André Jerenz

Revenue Management and Survival Analysis in the Automobile Industry

GABLER EDITION WISSENSCHAFT

André Jerenz

Revenue Management and Survival Analysis in the Automobile Industry

With a foreword by Prof. Dr. Ulrich Tüshaus

GABLER EDITION WISSENSCHAFT

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

Dissertation Helmut Schmidt-Universität Hamburg, 2008

Gedruckt mit Unterstützung der Helmut Schmidt-Universität,
Universität der Bundeswehr Hamburg.

1st Edition 2008

Editorial Office: Frauke Schindler / Nicole Schweitzer

Gabler Verlag is part of the specialist publishing group Springer Science+Business Media.
www.gabler.de

Cover design: Regine Zimmer, Dipl.-Designerin, Frankfurt/Main
Printed on acid-free paper
Printed in Germany

ISBN 978-3-8349-1037-0

Foreword

Revenue management has been successfully applied to service-oriented industries for a long time. In the more recent past, besides these classical application areas, it has been introduced to other production and logistics processes as well.

For the automobile industry so far, only a few revenue management models have been developed but practically none for its used car sector. Being a sector with suitable prerequisites and a low profit margin, this is a promising application area for price-based revenue management. As used cars are "individual" and "durable" goods – unlike seat or room bookings –, a different approach is necessary with "dynamic pricing" as the main control strategy. A somewhat similar problem can only be found in the real estate sector.

Thus, a conceptual framework of an appropriate revenue management model based on dynamic pricing must be developed. Using current and historical market data, the price-response function has to be estimated which then can serve as the basis for determining the optimal dynamic pricing strategy.

For two central components of this framework, optimization and estimation, innovative approaches are proposed.

Based on results from Control Theory, different possible models are suggested and extensively evaluated. Finally, a stochastic discrete-time model is identified as the most appropriate. With this, it is possible to develop iterative algorithms to determine the optimal pricing strategy even for problems without closed-form solution.

For estimating the specific demand function for a specific type of used car, methods from Survival Analysis are introduced, based on available market data and evaluated, with an accelerated failure time model resulting as the most fitting.

The conceptual work, the development of the model, the software implementation, and its evaluation based on practical data provide various valuable results for this area of research.

The proposed comprehensive dynamic pricing model (data, demand, estimation, optimization, pricing etc.) provides new theoretical insight for the revenue management in the used car sector.

But besides these contributions to theory, the results are also of great practical value. Being developed in close cooperation with experts from practice, this framework has already been implemented in software, thoroughly evaluated and successfully applied to real-world data.

With its novel modeling and algorithmic concepts as well its strong economic results, this contribution will be of interest to both researchers and practitioners alike.

Prof. Dr. Ulrich Tüshaus

Acknowledgements

Completing a dissertation is not an individual project. Therefore, I would like to thank the following for their guidance, assistance, support and friendship during the course of my doctoral studies.

First, I would like to express my sincere gratitude to my advisor, Prof. Dr. Ulrich Tüshaus, for his commitment of time, trusted conscientious advice, constant support and encouragement during this research work. I really appreciate his insight on both research and non-research related matters.

I would like to express my deep appreciation to my industry colleagues and research collaborators. In particular, I am thankful for Dr. Stefan Gnutzmann's support of this work, not only with his scientific expertise in modeling and programming questions, but especially for his guidance, intellectual support and mentoring. My sincere thanks also go to Hans-Christian Winter for providing access to the necessary vehicle data from the used car sector and for his constructive criticism and insight at different stages of my research. I feel grateful to both of them for reviewing a previous draft of this dissertation and committing time to this exciting project. Without their help the research may not have been possible.

I want to thank the thesis advisory committee members, Prof. Dr. Claudia Fantapié Altobelli and Prof. Dr. Wolf Krumbholz, for their precious time and valuable advice on the dissertation. Thanks are also due to Prof. Dr. Ronald C. Lasky and Jonathan Lasky for their help in proofreading and editing the English text of this thesis.

There are many other people who accompanied me during the course of my doctoral studies. I am deeply indebted to my friends and colleagues in the computing center at Helmut Schmidt University / University of the German Armed Forces in Hamburg, who make this an intellectual working environment. My sincere thanks go to Gerd Bartsch, Peter Cohrs, Peter Canis, Inge Duschek, Stefan Grabert, Axel Jahr, Monika Lange, Sonia Lazarova, Dirk Richter, Andreas Rohr, Thomas Rüegg, and Olaf Ziemer. Guido Gravenkötter deserves my thanks for many cups of espresso, for his encouragement during every stage of my academic career, but most of all for his valued friendship.

Finally, I owe my greatest debts to my family for their continuous love, encouragement and support of my academic pursuit. Most of all, I would like to thank Simone for her patience, understanding and love throughout my studies. Without her, this dissertation would never have been completed. Her encouragement was ultimately what made this dissertation possible.

André Jerenz

Contents

List of Figures

List of Tables

Nomenclature

Symbols for Optimal Control Theory

$\Omega(t)$	Constraint set of values of the control variable $u(t)$ at time t
$\Omega_p(t)$	Constraint set of values of price $p(t)$ for the optimal control problem of a used car retailer
$d(p,t)$	Demand at time t representing the state equation
$E(X)$	Expected value of a discrete random variable X
$f(x,u,t)$	Instantaneous profit rate for the deterministic control problem
$g(x,u,t)$	State equation describing the evolution of the state variable
$J(n,u,t)$	Objective function for the stochastic optimal control problem
$J(x,u,t)$	Objective function for the deterministic control problem
$n(t)$	State variable at time t for the stochastic control problem
$p(t)$	Price at time t representing the control variable for the optimal control problem of a used car retailer
T	Terminal time of a system for the optimal control problem
$u(t)$	Control variable at time t for the optimal control problem
$V(n,t)$	Value function for the stochastic optimal control problem
$V(x,t)$	Value function for the deterministic optimal control problem
$x(t)$	State variable at time t of a system for the deterministic optimal control problem

Symbols for Survival Analysis

β	Vector of covariates
x	Vector of regression coefficients
σ	Scale parameter on error term W in AFT models
d_i	Number of events at time $t_{(i)}$

