models are proposed to forecast not only the call volume for future days, but also the updating of expected call volumes for future time periods within the day. In Weinberg et al. (2007) a multiplicative Gaussian time series model, with a Bayesian Markov chain Monte Carlo algorithm for parameter estimation and forecasting, is proposed. Shen and Huang (2008) suggest a competitive updating method which requires less computation time. Their method consists of a dimensionality reduction of the forecasting problem and a penalized least square procedure to adjust the time series forecast to observed realizations.
In this chapter we are adapting the ideas of Shen and Huang (2008) for call center forecasting to the RM context of hotel reservation forecasting. The equivalence to the within-day periods for which the forecast is updated is the booking horizon in the RM setting. In contrast to the call center case, booking horizons for different product instances are overlapping and correlated in their booking pattern and behavior. Another important difference is the level of forecast data. The call volume in call centers is generally very large, compared to often small demand numbers in the revenue management case. In RM problems a forecast on disaggregated level is required, since booking control actions are applied daily and on product level. A detailed description of the hotel reservation forecasting problem and its characteristics with a comparison of basic forecasting methods is given in Weatherford and Kimes (2003). A more advanced model is presented by Rajopadhye et al. (2001), in which they propose long-term forecasting of total reservations by the Holt-Winters method with a combination of booking curve pickup methods for short-term forecasts. More recently, Zakhary et al. (2011) presented a probabilistic hotel arrival forecasting method based on Monte Carlo simulation. All approaches aim to forecast the final reservation numbers, rather than the booking process, which is the focus of this chapter and Haensel and Koole (2011a).

The chapter is organized as follows: First, in Section 6.1, we introduce and analyze the data. Next, in Section 6.2, the forecasting methods are explained, followed by the introduction of the forecast updating procedure and methods in Section 6.3. Finally, in Section 6.4, numerical results are presented before we summarize our findings in Section 6.5.

6.1 Available Dataset

For our forecasting analysis we are able to work with real sales data, as in the previous chapter provided by the company Bookit B.V. The data is extracted from three regions A,B and C. The selected regions have very different characteristics such as reservation volume, city or countryside location and distance from major market. A hotel product is a combination of region, arrival day of week (DOW) and length of stay. For the analysis we consider a booking horizon of four weeks prior to arrival at the hotel, thus the 28^{th} day coincides with the arrival day. The datasets consist of all reservations made for a particular hotel product gathered over 3 years and multiple comparable hotels over all regions. The hotel products are not divided into different price classes, since hotels are interchangeable and products are not distinguished by specific hotels, but by location and hotel standard. To better illustrate the method, we will restrict this analysis to a fixed arrival DOW and length of stay combination. This separation of the forecasting problem into DOWs is widely common in practice, since the reservation patterns and volumes vary significantly for different arrival DOWs, compare for a discussion with Weatherford and Kimes (2003). In research and practice it is common to work on accumulated reservations, i.e., booking curves, rather than on individual reservations per booking horizon day. However, we see two reasons to prefer the latter. First, as a large reservations agency, one rarely runs out of stock. So the primal goal is to maximize the daily number of reservations. Therefore the second visualization form gives a more usable view as to which product to promote. Second, as also stated in van Ryzin (2005), the current major direction in revenue management research is to incorporate customer choice behavior under offered alternatives. Thus, it is more important to know the expected customer group demand per individual booking day rather than the aggregated totals. We will work with both visualizations of the booking process, using “Acc” and “Ind” to abbreviate the accumulated and individual reservations respectively. Hence, we obtain six datasets on the three regions: A-Acc, A-Ind, B-Acc, B-Ind, C-Acc and C-Ind. All datasets are given in form of a $n \times m$ reservation data matrix X, with $n = 155$ product instances (as rows) and their associated $m = 28$ booking horizon days (as columns). In our case, the product instances correspond to the successive arrival weeks of our hotel products, fixed DOW and length of stay. For clarity, the $X_{i,j}$ entry denotes the number of bookings made for product instance i (arrival week i) at the j^{th} day in the booking horizon. The first 130 product instances/ rows are used for data analysis, testing and parameter estimation. The last 22 instances/ rows are used in Section 5 for evaluation of the proposed forecast updating methods. There is a gap of four weeks between the estimation and evaluation sample, caused by the time structure in the dataset: At arrival of product i, the realization of the first booking horizon week of product instance $i+3$ is known. The booking behavior of the first three instances in A-Acc and A-Ind, i.e., rows of X, are shown in Figure 6.1. The total aggre-

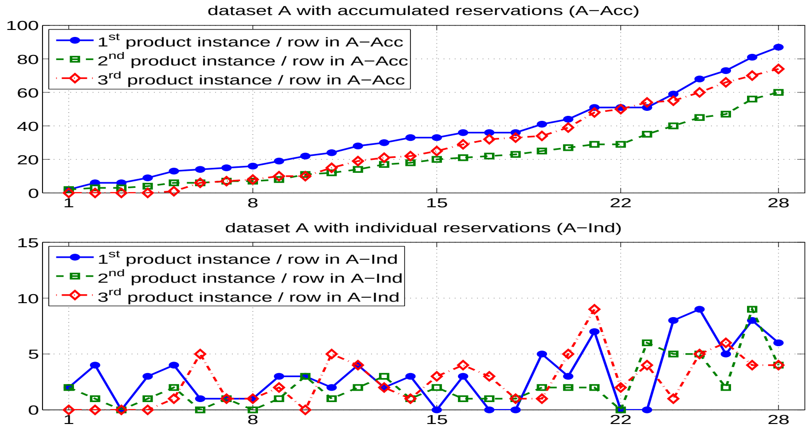

Figure 6.1. Example plot of first three product instances / first three rows of A-Acc and A-Ind.

gated numbers of reservations received for the first 130 product instances of all regions are shown in Figure 6.2. Note that the time between the product

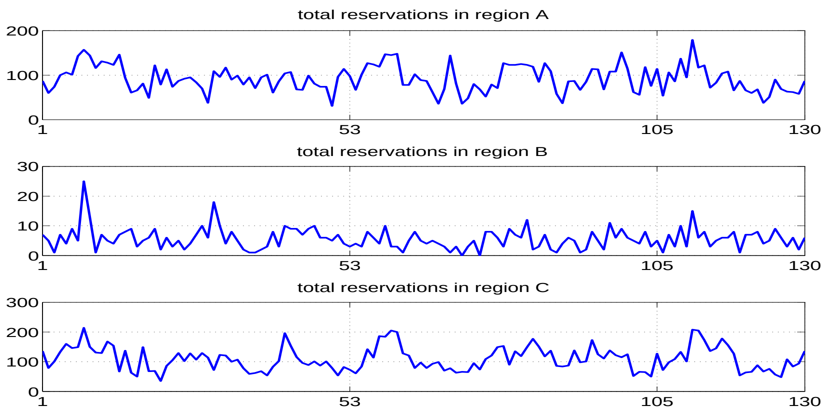

Figure 6.2. Number of total aggregated reservations for the first 130 product instances of all three regions.

instances, one week, is much smaller than the booking horizon of four weeks. The mean and variance of the booking behaviors within the booking horizon for all datasets are shown in Figure 6.3. We observe that the variance is not constant (heteroscedasticity) and that the variance is greater than the mean (overdispersion). In order to stabilize and reduce the variance, we will work

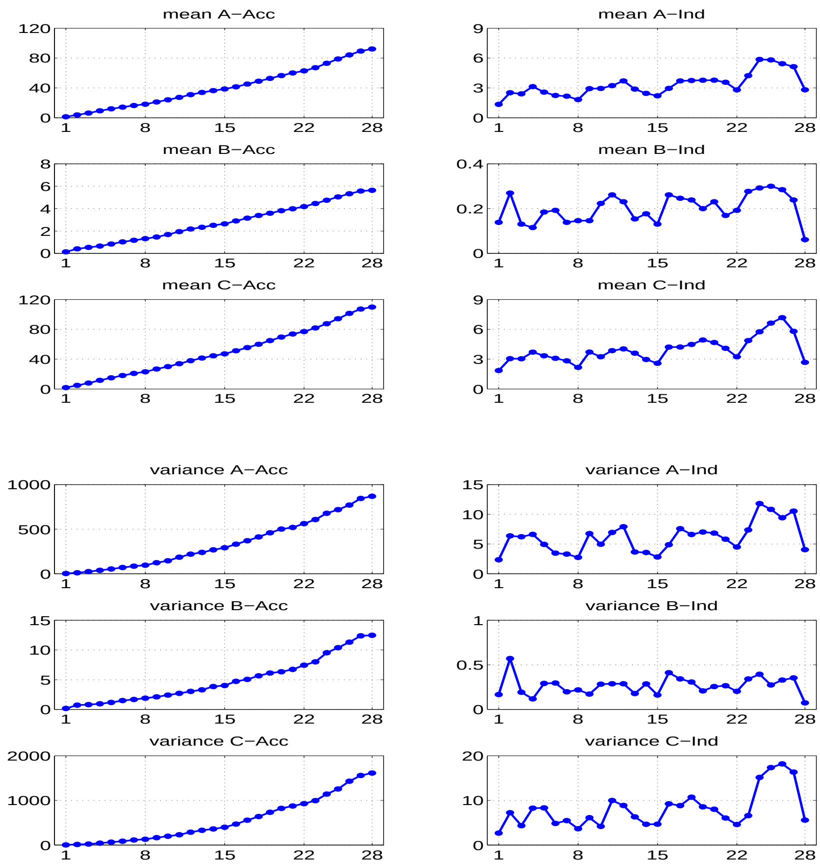

Figure 6.3. Mean and variance of booking process of all datasets.

on the logarithmic transformed data. Let x denote the number of reservations. Set $y = \log(x + 1)$, one is added because the dataset contains many booking days with zero reservations. The forecast of y is denoted by $\hat{y}$ and the forecast of x is then given by $\hat{x} = \exp(\hat{y}) - 1$. The following forecasting methods are working on the transformed data, and the forecast error analysis in Section 5 is made on the back transformed data.

An important property in the data structure is the shifted realization time, which means that parts of different product instances are realized at the same time. For example, suppose we select a product instance i, i.e., the i^{th} row in X, and consider the corresponding time as the current time. All information up to instance i plus the first three weeks of the booking horizon of the following instance $i+1$, the first two weeks of $i+2$ and the first week of instance $i+3$ are known at our current time. In other words, fixing an arrival week in our data set as a time point enables us to know the demand realization for the first three weeks of the booking horizon for the next-week-arrival product. The same is true for the in-two-weeks arrival and in-three-weeks arrival products, where we know the realization of the first two and first week of the booking horizon, respectively. This paper is concerned with the question of how to update reservation forecasts when the realizations of earlier time stages in the booking horizon become known. Therefore we analyze the correlation between reservations made at different moments of the booking horizon.

In Figure 6.4, the correlation coefficients between early (long in advance) and late reservations (close to arrival at hotel) are plotted as a function of the day in the booking horizon that is the frontier between early and late. The correlation function for split day k is defined on the accumulated reservation dataset X by

$$C(k) = corr\left(X_{\cdot,k}, (X_{\cdot,28} - X_{\cdot,k})\right) \qquad k = 1, \ldots, 27, \tag{6.1}$$

where $corr(a, b)$ is a function returning the linear correlation coefficient between the vectors a and b. The correlation is found to be very different for all three regions. The correlation is highest for region C and lowest for region B. Consequently, the benefit of dynamic forecast updating is assumed to be most beneficial for the datasets of region C. Also the shape of the correlation function differs between the regions. The correlation in region A decreases slightly over the booking horizon, in contrast to regions B and C where the maximum is attained around day 15 (half of the considered booking horizon).

Now consider the correlation between bookings in different weeks. Define $w_i = \{7(i-1)+1, \ldots, 7(i-1)+7\}$, the set of days in week i. The correlation function defined on booking weeks w_i and w_j and a dataset X consisting of individual reservations per booking day, is given by

$$C(w_i, w_j) = corr\left(\sum_{d \in w_i} X_{\cdot,d}, \sum_{d \in w_j} X_{\cdot,d}\right). \tag{6.2}$$

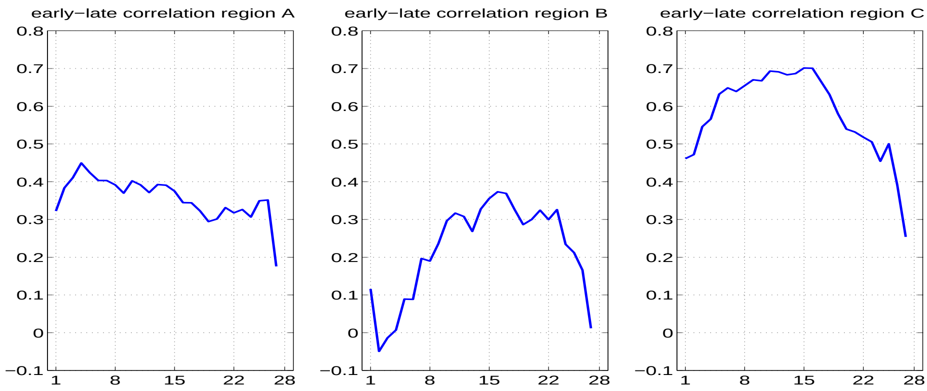

Figure 6.4. Correlation coefficients $C(k)$ for early-late booking split days $k = 1, \ldots, 27$.

The correlations are shown in Table 6.1. Multiple subscripts represent the aggregation over multiple weeks, e.g., $w_{1,2,3}$ stands for the aggregated reservations made in week 1, 2 and 3. It illustrates again the dependence between early and late bookings.

Region	$C(w_1, w_2)$	$C(w_2, w_3)$	$C(w_3, w_4)$
A	0.59	0.59	0.42
B	0.00	0.10	0.20
C	0.65	0.67	0.46
	$C(w_1, w_{2,3,4})$	$C(w_2, w_{3,4})$	$C(w_{1,2,3}, w_4)$
A	0.41	0.44	0.27
B	0.16	0.24	0.33
C	0.64	0.68	0.53

Table 6.1. Correlation $C(\cdot, \cdot)$ of aggregated reservations between specific booking weeks.

6.2 Forecasting Method

The reservation matrix X contains all reservations per product instance and day in the booking horizon. Our aim is to forecast the number of future reservations to expect for the forthcoming product instances in the next four weeks. For a company that must choose each day which products to offer, advertise or promote, it is very important to know the expected number of

reservations per day in the booking horizon. Therefore the forecast is chosen to work on individual numbers of reservations per booking day ("Ind" datasets) as well as on the accumulated reservations ("Acc" datasets). Each of the next four product instances (arrival weeks) has a booking horizon of $m = 28$ days. The i^{th} row of X, $x_i = (x_{i,1}, \dots, x_{i,m})^\top$, represents all reservations per day in the booking horizon of instance i. As in Shen and Huang (2008) we are applying singular value decomposition (SVD) to reduce the forecasting dimension. The procedure works as follows:
We are interested in computing a small number of base vectors $f_1, \dots, f_K$ with which the time series $\{x_t\}$ can be reasonably well approximated. The decomposition is given by

$$x_i = \gamma_{i,1} f_1 + \dots + \gamma_{i,K} f_K + \epsilon_i \qquad i = 1, \dots, n, \tag{6.3}$$

where $\gamma \in \mathbb{R}^{n \times K}$ is the weight matrix, $f_1, \dots, f_K \in \mathbb{R}^m$ are the base vectors and $\epsilon_1, \dots \epsilon_n \in \mathbb{R}^m$ are the error terms. We suppose that the $x_i's$ can be well approximated by a linear approximation of the base vectors, so that the error terms are reasonably small. This leads to the following optimization problem

$$\min_{\substack{\gamma_{1,1}, \dots, \gamma_{n,K} \\ f_1, \dots, f_K}} \quad \sum_{i=1}^{n} \|\epsilon_i\|, \tag{6.4}$$

for a fixed value K. This problem can be solved by applying SVD to matrix X as follows. Matrix X can be rewritten as

$$X = USV^\top, \tag{6.5}$$

where S is a $m \times m$ diagonal matrix, U and V are orthogonal matrices with dimension $n \times m$ and $m \times m$ respectively. The diagonal elements of S are in decreasing order and nonnegative, $s_1 \geq \dots \geq s_r > 0$, with $r = rank(X)$ and $s_k = 0$ for all $r + 1 \leq k \leq m$. From (6.5) we follow now

$$x_i = s_1 u_{i,1} v_1 + \dots + s_r u_{i,r} v_r, \tag{6.6}$$

where v_k denotes the k^{th} column of matrix V. The K-dimensional approximation is obtained by keeping the largest K singular values ($K < r$), since S is ordered decreasingly the largest are equivalent with the first K values,

$$x_i \approx s_1 u_{i,1} v_i + \dots + s_K u_{i,K} v_K. \tag{6.7}$$

Setting now $\gamma_{i,k} := s_k u_{i,k}$ and $f_k := v_k$, for all $i = 1, \ldots, n$ and $k = 1, ..., K$, we have found an optimal solution of (6.4). The mean squared estimation error (MSEE) of product instance i and fixed K is computed by

$$\text{MSEE}_i = \frac{1}{m} \sum_{j=1}^{m} \left(x_{i,j} - \big(\sum_{k=1}^{K} \gamma_{i,k} f_k \big)_j \right)^2 . \tag{6.8}$$

Figure 6.5 shows the empirical distribution function of the MSEE, computed over the first 130 product instances, for different values of K. We find reason-

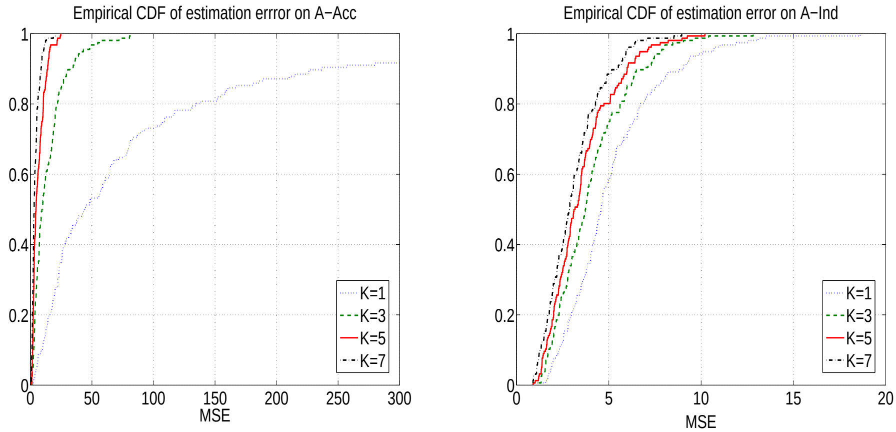

Figure 6.5. Empirical distribution function of mean squared estimation errors for different numbers of base vectors K.

ably small errors for $K = 3$. These values are still outperformed by $K = 5$ or 7, but for computational reasons we try to keep the dimension small. In the numerical results, where we use $K = 3$ and 5, we will see that $K = 3$ will produce reasonably good forecasting results. The resulting three base vectors in the case of $K = 3$ and their weights computed over the first 130 instances of the datasets A-Acc and A-Ind are shown in Figure 6.6. The base vectors represent the data characteristics in decreasing importance, i.e., the first base vector in A-Ind represents the strong weekly pattern and the first base vector in A-Acc represent the general increasing booking curve pattern. In fact, base vector f_1 in A-Acc is negative and decreasing, but since the corresponding weights time series γ_1 takes negative values, the represented booking pattern is increasing. Remember that the singular value decomposition is applied to the transformed data, when comparing with Figure 6.3. The forecasting method will work on the time series of $\gamma_{i,k}$ values. The base vectors $f_1, \ldots, f_K$ are calculated on the historical data and are kept fixed during the forecasting process. Due to the

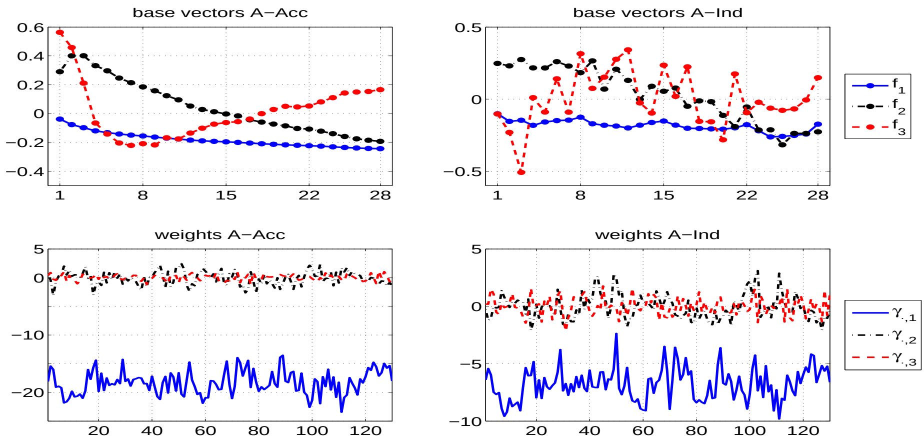

Figure 6.6. Base vectors f_k and weights $\gamma_{i,k}$ for $K = 3$.

construction of the weights series $\gamma_1, \ldots, \gamma_K$ out of the columns of U, we have that vectors γ_k and γ_l are orthogonal for $k \neq l$. Hence the cross correlation between different weight series can be assumed to be small.