D_{xy}	Somers' rank correlation index for assessing the discriminatory power
$F(t)$	Cumulative distribution function
$f(t)$	Probability distribution function
$H(t)$	Cumulative hazard function
$h(t)$	Hazard function
$L(\beta)$	Likelihood function for the accelerated failure time model
$l(\beta)$	Natural log of the likelihood function for AFT models
$PL(\beta)$	Partial likelihood function for the Cox proportional hazards survival model
$pl(\beta)$	The natural logarithm of the partial likelihood function
r_i	Total number of individuals exposed to risk at time $t_{(i)}$
R_N^2	Nagelkerke's index for assessing the discriminatory power
$S(t)$	Survival function
W	Error term in the accelerated failure time model

Abbreviations

AFT	Accelerated failure time
AIC	Akaike information criterion
ASEAN	The Association of Southeast Asian Nations
BIC	Bayesian information criterion
DDP	Deterministic dynamic program
DOP	Degree-of-overpricing
DSDP	Discrete-time stochastic dynamic program
HJB	Hamilton-Jacobi-Bellman equation
KM	Kaplan-Meier
OCP	Optimal control problem
PO	Number of previous owners
SDP	Stochastic dynamic program
TOM	Time-on-market
VDA	Verband der Automobilindustrie

Chapter 1
Introduction

Economic questions involve thousands of complicated factors which contribute to a certain result. It takes a lot of brain power and a lot of scientific data to solve these questions.

THOMAS EDISON
(1847–1931)

1.1 Motivation

In recent years, revenue management has experienced tremendous growth in a number of different industries, most notably in the aviation, hospitality and rental car industries. With its expansion from service-oriented industries to other sectors, the development and application of associated techniques widened as well from capacity control and dynamic pricing to more sophisticated approaches. The implementation of revenue management techniques has prevented bankruptcy at a number of companies and has changed the business process in whole branches.

In contrast, the use of revenue management within the automobile industry is still in its infancy, with very few exceptions. This fact is surprising since the automobile industry can be characterized as a low profit margin sector, where small improvements in revenue can lead to significant changes in profit. In addition, it is one of the most important industries in terms of contributing to the total global turnover and regarding the number of jobs, where in Germany, one out of seven jobs depends on this sector. Analyzing the current economic state of the automobile industry in the developed countries however, it must be ascertained that the branch is faced by overcapacity, cost pressure, and low profitability. The used car sector is affected the most, where the global tendencies are amplified by excess supply in day registrations, car rental buy-backs, and by the impact of the rebate war with new vehicles.

In this context, the primary objective of this book is to develop a framework to optimize the profit of a used vehicles retailer. Besides the development of optimization algorithms, there is a particular emphasis on the task of estimating individual price response functions for durable goods such as used vehicles. The objective of this research is multifaceted. First, the automobile industry is fundamentally assessed according to the adaptability of revenue management techniques. Second, professional price management within the used car sector is a major challenge for retailers wanting to increase their profitability. In this regard, different algorithms are necessary to determine optimal pricing strategies for the inventory of used vehicles. And third, a framework estimates individual demand by applying survival analysis, as the demand function plays a decisive role in any revenue management optimization algorithm.

1.2 Outline and Research Contribution

The remainder of this work is organized as follows. In chapter 2, the concept of revenue management is introduced in general. The automobile industry, with the used car sector in particular, is fundamentally assessed for the adaptability of revenue management techniques. Then, the used car sector is analyzed in more detail for the development of sector-specific revenue management techniques, since it promises the most potential for profit enhancement. This chapter concludes by presenting an integrated revenue management framework to determine optimal pricing strategies for used car retailers.

The revenue management module contains two core components: optimization and demand forecasting. In chapter 3, the optimization module is addressed, where pricing models are derived from stochastic control theory with the used car retailer in mind. There are several approaches differing between continuous-time and discrete-time models as well as between continuous price regions and finite price sets. This chapter concludes by extending the dynamic pricing model in order to address requirements of retailers in practice.

Survival analysis is proposed as an approach to estimate individual price response functions in chapter 4. The concept of time duration market data is introduced to obtain information about the quoting history of individual used vehicles displayed by a retailer.

The conclusions from chapter 4 regarding survival analysis become apparent in the next chapter by model identification and selection on a used car dataset obtained from a market study conducted within the German used car sector.

Chapter 6 merges the different modules and demonstrates the concept of the integrated revenue management program on a selected example. In addition, the monetary and organizational aspects are explained.

In chapter 7, the book concludes with possible areas of future research.

Chapter 2
Revenue Management and the Automobile Industry

I believe that yield management is the single most important technical development in transportation management since we entered the era of airline deregulation in 1979.

ROBERT CRANDALL
Former president of American Airlines

In this chapter, the revenue management concept is introduced with regard to the automobile industry. Its application within the used car sector provides a framework for retailers who wish to determine optimal pricing strategies.

Introducing the general concepts of revenue management in section 2.1, section 2.2 presents an analysis of the German automobile industry and identifies the used car sector as the segment with the most potential for profit enhancement. In section 2.3, revenue management is analyzed in the context of the automobile industry. Section 2.4 discusses challenges of price-based revenue management in the used car sector. This chapter concludes with determining profit-maximizing pricing strategies for used car retailers.

2.1 Concept of Revenue Management

2.1.1 Definition of Revenue Management

In recent years, an increasing number of companies have recognized the importance of revenue management for its ability to enhance sales and profitability. However, it is difficult to define the term *revenue management* and its understanding in the academic world. An extensive number of nomenclatures can be found dealing with the concept of revenue management. These concepts possess certain nuances in their meaning and positioning to each other or are explicitly used in certain areas. The terms ‘yield management’ (the traditional term in the aviation industry), ‘pricing and revenue management’, ‘pricing and revenue optimization’, ‘revenue process optimization’, ‘demand management’ and ‘perishable-asset revenue management’ are considered synonymous. Furthermore, one can observe a chronological development of the usage and definition of these terms and their underlying concepts.

Revenue management has its origins in the aviation industry during the late 1960s and early 1970s. There, the term *yield management* referred to the yield or return per seat in an airplane. Various academic papers related to the aviation industry show this development:

> The objective of yield management is to maximize passenger revenue by selling the right seats to the right customers at the right time (yield management definition by American Airlines in the American Airlines Annual Report 1987, p. 22).

> Whether an airline calls it yield management or, more appropriately, revenue management, efforts to manage the revenue mix of passengers carried involve both pricing and seat inventory control (Belobaba 1989, p. 183).

Belobaba (1987, p. 63) reasoned that yield management involves two major components, pricing and inventory control, since the yield from the airline perspective is a function of the price that the airline charges for various service options, and the number of seats sold at each price. With the expansion of yield management techniques into other industries, the concept of yield management itself evolved, and the new term ‘revenue management’ was used instead. Wirtz, Kimes, Theng, and Patterson (2003, p. 217) define revenue management as a sophisticated form of supply and demand management that balances pricing strategies with inventory management. And Cross (1997) stated that it focuses the organization on maximizing profitability by forecasting consumer behavior at the micro-level and control inventory availability at each price level.

However, the term revenue management is misleading, since it encompasses far more than revenue as Pinchuk (2002, p. 283) pointed out. Instead, revenue management is more accurately reflected by the term *profit optimization* since the objective is to optimize profit over just increasing revenue or lowering cost. Yet, to be consistent with the academic terminology, we adhere though to the commonly accepted notion of revenue management as described by Talluri and van Ryzin (2004):

Definition 2.1 (Revenue Management). Revenue management is a technique to determine the optimal price of products and services, with the objective to maximize associated expected profits generated by sale. It involves detailed forecasting of demand behavior and sophisticated mathematical modeling.

From a strategic point of view, revenue management is concerned with different categories of demand-management decisions[1], namely structural, pricing and

[1] Talluri and van Ryzin (2004, p. 2) refer to demand-management decisions as the situation where demand and its characteristics are estimated as a basis managing demand by price and capacity control.

quantity decisions. In addition, it can be distinguished between *quantity-based* and *price-based revenue management*, where the former uses inventory-allocation decisions for managing demand, and the latter applies prices as the primary control variable.

Recapitulating, due to the expansion of revenue management to sectors next to the aviation industry, it is getting more difficult to find a general definition for this management science, as Stuhlmann (2000, p. 222) pointed out. In addition, Kimms and Klein (2005, p. 5) note that it is even more difficult to assess for which framework or industry a certain definition can be applied. Consequently, in their article they present and discuss specific frameworks and instruments for certain industries.

This section gives a brief introduction to the science of revenue management. For an in-depth discussion we refer the reader to the following research literature. McGill and van Ryzin (1999) provide a survey reviewing the history of revenue management and a bibliography of relevant academic publications. Tscheulin and Lindenmeier (2003) present key elements of the revenue management problem and give a detailed summary in terms of the bibliography of work in this research area. In the book *The Theory and Practice of Revenue Management*, Talluri and van Ryzin (2004) provide an excellent presentation of the technical aspects of revenue management with mathematical and modeling approaches. Lastly, Kimms and Klein (2005) form a general and industry-independent definition of the prerequisites of revenue management by analyzing several industries.

2.1.2 The History of Revenue Management

Revenue management origins from the aviation industry of the late 1960s and early 1970s through the research of capacity management decisions (cf. Lieberman 2004, p. 92). Up to that time, research into reservations control was focused on overbooking, in which the calculations depended on the probability of the number of passengers who showed up for boarding at flight time. In this context, the work of Rothstein (Rothstein 1971, 1974) should be mentioned.

However, the beginning of revenue management (or yield management as it was called in the aviation industry at that time) was marked by Littlewood's rule. In the early 1970s, airlines began to offer restricted airline fares to gain revenue from seats that would otherwise be empty. The problem was to determine the number of seats that should be reserved for late full fare bookings versus discount fare bookings. In 1972, Kenneth Littlewood, an employee of the British Overseas Airways Corporation (a predecessor of British Airways) described in an article (Littlewood 1972) the application of mathematical models for this problem. He proposed that discount fare bookings should be accepted as long as their revenue value exceeded the expected revenue of future full fare bookings (cf. McGill and van Ryzin 1999, p. 233). He laid the groundwork for a number of controlling models through the description of forecast procedures and turnover control.

One of the first companies that applied yield management techniques was British Overseas Airways Cooperation by offering early bird rebates for customers who booked tickets at least twenty-one days before takeoff. Several other airlines followed, but the most noticeable and successful adoption has to be credited to American Airlines. Prior to 1978, the aviation industry in the United States was heavily regulated by the U.S. Civil Aeronautics Boards, which kept fares sufficiently high in order to guarantee the airlines a reasonable return on their investment. With the passing of the Airline Deregulation Act in 1978, the board loosened control over commercial aviation and thus, the passenger airline industry was exposed to the free market. One of the first airlines to be founded after the deregulation was PeopleExpress. Based on a significant lower cost structure, PeopleExpress was able to offer fares up to 70 percent below the prices of major airlines, attracting a significant number of price-sensitive customers and thereby damaging the profits of major airlines. American Airlines itself reacted to the Airline Deregulation Act by launching a new pricing scheme in 1978 with American Super-Saver fares. American Airlines restricted the number of discount seats sold on each flight by a fixed amount with a combination of purchase restricted and capacity-controlled fares. Through this new pricing scheme, they revised the program according to the demand for different flights on different days, requiring a different allocation of discounted fares. Thus in January 1985, American Airlines announced its Ultimate Super-Saver Fares program and within one year, PeopleExpress went bankrupt and was sold to Texas Air for less than 10 percent of the market value it had been worth a year before (Phillips 2005, p. 121).

Up to that time however, yield management was still defined rather narrowly and its applications concentrated on capacity management and overbooking, and little was done in the area of dynamic pricing. Still, prices for single classes were generally assumed as fixed and the managers had to decide when to open or close a certain class depending on the demand, but they did not yet have the control to set dynamic prices for these tickets. But with the work of Belobaba (Belobaba 1987, 1989) and the success of American Airlines among others, interest in quantity-based yield management expanded and spread to other industries. With the increase of interest beyond the aviation industry, the term's use changed and the expression 'revenue management' was established, since it was more acceptable to executives outside the airline industry. At the same time, the scope of the research expanded and began to include price-based decisions as well as inventory control-based decision supports. During the 1990s, the expansion of interest in revenue management led to a broadening of the theory to areas where the opportunities to turn theory into practice faced greater implementation challenges. Here, worth mentioning is the development of incorporating dynamic pricing strategies, especially in form of price markdowns within the retail industry.