We initially choose as a forecasting method the univariate exponential smoothing with trend and seasonality, i.e., the Holt-Winters (HW) method developed by Holt (1957) and Winters (1960). Holt-Winters is a commonly used method in similar problems, and in practice known to be reasonably accurate, robust and easy to implement. Two seasonal models, additive and multiplicative, are distinguished and both are tested on the three datasets. The additive Holt-Winters (AHW) h-step ahead forecast of $\gamma_{i+h,k}$, for fixed $k = 1, \ldots, K$ is

$$\hat{\gamma}_{i+h,k}^{AHW} = a(i) + h \cdot b(i) + c\big((i+h) \bmod P\big), \tag{6.9}$$

where $a(i)$, $b(i)$ and $c(i)$ are given by

$$\begin{aligned} a(i) &= \alpha \cdot \big(\gamma_{i,k} - c(i-p)\big) + (1-\alpha) \cdot \big(a(i-1) + b(i-1)\big), \\ b(i) &= \beta \cdot \big(a(i) - a(i-1)\big) + (1-\beta) \cdot b(i-1), \\ c(i) &= \delta \cdot \big(\gamma_{i,k} - a(i)\big) + (1-\delta) \cdot c(i-p). \end{aligned}$$

In contrast, the multiplicative Holt-Winters (MHW) h-step ahead forecast of $\gamma_{i+h,k}$ is

$$\hat{\gamma}_{i+h,k}^{MHW} = \Big(a(i) + h \cdot b(i)\Big) \cdot c\big((i+h) \bmod P\big), \tag{6.10}$$

where $a(i)$, $b(i)$ and $c(i)$ are computed by

$$\begin{aligned}
a(i) &= \alpha \cdot \frac{\gamma_{i,k}}{c(i-p)} + (1-\alpha) \cdot \big(a(i-1) + b(i-1)\big),\\
b(i) &= \beta \cdot \big(a(i) - a(i-1)\big) + (1-\beta) \cdot b(i-1),\\
c(i) &= \delta \cdot \frac{\gamma_{i,k}}{a(i)} + (1-\delta) \cdot c(i-p).
\end{aligned}$$

The period length is one year and because the product instances are weekly $p = 52$. The initial values of a, b and c are derived from a simple decomposition in trend and seasonal component using moving averages (averaging for each time unit over all periods). The decomposition is performed by the R function Decompose from the R-stats library. Optimal α, β and δ values are found by minimizing the squared one-step prediction error, evaluated over historical values. Since the weights time series γ_k take negative and positive values, a positive constant is added in the MHW calculation to ensure positivity and subtracted from the forecasts before being processed further. The Holt-Winters forecast for both seasonal models, of the future booking horizon $\hat{x}^{HW}_{i+h} = (\hat{x}^{HW}_{i+h,1}, \dots, \hat{x}^{HW}_{i+h,m})$ is computed by

$$\hat{x}^{HW}_{i+h} = \hat{\gamma}^{HW}_{i+h,1} \cdot f_1 + \cdots + \hat{\gamma}^{HW}_{i+h,K} \cdot f_K. \tag{6.11}$$

The forecast accuracy for both seasonal models is tested on the sample of the 1-15 step/weeks ahead forecasts, starting at instance 110 (within the estimation sample) and computed on all datasets. The mean squared errors between the actual γ_k and forecasted $\hat{\gamma}^{HW}_k$ are shown in Table 6.2 and abbreviated with ϵ_k, for k=1,2,3. No seasonal model outperforms the other and both models

Dataset	ϵ_1 AHW	ϵ_1 MHW	ϵ_2 AHW	ϵ_2 MHW	ϵ_3 AHW	ϵ_3 MHW
A-Ind	2.28	2.30	1.23	1.23	1.12	1.12
A-Acc	5.18	5.18	0.79	0.79	0.80	0.80
B-Ind	0.41	0.40	0.29	0.29	0.51	0.51
B-Acc	18.82	19.06	1.42	1.42	2.34	2.78
C-Ind	7.03	7.05	1.87	1.87	0.34	0.40
C-Acc	9.93	10.81	0.90	0.90	0.11	0.11

Table 6.2. Holt-Winters mean squared errors between the actual and forecasted $\gamma_{\cdot,k}$.

produce generally the same forecast error. We will further continue only with the additive seasonal model for the Holt-Winters forecasting method.

Our second forecasting approach is to decompose the γ time series into seasonal, trend and remainder components and to apply an auto-regressive (AR) time series model on the remainder. The additive seasonal models seem to give a good approximation. Therefore we apply a decomposition procedure based on LOESS, i.e., local polynomial regression fitting, as described by Cleveland et al. (1990). The decomposition is performed by the R function STL (seasonal decomposition of time series by loess) from the R-stats library. The γ time series are separately decomposed into additive seasonal, trend and remainder components, see Figure 6.7 for the case of dataset A-Acc and $K = 3$. The decomposition equation is

$$\gamma_{i,k} = s_{i,k} + d_{i,k} + r_{i,k} \qquad \text{for all } i = 1, \dots, n \text{ and } k = 1, \dots, K, \tag{6.12}$$

where s denotes the seasonal, d the trend and r the remainder component of the time series γ. The auto-regressive model of $r_{\cdot,k}$ of order p is given by

$$r_{i,k} = \nu_k + a_1 \cdot r_{i-1,k} + \dots + a_p \cdot r_{i-p,k} + u_{i,k}, \qquad k = 1, \dots, K, \tag{6.13}$$

where $a_1, \dots, a_p \in \mathbb{R}$ represents fixed coefficients, u_k a zero mean white noise process and ν_k the intercept. When μ_k denotes the mean of $r_{\cdot,k}$, the intercept is defined as $\nu_k = (1 - \sum_{t=1}^{p} a_t)\mu_k$. Let $\hat{r}_{i,k}$ denote the forecast of $r_{i,k}$, the h-step ahead forecast of the weights time series γ_k at instance i is then computed by

$$\hat{\gamma}^{AR}_{i+h,k} = s_{i+h,k} + d_{i+h,k} + \hat{r}_{i+h,k}, \tag{6.14}$$

where the trend component d is computed by a linear extrapolation of the last five known trend values $d_{i-4,k}, \dots, d_{i,k}$ and the respective seasonal components s are obtained from the LOESS decomposition. The Akaike Information Criterion (AIC) is used to find the optimal model order p. We test the AIC for orders $p = 0, \dots, 6$ on all datasets. The AIC results were generally the best for $p = 1$ and we will continue to use the AR(1) model for all six datasets. A further interesting observation is that the dependency, measured by the pairwise correlation coefficients, among the remainder $r_{\cdot,k}$ time series after the decomposition slightly increases compared to the dependency between the original weights series $\gamma_{\cdot,k}$. Therefore, we will compare the univariate AR(1) model with the vector auto-regressive (VAR) model, see Lütkepohl (2005) and Box et al. (1994), on the joint remainder time series r. The VAR(1) model of the remainder $r(i) = (r_{i,1}, \dots, r_{i,3})$ has the following form

$$r(i) = \nu + A \cdot r(i-1) + u_i, \tag{6.15}$$

where $A \in \mathbb{R}^{K \times K}$ represents a fixed coefficient matrix, u a zero mean white noise process and ν the intercept. When μ denotes the mean of r, the intercept is equivalently defined as $\nu = (I - A)\mu$. Let $\hat{r} = (\hat{r}_{i,1}, \dots, \hat{r}_{i,K})$ denote the

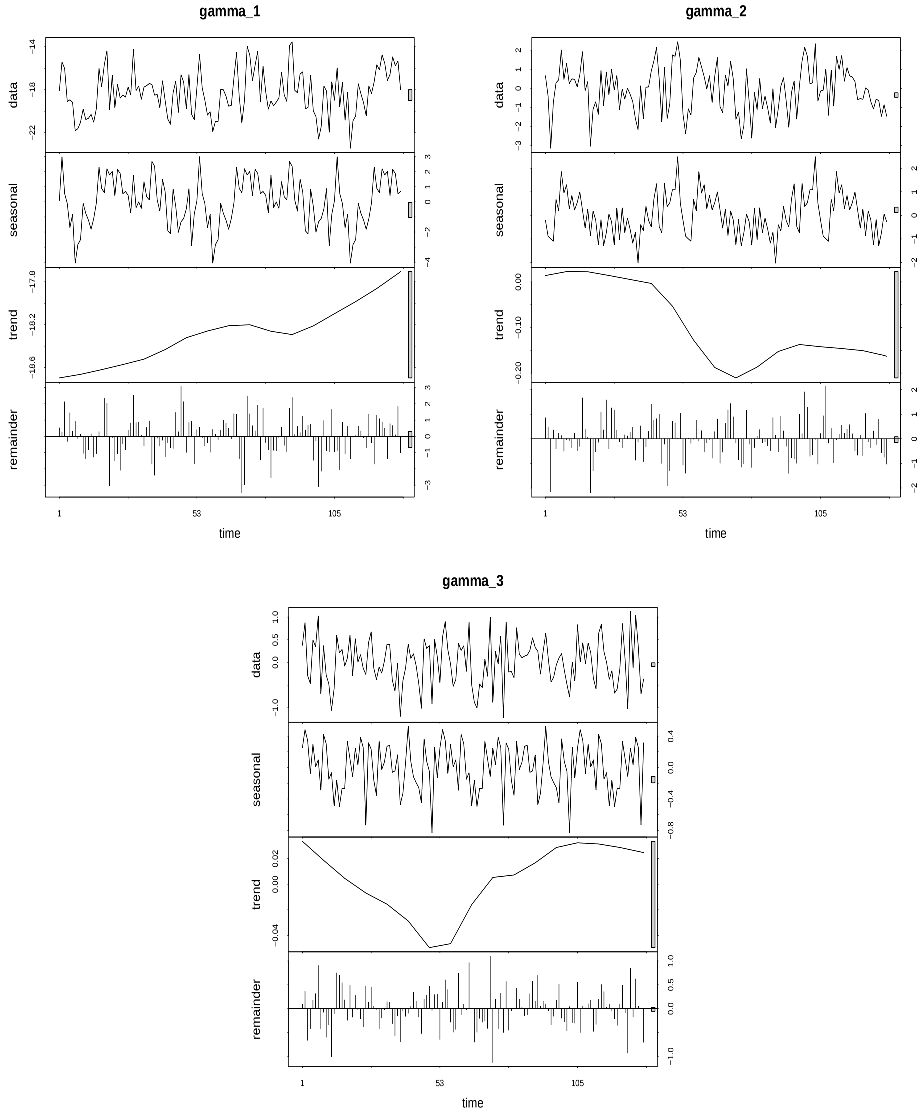

Figure 6.7. Decomposition of γ time series of the dataset A-Acc into seasonal, trend and remainder for $K = 3$.

VAR forecast of r, the h-step ahead forecast of the weights time series γ at instance i is equivalent to the AR forecast computed by

$$\hat{\gamma}^{VAR}_{i+h,k} = s_{i+h,k} + d_{i+h,k} + \hat{r}_{i+h,k} \qquad \text{for all } k = 1, \dots, K, \tag{6.16}$$

where the trend component d is again computed by a linear extrapolation of the last five known trend values. As in the comparison of both HW models, the forecast accuracy of the AR and the VAR forecast is tested on the sample of the 1 till 15 step ahead forecast, starting at instance 110 and computed on all datasets. The mean squared forecast errors ϵ_k in the weights time series γ_k are shown in Table 6.3. Both forecasts generate roughly the same forecast errors,

Dataset	ϵ_1 AR	ϵ_1 VAR	ϵ_2 AR	ϵ_2 VAR	ϵ_3 AR	ϵ_3 VAR
A-Ind	2.14	2.12	0.60	0.58	1.09	1.09
A-Acc	2.97	2.93	0.50	0.49	0.42	0.42
B-Ind	0.31	0.30	0.34	0.34	0.33	0.33
B-Acc	11.20	11.15	2.57	2.55	1.09	1.09
C-Ind	1.67	1.69	2.40	2.41	0.31	0.30
C-Acc	2.24	2.27	0.81	0.79	0.37	0.38

Table 6.3. AR and VAR mean squared errors between the actual and forecasted $\gamma_{\cdot,k}$.

but a closer look shows that the VAR produces slightly smaller errors. More interesting is to compare the accuracy results of the Holt-Winters with the auto-regression forecasts, i.e., Table 6.2 with Table 6.3. The combination of seasonal-trend decomposition and auto-regression forecasts on the remainder increases the forecast accuracy significantly. In the remainder of the paper, we will work with the VAR(1) model as the second forecasting method as opposed to the AHW. The VAR forecast of the future booking horizon $\hat{x}_{i+h}^{VAR} = (\hat{x}_{i+h,1}^{VAR}, \ldots, \hat{x}_{i+h,m}^{VAR})$ is obtained by

$$\hat{x}_{i+h}^{VAR} = \hat{\gamma}_{i+h,1}^{VAR} \cdot f_1 + \cdots + \hat{\gamma}_{i+h,K}^{VAR} \cdot f_K. \tag{6.17}$$

6.3 Forecast Updating

With the previously described method, the forecaster is able to compute a forecast of an entire booking horizon, i.e., a forecast of the accumulated booking curve or the estimated incoming reservations for each day in the booking horizon. The forecasting methods work only on completed booking horizons. This means that we are not updating a forecast for future weeks when the realization of week 1,2 or 3 become known. Furthermore, the information of the booking realizations for 1,2 or 3 weeks prior to the forecast date are not used in the computation, because their booking horizon is not completed yet. Therefore we propose the following general updating procedure, which includes one of the forthcoming updating methods.

The procedure is as follows:

1. Forecast the next booking horizon based on data of all completed product instances (of which all realizations are known).

2. If realizations of the forecasted booking process are known, update the future part of the horizon accordingly.

3. If a further forecast is required, regard the forecasted (and updated) horizon as completed and go back to 1.

The application of the procedure to our datasets (being at arrival of product instance i) is as follows:

(1) A one week ahead forecast of the complete booking horizon for instance $i + 1$ (the next week arrival) is generated based on data of all completed booking horizons (instances $1, \dots, i$). Realizations of the first three weeks in the booking horizon are already known and the forecast of reservations in the last week for instance $i + 1$ are adjusted based on this.

(2) The adjusted forecast of $i+1$ is further regarded as the completed booking horizon of instance $i + 1$.

(3) A one week ahead forecast of the complete booking horizon for instance $i+2$ is computed based on the data of all instances $1, \dots, i+1$. Realizations of the first two weeks are already known and the forecast for the last two weeks of the booking horizon is adjusted.

(4) The adjusted forecast of $i+2$ is further regarded as the completed booking horizon of instance $i + 2$.

(5) A one week ahead forecast of the complete booking horizon for instance $i+3$ is computed based on the data of all instances $1, \dots, i+2$. The realization of the first week is already known and the forecast for the following three weeks is adjusted.

(6) The adjusted forecast of $i+3$ is further regarded as the completed booking horizon of instance $i + 3$.

(7) Finally, a one week ahead forecast of the complete booking horizon for instance $i + 4$ is computed based on the data of all instances $1, \dots, i + 3$. Because we are considering a booking horizon of four weeks, no bookings for this instance are known and the forecast can not be adjusted.

In the following we will discuss two forecast updating methods, as described in Shen and Huang (2008). Let us concentrate on instance $i+1$, the forecast of (6.11) or (6.17) can be written as

$$\hat{x}_{i+1} = F\hat{\gamma}_{i+1}, \tag{6.18}$$

where $F = (f_1, \ldots, f_K)$ denotes a $m \times K$ matrix formed by the base vectors and $\hat{\gamma}_{i+1} = (\hat{\gamma}_{i+1,1}, \ldots, \hat{\gamma}_{i+1,K})^\top$ represents a column vector. Let the superscript a denote that we only consider the first a columns of a matrix or components of a vector. When x^a_{i+1} becomes known, we can compute the forecast error ϵ^a_{i+1} by

$$\epsilon^a_{i+1} = x^a_{i+1} - \hat{x}^a_{i+1} = x^a_{i+1} - F^a\hat{\gamma}_{i+1}. \tag{6.19}$$

The direct least squares (LS) method would now try to solve the problem

$$\hat{\gamma}^{LS}_{i+1} = \operatorname{argmin}_{\hat{\gamma}_{i+1}} \left\|\epsilon^a_{i+1}\right\|^2, \tag{6.20}$$

to find the γ_{i+1} values for which the forecast of the first a days fits the actual bookings x^a_{i+1} best. The LS solution can be obtained by

$$\hat{\gamma}^{LS}_{i+1} = \left((F^a)^\top F^a\right)^{-1}(F^a)^\top x^a_{i+1}. \tag{6.21}$$

To uniquely define $\hat{\gamma}^{LS}_{i+1}$, we need of course that $a \geq K$. In our case $K = 3$ or 5, the booking horizon is further in days and the forecast updates are made weekly, $a = 7, 14$ and 21. The idea is to apply the solution of (6.20) in (6.18) to obtain the direct least squares forecast update $\hat{x}^{LS}_{i+1}$ by

$$\hat{x}^{LS}_{i+1} = F\hat{\gamma}^{LS}_{i+1}. \tag{6.22}$$

Clearly this is a very volatile updating method and the forecast update will not be too reliable for small a values compared to m, length of the whole booking horizon. Therefore we suggest the penalized least squares method (PLS), which works as the LS method but it penalizes large deviations from the original time series (TS) forecast. The optimization problem (6.20) is altered with the parameter λ to

$$\hat{\gamma}^{PLS}_{i+1} = \operatorname{argmin}_{\hat{\gamma}_{i+1}} \left\|\epsilon^a_{i+1}\right\|^2 + \lambda \left\|\hat{\gamma}_{i+1} - \hat{\gamma}^{TS}_{i+1}\right\|^2, \tag{6.23}$$

where $\hat{\gamma}^{TS}$ denotes the original time series forecast. We observe that if $\lambda = 0$, $\hat{\gamma}^{PLS} = \hat{\gamma}^{LS}$, and for $\lambda \to \infty$, $\hat{\gamma}^{PLS} = \hat{\gamma}^{TS}$. As shown in Shen and Huang (2008), the PLS updated forecast can be computed with

$$\hat{\gamma}^{PLS}_{i+1} = \left((F^a)^\top F^a + \lambda I\right)^{-1}\left((F^a)^\top x^a_{i+1} + \lambda\hat{\gamma}^{TS}_{i+1}\right). \tag{6.24}$$

And finally, the PLS updated forecast of the future booking horizon $\hat{x}_{i+1}^{PLS}$ is obtained by

$$\hat{x}_{i+1}^{PLS} = \hat{\gamma}_{i+1,1}^{PLS} f_1 + \cdots + \hat{\gamma}_{i+1,K}^{PLS} f_K. \tag{6.25}$$

One other more intuitive updating approach is the historical proportion (HP) method. The accuracy of the forecast is simply computed by the ratio of already observed realization and their forecasted values. Suppose we are at updating point a, i.e., realizations of the first a days in the booking horizon are known. The ratio R is given by

$$R = \frac{\sum_{j=1}^{a} x_{i+1,j}}{\sum_{j=1}^{a} \hat{x}_{i+1,j}}, \tag{6.26}$$

keeping in mind that $\hat{x}_{i+1}$ denotes the time series based forecast of x_{i+1}. The HP updated forecast for the remaining booking days is the with R scaled $\hat{x}_{i+1}$,

$$\hat{x}_{i+1,j}^{\mathrm{HP}} = R \cdot \hat{x}_{i+1,j} \qquad\qquad j = a+1, \ldots, m. \tag{6.27}$$

In the following section we will compare the PLS and HP updating method with the forecast results that are not updated.

6.4 Numerical Results

In this section we will compare all combinations of the additive Holt-Winters and the vector auto-regressive forecasts with the two previously proposed updating methods penalized least squares and historical proportion, as well as with the not updated forecasts (NU). The number behind the abbreviation of the forecasting method (AHW or VAR) denotes the number of base vectors K used in the singular value decomposition. In our test case we are working with $K = 3$ or 5. The evaluation set consists of the last 22 instances, i.e., arrival weeks, of our six datasets (instances 134-155). Thus, the evaluation is made in a time frame of five months. As measures of forecast accuracy the mean squared error (MSE) and the mean week relative absolute error (MWRAE) are computed for the four booking horizon weeks. The squared error (SE) and the week relative absolute error (WRAE) are defined for instance i and weeks

$w = 1, \ldots, 4$ by

$$\mathrm{SE}(i, w=1) = \sum_{k=1}^{7} (x_{i,k} - \hat{x}_{i,k})^2 \qquad \mathrm{WRAE}(i, w=1) = \frac{|\sum_{k=1}^{7} x_{i,k} - \sum_{k=1}^{7} \hat{x}_{i,k}|}{\sum_{k=1}^{7} x_{i,k}}$$

$$\mathrm{SE}(i, w=2) = \sum_{k=8}^{14} (x_{i,k} - \hat{x}_{i,k})^2 \qquad \mathrm{WRAE}(i, w=2) = \frac{|\sum_{k=8}^{14} x_{i,k} - \sum_{k=8}^{14} \hat{x}_{i,k}|}{\sum_{k=8}^{14} x_{i,k}}$$

$$\mathrm{SE}(i, w=3) = \sum_{k=15}^{21} (x_{i,k} - \hat{x}_{i,k})^2 \qquad \mathrm{WRAE}(i, w=3) = \frac{|\sum_{k=15}^{21} x_{i,k} - \sum_{k=15}^{21} \hat{x}_{i,k}|}{\sum_{k=15}^{21} x_{i,k}}$$

$$\mathrm{SE}(i, w=4) = \sum_{k=22}^{28} (x_{i,k} - \hat{x}_{i,k})^2 \qquad \mathrm{WRAE}(i, w=4) = \frac{|\sum_{k=22}^{28} x_{i,k} - \sum_{k=22}^{28} \hat{x}_{i,k}|}{\sum_{k=22}^{28} x_{i,k}}.$$

The MSE and the MWRAE are computed by averaging the SE and WRAE over our 22 evaluation instances. Both error measures are the ones used by the company, Bookit. The MSE gives insight into the accuracy on daily level, while the MWRAE provides the proportional absolute difference in week totals. The

Forecast	Dataset	λ_1	λ_2	λ_3	Dataset	λ_1	λ_2	λ_3
VAR 3	A-Acc	0.0114	0.0452	0.1691	A-Ind	0.1676	0.2893	2.0623
VAR 5		0.0040	0.0182	0.0615		0.2617	0.8529	0.9294
AHW 3		0.0029	0.0213	0.0411		0.0736	0.1117	0.4801
AHW 5		0.0010	0.0072	0		0.1511	0.6909	0.3270
VAR 3	B-Acc	0.0206	0.0089	0.0594	B-Ind	148.6172	0	0.2699
VAR 5		0.0090	0.0093	0.0289		177.9697	0.0920	0.3775
AHW 3		0.0141	0.0099	0.0322		0.5054	0.1471	0.2225
AHW 5		1.1320	2.0711	0.0123		0.4020	0.3317	0.2757
VAR 3	C-Acc	0.0276	0.0452	0.0962	C-Ind	0.0236	0.4586	0.0635
VAR 5		0.0084	0.0275	0.0583		11.1344	0.2037	0.1561
AHW 3		0.0033	0	0.0015		0.0421	0	0.1801
AHW 5		0.0011	0	0.0187		0.0761	0	0.1431

Table 6.4. λ parameter for PLS updating methods, respectively for dataset and booking week.

optimal λ parameter for the PLS are found by minimizing the MSE updating error at the last 22 instances of the testing and estimation sample (instances 109 till 130), see Table 6.4 for the final values. λ_1, λ_2 and λ_3 are respectively used in the updating in booking horizon weeks 1,2 and 3. The best forecast and updating method combinations for each dataset, which minimize the forecast error per booking horizon week, are given in Table 6.5. All generated MSE and MWRAE error results are shown in Tables 6.8-6.10, the smallest errors for dataset and booking week combination are highlighted by an asterisk (*). For the MSE we find in weeks 2, 3 and 4 the smallest values for the PLS updated forecasts. Except for the C-Ind dataset in week 4, there the VAR 5 PLS value exceeds the VAR 3 HP value by 12, which only corresponds to a

Dataset	MSE			
	1^{st} week	2^{nd} week	3^{rd} week	4^{th} week
A-Ind	VAR 5 NU & PLS	VAR 3 PLS	VAR 3 PLS	VAR 3 PLS
A-Acc	VAR 3 NU	VAR 3 PLS	VAR 3 PLS	VAR 3 PLS
B-Ind	VAR 3 NU & PLS	VAR 3 NU & PLS	VAR 3 PLS	VAR 3 PLS
B-Acc	AHW 3 NU	AHW 3 PLS	VAR 3 PLS	VAR 3 PLS
C-Ind	VAR 3 HP & PLS	VAR 3 NU & 5 PLS	VAR 3&5 PLS	VAR 3 HP
C-Acc	VAR 3 NU	VAR 3 PLS	VAR 5 PLS	VAR 5 PLS
	MWRAE			
	1^{st} week	2^{nd} week	3^{rd} week	4^{th} week
A-Ind	VAR 3 NU	VAR 3 PLS	AHW 3 PLS	AHW 5 PLS
A-Acc	VAR 3 NU	VAR 3 PLS	VAR 5 PLS	VAR 3 PLS
B-Ind	VAR 3 PLS	VAR 3 PLS	VAR 3 HP	VAR 3 NU
B-Acc	AHW 3&5 NU	AHW 3 HP	VAR 3 PLS	VAR 5 PLS
C-Ind	VAR 3 HP & PLS	VAR 3 NU	VAR 5 NU	AHW 3 HP
C-Acc	VAR 5 NU & PLS	VAR 3 PLS	VAR 5 PLS	VAR 5 PLS

Table 6.5. Best forecast and updating methods respectively to minimization of MSE or MWRAE per booking week and dataset.

mean absolute daily error of 0.5. We also observe that the VAR outperforms the AHW forecast, except for the B-Acc dataset in week 1 and 2, but here we find again an insignificant increase of the MSE by only 3 and 7 compared with the VAR 5 PLS results. Considering the MWRAE we initially observe that the accuracy increases with the amount of reservations contained in a dataset. Consequently the C datasets have the lowest MWRAE, as they also hold the most reservations. The VAR 5 forecast of week 4 in C-Acc without updating already has a MWRAE of 0.18, but can still be decreased to 0.08 by the PLS updating method. Looking at the B-Acc dataset, again week 4 and VAR 5, the not updated forecast has a MWRAE of 0.73 which can be significantly decreased by the PLS updating to 0.2. At first glance at Table 6.5 in the MWRAE area, the PLS updating is still the best, but with less dominance compared to the MSE part. With a closer look at Tables 6.8-6.10, we find that the PLS error values are very close to the best performing methods; compare for example the VAR 3 PLS values for A-Ind, B-Ind and C-Ind. Note that approximately 80% of all bookings are made within this last three weeks of the booking horizon and still 60% within the last two booking weeks. Therefore a forecast accuracy increase in the later part of the booking horizon is more important than in the early stages. Comparing the two forecasting methods additive Holt-Winters and vector auto-regression, we observe that the mean values of the VAR outperform the AHW forecast. This shows that the correlation between the base vectors should not be neglected.

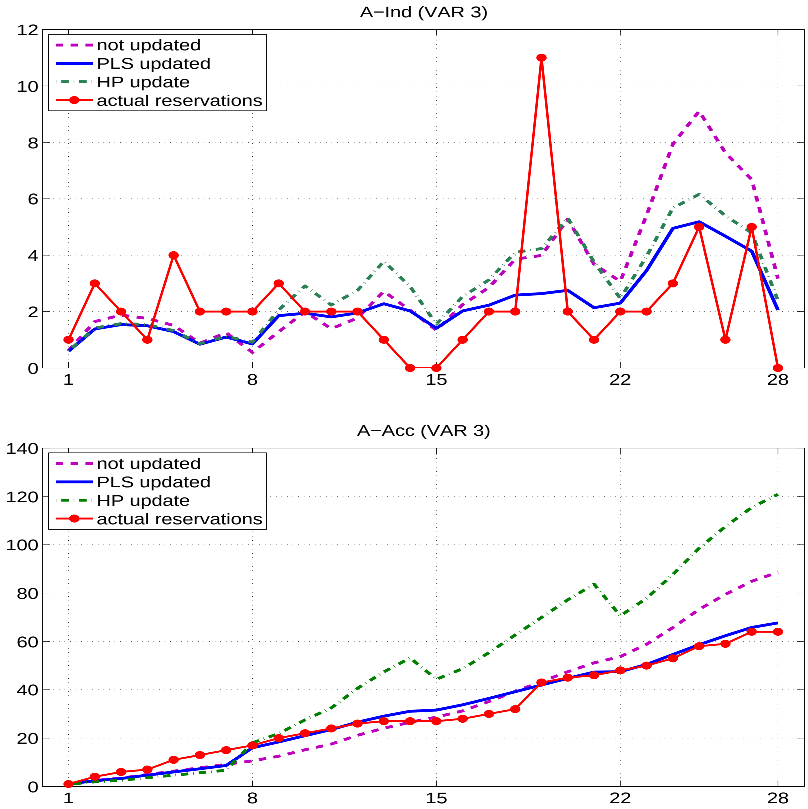

Figure 6.8. Example of reservation actuals vs. updated and not updated forecasts for one instance of datasets A-Acc and A-Ind.

The results of the different updating methods are graphically illustrated in Figure 6.8, for one instance of the evaluation set for datasets A-Ind and A-Acc and base forecast VAR 3. For practitioners, the accurate forecast of the overall number of reservations is as important as the forecast of the booking curve or individual reservations per booking day. The total relative absolute error (TRAE) for "Ind" datasets is then defined by

$$\mathrm{TRAE}(i) = \frac{|\sum_{k=1}^{28} \hat{x}_{i,k} - \sum_{k=1}^{28} x_{i,k}|}{\sum_{k=1}^{28} x_{i,k}}.$$

In case of accumulated "Acc" datasets the TRAE is simply computed by the last column values

$$\mathrm{TRAE}(i) = \frac{|\hat{x}_{i,28} - x_{i,28}|}{x_{i,28}}.$$

In Table 6.6 the different mean and median TRAE values are shown for all datasets and updating methods applied to a VAR 3 forecast. The PLS is

Dataset	Mean TRAE			Median TRAE		
	NU	PLS	HP	NU	PLS	HP
A-Ind	0.28	0.21	0.27	0.24	0.16	0.28
B-Ind	0.46	0.32	0.40	0.44	0.33	0.46
C-Ind	0.22	0.17	0.21	0.23	0.16	0.20
A-Acc	0.22	0.11	0.28	0.21	0.09	0.24
B-Acc	0.54	0.30	0.92	0.41	0.26	0.66
C-Acc	0.18	0.11	0.27	0.16	0.10	0.25

Table 6.6. Mean and median of total relative absolute error (TRAE) for all updating methods and datasets with VAR 3 as base forecast.

completely outperforming the not updating and HP updating in all cases. Further, the TRAE for datasets of the same region is always minimized on the accumulated dataset. Two interesting observations are: First, the HP updating is only better than not updating in the case of “Ind” datasets, and second that PLS updating on “Ind” datasets results in lower TRAE than not updating on “Acc” datasets. Finally, we are interested in testing the

	MSE			
	1^{st} week	2^{nd} week	3^{rd} week	4^{th} week
A-Ind λ_{opt}	16	22	29	58
A-Ind λ_{est}	16	24	31	58
A-Acc λ_{opt}	98	99	219	338
A-Acc λ_{est}	89	137	267	334
	MWRAE			
	1^{st} week	2^{nd} week	3^{rd} week	4^{th} week
A-Ind λ_{opt}	0.27	0.25	0.33	0.35
A-Ind λ_{est}	0.28	0.27	0.38	0.34
A-Acc λ_{opt}	0.47	0.14	0.11	0.07
A-Acc λ_{est}	0.47	0.15	0.14	0.07

Table 6.7. MSE and MWRAE values of VAR 3 PLS with optimal (opt) and estimated (est) λ values.

sensitivity of the PLS to its λ values. Therefore, we compare the MSE and MWRAE generated using λ values estimated on historical data versus the error values generated by optimal λ values, i.e., estimated by minimizing the MSE over the evaluation set. In the following the VAR 3 is used as base forecast method. The optimal λ_1, λ_2 and λ_3 values are $0.4991, 0.0841$ and 5.5078 for

A-Ind and $0.0372, 0.0039$ and 0.1789 for A-Acc, which are deviating from the so far used values (see Table 6.4). The generated MSE and MWRAE results of the optimal λ's are given in Table 6.7. As expected, using the optimal λ in the PLS updating results in an accuracy increase. But, we also find the error decrease by using the optimal lambda values to be considerably small. These results support our positive validation on the robustness of the PLS method.

			MSE				MWRAE			
			1^{st} week	2^{nd} week	3^{rd} week	4^{th} week	1^{st} week	2^{nd} week	3^{rd} week	4^{th} week
A-Ind	VAR 3	NU	16	23	35	64	0.26*	0.32	0.45	0.37
		PLS	16	24*	31*	58*	0.28	0.27*	0.38	0.34
		HP	17	29	32	74	0.36	0.40	0.37	0.46
	VAR 5	NU	15*	25	38	66	0.30	0.32	0.46	0.37
		PLS	15*	27	34	58	0.34	0.29	0.39	0.35
		HP	16	30	38	75	0.34	0.42	0.40	0.46
	AHW 3	NU	32	47	47	98	1.07	0.91	0.65	0.45
		PLS	26	27	29	67	0.78	0.29	0.34*	0.34
		HP	26	58	30	205	0.78	0.52	0.35	0.66
	AHW 5	NU	34	50	51	101	1.05	0.97	0.65	0.44
		PLS	36	31	44	64	0.90	0.35	0.49	0.33*
		HP	36	54	37	295	0.90	0.55	0.36	0.68
A-Acc	VAR 3	NU	85*	282	929	2,047	0.45*	0.32	0.32	0.25
		PLS	89	137*	267*	334*	0.47	0.15*	0.14	0.07*
		HP	193	1,148	1,467	2,696	0.60	0.54	0.30	0.24
	VAR 5	NU	100	284	873	2,206	0.50	0.31	0.31	0.26
		PLS	135	288	272	404	0.56	0.19	0.13*	0.08
		HP	167	1,094	1,443	2,643	0.59	0.54	0.30	0.22
	AHW 3	NU	373	1861	6,189	19,527	1.27	0.93	0.84	0.67
		PLS	325	642	341	517	0.99	0.27	0.15	0.09
		HP	325	16,320	5,203	207,650	0.99	0.74	0.41	0.67
	AHW 5	NU	353	2,029	6,454	18,760	1.21	0.98	0.86	0.64
		PLS	345	561	477	785	0.98	0.27	0.18	0.11
		HP	345	34,193	5,153	100,050	0.98	0.89	0.41	0.57

Table 6.8. Region A: Mean squared errors (MSE) and mean week relative absolute errors (MWRAE) for all forecast combinations - the asterisk highlights the smallest errors for each dataset and booking week combination.

			MSE				MWRAE			
			1^{st} week	2^{nd} week	3^{rd} week	4^{th} week	1^{st} week	2^{nd} week	3^{rd} week	4^{th} week
B-Ind	VAR 3	NU	1.5*	3.1*	5.4	2.7	0.61	0.69	0.74	0.50*
		PLS	1.5*	3.1*	5.3*	2.6*	0.60*	0.67*	0.61	0.54
		HP	1.5*	3.4	5.4	2.7	0.62	0.95	0.56*	0.59
	VAR 5	NU	1.6	3.1	5.4	2.9	0.62	0.73	0.66	0.62
		PLS	1.6	3.2	5.4	2.7	0.63	0.75	0.77	0.62
		HP	1.6	3.4	5.6	2.9	0.65	0.94	0.73	0.70
	AHW 3	NU	1.6	3.6	5.5	3.1	0.72	0.88	0.77	0.69
		PLS	1.5*	3.5	5.3*	2.7	0.75	0.69	0.68	0.63
		HP	1.5*	3.7	5.4	2.9	0.75	0.93	0.72	0.71
	AHW 5	NU	1.7	3.6	5.5	3.3	0.76	0.94	0.73	0.77
		PLS	1.5*	3.7	5.3*	2.9	0.77	0.74	0.86	0.73
		HP	1.5*	3.7	5.4	4.6	0.77	0.94	0.74	1.07
B-Acc	VAR 3	NU	13	47	85	93	3.37	3.71	1.65	0.74
		PLS	13	41	43*	33*	3.01	1.89	0.76*	0.28
		HP	14	282	764	434	3.47	0.86	1.11	0.80
	VAR 5	NU	13	54	75	92	3.49	3.85	1.54	0.73
		PLS	12	36	75	43	2.99	2.53	1.10	0.20*
		HP	15	284	1,638	361	3.90	0.87	1.23	0.80
	AHW 3	NU	9*	36	87	128	1.97*	3.70	3.23	1.08
		PLS	14	29*	58	48	3.65	1.50	1.95	0.32
		HP	14	39	574	1245	3.65	0.68*	1.15	0.51
	AHW 5	NU	10	36	92	125	1.97*	3.63	3.43	1.06
		PLS	15	36	201	63	3.68	3.43	6.60	0.27
		HP	15	39	492	205	3.68	0.71	1.11	0.55

Table 6.9. Region B: Mean squared errors (MSE) and mean week relative absolute errors (MWRAE) for all forecast combinations - the asterisk highlights the smallest errors for each dataset and booking week combination.

			MSE				MWRAE			
			1^{st} week	2^{nd} week	3^{rd} week	4^{th} week	1^{st} week	2^{nd} week	3^{rd} week	4^{th} week
C-Ind	VAR 3	NU	33	25*	43	85	0.42	0.20*	0.25	0.37
		PLS	32*	27	37*	96	0.38*	0.29	0.23	0.38
		HP	32*	31	46	80*	0.38*	0.34	0.27	0.37
	VAR 5	NU	37	26	43	87	0.51	0.23	0.21*	0.38
		PLS	35	25*	37*	92	0.47	0.21	0.25	0.40
		HP	35	35	49	81	0.45	0.38	0.30	0.38
	AHW 3	NU	52	65	122	265	0.81	0.62	0.72	0.76
		PLS	41	26	43	109	0.70	0.26	0.28	0.47
		HP	41	33	42	94	0.70	0.35	0.27	0.34*
	AHW 5	NU	62	67	124	281	0.93	0.64	0.74	0.77
		PLS	49	42	61	136	0.77	0.39	0.39	0.59
		HP	49	48	40	99	0.77	0.47	0.23	0.36
C-Acc	VAR 3	NU	156*	437	886	2,629	0.50	0.25	0.18	0.17
		PLS	160	112*	371	963	0.47	0.09*	0.09	0.09
		HP	233	2,513	2,468	4,900	0.54	0.44	0.29	0.25
	VAR 5	NU	180	505	990	2,920	0.46*	0.27	0.19	0.18
		PLS	188	128	286*	723*	0.46*	0.11	0.08*	0.08*
		HP	255	2,781	2,762	5,981	0.56	0.43	0.29	0.25
	AHW 3	NU	728	4,236	12,444	31,035	0.96	0.82	0.78	0.75
		PLS	445	646	573	996	0.75	0.20	0.10	0.10
		HP	445	4,924	1,286	9,693	0.75	0.57	0.22	0.30
	AHW 5	NU	831	4,840	12,116	27,654	0.98	0.90	0.77	0.70
		PLS	505	564	481	770	0.76	0.26	0.11	0.08
		HP	505	5,622	1,149	15,958	0.76	0.63	0.21	0.30

Table 6.10. Region C: Mean squared errors (MSE) and mean week relative absolute errors (MWRAE) for all forecast combinations - the asterisk highlights the smallest errors for each dataset and booking week combination.