Today, revenue management techniques are applied in many different industries, such as automobile rental, lodging and gastronomy, passenger railways, internet service providers and cruising lines. A more extensive discussion of the history of revenue management can be found in Smith, Leimkuhler, Darrow, et al. (1992), Cross (1995), McGill and van Ryzin (1999), and Talluri and van Ryzin (2004), among others.

2.1.3 Industry-Specific Applications of Revenue Management

As already mentioned, the success of revenue management in the aviation industry stimulated the development of similar systems in other areas of the service industry and even beyond. This subsection gives an overview of its adoption in specific industries. Table 2.1 lists industries to which revenue management methods are applied and gives examples of corresponding research articles.

Two industries which incorporate revenue management most heavily are aviation and lodging. Through the use of specialized software, airlines monitor seat reservations and react accordingly, for example by offering discounts when the seats would otherwise be vacant. Hotels apply revenue management in the same way by calculating rates and sales restrictions to maximize the return for the property. Another successful use of revenue management is the car rental industry. First applied in the early 1990s by major companies in the United States, the car rental industry possesses similar characteristics as the airline industry. Retailing has been a recent adopter of revenue management techniques. The industry is characterized by volatile and uncertain demand, short selling periods according to the season, relatively long production and distribution lead times, and relatively inflexible supply (cp. Talluri and van Ryzin 2004, p. 16). There, the consumer electronics, fashion apparel, and toy retail industries are particularly suitable for the application of revenue management.

In sum, revenue management has experienced tremendous growth since its introduction in the aviation industry, and its techniques and methods are applied to an increasing set of business areas. In the future, many industries and sectors will be candidates for revenue management, as the discussion of beneficial conditions in the next section will confirm.

2.1.4 The Revenue Management Framework

This subsection describes a general revenue management framework by introducing its key components. The basic revenue management process consists of four com-

Industry	Sector	References	Example
Aviation	Passage	Smith, Leimkuhler, Darrow, et al. (1992), Cross (1995)	Yield management at American Airlines
	Air Cargo	Kasilingam (1996)	Implementation of cargo revenue management at American Airlines
Tourism	Lodging	Kimes (1989), Bitran and Mondschein (1995), Bitran and Gilbert (1996), Donaghy, McMahon-Beattie, and McDowell (1997), Jones (1999)	Implementation at Marriott International Inc.
	Gastronomy	Kimes, Chase, Choi, Lee, and Ngonzi (1998), Kimes, Barrash, and Alexander (1999), Kimes (1999)	Implementation at 100-seat restaurant.
	Cruising Lines	Ladany and Arbel (1991)	
	Car Rental	Carroll and Grimes (1995), Geraghty and Johnson (1997)	Adoption at Avis Rent A Car, National Car Rental, and Hertz Car Rentals
	Ticketing	Courty (2000)	
Others	Retailing	Shoemaker and Subrahmanyan (1996), Smith and Achabal (1998), Mantrala and Rao (2001), Heching, Gallego, and van Ryzin (2002)	Pricing optimization at Northern Group Retail and Old Navy
	Railways	Strasser (1996), Ciancimino, Inzerillo, Lucidi, and Palagi (1999), Krämer and Luhm (2002)	Implementation at Deutsche Bahn
	Manufacturing	Harris and Pinder (1995), Elimam and Dodin (2001), Defregger and Kuhn (2003)	Adoption in steel and aluminum industry
	Energy Sector	Valkov and Secomandi (2000)	Adoption in natural gas industry

Table 2.1: Industry-specific adoptions of revenue management

ponents: data collection, demand forecasting, optimization, and controlling. Figure 2.1 outlines the framework and the process flow within a general system. In step one, business data is collected to extract information on historical prices and demand, customer purchases as well as internal and external conditions. In this stage it may be important to collect information about demand that was not satisfied. This case can occur when demand is sporadic and only a small number of products are sold (see Orkin 1988, p. 56). In stage two, demand is forecasted by estimating parameters for a specific demand model. Here, a widespread approach is to apply a time-series forecast to determine demand based on known past events. In addition, this component identifies customer segments as well as the potential impact by com-

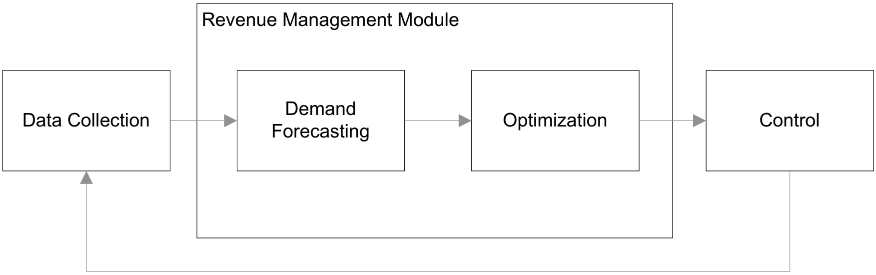


Fig. 2.1: Elementary components of the revenue management system, following Tscheulin and Lindenmeier (2003, p. 631)

petitors. Based on this information, sub-markets can be defined according to their size. Specific characteristics and mechanisms also have to be established to prevent arbitrage across segments (cp. Harris and Peacock 1995, p. 39). The optimization in stage three represents the core component of each revenue management system. Here, the aim is to find the best set of controls for price and capacity. Based on the information gathered in the demand estimation stage, prices and the allocation of capacity can be determined for each sub-market by applying optimization models. In the last step, responses to the optimized control are measured systematically using a transaction-processing system.

2.1.5 Conditions for the Application of Revenue Management

In the following subsection, we discuss beneficial characteristics for the successful adoption of revenue management. These conditions can be classified into three separate categories: motivating and demand characteristics, defining the requirements for a specific industry or branch of industry, and the category of company-specific conditions. A more detailed discussion of the conditions for adopting revenue management techniques can be found in Kimes (1989), Weatherford and Bodily (1992), Harris and Peacock (1995), Friege (1996), Klein (2001), Wirtz, Kimes, Theng, and Patterson (2003), Talluri and van Ryzin (2004), and Kimms and Klein (2005).