6.5 Summary

In this chapter, we have analyzed forecast updating methods using actual hotel reservation data from three different regions. Statistical tests have shown that there is a significant correlation between early and late bookings. Consequently not updating the demand forecast when early realizations in the booking horizon become available means ignoring important information that can dramatically affect the end result. But, also for datasets with low correlation between early and late bookings (region B in our data set) we observe a significant accuracy increase by updating. The forecast updating is performed dynamically when new demand realizations become available, in our case weekly. The initial forecast results are then updated using either the penalized least squares (PLS) or the historical proportion method. In the case of a multi step ahead forecast, the base forecast produces multiple one step ahead forecasts on historical data and previous updated forecasts. We find that dynamic updating reservation forecasts using PLS is very beneficial in most situations with low and high correlation between different parts of the booking horizon, and is never significantly harmful compared to not updating. Also computationally the method is very fast and therefore feasible for use by practitioners in larger forecasting problems. Singular value decomposition is applied to reduce the dimensionality of the forecasting problem and the results show its effectiveness. As base forecasts we use a multivariate vector autoregressive model and a univariate Holt-Winters model on the reduced forecasting problem. The results show that the VAR outperform the AHW. Thus, the dependency between base vectors after the SVD should not be ignored. In addition, an increase of base vectors from 3 to 5 generally does not result in lower error values. Overall, the VAR 3 forecast method seems to be the best base forecast for our datasets.

Chapter 7

Time Dependent Bid Prices Models

This chapter is in part based on the paper Bijvank, Haensel, L'Ecuyer and Marcotte (2011).

In the following, we consider a network revenue management problem. So the firm is assumed to have different resources, with specific capacities, and offers products which are compositions of resource units. The booking control is made by a bid price policy, originally studied by Simpson (1989) and Williamson (1992). Talluri and van Ryzin (1998) precisely define a bid price control as a set of bid prices for each resource in the network at each point in time and each capacity, such that a request for a particular product is accepted if and only if there is available capacity and the price of the product exceeds (is greater or equal to) the sum of bid prices for all units of resources used by the product. The idea behind it is that a request is only accepted when the price exceeds the marginal costs of the capacity reduction in case of a sale. This idea is the reason why bid prices are usually directly or indirectly derived by duality, the bid prices are associated with the dual variables, shadow prices, related to capacity constraints in the primal problem. Talluri and van Ryzin (1998) show that bid price controls are asymptotically optimal, when capacities and demands are large. Further, there is no one-to-one relation between optimal bid prices and opportunity costs. There are situations, where a set of bid prices generates optimal accept-reject decisions, but is a poor estimate of the actual marginal value of capacity. The challenge is, according to Talluri and van Ryzin (1998), to construct bid prices resulting in close to optimal acceptance decisions and provide at the same time good estimates of the opportunity costs, such that special events like group requests can be optimally evaluated. Traditional approaches to compute the bid prices are based on deterministic demand and therefore ignore distributional demand information. Also, bid prices are static whereas the actual system is dynamic. The dynamic behavior can be approximated by re-solving the static models.

But, re-solving does not always improve the control, we refer to Secomandi (2008) for a detailed study on re-solving. Dynamic bid prices are the focus of recent research. Adelman (2007) studies the dynamic program (DP) of the problem and assumes an affine value function, resulting in a linear program (LP) formulation. A different approach is proposed by Topaloglu (2008), who associates Lagrange multipliers with the capacity constraints in the DP formulation and decomposes the network problem into its resource subproblems. A similar method is proposed by Kunnumkal and Topaloglu (2010), but with a different relaxation. All these models are based on the DP formulation and assume at most one customer request per time period. This is unrealistic in situations, when group request are present or when bid prices are not changed at very high frequencies. For example, some airlines set their availabilities on daily level and normally do not update these values during the day. Figure 7.1 shows an actual sales trajectory for a single price class. The availability is

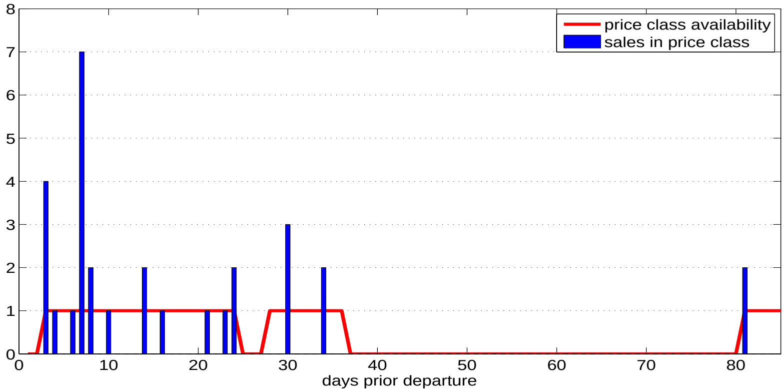

Figure 7.1. Actual airline sales trajectory for a single price class.

illustrated by the 0-1 curve, one indicating that the price class was available/ open. We observe that the open and close decision is on day level. Also, we have many days without sales, but if there is a sale, we have often multiple bookings. So the assumptions are not met in practice. An additional disadvantage of the previous models is their mathematical and computational complexity.

The chapter continues with a description of the traditional bid price approach. In Section 7.2, we introduce a new approach to compute time dependent bid prices. Our TDB model assigns capacities to time varying demand and the corresponding bid prices are computed directly in the primal optimization model and are not derived by duality. Section 7.3 proposes extensions of the TDB to incorporate stochastic demand information. Section 7.4 extends the models proposed in Bijvank et al. (2011) with formulations to incorporate the customer's choice behavior under offered alternatives, namely the considering of choice-set demand as input. The chapter is summarized in Section 7.5.

7.1 Traditional Bid Price Approach

Let us start with some basic notation of the network RM problem. Consider a network with $M \in \mathbb{N}$ resources, each with initial capacity $c \in \mathbb{R}^M$. There are $N \in \mathbb{N}$ products and $r \in \mathbb{R}^N$ denotes the product prices. The resource-product matrix $A \in \{0,1\}^{M \times N}$ defines the structure of the network, such that $a_{i,j}$ equals one if product j utilizes resource i and zero otherwise. The j^{th} column of A, denoted by A_j, represents the subset of resources utilized by product j. Respectively, the i^{th} row of A, denoted by A^i, represents the subset of products consuming resource i. We consider a discrete booking horizon with time stages $t = 1, \dots, T$, with $t = 1$ denoting the usage time of the product and $t = T$ the first period in the booking horizon. The demand $D \in \mathbb{R}^{N \times T}$ is assumed to be deterministic and given per product and time stage.

Generally, bid prices are derived by duality and are interpreted as the marginal price of capacity. The standard approach is by the deterministic linear program (DLP), described in Talluri and van Ryzin (2004b), which maximizes the total revenue based on the expected demand. The primal optimization problem at each time stage t with available inventory level $x \leq c$ is given by

$$V_t^{DLP_P}(x) = \max_u \; r^\top \cdot u \tag{7.1}$$

$$s.t. \quad u \leq \sum_{\tau \in \mathcal{T}_t} D_\tau, \tag{7.2}$$

$$A \cdot u \leq x, \tag{7.3}$$

$$u \geq 0, \tag{7.4}$$

with $\mathcal{T}_t$ denoting the set of remaining time stages $t, \dots, 1$. The actual bid prices $\pi \in \mathbb{R}^M$ are the dual solution associated with the capacity constraints (7.3) in the primal problem. The dual problem is expressed by

$$V_t^{DLP_D}(x) = \min_{\gamma,\pi} \left(\sum_{\tau \in \mathcal{T}_t} D_\tau \right)^\top \cdot \gamma + x^\top \cdot \pi \tag{7.5}$$

$$s.t. \quad I_N \cdot \gamma + A^\top \cdot \pi = r, \tag{7.6}$$

$$\gamma, \pi \geq 0, \tag{7.7}$$

with I_N denoting the N-dimensional identity matrix. Theoretically, bid prices must be updated whenever any of the input parameters, such as expected demand or free inventory, change. So ideally, problem (7.5) would need to be updated after each sale or customer arrival. For large RM networks, as for airlines, this is computationally infeasible. Therefore, in most practical settings, bid prices are updated on a daily or weekly schedule. In between, all product requests, which fulfill the bid price condition, are accepted.

7.1.1 Small Example Network

Let us illustrate the results of the different optimization models on a small airline network consisting of three airports A, B and C, as shown in Figure 7.2. The airline offers two price classes, high (H) and low (L), at each flight

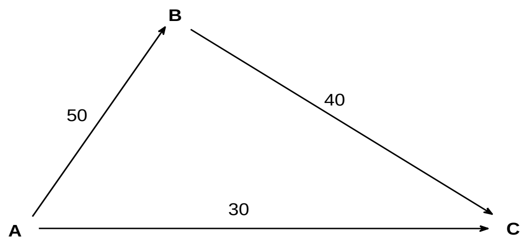

Figure 7.2. Example network with resource capacities.

and the booking horizon consists of 5 time stages, i.e., re-optimization steps. Table 7.1 shows all products with prices and deterministic demand.

Product	Route	Class	Price	$D_{\cdot,t=1}$	$D_{\cdot,t=2}$	$D_{\cdot,t=3}$	$D_{\cdot,t=4}$	$D_{\cdot,t=5}$
1	AB	H	150	4	4	3	2	2
2	AB	L	110	2	2	2	3	3
3	BC	H	150	4	4	3	2	2
4	BC	L	110	2	2	2	3	3
5	ABC	H	280	4	4	3	2	2
6	ABC	L	230	3	3	3	4	4
7	AC	H	350	5	5	3	3	2
8	AC	L	300	3	3	3	4	4

Table 7.1. Example airline products.

The theoretic optimal obtainable revenue for this RM network is 22000. If we solve the problem (7.5) at the beginning of each time stage we only generate a revenue of 20680, which is equivalent to 94% of the optimal revenue. The computed bid prices and revenues per time stage are shown in Table 7.2.

Time stage	Revenue	π_{AB}	π_{BC}	π_{AC}
5	1860	110	120	300
4	2240	80	150	300
3	4600	0	110	0
2	5880	0	110	0
1	6100	0	0	0
total	20680	-	-	-

Table 7.2. DLP results on the example network.

7.2 Time Dependent Bid Price Model (TDB)

The DLP may be a poor model of the real underlying problem, but it is appealing, because it's easy understandable and computational very efficient even for large scale problems. Our proposed TDB model is to a certain degree extending the DLP. Both models allocate product demand to resources. But the TDB considers expected sales realizations per times t, which are captured in $Y_t \in \mathbb{R}^N$. The idea is to compute a set of bid prices $\pi_t \in \mathbb{R}^M$, to be used during time stage $t = T, \ldots, 1$, which produce near optimal sales realization. Remember when we use a bid price control, a product j in time period t will be accepted when $r_j \geq \sum_{i \in A_j} \pi_{i,t} = A_j^\top \pi_t$. We define $Y_{j,t} \in \{0, D_{j,t}\}$, i.e., we accept all requests for product j at time t or none. Next, we need to introduce assignment variables $Z \in \{0,1\}^{N \times T}$, with $Z_{j,t} = 1$ denoting that product j is available at time t and zero otherwise. The bid price control at time t is modeled by the following equation

$$r = A^\top \pi_t + U_t - V_t. \tag{7.8}$$

The additional slack variables $U_t \in \mathbb{R}_+^N$ represent a positive difference between product price and the sum of bid prices

$$U_{j,t} = \begin{cases} r_j - A_j^\top \pi_t & \text{, if } Z_{j,t} = 1 \\ 0 & \text{, else} \end{cases}, \tag{7.9}$$

and $V_t \in \mathbb{R}_+^N$ represent the respective negative difference

$$V_{j,t} = \begin{cases} A_j^\top \pi_t - r_j & \text{, if } Z_{j,t} = 0 \\ 0 & \text{, else} \end{cases}. \tag{7.10}$$

The complete TDB model at time stage t and available capacity level x takes the following form

$$V_t^{TDB}(x) = \max_{Y,\pi} \sum_{\tau \in \mathcal{T}_t} r^\top \cdot Y_\tau \tag{7.11}$$

$$\text{s.t.} \quad A \cdot \left(\sum_{\tau \in \mathcal{T}_t} Y_\tau \right) \leq x \tag{7.12}$$

$$Y_\tau = D_\tau \cdot Z_\tau \qquad \forall \tau \in \mathcal{T}_t \tag{7.13}$$

$$r = A^\top \pi_\tau + U_\tau - V_\tau \qquad \forall \tau \in \mathcal{T}_t \tag{7.14}$$

$$U_\tau \leq \kappa \cdot Z_\tau \qquad \forall \tau \in \mathcal{T}_t \tag{7.15}$$

$$V_\tau \leq \kappa \cdot (1 - Z_\tau) \qquad \forall \tau \in \mathcal{T}_t \tag{7.16}$$

$$\epsilon \leq Z_\tau + V_\tau \qquad \forall \tau \in \mathcal{T}_t \tag{7.17}$$

$$Z_\tau \in \{0,1\}^N \qquad \forall \tau \in \mathcal{T}_t \tag{7.18}$$

$$U_\tau, V_\tau \geq 0 \qquad \forall \tau \in \mathcal{T}_t \tag{7.19}$$

with κ denoting a sufficient large constant and ϵ a small but strictly positive constant. Constraint (7.12) defines the capacity limit, whereas the demand allocation is defined in constraint (7.13). The bid price control is modeled in constraints (7.14)-(7.19). Constraint (7.17) ensures that $V_{j,\tau}$ and $Z_{j,\tau}$ can not simultaneously equal zero, according to the bid price policy a product is available if there is no negative difference between product price and the sum of utilized bid prices.

We solved the example network from Section 7.1.1 with the TDB model, generating a revenue of 21700 which equals 98.6% of the optimal revenue and improves the DLP revenue by 4.9%. The computed bid prices per time stage and the respective revenues are shown in Table 7.3.

Time stage	Revenue	π_{AB}	π_{BC}	π_{AC}
5	3390	81	150	0
4	4660	0	150	0
3	3700	0	150	301
2	4980	0	150	301
1	4970	111	150	0
total	21700	-	-	-

Table 7.3. TDB results on the example network.

So far, we assumed deterministic demand and disregarded the stochastic nature of the problem.

7.3 Stochastic TDB Extensions

The current section proposes scenario based stochastic programming (SP) extensions of the TDB model. Demand samples/ scenarios are used to approximate the demand uncertainty. The overall objective is to compute a single set of bid prices π, which maximizes the expected revenue over all scenarios. A scenario based SP approach is successfully applied to the car rental RM problem, see Haensel et al. (2012). This framework also allows to change the expected revenue objective into more risk averse formulation, such as maximizing the conditional value-at-risk (CVaR) revenue realization, to obtain more robust solutions. A related risk averse CVaR application to the problem of evaluating extraordinary large requests in a RM network is found in Haensel and Koole (2011c). The scenario approach allows to model complex stochastic problems, such as the scheduling in queueing systems as in Haensel et al. (2011a), which are often not tractable by Markov decision processes for realistic sized instances due to the large state space of the models. Let us return to our bid price RM problem. The demand scenarios $D^{(s)} \in \mathbb{R}^{N \times T}$ with $s = 1, \ldots, S$ are generated by means of simulation and each has a probability of $\frac{1}{S}$. We are aware of scenario reduction algorithms, e.g., as proposed by Heitsch and Römisch (2009), but our focus is on the SP formulation and not on advanced scenario generation techniques. The resulting sales realization are likewise scenario specific and similar denoted by $Y^{(s)}$. We want to compute an optimal set of bid prices π_τ, for all $\tau \in \mathcal{T}_t$, which maximizes the expected revenue

$$\frac{1}{S}\left(r^\top \sum_{s\in\mathcal{S}} \sum_{\tau\in\mathcal{T}_t} Y_\tau^{(s)}\right), \tag{7.20}$$

with $\mathcal{S}$ denoting the set of all scenarios $\{1, \ldots, S\}$. The problem with the consideration of stochastic demand is the capacity limiting constraint. Namely, the question on how to deal with potential overbookings in high demand scenarios resulting from the bid price policy, which accepts all requests for products with prices exceeding the sum of utilized bid prices. The classical definition of a bid price control assumes total knowledge of available capacities. We believe that this assumption is not always the case, especially when a firm works with a global network of sales channels and multiple sales agents simultaneously. In such cases, real time full information sharing can be a problematic issue. Hence, we will propose a model, the virtual overbooking model, which assumes this full capacity information and two other approaches, which assume that during a time stage the accept/reject decision is purely made according to the bid prices. Moreover, overbooking is a common RM practice and is usually done by simply inflating the capacity limit to some predefined overbooking level. The following three approaches allow a more sophisticated treatment.

7.3.1 Minimized Overbookings with Dynamic Bid Prices (MOD)

The idea is to correct for overbookings directly in the objective function. We introduce penalty costs λ_i for every sold unit of resource i exceeding the capacity. Therefore, we need to introduce scenario specific overbooking variables $W^{(s)} \in \mathbb{R}_+^M$. The resulting MOD formulation takes the following form

$$
\begin{aligned}
V_t^{MOD}(x) = \max_{Y,\pi} \quad & \frac{1}{S}\left(r^\top \sum_{s\in\mathcal{S}} \sum_{\tau\in\mathcal{T}_t} Y_\tau^{(s)} - \lambda^\top \sum_{s\in\mathcal{S}} W^{(s)} \right) && (7.21)\\
s.t. \quad & A \cdot \left(\sum_{\tau\in\mathcal{T}_t} Y_\tau^{(s)} - W^{(s)} \right) \le x && \forall s \in \mathcal{S}\\
& Y_\tau^{(s)} = D_\tau^{(s)} \cdot Z_\tau && \forall \tau \in \mathcal{T}_t, \forall s \in \mathcal{S}\\
& r = A^\top \pi_\tau + U_\tau - V_\tau && \forall \tau \in \mathcal{T}_t\\
& U_\tau \le \kappa \cdot Z_\tau && \forall \tau \in \mathcal{T}_t\\
& V_\tau \le \kappa \cdot (1 - Z_\tau) && \forall \tau \in \mathcal{T}_t\\
& \epsilon \le Z_\tau + V_\tau && \forall \tau \in \mathcal{T}_t\\
& Z_\tau \in \{0,1\}^N && \forall \tau \in \mathcal{T}_t\\
& U_\tau, V_\tau \ge 0 && \forall \tau \in \mathcal{T}_t\\
& W^{(s)} \ge 0 && \forall s \in \mathcal{S}.
\end{aligned}
$$

7.3.2 Restricted Overbookings with Dynamic Bid Prices (ROD)

Risk awareness is an increasing trend in corporations and has lead to a growing number of risk dependent operations research applications. As found for example in Haensel and Koole (2011c), on the problem of evaluating extraordinary large requests in network RM, or Haensel and Laumanns (2012), who study a problem of non-compliance risk in production planning. A well known risk measure from portfolio optimization is the conditional value-at-risk (CVaR), a coherent risk measure proposed by Artzner et al. (1999). It was proposed as an alternative risk measure to the widely used value-at-risk (VaR), which is defined for a continuous distributed random variable X by

$$
\mathrm{VaR}_{1-\alpha} = F^{-1}(\alpha) = \inf\{x | P(X \le x) \ge \alpha\}, \tag{7.22}
$$

where α denotes the risk level and $F(\cdot)$ the cumulative distribution function. Hence, the $\mathrm{VaR}_{1-\alpha}$ is simply the α-Quantile. Coherent risk measures satisfy the following four properties: translation invariance, sub-additivity, positive homogeneity and monotonicity. The VaR is generally not sub-additive, i.e., for two random variables X and Y it does not generally hold that $\mathrm{VaR}(X + Y) \leq \mathrm{VaR}(X) + \mathrm{VaR}(Y)$. The CVaR is a coherent risk measure and is for a continuous distributed random variable X defined by

$$\mathrm{CVaR}_{1-\alpha}(X) = \mathbb{E}[X | X \geq \mathrm{VaR}_{1-\alpha}]. \tag{7.23}$$

A graphic illustration of the VaR and CVaR of a random variable, representing loss realizations, is given in Figure 7.3. A second advantage of the CVaR, also

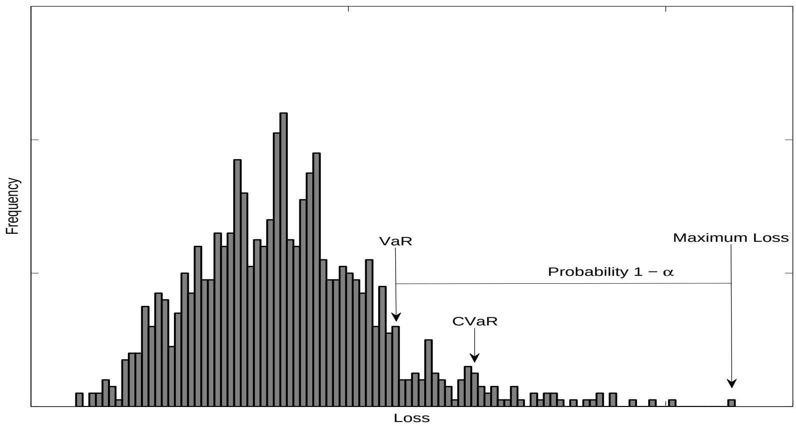

Figure 7.3. Example loss distribution with VaR and CVaR.

called mean shortfall, is that it provides an estimate of the expected outcome in the $(1-\alpha)\%$ extreme cases. Whereas the VaR only indicates that the $(1-\alpha)\%$ extreme outcomes will take values above the VaR, but without any additional information on the potential impact of the extreme realizations. Rockafellar and Uryasev (2002) show that the $\mathrm{CVaR}_{1-\alpha}$ for general, even discrete, distributed random variables X, can be efficiently computed by

$$\min_{w} \quad w + \frac{1}{1-\alpha}\mathbb{E}[\max\{0, X - w\}], \tag{7.24}$$

the solution of w provides the corresponding $\mathrm{VaR}_{1-\alpha}$ of the underlying distribution. In contrast to the MOD approach, our objective is not to penalize overbookings in the objective function, but to bound the expected overbookings at a certain risk level α by some predefined overbooking limit b. In other words we introduce a CVaR constraint on the overbooking realizations of the demand scenarios under a bid price control π. The LP formulation of equation (7.24) takes the following form

$$W + \frac{1}{1-\alpha}\frac{1}{S}\sum_{s\in\mathcal{S}} \eta^{(s)} \leq b \tag{7.25}$$

$$A \cdot \left(\sum_{\tau=1}^{T} Y_t^{(s)}\right) - c - W \leq \eta^{(s)} \tag{7.26}$$

$$0 \leq \eta^{(s)}, \tag{7.27}$$

with $W \in \mathbb{R}^M$ denoting the resource dependent VaR variable , $b \in \mathbb{R}^M$ the predefined overbooking resource limits and $\eta^{(s)} \in \mathbb{R}^M$ represents a scenario dependent slack variable. The complete ROD model is written as

$$\begin{aligned}
V_t^{VOD}(x) = \max_{Y,\pi} \quad & \frac{1}{S}\left(r^\top \sum_{s\in\mathcal{S}}\sum_{\tau\in\mathcal{T}_t} Y_\tau^{(s)}\right) && (7.28)\\
s.t. \quad & A \cdot \left(\sum_{\tau\in\mathcal{T}_t} Y_t^{(s)}\right) - x - W \leq \eta^{(s)} && \forall s \in \mathcal{S}\\
& W + \frac{1}{1-\alpha}\frac{1}{S}\sum_{s\in\mathcal{S}} \eta^{(s)} \leq b &&\\
& Y_\tau^{(s)} = D_\tau^{(s)} \cdot Z_\tau && \forall \tau \in \mathcal{T}_t, \forall s \in \mathcal{S}\\
& r = A^\top \pi_\tau + U_\tau - V_\tau && \forall \tau \in \mathcal{T}_t\\
& U_\tau \leq \kappa \cdot Z_\tau && \forall \tau \in \mathcal{T}_t\\
& V_\tau \leq \kappa \cdot (1 - Z_\tau) && \forall \tau \in \mathcal{T}_t\\
& \epsilon \leq Z_\tau + V_\tau && \forall \tau \in \mathcal{T}_t\\
& Z_\tau \in \{0,1\}^N && \forall \tau \in \mathcal{T}_t\\
& U_\tau, V_\tau \geq 0 && \forall \tau \in \mathcal{T}_t\\
& \eta^{(s)} \geq 0 && \forall s \in \mathcal{S}.
\end{aligned}$$

It is also possible to extend the ROD model with penalty costs for overbookings, so that the CVaR on the costs associated with overbookings is bounded by some predefined value. An important question in the ROD approach is of course the choice of α and b values to be used in the computation. The "optimal" choice is problem specific and requires additional analysis of the actual problem case.

7.3.3 Virtual Overbookings with Dynamic Bid Prices (VOD)

We now assume that the booking control has always full information of the available capacities. So an actual overbooking can practically not happen. Still, we need to deal with high demand scenarios in our optimization model. Problems arise, when we have a scenario s and an available product j, i.e., $r_j \geq \sum_{i \in A_j} \pi_{i,t}$, and there is not enough capacity left, i.e, for some resource i the $(A^i)(\sum_{\tau > t} Y_\tau^{(s)} + D_{j,t}^{(s)}) > c_i$. For such cases, we introduce the notion of virtual overbookings $W_{j,t}^{(s)}$ on a product j at time t in demand scenario s, such that $Y_{j,t}^{(s)} + W_{j,t}^{(s)} = D_{j,t}^{(s)} Z_{j,t}$ with $(A^i) \sum_{\tau \geq t} Y_\tau^{(s)} \leq c_i$. The complete VOD formulation is given by

$$
\begin{aligned}
V_t^{VOD}(x) = \max_{Y,\pi} \quad & \frac{1}{S}\left(r^\top \sum_{s \in \mathcal{S}} \sum_{\tau \in \mathcal{T}_t} Y_\tau^{(s)} \right) && && (7.29)\\
s.t. \quad & A \cdot \left(\sum_{\tau \in \mathcal{T}_t} Y_\tau^{(s)} \right) \leq x && \forall s \in \mathcal{S} \\
& Y_\tau^{(s)} + W_\tau^{(s)} = D_\tau^{(s)} \cdot Z_\tau && \forall \tau \in \mathcal{T}_t, \forall s \in \mathcal{S} \\
& W_\tau^{(s)} \leq \kappa A^\top \bar{W}_\tau^{(s)} && \forall \tau \in \mathcal{T}_t, \forall s \in \mathcal{S} \\
& \bar{W}_{\tau-1}^{(s)} \geq \bar{W}_\tau^{(s)} && \forall \tau \in \mathcal{T}_t, \forall s \in \mathcal{S} \\
& AY_{\tau-1}^{(s)} \leq \kappa(1 - \bar{W}_\tau^{(s)}) && \forall \tau \in \mathcal{T}_t, \forall s \in \mathcal{S} \\
& r = A^\top \pi_\tau + U_\tau - V_\tau && \forall \tau \in \mathcal{T}_t \\
& U_\tau \leq \kappa \cdot Z_\tau && \forall \tau \in \mathcal{T}_t \\
& V_\tau \leq \kappa \cdot (1 - Z_\tau) && \forall \tau \in \mathcal{T}_t \\
& \epsilon \leq Z_\tau + V_\tau && \forall \tau \in \mathcal{T}_t \\
& Z_\tau \in \{0,1\}^N && \forall \tau \in \mathcal{T}_t \\
& \bar{W}_\tau^{(s)} \in \{0,1\}^M && \forall \tau \in \mathcal{T}_t, \forall s \in \mathcal{S} \\
& U_\tau, V_\tau \geq 0 && \forall \tau \in \mathcal{T}_t \\
& W_\tau^{(s)} \geq 0 && \tau \in \mathcal{T}_t, \forall s \in \mathcal{S}.
\end{aligned}
$$

Customer demand exceeding the capacity is only virtually assigned and we do not gain revenue from it nor does it reduce available resources. When a resource i is virtually overbooked, i.e., there is a scenario s and time t with $(A^i) \sum_{\tau \geq t} D_\tau^{(s)} Z_\tau > c_i$, we can not allocate any additional demand to it.

Therefore we introduce a indicator variable $\bar{W}^{(s)} \in \{0,1\}^{M \times T}$ with

$$\bar{W}^{(s)}_{i,t} = \begin{cases} 1 & \text{, if } (A^i) \sum_{\tau \geq t} D^{(s)}_{\tau} Z_{\tau} \leq c_i \\ 0 & \text{, else.} \end{cases} \tag{7.30}$$

Once a resource is virtually overbooked, it retains in the status for the rest of the booking horizon.

For a detailed comparison study of all above mentioned methods, also with a reformulated Lagrangian relaxation method based on Kunnumkal and Topaloglu (2010), we refer to Bijvank et al. (2011). The focus of this book is on a new choice-base demand model, namely the choice-set approach. An extension of the TDB for choice-set demand is presented in the next section.

7.4 Choice-set based TDB Model (CTDB)

Up until now, all models in this chapter did not consider the customer's choice behavior under offered alternatives and assumed independent product demand. This section extends the TDB model to incorporate choice behavior. Demand information is given in the form of choice-sets, as introduced in Chapter 2, i.e., the demand matrix $D \in \mathbb{N}^{C \times T}$ is now given per choice-sets and time stages, with C denoting the number of choice-sets. We further assume the market to have customer segments with hierarchical preference. A small example will illustrate this assumption: Let us consider a flight with two fare classes F_1 and F_2, with price of F_1 being less than price of F_2. Remember that choice-sets are sets of substitutable products with a strict decreasing preference order, representing interest and the willingness to buy of a certain customer group. All possible choice-sets in our example are: $\{F_1\}, \{F_2\}, \{F_1, F_2\}$ and $\{F_2, F_1\}$. In a market with hierarchical customer preference, we assume that the customer group represented by choice-set $\{F_2, F_1\}$ does not exist or can be neglected from the consideration. In other words, customers can only have hierarchical preferences in one direction, i.e. you can prefer the low fare class over the high fare class, but not vice versa.

Choice-sets are represented by a binary matrix $CS \in \{0,1\}^{C \times \hat{N}}$, with $\hat{N} = N + 1$ and entries of the matrix are interpreted as follows

$$\mathrm{CS}_{i,j} = \begin{cases} 1 & \text{, choice-set } i \text{ contains product } j \\ 0 & \text{, choice-set } i \text{ does not contain product } j. \end{cases}$$

The number of products is increased by one to account for the non-purchase decision, the additional virtual product has a zero utility for the customer and represents the least favored choice. The product price vector r is also increased by one dimension and $r_{\hat{N}} = 0$, since the non-purchase decision generates no revenue. Also, we need to expand the resource product matrix with a zero column for $A_{\hat{N}}$ as the non-purchase decision consumes no resource capacities.

The hierarchical preference assumption allows us to order our products such that the CS matrix is in a very intuitive form. Namely, the preference order of the products within a choice-set can be read in decreasing order from left to right. In mathematical terms, there exists a product ordering, with indexes given by the one-to-one mapping $ind(\cdot)$, such that, if products k and l are contained in a choice-set i and product k is preferred over l, we have that $\mathrm{CS}_{i,k} = 1$ and $CS_{i,l} = 1$ and $ind(k) < ind(l)$. The CS matrix of our initial two fare class example is shown in Table 7.4, with $\emptyset$ denoting the non-purchase decision.

	F_1	F_2	$\emptyset$
$\{F_1, \emptyset\}$	1	0	1
$\{F_2, \emptyset\}$	0	1	1
$\{F_1, F_2, \emptyset\}$	1	1	1

Table 7.4. CS matrix of initial two fare class example with hierarchical preferences.

Please note that the hierarchical preference assumption is not a necessary condition, more a nice technical assumption, which is often found in practical problems. The assumption can be relaxed and we can still write the CS matrix in its intuitive decreasing preference from from left to right. But the CS definition is not as straightforward and we need to increase the product set with duplicate entries, as shown in Table 7.5.

	F_1	F_2	F_1	$\emptyset$
$\{F_1, \emptyset\}$	1	0	0	1
$\{F_2, \emptyset\}$	0	1	0	1
$\{F_1, F_2, \emptyset\}$	1	1	0	1
$\{F_2, F_1, \emptyset\}$	0	1	1	1

Table 7.5. CS matrix of initial two fare class example without hierarchical preferences.

An ordered product set for our example network, is given by

$$\begin{aligned}\mathcal{P} = \{&A \rightarrow B \quad L, \quad A \rightarrow B \quad H,\\ &B \rightarrow C \quad L, \quad B \rightarrow C \quad H,\\ &A \rightarrow B \rightarrow C \quad L, \quad A \rightarrow B \rightarrow C \quad H,\\ &A \rightarrow C \quad L, \quad A \rightarrow C \quad H, \quad \emptyset\},\end{aligned}$$

The low fare class is always preferred to the high fare class and the last product represents the non-purchase decision. An example choice-set matrix CS with nine choice-sets is shown in Table 7.6.

	AB L	AB H	BC L	BC H	ABC L	ABC H	AC L	AC H	$\emptyset$
$c = 1$	1	0	0	0	0	0	0	0	1
$c = 2$	1	1	0	0	0	0	0	0	1
$c = 3$	0	0	1	0	0	0	0	0	1
$c = 4$	0	0	1	1	0	0	0	0	1
$c = 5$	0	0	0	0	1	0	0	0	1
$c = 6$	0	0	0	0	1	1	0	0	1
$c = 7$	0	0	0	0	1	1	1	0	1
$c = 8$	0	0	0	0	0	0	1	0	1
$c = 9$	0	0	0	0	0	0	1	1	1

Table 7.6. CS matrix of the example network.

Choice-set 7, with its equivalent notation $\{A \rightarrow B \rightarrow C \quad L, \quad A \rightarrow B \rightarrow C \quad H, \quad A \rightarrow C \quad L, \quad \emptyset\}$, represents customers who are interested in a flight from A to C and are mainly price sensitive with a maximum willingness to pay strictly less than 350. They prefer the low price connecting flight over the high priced connecting flight, the latter is preferred over the low priced direct flight and if none of the three is available these customers do not buy any product.

Next, we need to model the allocation of demand per choice-set into sales per products, according to the booking control. This can be referred to as the constraining of demand, whereas the unconstraining in Chapter 2-4 denotes the process of retrieving demand information from given product sales data.

The dimension of the binary assignment variable Z is extended to account for the choice-set dimension, i.e., $Z \in \{0,1\}^{C \times \hat{N} \times T}$. Z indicates the product each customer will choose, according to the corresponding choice-set and the booking control. The product allocation constraint (7.13) in the TDB model, is rewritten into

$$Y_{p,\tau} = \sum_{c \in \mathcal{C}} D_{c,\tau} \cdot \mathrm{CS}_{c,p} \cdot Z_{c,p,\tau} \qquad \forall p \in \mathcal{P}, \forall \tau \in \mathcal{T}_t, \tag{7.31}$$

with $\mathcal{C} = 1, \ldots, C$ denoting the set of all choice-sets and $\mathcal{P} = 1, \ldots, \hat{N}$ the set of all products. This explains why we had to add an additional virtual product representing the non-purchase decision. Customers can not be lost in the way that they make no decision nor can they make multiple simultaneous choices, i.e., they are always assumed to buy exactly one of the offered products or to make the non-purchase decision. The according constraints are

$$\sum_{p \in \mathcal{P}} Z_{c,p,\tau} = 1 \qquad \forall c \in \mathcal{C}, \forall \tau \in \mathcal{T}_t \tag{7.32}$$

$$\sum_{c \in \mathcal{C}} Z_{c,p,\tau} \leq \hat{N} \cdot \hat{Z}_{p,\tau} \qquad \forall p \in \mathcal{P}, \forall \tau \in \mathcal{T}_t, \tag{7.33}$$

the variable $\hat{Z} \in \{0,1\}^{\hat{N} \times T}$ indicates the availability or non-availability of products per time stages, denoted by one or zero respectively. Constraint (7.32) ensures that each choice-set is assigned to exactly one product and constraint (7.33) assures that $\hat{Z}_{p,t} = 1$ if at least one choice-set is assigned to product p at time period t.

The strict preference structure of the choice-sets is modeled with the help of matrix $B \in \{-1,0,1\}^{\hat{N} \times \hat{N}}$, having the following structure

$$B_{i,j} = \begin{cases} 1 & \text{, if } i = j \\ -1 & \text{, if } i > j \\ 0 & \text{, if } i < j. \end{cases}$$

Using matrix B, we can ensure that a choice-set is always assigned to the available product with the highest preference within the choice-set. For all combinations of choice-set c, product p and time stages τ we need to force

$$\sum_{q \in \mathcal{P}} \hat{Z}_{q,\tau} \cdot B_{p,q} \cdot \mathrm{CS}_{c,q} \leq Z_{c,p,\tau}. \tag{7.34}$$

The complete CTDB model takes the following form:

$$
\begin{aligned}
V_t^{CTDB}(x) = \max_{Y,\pi} \quad & \sum_{\tau \in \mathcal{T}_t} r^\top \cdot Y_\tau && && (7.35)\\
s.t. \quad & A \cdot \left(\sum_{\tau \in \mathcal{T}_t} Y_\tau \right) \leq x \\
& Y_\tau = \sum_{c \in \mathcal{C}} D_{c,\tau} \cdot \mathrm{CS}_c \cdot Z_{c,\cdot,\tau} && \forall \tau \in \mathcal{T}_t \\
& \sum_{p \in \mathcal{P}} Z_{c,p,\tau} = 1 && \forall c \in \mathcal{C}, \forall \tau \in \mathcal{T}_t \\
& \sum_{c \in \mathcal{C}} Z_{c,p,\tau} \leq \hat{N} \cdot \hat{Z}_{p,d} && \forall p \in \mathcal{P}, \forall \tau \mathcal{T}_t \\
& \sum_{q \in \mathcal{P}} \hat{Z}_{q,\tau} \cdot B_{p,q} \cdot \mathrm{CS}_{c,q} \leq Z_{c,p,\tau} && \forall c \in \mathcal{C}, \forall p \in \mathcal{P}, \forall \tau \in \mathcal{T}_t \\
& r_p = A_{p,\cdot}^\top \cdot \pi_d + U_{p,\tau} - V_{p,\tau} && \forall p < \hat{N}, \forall \tau \in \mathcal{T}_t \\
& U_\tau \leq \kappa \cdot \hat{Z}_\tau && \forall p < \hat{N}, \forall \tau \in \mathcal{T}_t \\
& V_\tau \leq \kappa \cdot (1 - \hat{Z}_\tau) && \forall p < \hat{N}, \forall \tau \in \mathcal{T}_t \\
& \epsilon \leq \hat{Z}_\tau + V_\tau && \forall p < \hat{N}, \forall \tau \in \mathcal{T}_t \\
& 0 \leq U_\tau, V_\tau && \forall \tau \in \mathcal{T}_t \\
& Z \in \{0,1\}^{C \times \hat{N} \times t} \\
& \hat{Z} \in \{0,1\}^{\hat{N} \times t}
\end{aligned}
$$

The new bid price constraints are only working on the real products $p = 1, \ldots, N$, the non-purchase option is by default available if the bid price control closes all real products.

Similar to the TDB, we can also extend the CTDB for stochastic demand input per choice-set.

An interesting question is of course, how the CTDB performs compared to the TDB, when the choice-sets only consist of single products. So products are no substitutes and there are no buy-up possibilities in the market. The CTDB should then generate the same revenue as the the TDB. The according choice-set definition matrix CS is given in Table 7.7. The respective demand matrix is simply the product demand matrix, since choice-sets are in this case equivalent to products.

	AB L	AB H	BC L	BC H	ABC L	ABC H	AC L	AC H	$\emptyset$
$c = 1$	1	0	0	0	0	0	0	0	1
$c = 2$	0	1	0	0	0	0	0	0	1
$c = 3$	0	0	1	0	0	0	0	0	1
$c = 4$	0	0	0	1	0	0	0	0	1
$c = 5$	0	0	0	0	1	0	0	0	1
$c = 6$	0	0	0	0	0	1	0	0	1
$c = 7$	0	0	0	0	0	0	1	0	1
$c = 8$	0	0	0	0	0	0	0	1	1

Table 7.7. CS matrix of the example network without substitution.

As expected, the CTDB generates a revenue of 21700, same as the TDB. The results are shown in Table 7.8. The CTDB bid prices differ from the TDB ones, but this is only because the example does not have a unique bid price solution.

Time stage	Revenue	π_{AB}	π_{BC}	π_{AC}
5	3390	110	121	0
4	4660	0	111	0
3	2790	111	150	301
2	4980	0	150	301
1	5880	0	150	0
total	21700	-	-	-

Table 7.8. CTDB results on the example network with no substitutable products.

7.5 Summary

In this chapter, we developed a new time dependent bid price approach, which allows for an optimal acceptance control of multiple arriving customers per time stage. The deterministic TDB model is expanded with three stochastic programming approaches, to incorporate stochastic demand information in the form of scenarios in the optimization problem. The different approaches reflect different treatments of potential overbookings in the sales process. The time dependent bid price model is further extended to consider the customer's choice behavior and to account for customer preferences. More precisely, we extended the TDB model to work with choice-set demand. A simulation test of the resulting CTDB on markets with different choice behavior, i.e., different degrees of buy-up/ substitution possibilities, is presented in following chapter.