The application of revenue management techniques is appropriate when the following *motivating characteristics* are fulfilled by an industry:

Lack of flexibility in capacity The flexibility of capacity is limited given that companies cannot add additional capacity to satisfy demand in the short run. More precisely, since capacity can in most cases be increased, high costs are associated with the addition of an incremental unit of capacity. Often the term ‘relatively fixed capacity’ is found in the literature, but as Kimms and Klein (2005,

p. 7) notice, the usage of this abstract term might be problematic. An example of the lack of flexibility in capacity is illustrated by the hotel industry. Once all rooms in a hotel are occupied, additional rooms cannot be added without generating substantial costs and time.

High ratio of fixed to variable costs The marginal costs of selling an additional unit of inventory are small in relation to the marginal costs associated with an increase of capacity. Thus, adding capacity is expensive for a company, but selling another unit of inventory is relatively inexpensive (cf. Kimes 1989, p. 350). A high ratio of fixed to variable costs yields large contribution margins and entices companies to decrease prices, attracting additional incremental business (cp. Harris and Peacock 1995, p. 40). Marginal costs are relatively low for selling a hotel room that is available, but the costs of increasing the capacity of the hotel are substantially higher.

Perishable inventory In the literature, inventory perishability is noted as a key factor distinguishing service from manufacturing industries. In case of the former, where capacity is represented by services, inventories are completely perishable. Capacity is only available within a certain time frame. After the expiration of the time period, the salvage value of capacity is substantially smaller or even zero. For example, if seats in an airplane or rooms in a hotel are not sold until the date of service, potential revenue is lost.

Following Harris and Peacock (1995, p. 40), these motivating characteristics encourage a company to apply revenue management, but external *demand conditions* and internal conditions also enable a company to apply revenue management.

Customer heterogeneity A company is able to divide a specific market based on the heterogeneity of customers' purchase behavior, their willingness to pay, and variations in their preferences for different products as well as different purchase behaviors over time. Consequently, price discrimination can be applied. Thus, an airline can discriminate between a time-sensitive business traveler and a price-sensitive leisure customer by asking higher prices as more time elapses.

Fluctuating, uncertain demand Fluctuating and uncertain demand can depend on seasonality or market shocks. It becomes more difficult for a company to make pricing and capacity decisions when demand varies over time and future demand is uncertain. In such circumstances, it is even more important to apply sophisticated revenue management techniques since the potential for losses based on inferior decisions increases. For example, during low demand periods, the number of discounted air fares increases, and during periods of high demand the number of discount tickets will be limited in order to increase profitability.

Management can support the adoption of revenue management techniques by *company-specific conditions*:

Data and Information System Infrastructure Talluri and van Ryzin (2004, p. 15) conclude that revenue management techniques tend to be better suited to industries where data information systems are already employed. This requires the use of historical sales databases to collect, store and process demand data. For example, in 1985, PeopleExpress was not capable of confirming seat requests from late-booking business travelers due to inadequate reservation and capacity-allocation systems.

Staff and Management Commitment For successful implementations of a revenue management system it is essential to have the commitment of the staff involved with these techniques. This can be achieved by careful planning and training of the employees. Furthermore, the incentive system has to be adjusted to incorporate these methods. For instance, in the past sales-persons were rewarded by the number of sales they made, but with the implementation of a revenue management system, they would be rewarded by the amount of profit they generate. Even more important is the support of top management. They have to communicate the necessity and relevance of these new systems so that all employees understand their importance to the success of the company.

In summary, there are three categories of conditions beneficial for the adoption of revenue management. The first two categories, namely motivating characteristics and demand characteristics, deal with the industry-specific level, whereas the latter category defines company-specific conditions suitable for the use of revenue management techniques.

Certain industry branches are better suited than others, especially in the aviation and tourism industry. Among the existing industries where revenue management is already applied, one can observe that no prototypical model for the application and implementation exists. Rather, as Lieberman states, the root concepts are the same, but the applications and the techniques differ widely (Lieberman 1993, p. 34). Thus, the decision to apply revenue management techniques depends on the specific circumstances of the company and the industry branch in which it is found. In the end, it comes down to a cost-benefit analysis for each individual company, where management assesses potential benefits versus potential costs and risks associated with the process. Following Talluri and van Ryzin (2004, p. 17), once the technology and the methodology mature within a specific industry branch, the majority of companies will benefit from revenue management.

The automobile industry, although an important global sector, is still in an early stage of applying revenue management systems. Thus, the objective of the following sections is to assess the automobile industry as a potential sector for the application of revenue management techniques, focusing the examination on the German used car sector and analyzing it with regard to the adaptability of revenue management.

2.2 The Automobile Industry

2.2.1 State of the Global Automobile Industry

Today, the automobile industry, often thought of as one of the most global of all industries, is in the middle of a dramatic transformation. The automobile industries in the Triad regions (the United States of America and Canada, Japan, and Western Europe) are faced with overcapacity, cost pressure and declining profitability. The stagnating production and sale in these regions is a contrast to the growth of the industry in the rest of the world, with a considerable part of this rapid growth concentrated in a small number of developing countries, including Latin America (mainly Brazil and Mexico), the Association of Southeast Asian Nations (ASEAN)[2], Eastern Europe, China and India. These regions are not only sources of low-cost labor, but also represent growing markets themselves.

There are several major trends in the global automobile market. The world's largest automobile manufacturers continue to invest into production facilities in emerging markets to reduce production costs. In addition, a trend to expand in overseas markets and establish global alliances such as Renault with Nissan can be seen. And lastly, the increasing global competition among the global manufacturers has led to an accelerated process of consolidation within the industry (cp. Hiraoka 2001, p. 15).

Considering these trends, the automobile industry faces several challenges. Global overcapacity is thought by many executives to exceed ten percent, ranging from six to eight million vehicles in total. The globalization trend exacerbates this problem on an even larger scale. Furthermore, the market appears to be fragmenting into ever finer segments and vehicle manufacturers must offer products that meet the needs of these niches. Finally, cost pressure, declining profitability and the expense of various products have forced manufacturers to look downstream for new ways to create and capture value, that is by participating more extensively in the stream of post-assembly transactions relating to the vehicle (cf. Brandstad, Williams, and Rodewig 1999, p. 3).

The leading issue in the automobile industry is product quality, but the secondary issue is cost reduction. With competitive pressure felt across the global industry, companies are beginning to think about opportunities for future cost savings. In a survey of KPMG, innovation in manufacturing and materials as well as outsourcing are seen to be the greatest source of cost savings (see Achterholt and Schmid 2007,

[2] The Association of Southeast Asian Nations or ASEAN was established on 8 August 1967 in Bangkok and today consists of the five original Member Countries, namely, Indonesia, Malaysia, Philippines, Singapore, and Thailand as well as Brunei Darussalam, Vietnam, Lao PDR, Myanmar, and Cambodia.

p. 7)[3]. According to executives however, computer modeling is the third most important factor in the future. Following this estimate, these kinds of computer models are developed in the subsequent chapters of this work to approach the challenges in the automobile industry.

2.2.2 The German Automobile Industry

The automobile industry has become Germany's most important sector after more than doubling its revenue over the last ten years, and increasing its global market share from 12 percent to 19 percent. In 2005, it contributed 18 percent to the total turnover of German's industries (with 236 billion euro of 1.2 trillion euro) and has increased the number of employees by 21 percent over the last ten years, with 1.4 million employees. In total, one out of seven industry jobs is part of the automobile sector in Germany.

The car passenger market is, by the number of new and used registrations, the most important branch of the German automobile industry (the other is the commercial vehicle sector). In 2005, the car passenger market slowed for the fifth year in succession, largely caused by the private passenger car segment which decreased by 27 percent since 1999. Operative margins of German manufacturers are only 2.5 percent, which is significantly lower compared to Asian manufacturers with 5.7 percent. In addition, the sector in Germany is faced with cost constraints, in particular driven by a significant increase in prices for raw materials. For instance, prices for important metals such as aluminum, copper or lead showed increases of more than 10 percent from 2004 to 2005, and even more than 50 percent for steel products (cp. VDA 2006, p. 36).

In summary, the automobile industry, with the passenger market as the most important segment, is Germany's key industrial sector. Being the third largest manufacturer worldwide, the German passenger market is faced with overcapacity, cost pressure, and low profitability.

2.2.3 Segmentation of the German Passenger Car Market

The German car passenger market can be divided into four sections: new car sales, financial services, parts sales, and the used car market. In 2004, it generated 160 billion euro in revenue with 4.1 billion euro in profit. This corresponds to a comparably low profit margin of approximately 2.5 percent. For the last seven years, the

[3] Findings of KPMG's annual automobile survey, which was based upon the interviews of 140 senior executives at vehicle manufacturers and automobile suppliers worldwide.

German car market has been characterized by shrinking net profit margins, since excess manufacturing capacity is being forced into the market globally, leading to expensive rebate wars. Thus, in 2004, the average list price discount was almost 16 percent and even 30 percent for the premium sector, matching U.S. discount levels for the first time (cf. Mercer 2005, p. 4).

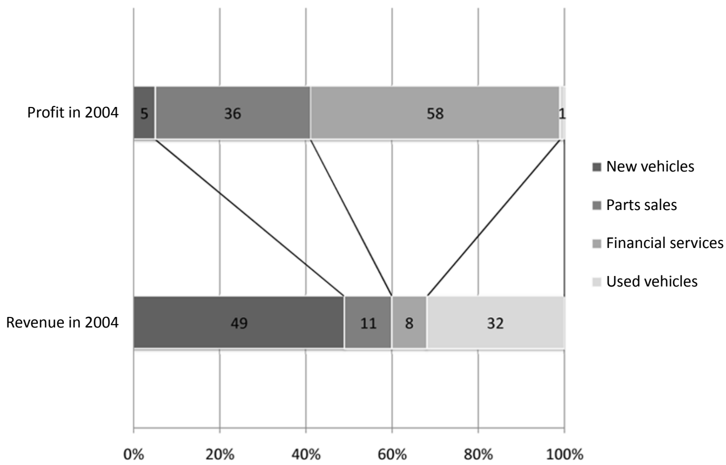


Fig. 2.2: Comparison of revenue and profit in the German automobile industry for the year 2004, source: Mercer (2005)

A comparison of revenue and associated profit margin uncovers the problems of the German car market, shown in Figure 2.2. The new vehicles segment accounts for almost half of the German automobile industry's revenue, but is responsible for only approximately five percent of the industry‘s profits in 2004, translating to profit margins for this segment of less than one percent (according to Mercer (2005), revenue for the new vehicle sector was 79 billion euro in 2004 with 0.2 billion euro in profit). On the other hand, financial services and parts sales together account for 19 percent of the industry's revenue (29 billion euro in 2004), but with 3.9 billion euro they are responsible for 94 percent of the industry's profits. The weakest business segment however, with the most potential for profit enhancement, is the used car market. Although accounting for 32 percent of the industry's revenue (52 billion euro), only a marginal profit of 0.03 billion euro was generated in 2004.

To avoid further operating losses and even financial bankruptcies, retailers must find ways to generate revenue growth in conjunction with significant profit margins of at least two to four percent within the used car business.

In the following sections, the focus is on the used car sector, since it shows the most potential for profit enhancement. Current price management approaches, potential for improvement, and better pricing mechanism are discussed.

2.2.4 The German Used Car Market

2.2.4.1 General Conditions of the German Used Car Market

The used car market covers all sales of second-hand cars, including private sales, rental and leasing disposals, manufacturer and corporate disposals, and other remarketed sales[4] (Datamonitor 2005, p. 2).

The German used car market is highly fragmented, with private sellers accounting for nearly half of all used car sales. After a five-year period of negative compound growth rates from 2000 to 2004, the number of sold used vehicles remained stable in 2005, but revenue increased significantly by 6.2 percent, amounting to 55.4 billion euro (Schönleber 2006, p. 40).

Historically, the German used car market has consisted of three business subsegments: new car dealers, used car dealers, and the private market. The private segment was the most active in 2005 with 3.1 million transactions (47 percent). New car retailers generated 2.5 million transactions (27 percent), and used car retailers accounted for a further 26 percent.

Focusing on the professional dealer, there are three different means through which to sell vehicles within the used car sector. The first segment consists of rental and leasing disposals, which have shown a consistent growth of volume in recent years. There, the amount of vehicles and prices are fixed and predictable, since dealers are generally committed to taking back leased vehicles for a fixed price and at a fixed date. Secondly, car dealers accept trade-ins in order to sell a new car. The price for the trade-ins is negotiable, though in many cases it is inflated for stimulating sales. The number of cars can be assumed to be fixed however, since a trade-in occurs isochronously with the disposal of a new car. And lastly, a dealer can buy according to the market. Here, the number of vehicles and prices can both be determined by the dealer.