Chapter 8

A Combined Simulation Test

The objective of this chapter is to test the combination of all three aspects of an RM system (unconstraining, forecasting and optimization) in a simulation study. In the previous parts, the focus was always only on one aspect at a time. Naturally, revenue management is an integrated and dynamic process where unconstraining, forecasting and optimization are simultaneously performed. The goal of the chapter is first to provide an example of integrating choice-set demand into all RM aspects, and second to show the performance potential of a choice-set based revenue management system. A simulation study is chosen to really evaluate the behavior and performance of the proposed choice-set demand model. In case studies on real world data, we never know the exact underlying choice and demand process and thus there is always the question if the observed performance is due the tested model or just by chance. In a simulation study, we know the underlying choice and demand process. So, the estimation accuracy can really be measured, and potential revenue gains for different optimization models can be assessed. Another advantage of a simulation study is to test the performance on markets with different characteristics, e.g., markets with high or low buy-up potential.

8.1 Simulation Setup

This section will explain the structure and steps of the simulation tool and the optimization models used to compute the booking controls. The simulation consist of a loop over the following steps:

1. generate demand per choice-set from known parameter,
2. bid price booking control to obtain availability and sales data,
3. unconstrain the sales data to obtain information on choice-set demand.

The objective is twofold: first, to measure how well the unconstraining method estimates the underlying demand process, and second, to estimate the potential revenue gain by implementing a choice-set based RM system.

The optimization models, which compute the bid prices for the booking control to obtain sales data, are based on demand forecasts on choice-set level. But the demand forecast is made on the basis of historical demand data and this demand data is obtained by unconstraining historical sales data. Thus, we have a knowledge problem: we need sales data as input in the unconstraining, but to obtain sales data we need some information on the demand. Giving the optimization models a priori perfect knowledge of the demand process will adulterate the test results. Applying a very stupid booking control on the other hand, e.g., offer only one product or even offer nothing, will result in sales data which does not contain significant information on the specific demand process. Also, current booking controls in practice are already quite sophisticated. So, to overcome this information problem, we decided to divide the simulation test into two periods, namely the learning and the testing period.

In the learning period, we generate sufficient historical sales to base the forecast of unconstrained choice-set demand on a good foundation. To obtain sales realizations, we need an initial booking control to constrain the choice-set demand into sales per products. Within the learning period, we apply the deterministic linear program (DLP) to obtain the bid prices. The DLP, as described in Talluri and van Ryzin (2004b), takes the following form

$$\max_x \; r^T x \tag{8.1}$$
$$\text{s.t.} \quad Ax \leq c \tag{8.2}$$
$$0 \leq x \leq \mathrm{E}[D], \tag{8.3}$$

where r denotes the revenue vector, A the resource-product matrix, c the resource capacity vector, D the demand per product and x represents the optimal capacity allocation or booking limits per product. The bid prices are the dual variables corresponding to constraint (8.2). The DLP does not consider choice behavior of customers, it assumes the demand to be independent between products. We want to give the model some information on the choice behavior. Hence, the expected product demand for the DLP is therefore calculated such that 50% of the choice-set demand goes to the product with the highest preference and the remaining 50% are added to the product demand of the second preferred product. The DLP has therefore some buy-up knowledge of the demand. In case the choice-set consists only of one product, the whole demand is added to this product. Within the learning period we do not forecast demand. The DLP is solved based on the expected demand information, as described above. Sales and availability data is generated and further unconstrained to obtain estimates of the choice-set demand curves.

In the test period, we forecast choice-set demand based on historic data. The forecasted demand information is used in the optimization models to compute the bid prices for the booking control. So, the test period represents a practical situation, when we have historic sales data on hand. The question is how well a choice-set based RM system performs. The iterations in the test period consist of the following steps:

1. forecast the demand curve for each choice set,
2. compute bid prices based on the forecast,
3. sales realizations are generated based on the bid price acceptance control,
4. choice-set demand curves are estimated from sales and availability data.

The forecasting is performed as described in Chapter 6. Although, the input is not a booking curve of reservation data, but booking horizons of demand estimates of independent choice-sets. Hence, the methodology is essentially the same, we only apply it separately for each choice-set. All previous estimated demand curves are combined to a data matrix, such that each row corresponds to a booking horizon, compare with Chapter 6. Having this matrix, with the dimension of observed booking horizons times days in the booking horizon, we can perform a dimension reduction by singular value decomposition (SVD). SVD provides us with a set of base vectors and a time series of weights, such that the linear combination of the base vectors with the corresponding weights gives a reduced dimensional approximation of the demand matrix. We will work with $K = 1$, i.e., only one base vector. The actual forecast is obtained by applying Holt-Winters method on the weights time series. The forecast of the entire future booking horizon is then obtained by multiplying the forecasted weights with the base vector. We do not perform forecast updating within not completed booking horizons for several reasons. First, the impact of forecast updating is studied in Chapter 6. Second, the simulated demand quantities within different parts of the booking horizon are uncorrelated in the simulation. And third, the focus of the simulation test is on the unconstraining accuracy and the potential revenue gains of the choice-based optimization models and not on an explicit forecasting method.

Figure 8.1 illustrates the two simulation parts and the components of the complete simulation process.

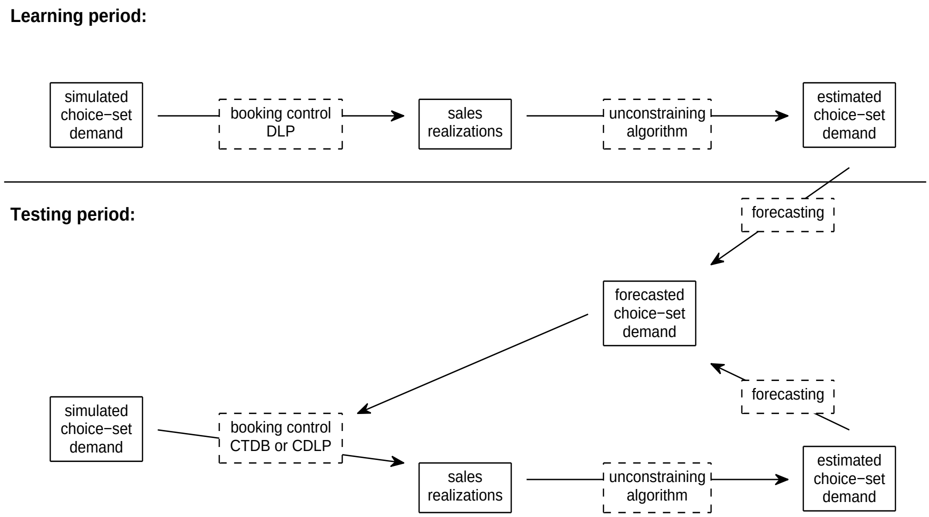

Figure 8.1. Simulation process diagram.

Our proposed optimization model is the choice-set based time dependent bid price model (CTDB), as introduced in Chapter 7. The performance of the CTDB is compared by the mean generated revenue improvements over the DLP and the choice based extension of the DLP (CDLP), which is described in van Ryzin and Liu (2008). Let $\mathcal{N}$ denote the set of all products, thus we can denote a set of offered products by $S \subseteq \mathcal{N}$. Instead of computing booking limits per products, as in the DLP case, the CDLP aims to compute availability times $t(S)$ per considered offer-set $S \subseteq \mathcal{N}$. The customer arrival rate per time stage is denoted by λ. Each offer-set S is associated with a probability vector $P(S) \in [0,1]^{|\mathcal{N}|}$, referring to the sales probabilities per product under a specific offer-set S. Based on this we can compute the expected capacity consumptions of offer-set S by

$$Q(S) = AP(S), \tag{8.4}$$

and the expected revenue generated by set S

$$R(S) = \sum_{j \in \mathcal{N}} r_j P_j(S). \tag{8.5}$$

The CDLP takes now the following form

$$\begin{aligned}
\max{}_t \; & \sum_{S \subseteq \mathcal{N}} \lambda R(S) t(S) && (8.6)\\
\text{s.t.} \; & \sum_{S \subseteq \mathcal{N}} \lambda Q(S) t(S) \leq c && (8.7)\\
& \sum_{S \subseteq \mathcal{N}} t(S) \leq T && (8.8)\\
& t(S) \geq 0 \qquad \forall S \subseteq \mathcal{N}. && (8.9)
\end{aligned}$$

The solution $\hat{t}(S)$ provides us with the optimal time fraction set S should be offered. The bid price, i.e., the marginal price of capacity, is obtained by solving the dual problem

$$\begin{aligned}
\min{}_{\pi,\sigma} \; & \pi^T c + T\sigma && (8.10)\\
\text{s.t.} \; & \lambda \pi^T Q(S) + \sigma \geq \lambda R(S) \qquad \forall S \subseteq N && (8.11)\\
& \pi, \sigma \geq 0, && (8.12)
\end{aligned}$$

with π being the dual variable referring to the capacity constraint (8.7) and σ refers respectively to the time constraint (8.8). So π is used as the bid price. The number of possible offer-sets growth exponentially in the number of products. Talluri and van Ryzin (2004a) and van Ryzin and Liu (2008) propose to consider only *efficient sets* in the optimization model. Offer-sets are efficient, if there exist no mixture of other sets which produces a strictly greater revenue and consumes in expectation less or equal capacity. The number of efficient sets is considerably smaller than the number of all possible sets. Further, van Ryzin and Liu (2008) propose a column generation approach in order to make their model applicable for practical problem settings. Remember that choice-sets are represented by a binary matrix $CS \in \{0,1\}^{C \times (N+1)}$, the entries of the matrix are interpreted as follows

$$\mathrm{CS}_{i,j} = \begin{cases} 1 & \text{, choice-set } i \text{ contains product } j \\ 0 & \text{, choice-set } i \text{ does not contain product } j \end{cases},$$

with the $(N+1)^{th}$ product denoting the non-purchase decision. Offer-sets are similar defined by a binary matrix $O \in \{0,1\}^{(N+1) \times N}$, meaning

$$O_{i,j} = \begin{cases} 1 & \text{, offer-set } i \text{ contains product } j \\ 0 & \text{, offer-set } i \text{ does not contain product } j \end{cases}.$$

Since we implement a bid price based sales control, we only have to consider all bid price feasible sets to offer. Remember that a bid price control means that prices are associated with resources and products are available if their price exceeds the sum of bid prices over resources utilized by this product. So the set of all bid price feasible offer-sets ranges from the full offer-set, all products are available, to the empty offer-set, no products are available. Such bid price feasible sets can in many case be straightforward constructed and their number is considerably small. For example, in a single resource case with N products ordered by increasing prices, we have $N+1$ bid price feasible offer-sets. The matrix O takes then the following form

$$O = \begin{pmatrix} 1 & 1 & 1 & \cdots & 1 \\ 0 & 1 & 1 & \cdots & 1 \\ 0 & 0 & 1 & \cdots & 1 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & & \cdots & 0 & 1 \\ 0 & & \cdots & & 0 \end{pmatrix}.$$

The rows of O correspond to the offer-sets, starting with the set of offering all products and then closing the cheapest available in the next offer set. Until we end up with the empty offer-set, i.e., no products are offered.

8.2 Test Cases

The simulation test is performed on a general problem setting with one resource, e.g., one hotel or one flight, and five products/ classes. The prices of the products (Prod1, ..., Prod5) are fixed for all test cases and shown in Table 8.1. We further consider a three month booking horizon, i.e., 84 booking

Prod1	Prod2	Prod3	Prod4	Prod5
25	30	55	80	110

Table 8.1. Prices of products.

days. The horizon is split into 12 time stages, representing the booking weeks. Booking week 1 corresponds to the last booking week, i.e., the week including the usage time of the product. Respectively, booking week 12 represents the first booking week, i.e., three month prior usage of the product. The booking intensity is not constant over all weekdays and all booking weeks, the patterns are shown in Figure 8.2. The usage day of our products is a Monday, e.g.,

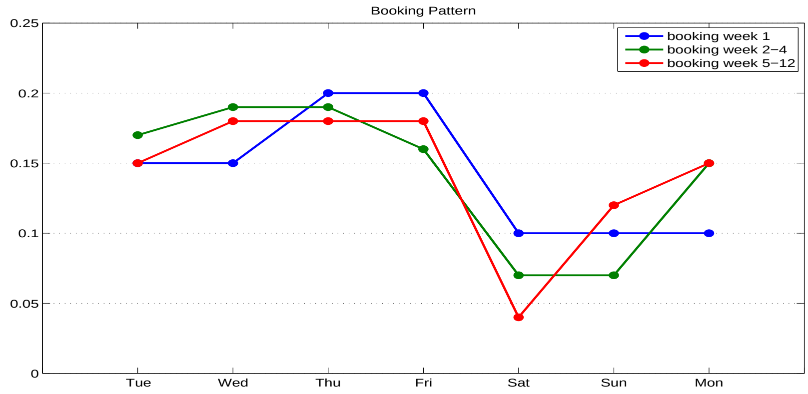

Figure 8.2. Final estimated demand curves per choice-set.

arrival at the hotel, and thus the booking week goes from Tuesday till Monday. Up on this basic problem structure, we test three different test cases. Each test case represents a certain market with different buy-up possibilities and different demand groups. Therefore, the choice-sets and their parameters are specific per test cases. The demand rates for choice-set with low priced products have always a flatter slope than choice-sets covering the higher priced segment. This reflects the observed behavior in practice that request for higher priced products occur closer to the usage time of the product, e.g., arrival at the hotel or departure of the airplane. Capacity is set to 170 in all test cases and the resulting load factor is always around 1.2.

Test case 1:
The first test case represents a market made up of five choice-sets with moderate buy-up possibilities. All choice-sets, except the first choice-set (CS1), cover two products and have thus one buy-up possibility. The choice-sets and the corresponding demand rate parameter are shown in Table 8.2. The un-

	Prod1	Prod2	Prod3	Prod4	Prod5	α	β
CS1	1	0	0	0	0	-0.05	6
CS2	1	1	0	0	0	-0.05	4
CS3	0	1	1	0	0	-0.10	10
CS4	0	0	1	1	0	-0.20	8
CS5	0	0	0	1	1	-0.30	6

Table 8.2. Test case 1: choice-sets definition and demand rate parameter.

constraining algorithm has full information on the choice-set definitions.

Test case 2:
In the second test case, we model a market with no buy-up possibilities, the choice-sets are shown in Table 8.3. The unconstraining algorithm has also information that there may be customers with buy-up possibilities in the market, choice-sets six and seven. We are interested to see if the estimation algorithm provides a good approximation of the real market characteristics, or if it puts weight on the dummy choice-set covering multiple products. The dummy choice-sets, i.e., the ones with buy-up, have almost zero demand rates.

	Prod1	Prod2	Prod3	Prod4	Prod5	α	β
CS1	1	0	0	0	0	-0.10	12
CS2	0	1	0	0	0	-0.15	10
CS3	0	0	1	0	0	-0.20	9
CS4	0	0	0	1	0	-0.30	8
CS5	0	0	0	0	1	-0.40	7
CS6	1	1	1	0	0	-0.05	0.01
CS7	1	1	1	1	1	-0.05	0.01

Table 8.3. Test case 2: choice-sets definition and demand rate parameter.

Test case 3:
In the third test case, we test the counter part to test case 2. Namely, we model a market with buy-up possibilities and add dummy choice-sets covering only single products. We want to test if the estimation algorithm approximates the characteristics right and puts no weight in the dummy choice-sets. The choice-sets are shown in Table 8.4.

	Prod1	Prod2	Prod3	Prod4	Prod5	α	β
CS1	1	1	0	0	0	-0.10	15
CS2	1	1	1	0	0	-0.15	13
CS3	0	1	1	1	0	-0.22	6
CS4	0	0	0	1	1	-0.30	5
CS5	0	0	1	0	0	-0.05	0.01
CS6	0	0	0	1	0	-0.05	0.01
CS7	0	0	0	0	1	-0.05	0.01

Table 8.4. Test case 3: choice-sets definition and demand rate parameter.

8.3 Numerical Results

The simulation tool is implemented in MATLAB R2011b and the optimization models are solved with Ilog CPLEX version 12.2. The computation is performed on a duo core (2.3 GHz) Windows machine with 3 GB of memory. We compute the bid prices at the beginning of each booking week and they are not updated during the week, i.e., the optimization models are solved 12 times per iteration. The CDLP and the CTDB are tested on the same demand realization, to ensure a fair comparison. The unconstraining and the forecasting of the choice-set demand is performed separately for both optimization models, since both models provide different bid prices and hence result in different availability and sales data. So the parameter estimation results of the unconstraining algorithm are given respectively for both optimization models. The DLP revenue results are obtained by averaging the generated revenue over the learning period. Further, the number of iterations in the learning and testing period is set to 100 each. The choice-set demand parameter are estimated by the algorithm proposed in Chapter 4 and the stopping criteria is set to maximum 100 iterations or a simultaneous satisfaction of the following tolerance bound on parameter changes: $\Delta\alpha$-tolerance 0.001, $\Delta\beta$-tolerance 0.01 and a Δ log-likelihood-tolerance of 0.1.

8.3.1 Results for test case 1

The demand estimation algorithm took in average 29 iterations and stopped after 11 seconds. The parameter estimates at each iteration in the test period are shown in Figures 8.3. We observe that the parameters are converging in all cases. Generally, the beta parameter seem to converge faster than the alpha parameter, but note that the scale of the betas is larger than for the alphas. For both optimization models, we observe that the parameters are not always converging to the real parameter. But even in cases when the estimated parameter values are generally not equal to the real demand parameter, the actual demand curve can be quite well approximated, as in the case of CS3. This observation coincides with our findings in Chapter 2 and Haensel and Koole (2011b) that multiple parameter can result in very similar demand curve. The last forecasted demand rates, i.e., forecasted at the last iteration in the testing period, are shown in Figure 8.4. The last forecasted curve contain all the so far gained information on the choice-set demand and are therefore a good measure for the learned demand behavior. The final demand curves have the same behavior for both optimization models, but are not equal. Especially

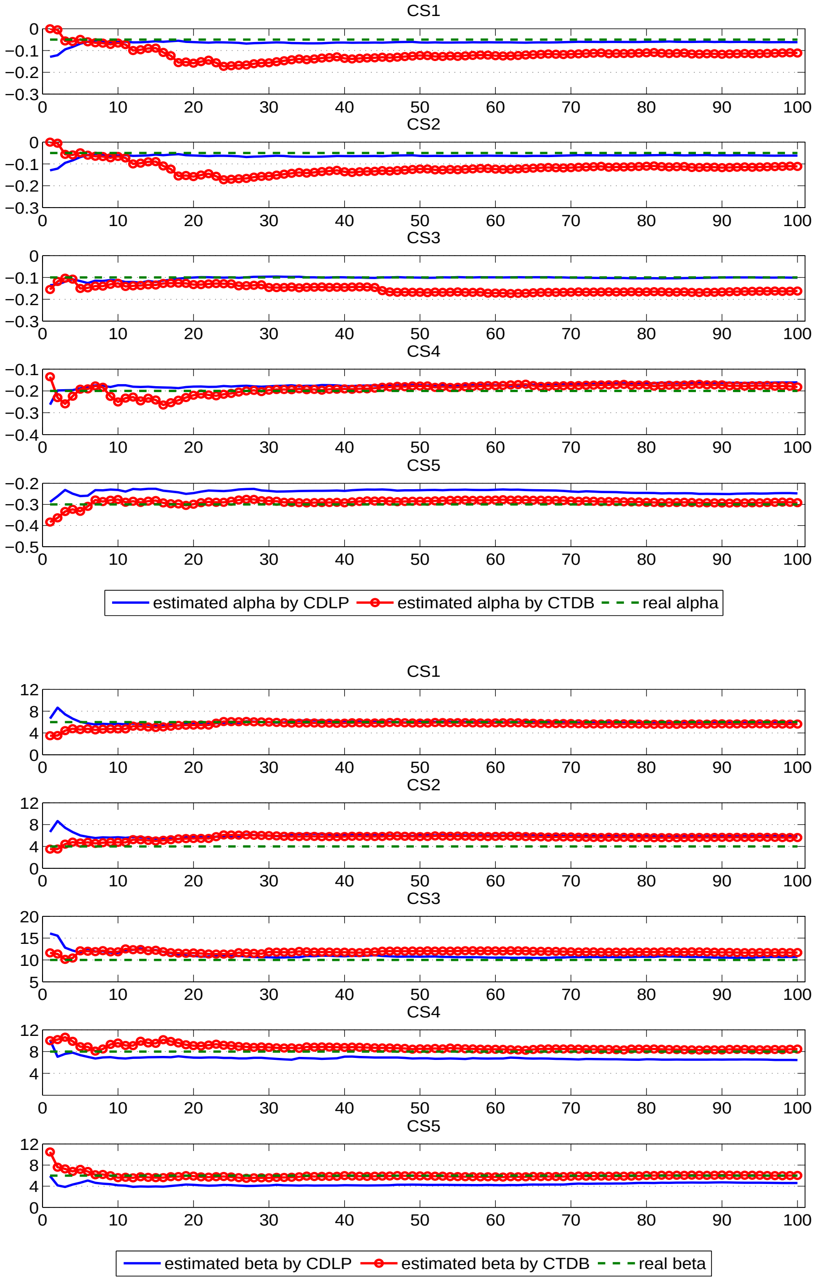

Figure 8.3. Test case 1: The averaged α and β estimates per choice-set and iteration.

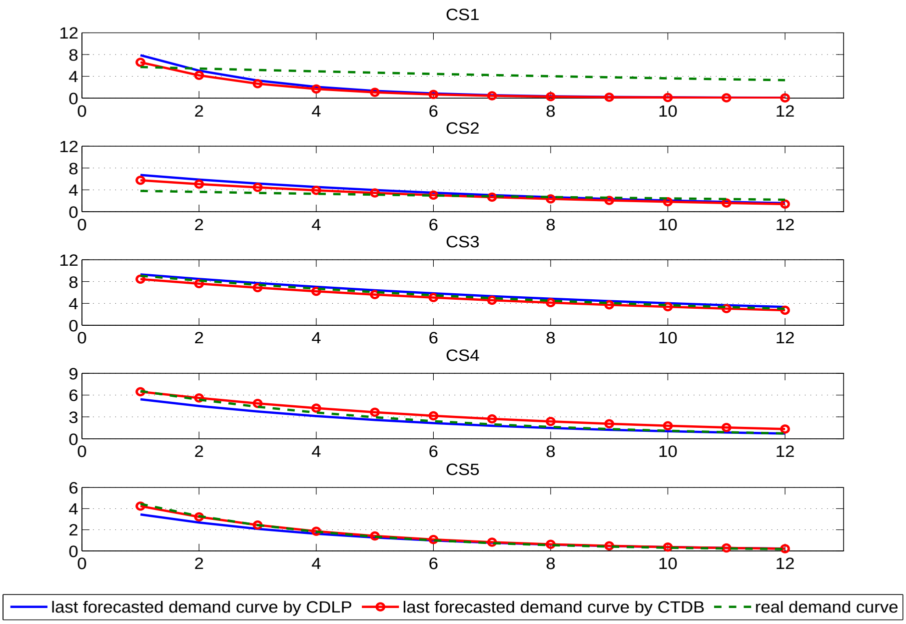

Figure 8.4. Test case 1: Final forecasted demand curves per choice-set.

for choice-sets CS4 and CS5 provide the estimated curves based on the CTDB data a better approximation. Comparing the demand curve approximations, we find that choice-sets CS3-CS5 are reasonably well approximated. CS2 is slightly overestimated towards the end of the booking horizon and CS1 is dramatically underestimated except for the last time stages. This is partly due to a lack of informative sales data. Table 8.5 shown the choice-set satisfaction ratios, i.e., the ratio between accepted and total customers per choice set. We

	CS1	CS2	CS3	CS4	CS5
DLP	0.61	0.89	0.92	0.91	0.88
CDLP	0.72	0.73	0.92	0.89	0.88
CTDB	0.29	0.28	1.00	0.99	1.00

Table 8.5. Test case 1: Choice-set satisfaction.

find that customer requests coming from CS1 and CS2, i.e., the lower segments, are more often rejected and are therefore more often unobserved. Whereas, customers of choice-sets CS3-CS5 are rarely rejected and the sales data is more informative about them. Both optimization models focus only on maximizing the expected revenue and do not consider a learning aspect. The combination of both is commonly known in Dynamic Pricing problems and referred to the struggle between exploration and exploitation. Our models, and most other network RM optimization approaches, only focus on the maximum revenue exploitation and ignore the demand exploration aspect. Let us now focus on the revenue generating performance of the optimization models. The mean obtained revenues and the percentage gains of the choice-based models over the initial DLP are shown in Table 8.6. The CDLP improves the DLP mean

	DLP	CDLP	CTDB
Mean revenue	6454	6694	6909
% gain over DLP	-	3.7	7.0
Capacity utilization	1	0.99	0.84

Table 8.6. Test case 1: Revenue performance of optimization models.

revenue by 3.7% and the CTDB even improves the mean DLP revenue by 7.0%. This shows that the revenue can be significantly improved by incorporating knowledge of the customers' choice behavior into the optimization model. A histogram of the generated revenues for the different optimization models is shown in Figure 8.5. The variance on the obtained revenue from the CTDB is

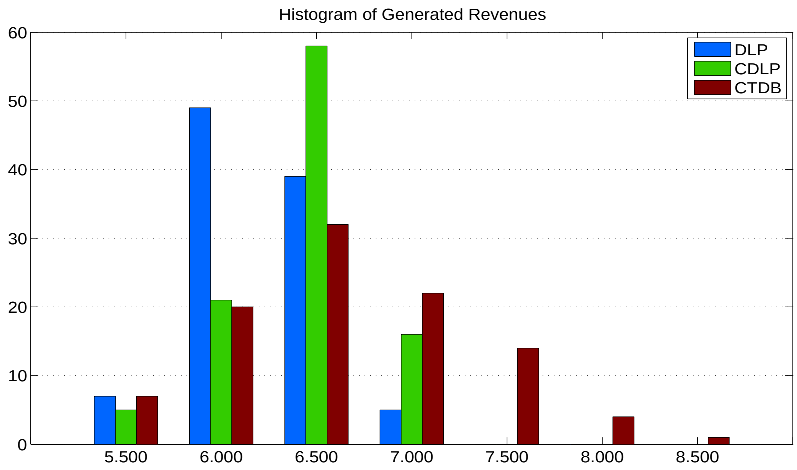

Figure 8.5. Test case 1: Histogram of generated revenues.

increased. But not in form of additional low revenue results, more in form of new realization with very high revenues. The mean capacity utilization, also given in Table 8.6, shows very different values for the CTDB and the other two models. The CTDB generates a significant lower capacity utilization, which is due to the advanced bid-prices. The CTDB knows that the bid-prices are not updated during a time stage, whereas the DLP and the CDLP only provide 'optimal' bid-prices for the current system state and do not consider that bid-prices are not updated within time stages. So the CTDB does a very good job with respect to the revenue maximization objective. It focuses on the customer segments at the higher price classes and rejects almost none of these customers. Companies do not want to reject customers interested in high priced product and since we usually have scarce capacity (load factor greater than one), the booking control should reject customers targeting the lower price classes more frequently.

8.3.2 Results for test case 2

For the second test case, a market with no buy-up possibilities, we observe in average that the unconstraining algorithm took 55 iterations and finished in 26 seconds. The last forecasted demand rates of the real contributing choice-sets are shown in Figure 8.6. CS1 is again underestimated, similar to the first test

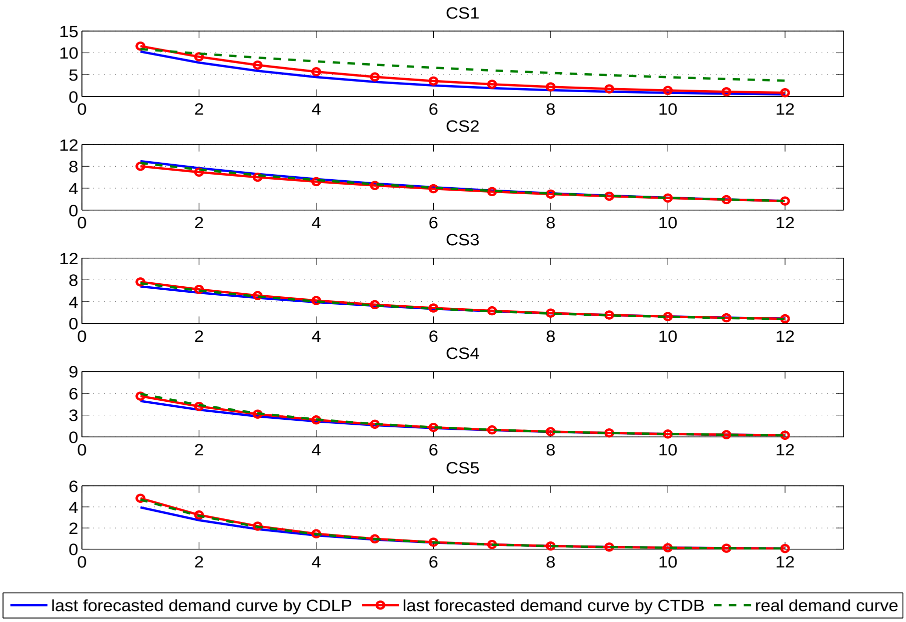

Figure 8.6. Test case 2: Final forecasted demand curves of real choice-sets.

case. The other four choice-sets, CS2 to CS5, are very close approximated by the unconstraining algorithm and the sales data based on both optimization models. The sales data from the CTDB produces slightly better approximation results for the choice-set CS4 and CS5. The final forecasted demand functions for the dummy choice-sets CS6 and CS7 are presented in Figure 8.7. Both

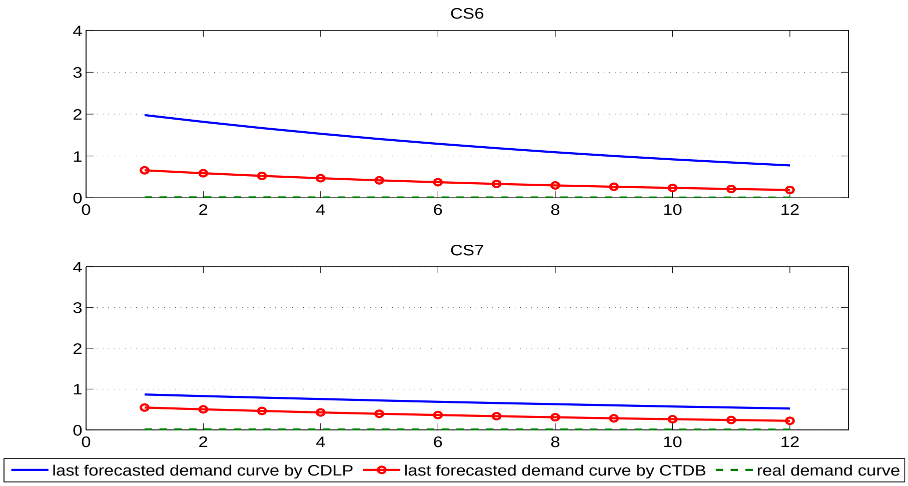

Figure 8.7. Test case 2: Final forecasted demand curves of dummy choice-sets.

choice-sets are estimated with small demand rates, but the estimates based on the CDLP data are significant larger than the ones based on CTDB sales data. Since the real contributing choice-sets are only consisting of one price, we expect that there is no gain in using a choice based optimization model, contrary to a market with buy-up, as in test case 1. This is precisely what we observe in the simulation test. The revenue results are shown in Table 8.7 and the histogram of generated revenues is displayed in Figure 8.8. We even find

	DLP	CDLP	CTDB
Mean revenue	7726	7594	7844
% gain over DLP	-	-1.7	1.5
Capacity utilization	0.997	0.998	0.947

Table 8.7. Test case 2: Revenue performance of optimization models.

that the CDLP generates lower revenue results than the DLP. This is probably caused by the overestimation of the dummy choice sets, which provides a wrong estimate of the demand in the market. The CTDB can increase the mean revenue by 1.5%. This revenue increase is due to the advanced time dependent bid price computation. Again, we observe that the capacity utilization with the CTDB is lowest for all three optimization model, which means that the CTDB is again most restrictive in the customer acceptance.

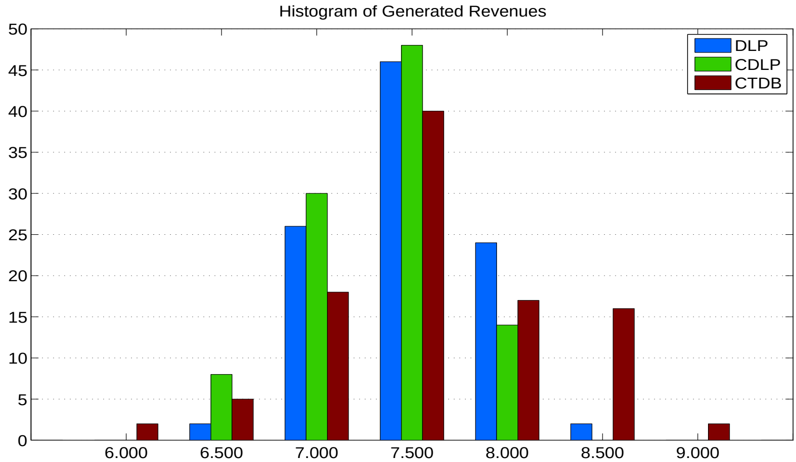

Figure 8.8. Test case 2: Histogram of generated revenues.

8.3.3 Results for test case 3

In the final case, we tested a market with significant buy-up and we added three dummy choice-sets covering only single products. The unconstraining algorithm took in average 52 iterations and 26 seconds. The last forecasted demand curves for the contributing choice-sets, i.e., the buy-up choice-sets CS1-CS4, are displayed in Figure 8.9. CS2 and CS3 are well approximated with the CDLP sales data. The estimation with the CTDB data provides only for the first and second choice-set a very close approximation. The final forecasted curves for the dummy choice-sets are displayed in Figure 8.10. Choice-sets CS5 and CS7 get no significant demand, they are at maximum approximated with 0.1. But CS6 is approximated with a significant demand rate. Comparing the results of CS4 and CS6, we find that the amount of underestimation of CS4 is almost identical with the demand estimated for CS6. Both choice-sets contain the forth product and the unconstraining algorithm wrongly accounts CS4 customers buying Prod4 to the dummy choice-set CS6. The obtained revenue results are shown in Table 8.8. With such a buy-up potential in the market, we expect the choice based optimization models to outperform the initial DLP.

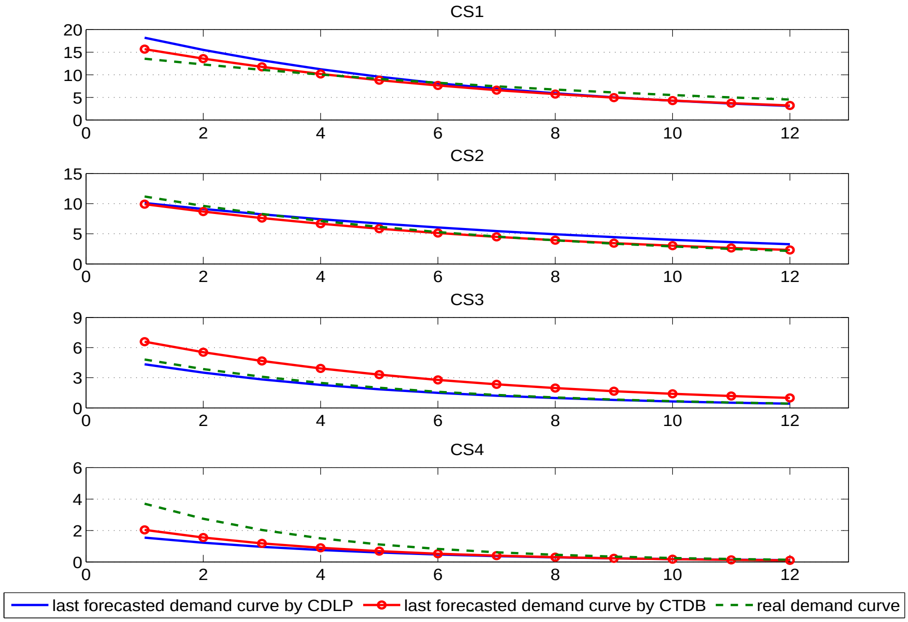

Figure 8.9. Test case 3: Final forecasted demand curves of real choice-sets.

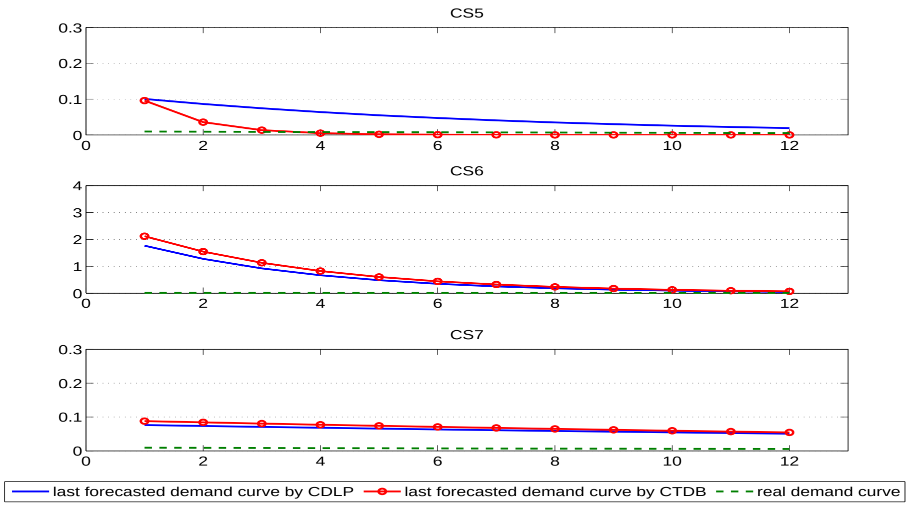

Figure 8.10. Test case 3: Final forecasted demand curves of dummy choice-sets.

	DLP	CDLP	CTDB
Mean revenue	5060	5526	6396
% gain over DLP	-	9.2	26.4
Capacity utilization	1.0	0.99	0.93

Table 8.8. Test case 3: Revenue performance of optimization models.

The CDLP improves the DLP results already by 9.2%, but the CTDB even improves the mean DLP revenue by 26.4%. The revenue histogram is shown in Figure 8.11. The CTDB worst case revenue is almost at the same level as best case DLP revenue results.

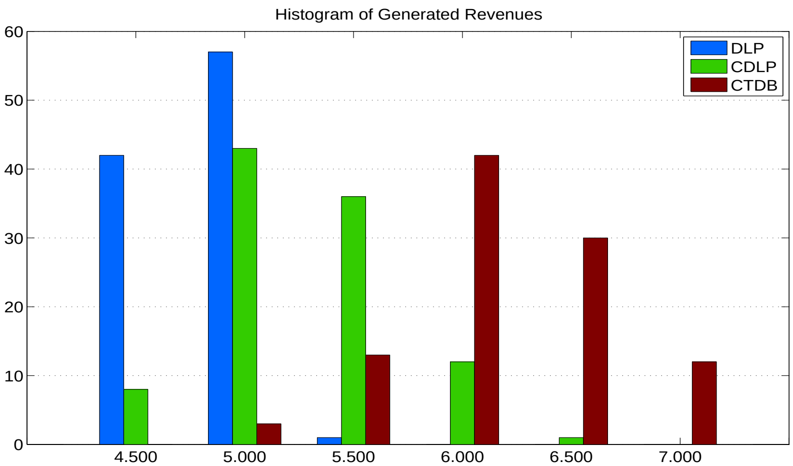

Figure 8.11. Test case 3: Histogram of generated revenues.

8.4 Conclusion

In this chapter, we tested the choice-set demand model in a combined simulation analysis of all three aspects in RM: unconstraining, forecasting and optimization. Let us start with the most relevant results for decision makers, the obtained revenues. The CTDB is the best revenue generating model in all three test cases. The CDLP model generates in test case one and three improved results over the pure DLP, but generates a slightly lower revenue in the second test case. Our general conclusion is that an application of a choice-based optimization model with consideration of existing buy-up possibilities

in the market results in large revenue improvements. The CTDB model generates on top of this choice consideration, a reasonable extra amount of revenue by using time dependent bid prices.
Coming to the choice-set demand estimation results, we find that the general demand pattern is overall well approximated. Even in cases with misspecified markets, i.e., by providing non existent dummy choice-set to the unconstraining algorithm, we observe that the dummy choice-sets do receive only very small demand rates. The unconstraining algorithm also converges relatively fast and is therefore applicable for practical problems. We find that the estimation results are dependent on the information contained in the sales data. Both optimization models compute different bid-prices, resulting in different sales realizations. In general, the demand rate approximations are better for choice-set which are more often accepted, and for which the sales data contains significant relevant information. If certain segments are very frequently rejected and are therefore often non-observable in the sales data, we can not hope to obtain very close approximations of the real demand curves. But this shortcoming is shared with any other estimation method, since an estimation method in always bounded by the relevant information contained in the underlying data. Optimization models are usually only aiming to maximize the obtainable revenue without any focus on the demand learning aspect. This is especially true for the third test case, where the CTDB improves the CDLP mean revenue by 16%. But the demand estimates in this case are slightly better for the CDLP sales data. Therefore, future research in the network RM optimization is needed to combine the trade-offs between exploration and exploitation of the demand in the market.