In summary, the used car sector has only limited options for controlling capacity according to the market. Practically, the size of the total market for cars is limited

[4] Remarketed sales are defined as sales of vehicles by a business which has utilized the vehicles as part of their business operations or used the vehicles to generate profit and therefore includes disposals and sales through franchised and independent dealers.

since a dealer has to trade in a used vehicle in order to sell a new one. Thus, the disposal of used cars through professional retailers has to be regulated by the price.

2.2.4.2 Challenges and Opportunities in the Used Car Market

As shown before, the used car market generates the lowest profit margins with nearly zero percent in 2004. The above mentioned global tendencies toward model variety and overproduction have affected the used car sector as well. The excess supply in day registrations and leasing disposals negatively impacts the residual value development of used vehicles. Residual value reductions lead to higher expenses with the return of leasing vehicles. The necessity of a re-evaluation of the vehicles results in higher future leasing rates. Furthermore, they influence the pricing for new cars (particularly in comparison to day-registrations) and tarnish the brand image in the medium and long run. Another major challenge within the German used car market is generated by competition. With the beginning of the block exemption regulation in October 2005, car manufacturers can no longer assign exclusive sales territories. Now, dealership chains and international dealers are able to expand into the German market, whereas traditional, authorized dealerships lose market shares. Last but not least, liquidity represents a serious problem for a large part of the traditional retailers. Their inventories hold a large number of used vehicles, which ties up a large amount of liquid assets. Due to tightened regulations regarding credits, this weakness leads to further financial problems for the retailer.

As mentioned before, leasing and rental disposals account for a significant number of the retailer's used car portfolio. Leasing and rental companies can order vehicle configurations that promise higher market values after disposal. The retailer's portfolio will then become more attractive, more used cars will sell and eventually the market value for used vehicles in general stabilizes, supporting the sale of new cars as well. Second, used vehicles have to be upgraded and maintained to increase their quality rating. Furthermore, used car retailers can achieve higher profit margins by offering attractive auxiliary services for their customers such as customized warranties on maintenance contracts. These increase customer loyalty.

A last and, in context of this work, central approach for improving profits is the professionalism of revenue management by applying more sophisticated pricing strategies to the sale process of used vehicles. Often, retailers use a static pricing strategy and do not change the original price afterward, even if it is significantly higher than the current market value and the used vehicle does not sell. Alternatively, asking prices are too far below the market price and the vehicles sell very quickly, but miss the potential profit of a higher asking price. Therefore, the next section examines the status of revenue management in the automobile industry in general, whereas in section 2.4 current pricing strategies applied by retailers are analyzed for potential in profit enhancement.

2.3 Revenue Management in the Automobile Industry

In this section, the concept of revenue management is introduced to the automobile industry. First, current applications in practice are described and, then, an analysis of desirable and beneficial conditions is carried out to determine whether the industry is suitable for revenue management techniques.

2.3.1 Current Applications of Revenue Management

The adoption of revenue management techniques is still in its infancy in the automobile industry compared to other branches, especially in service-driven industries such as aviation or lodging. In the practice-oriented literature, sporadically some examples can be found, but theoretical-focused papers are rare in this area.

This fact is all the more surprising since the automobile industry is a low profit margins industry, where small increases in revenue can lead to significant changes in profit. In this setting, pricing is one of the most important and complex marketing components, with a multitude of alternatives for differentiation. The brand of the vehicle has a significant value in the purchase decision process. This complexity offers potential for improvement as well as significant risks (refer to Al-Sibai, Möller, and Hofer 2004a, p. 351).

Only a few examples for the adoption of revenue management techniques into the automobile industry can be found in publications. Hofer, Ebel, and Al-Sibai (2004) describe the challenge of strategic price positioning of new models and propose a decision-support model for determining the initial list price in the introductory phase of a new vehicle. Engelke (2004) analyzes the pricing of options for a new vehicle and identifies significant potential for increased profits. He proposes conjoint measurements for determining individual willingness-to-pay as a basis for optimal pricing of options. Lastly, Al-Sibai, Möller, and Hofer (2004b) analyze the impact of customer incentive programs on the distribution of new vehicles. They conclude that in many cases, the enhancement of the standard equipment of a model affects the distribution more positively than a price reduction.

Although these documents describe the application of revenue management in the automobile industry, the most noticeable and comprehensive adoption can be found at the Ford Motor Company. Ford Motor Company was the first automaker to adopt revenue management in its industry sector. In late 2001, pricing specialist Lloyd E. Hansen was promoted to vice president at Ford, at that time the only one heading a revenue management unit. He announced a revenue management strategy as a vital part of the company's turnaround plans after Ford lost $6.4 billion over two years due to quality problems, productivity issues, outdated portfolio of models, and $25 billion in unfunded pension and health care liabilities (cp. Banham 2003, p. 72).

Hansen introduced a revenue management system consisting of three technology tools. First, a marketing response tool analyzed transaction prices per vehicle taking into account the range of different incentive programs offered. Thus, the company and its retailers were able to project the effect of different incentive programs on sales, determining the best program in each market. Second, a package optimizer was able to determine the car with the best selection of features most likely to appeal to consumers in a particular market. And third, Ford adopted a new ordering system for car dealers that optimizes inventory, based on profit margins, customer preferences, and customers' willingness to pay.

As a result, the new revenue management system contributed some $260 million to Ford's $896 million profit in the first quarter of 2003, according to Welch (2003, p. 38). Furthermore, retail incentive spending per vehicle was $900 less than at rival automakers in the U.S. market, and revenue per unit increased by $699. Ford's vice president of revenue management, Lloyd Hansen, summarized the application of these new strategies as follows (from Banham 2003, p. 69):

> 'Revenue Management has the most leverage in industries with low profit margins. That's what makes it so critical in the auto industry, where pretax profit margins have historically averaged only about three percent. If better pricing tools and processes can improve revenue by just one percent, and raise historical margins to four percent, button-line profits would grow by 33 percent.'