Chapter 9

Summary and Conclusion

The ambition of this thesis and the underlying research was to provide a practical feasible approach to estimate customer demand with information on the choice preferences, in order to understand the customer's decision process under offered alternatives. Beyond that, we aimed to develop and extend the major aspects of a RM system to use such advanced information on the customer demand to maximize the overall revenue or profit. Therefore, this thesis contributes mainly to the methodology of the revenue management areas demand modeling, forecasting and optimization; besides it contains also contributions to the general and theoretic body of RM.

Our proposed choice-set demand approach provides an intuitive and effective way to model the customer's choice behavior and preferences. The demand rate curves, which closely approximate the actual observed increasing demand characteristics, allow a straightforward demand estimation in non-observed periods. The proposed unconstraining algorithms require only data that is already practical available at companies, namely historical sales and offer data at some aggregated level. The unconstraining test on actual airline transaction data shows close approximations of the observed sales realizations and is therefore very promising. In the choice model comparison test on a hotel market, we find that the choice-set based model produces very good prediction results. After establishing and testing the demand unconstraining, we focused on the forecasting aspect. Usually in RM settings we have very long booking horizons and a common practical problem is the question of updating the forecast when new information becomes available, e.g., uncertainty becomes revealed by observed sales realizations. We propose a dynamic procedure for booking horizon forecasting, which is based on a forecast dimension reduction and an application of the penalized least squares method in the updating step. The test results on real hotel reservation data show a significant increase in forecast accuracy. To complete the considered revenue management process, we focused lastly on the optimization aspect. In particular, we propose a new approach to compute time dependent bid prices in a network RM setting. In our model, we allow for multiple arriving customers per time stage and assume that bid prices are kept constant per period. These assumptions are often more realistic than the usual assumed setting with at most one customer arrival per time stage and the possibility to adjust bid prices at each of this tiny periods. The initial proposed deterministic optimization model is extended to consider stochastic demand information in form of scenarios, resulting in stochastic programming formulations. We further extend the time dependent bid price model to work with choice-set demand input in order to consider the customer's choice behavior under offered alternatives. Finally, we combined all three aspects, i.e., unconstraining, forecasting and optimization, in a simulation study. The objective is to estimate and test the effectiveness and potential of incorporating the choice-set demand approach in a revenue management application. We find that in the presence of actual buy-up possibilities in the market, the consideration of choice behavior in the optimization results in large revenue improvements. In addition, the time dependent bid price model generates considerable extra revenue gains. The proposed unconstraining algorithm estimates the actual demand curves per choice-set reasonably well and therefore provides a good approximation of the underlying demand.

To conclude, the choice-set demand model and its unconstraining, forecasting and optimization approaches provide a coherent and promising framework to incorporate the consideration of customer choice behavior into a real world revenue management applications. All approaches are developed with the focus on a practical application and require only already available company data.

The proposed approaches are already tailored for practical use. But because of the general and scientific scope of the thesis, they are naturally not directly implementable into a complex corporate environment. The models need to be adjusted to comply with actual application and company specific requirements and restrictions. Moreover, the revenue management system, with the pricing and sales control, sits at the heart of the company and is absolutely crucial for the firm's success or failure. Therefore, further testing on more realistic sized and problem specific settings is needed to evaluate the explicit revenue potential and the actual computational requirements, as well as to identify and resolve potential risks. Turning to the theoretic side, we find that further research is needed toward a new perspective on optimization models in network

RM. The pure revenue maximization objective must be changed to incorporate the demand exploration aspect, and possibly also with a concentration on risk averse solutions instead of the plain expected revenue target.

Bibliography

Adelman, D. (2007). Dynamic bid prices in revenue management. *Operations Research*, 55:647–661.

Algers, S. and Beser, M. (2001). Modelling choice of flight and booking class - a study using stated preferences and revealed preference data. *Int. Journal Services and Technology Management*, 2(1/2):28–45.

Andersson, S. (1998). Passenger choice analysis for seat capacity control: A pilot study in Scandinavian airlines. *Int. Transactions in Operations Research*, 5(6):471–486.

Artzner, P., Delbaen, F., Eber, F., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9(3):203–228.

Belobaba, P. (1985). TWA reservation analysis: Project update - demand distribution patterns. Technical report.

Belobaba, P. (1989). Application of a probabilistic decision model to airline seat inventory control. *Operations Research*, 37:183–197.

Ben-Akiva, M. and Lerman, S. (1985). *Discrete Choice Analysis - Theory and Applications to Travel Demand.* MIT Press, Cambridge, MA.

Bijvank, M., Haensel, A., L'Ecuyer, P., and Marcotte, P. (2011). Robust time-dependent bid prices for multi-period network revenue management problems. *working paper.*

Bobb, L. and Veral, E. (2008). Open issues and future directions in revenue management. *Journal of Revenue and Pricing Management*, 7(3):291–301.

Box, G., Jenkins, G., and Reinsel, G. (1994). *Time Series Analysis: Forecasting and Control.* Prentice Hall, New Jersey, 3rd edition.

Chiang, W., Chen, J., and Xu, X. (2007). An overview of research on revenue management: Current issues and future research. *International Journal of Revenue Management*, 1(1):97–128.

Cleveland, R., Cleveland, W., McRae, J., and Terpenning, I. (1990). STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73.

Cooper, W., Homen-de-Mello, T., and Kleywegt, A. (2006). Models of the spiral-down effect in revenue management. *Operations Research*, 54(5):968–987.

Cross, R. (1997). *Revenue Management: Hard-core tactics for market domination*. Broadway books, New York.

Demidenko, E. (2004). *Mixed Models - Theory and Applications*. Wiley: Wiley Series in Probability and Statistics.

Dempster, A., Laird, N., and Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, B 39:1–38.

Deuflhard, P. and Hohmann, A. (2002). *Numerische Mathematik I: Eine algorithmisch orientierte Einführung*. de Gruyter, Berlin, 3$^{\text{rd}}$ edition.

Farias, V., Jagabathula, S., and Shah, D. (2012). A new approach to modeling choice with limited data. *Management Science*, forthcoming.

Geraghty, M. and Johnson, E. (1997). Revenue management saves national car rental. *Interfaces*, 27:107–127.

Greene, W. and Hensher, D. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B*, 37(8):681–698.

Haensel, A., Bhulai, S., and Koeleman, P. (2011a). A stochastic programming approach for scheduling in queueing systems: An application to the patient scheduling problem. *submitted*.

Haensel, A. and Koole, G. (2011a). Booking horizon forecasting with dynamic updating: A case study on hotel reservation data. *International Journal of Forecasting*, 27(3):942–960.

Haensel, A. and Koole, G. (2011b). Estimating unconstrained demand rate functions using customer choice-sets. *Journal of Revenue and Pricing Management*, 10(5):438–454.

Haensel, A. and Koole, G. (2011c). Evaluating extraordinarily large requests in a revenue management environment. *submitted.*

Haensel, A., Koole, G., and Erdman, J. (2011b). Estimating unconstrained customer choice-set demand: A case study on airline reservation data. *Journal of Choice Modelling*, 4(3):75–87.

Haensel, A. and Laumanns, M. (2012). A stochastic programming approach for decision dependent uncertainty in production planning under non-compliance risk. *submitted.*

Haensel, A., Mederer, M., and Schmidt, H. (2012). Revenue management in the car rental industry: A stochastic programming approach. *Journal of Revenue and Pricing Management*, 11(1):99–108.

Heitsch, H. and Römisch, W. (2009). Scenario tree modeling for multistage stochastic programs. *Mathematical Programming*, 118:371–406.

Holt, C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. *ONR Research Memorandum Carnigie Institute 52.*

Johnson, O. (2007). Log-concavity and the maximum entropy property of the Poisson distribution. *Stochastic Processes and their Applications*, 117.

Kunnumkal, S. and Topaloglu, H. (2010). Computing time-dependent bid prices in network revenue management problems. *Transportation Science*, 44:38–62.

Lee, A. (1990). *Airline Reservations Forecasting: Probabilistic and Statistical Models of the Booking Process.* PhD thesis, Massachusetts Institute of Technology.

Littlewood, K. (1972). Forecasting and control of passenger bookings. In *Proceedings of the 12th Annual AGIFORS Symposium*, Nathanya, Israel.

Lui, P., Smith, S., Orkin, S., and Carey, G. (2002). Estimating unconstrained hotel demand based on censored booking data. *Journal of Revenue and Pricing Management*, 1(2):121–138.

Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis.* Springer, Berlin.

Mahajan, S. and van Ryzin, G. (2001). Stocking retail assortments under dynamic consumer substitution. *Operations Research*, 49(3):334–351.

McFadden, D. and Train, K. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics*, pages 447–470.

McGill, J. (1995). Censored regression analysis of multiclass passenger demand subject to joint capacity constraints. *Annals of Operations Research*, 60:209–240.

McLachlan, G. and Krishnan, T. (1997). *The EM Algorithm and Extensions*. Wiley, New York.

O'Connor, M., Remus, W., and Griggs, K. (2000). Does updating judgmental forecasts improve forecast accuracy? *International Journal of Forecasting*, 16:101–109.

Pak, K. (2005). *Revenue Management: New Features and Models*. PhD thesis, Erasmus University Rotterdam.

Pindyck, R. and Rubinfeld, D. (2001). *Microeconomics*. Prentice Hall, 5th edition.

Pölt, S. (1998). Forecasting is difficult – especially if it refers to the future. In *Proceedings of the AGIFORS Reservations and Yield Management Study Group Meeting*.

Press, W., Teukolsky, S., Vetterling, W., and Flannery, B. (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press, 3rd edition.

Queenan, C., Ferguson, M., Kapoor, R., and Higbie, J. (2007). A comparison of unconstraining methods to improve revenue management systems. *Production and Operations Management*, 16(6).

Rajopadhye, M., Ben Ghalia, M., Wang, P., Baker, T., and Eister, C. (2001). Forecasting uncertain hotel room demand. *Information Science*, 132:1–11.

Ratliff, R., Rao, B., Narayan, C., and Yellepeddi, K. (2008). A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. *Journal of Revenue and Pricing Management*, 7(2).

Rockafellar, R. and Uryasev, S. (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking & Finance*, 26:1443–1471.

Secomandi, N. (2008). An analysis of the control-algorithm re-solving issue in inventory and revenue management. *Manufacturing & Service Operations Management*, 10:468–483.

Shen, H. and Huang, J. (2008). Interday forecasting and intraday updating of call center arrivals. *Manufacturing and Service Operations Management*, 10(3):391–410.

Silberston, A. (1970). Surveys of applied economics: Price behaviour of firms. *The Economic Journal*, 80(319):511–582.

Simpson, R. (1989). Using network flow techniques to find shadow prices for market and seat inventory control. MIT Fligth Transportation Laboratory Memorandum M89-1, Cambridge, MA.

Talluri, K. and van Ryzin, G. (1998). An analysis of bid-price controls for network revenue management. *Management Science*, 44(11):1577–1593.

Talluri, K. and van Ryzin, G. (2004a). Revenue management under a general discrete choice model of customer behavior. *Management Science*, 50(1):15–33.

Talluri, K. and van Ryzin, G. (2004b). *The Theory and Practice of Revenue Management*. Springer, New York.

Topaloglu, H. (2008). A stochastic approximation method to compute bid prices in network revenue management problems. *INFORMS Journal of Computing*, 20:596–610.

Train, K. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press, New York.

Train, K. (2006). *Matlab package: Mixed logit estimation by maximum simulated likelihood*.

van Ryzin, G. (2005). Future of revenue management - models of demand. *Journal of Revenue and Pricing Management*, 4(2):204–210.

van Ryzin, G. and Liu, Q. (2008). On the choice-based linear programming model for network revenue management. *Manufacturing and Service Operations Management*, 10(2):288–310.

van Ryzin, G. and Vulcano, G. (2011). An expectation-maximization algorithm to estimate a general class of non-parametric choice-models. *working paper*.

Vermunt, J. (2010). *International Encyclopedia of Education*, volume 7, chapter Latent Class Models, pages 238–244. Elsevier, Oxford.

Vinod, B. (2004). Unlocking the value of revenue management in the hotel industry. *Journal of Revenue and Pricing Management*, 3(2):178–190.

Vulcano, G., van Ryzin, G., and Chaar, W. (2010). Choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing and Service Operations Management*, 12(3):371–392.

Vulcano, G., van Ryzin, G., and Ratliff, R. (2011). Estimating primary demand for substitutable products from sales transaction data. *Operations Research*, forthcoming.

Weatherford, L. and Kimes, S. (2003). A comparison of forecasting methods for hotel revenue management. *International Journal of Forecasting*, 19:401–415.

Weatherford, L. and Pölt, S. (2002). Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues. *Journal of Revenue and Pricing Management*, 1(3):234–254.

Weinberg, J., Brown, L., and Stroud, J. (2007). Bayesian forecasting of an inhomogeneous Poisson process with applications to call center data. *Journal of the American Statistical Association*, 102(480):1185–1198.

Williamson, E. (1992). *Airline Network Seat Control*. PhD thesis, MIT, Cambridge, MA.

Winters, P. (1960). Forecasting sales by exponentially weighted moving averages. *Management Science*.

Wu, J. (1983). On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95–103.

Zakhary, A., Atiya, A., El-Shishiny, H., and Gayar, N. (2011). Forecasting hotel arrivals and occupancy using Monte Carlo simulation. *Journal of Revenue and Pricing Management*, 10(4):344–366.

Zeni, R. (2007). Can we really look to the past to forecast future demand? *Journal of Revenue and Pricing Management*, 6(4):312–314.

Samenvatting

Het doel van dit proefschrift en het achterliggende onderzoek was het verschaffen van een praktisch uitvoerbare manier om klantenvraag te schatten uit informatie over hun keuzevoorkeuren, om zo beter te begrijpen hoe een klant een keuze maakt uit verschillende beschikbare alternatieven. Daarnaast hebben we de belangrijkste delen van een RM systeem uitgebreid, zodat zulke geavanceerde informatie over klantenvraag gebruikt kan worden om de totale omzet of winst te maximaliseren. Derhalve draagt dit proefschrift vooral bij aan de methodologie van de revenue management (RM) gebieden: klantenvraag modelleering, voorspelling en optimalisatie; daarnaast bevat het ook bijdragen aan de algemene en theoretische kant van RM.

De choice-set-aanpak die we voorstellen biedt een intuïtieve en effectieve manier om het keuzegedrag en de voorkeuren van klanten te modelleren. De vraagcurven, die de werkelijk geobserveerde eigenschappen van de vraag goed benaderen, kunnen rechtstreeks gebruikt worden om de vraag te schatten in niet-geobserveerde periodes. De voorgestelde unconstraining-algoritmes hebben alleen gegevens nodig die in de praktijk al beschikbaar zijn bij bedrijven, namelijk historische gegevens over verkoop en aanbod op een bepaald geaggregeerd niveau. De unconstraining-test, met echte transactiegegevens van een vliegtuigmaatschappij, laat goede benaderingen zien van de geobserveerde verkoop, en is dan ook veelbelovend. Het choice-set-model produceert ook erg goede voorspellingsresultaten in een vergelijkingstest van verschillenden keuzemodellen, getest aan echte hotel reserveringen. Na het vaststellen en het testen van de vraag-unconstraining, hebben we ons gericht op het maken van voorspellingen. In RM is de boekingshorizon doorgaans erg lang, en een veelvoorkomend probleem uit de praktijk is de vraag hoe de voorspelling bijgesteld moet worden als nieuwe informatie beschikbaar komt, met andere woorden, als onzekere factoren bekend worden door geobserveerde, gerealiseerde verkooptransacties. We stellen een dynamische procedure voor om de boekingshorizon te voorspellen, gebaseerd op dimensiereductie en een toepassing van de penalized least squares methode bij de update stap. Testresulta-

ten op echte hotelreserveringsgegevens tonen een significante toename in de nauwkeurigheid van de voorspellingen. Tenslotte hebben we ons gericht op het optimalisatie-aspect, om zo het RM proces geheel te bestuderen. In het bijzonder stellen we een nieuwe benadering voor om tijdsafhankelijke biedprijzen in een RM netwerk te berekenen. In ons model mogen meerdere klanten per tijdsperiode aankomen, en we nemen aan dat de bidprices gedurende een periode constant worden gehouden. Deze aannames zijn doorgaans realistischer dan de gebruikelijke aannames dat hoogstens een klant per tijdsperiode aankomt, en dat biedprijzen in elke kleine periode kunnen worden aangepast. Het deterministische optimalisatiemodel dat we in eerste instantie voorstellen, wordt uitgebreid om stochastische vraaginformatie in de vorm van scenarios in beschouwing te nemen, wat resulteert in stochastic programming formuleringen. We breiden verder het tijdsafhankelijke biedprijsmodel uit zodat het kan werken met invoer van de choice-set vraag, om zo het keuzegedrag van klanten onder verschillende alternatieven mee te nemen. Tot slot hebben we alle drie aspecten, namelijk unconstraining, voorspellen en optimaliseren, gecombineerd in een simulatiestudie. Het doel is om de effectiviteit en het potentieel van het integreren van de choice-set-aanpak in een RM toepassing, te schatten en te testen. Het blijkt dat, als buy-up-mogelijkheden in de markt aanwezig zijn, het bestuderen van keuzegedrag in de optimalisatie leidt tot grote omzetstijgingen. Daarnaast genereert het tijdsafhankelijke biedprijsmodel aanzienlijk meer omzet. Het voorgestelde unconstraining-algoritme schat de daadwerkelijke vraagkrommen voor iedere keuzeverzameling redelijk goed, en biedt zo een goede benadering van de onderliggende vraag.

Samenvattend, het choice-set-model en de bijbehorende aanpak van unconstraining, voorspellen en optimaliseren, biedt een samenhangend en veelbelovend raamwerk om keuzegedrag van klanten in echte RM toepassingen te integreren. Alle methoden zijn ontwikkeld met de focus op een praktische toepassing en hebben alleen bedrijfsgegevens nodig die al beschikbaar zijn.

De voorgestelde methoden zijn al op maat gemaakt voor gebruik in de praktijk. Vanwege de algemene en wetenschappelijke insteek van dit proefschrift, zijn deze uiteraard niet direct implementeerbaar in een complexe, zakelijke omgeving. De modellen moeten nog aangepast worden om te voldoen aan de werkelijke toepassing en bedrijfsspecifieke eisen en beperkingen. Bovendien ligt het RM systeem, met de mogelijkheid om prijzen en verkopen te beïnvloeden, in het hart van het bedrijf, en is het absoluut cruciaal voor het succes of falen van de onderneming. Daarom zijn testen op problemen van realistische grootte nodig om expliciet een beeld te krijgen van het omzetpotentieel en de benodigde rekenkundige eisen, en om mogelijke risicos te identificeren en op te lossen. Aan de theoretische kant vinden we dat meer onderzoek nodig is naar een nieuw perspectief op optimalisatiemodellen in netwerk RM. Het doel van het puur de omzet maximaliseren moet veranderd worden, om het vraag-exploratie aspect mee te nemen, en mogelijk met meer aandacht op risicomijdende oplossingen in plaats van de verwachte omzet.

Curriculum Vitae

Alwin Haensel was born in 1982 in Osterburg (Altmark), Germany. In 2008, he received the Diplom in Mathematics from the Humboldt University of Berlin, Germany, with a thesis on "Stochastic Programming Approach for Network Capacity Controls in the Airline Industry". During his study he spend a year abroad in 2005-2006 as an exchange student at the Imperial Science College London, UK. In the summer 2006, he worked for three month as an intern in the Revenue Management department of Avis Europe Plc. in Bracknell, UK. There, he made his first contact with practical Revenue Management problems and especially with the challenging topic of accurate demand forecasting. In September 2008, Alwin joined the VU University Amsterdam, the Netherlands, as a PhD student with a part time assignment at the company Bookit B.V. in Amstelveen, the Netherlands. His research focused mainly on choice modeling and demand forecasting in the Revenue Management context of airlines and hotels, but simultaneously he continued to pursue his interest in Stochastic Programming. During that time he also worked on joint Revenue Management research projects with KLM Royal Dutch Airlines, Transavia Airlines C.V., and the Netherlands Philharmonic Orchestra. Alwin also enjoyed his teaching assignments and the supervision of student projects at the VU University. In spring 2011, he joined the Business Optimization group of the IBM Research Lab in Zurich, Switzerland, for two months to work on a different research project: risk management in production planning.

Alwin defends his PhD thesis at the VU University on June 27, 2012.