2.3.2 Assessment of the Automobile Industry

In the following section, the suitability for the adoption of revenue management techniques within the automobile industry is analyzed. Since the industry is made up of a number of fundamentally different branches, this assessment focuses on the used car sector, since this sector was identified as the one with the most potential for profit enhancement. Based on the characteristics given in section 2.1.5, the used car industry is ranked relative to its support for revenue management as low, medium or high on each of the conditions. The results are summarized in Table 2.2. Considering the first category of beneficial characteristics, namely motivating conditions, the used car sector provides a good environment for revenue management techniques. The capacity of a retailer is somewhat inflexible, since used vehicles are heterogeneous goods based on the various factors such as engine, mileage and age, as well as a combination of options. Thus, it is almost impossible for a retailer to obtain two identically used vehicles, only roughly similar ones. Used vehicles do not perish like an airplane seat that is lost after the flight has taken off, but their market value significantly decreases over time as age is an important characteristic for determining a car's selling price. From another perspective, profit decreases by

Category	Condition	Remark	Support
Motivating conditions	Lack of flexibility in capacity	Retailers offer only a limited number of identical vehicles. Furthermore, used cars are unique in terms of their characteristics such as mileage, age, and built in options.	high
	High ratio of fixed to variable costs	Depending on the model, the cost to add another vehicle of the same type can be significantly higher than the costs associated with the sale.	medium
	Perishable inventory	Vehicles do not become worthless like plane seats or hotel rooms. However, they age, tie up capital, and decrease in value since age is an important characteristic for the determination of the market value.	medium
Demand conditions	Customer heterogeneity	The used car sector can be segmented by different criteria such as customer, product, channel, and time.	high
	Fluctuating, uncertain demand	Demand can vary by season, by stage of model cycle, or by general economic conditions.	high
Company-specific conditions	Data and Information System	Given the existing enhanced information technology structure of major car retailer chains, the automobile industry provide an good environment for necessary databases.	medium
	Staff and Management Commitment	Given the fact that retailers are not accustomed to using algorithms for pricing their goods, there might a reluctance toward revenue management systems. Furthermore, reward systems are necessary.	medium

Table 2.2: Assessment of the suitability of revenue management for the used car sector as a branch of the automobile industry

costs associated with maintenance, servicing, and financing (the *opportunity cost of capital*). Therefore, longer display times of used vehicles result in decreasing profits (or increasing losses), whereas actual profit margins are already at a very low level, generating additional pressure on the retailer to sell the used vehicle within a short period of time.

The used car sector is characterized by customer heterogeneity and fluctuating, uncertain demand. This demand for used vehicles depends on the overall economic conditions and varies with the season and stage of the model cycle.

Last, internal conditions are indispensable for revenue management. Within the used car sector, it is important to get sufficient sponsorship and support by executive leadership. Given the fact that used car retailers do not use algorithms for pricing their vehicles, there might be reluctance toward a new revenue management system.

Recapitulating, the used car sector is a good area for the application of revenue management techniques. The capacity of retailers is often limited, demand is uncertain and variable, inventory ages, and the sector is segmented across different dimensions. Furthermore, the infrastructure of information systems as for collecting, analyzing, and applying data needed to forecast demand and optimize control sets is implemented in most cases.

2.4 Price-Based Revenue Management in the Used Car Sector

Section 2.2.4 identified the issue of professional price management as a major challenge for retailers of used vehicles. In this section, current approaches of used car pricing are described and analyzed to reveal potential for profit enhancement in the used car sector and, thus, provide the basis for the development of more sophisticated pricing algorithms. This section concludes with an outline of the approach of determining optimal pricing strategies for a used car retailer.

2.4.1 Selling Process of Used Vehicles

The sale of used vehicles is characterized by a complex relationship and a large number of variables. This and the present economic situation are responsible for the fact that only a few retailers can still earn a profit in this sector. In illustration 2.3, the individual steps of the process of selling used cars are displayed chronologically. First, a retailer begins with evaluating the used vehicle offered to him. Based on the

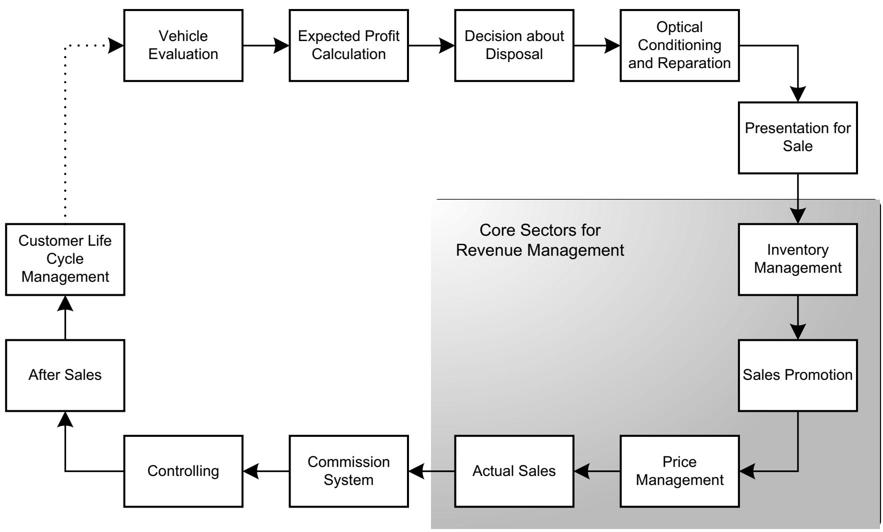


Fig. 2.3: Selling process in used car retailing, source: expert interviews with used car retailers

expected profit calculation, the dealer decides whether or not to buy the vehicle. As mentioned before, the dealer often is not in the position to decide, but has to accept vehicles (e.g. leasing disposals or trade-ins). Then, the car must be repaired and conditioned to take it in stock for sale. At this stage, the dealer must decide where and how the vehicle is displayed in his showroom or exhibition space. The following sub-processes are summarized under the concept of revenue management. They consist of inventory management, sales promotion (i.e. advertising), price management, and the actual sale. Lastly, the sales process is completed by the commission system, the controlling, the after sales segment, and the customer life cycle management to establish long-standing customer loyalty. This way it ensures follow-up business for the retailer.

The sub-processes inventory management, sales promotion, price management, and the actual sale form the main area for revenue management in the used car sector. They should not however be regarded separately from the total sales process. The used-car retailer interacts primarily at these stages with the customer through marketing instruments such as sales promotion and supports the customer with expert advice. In this context, the communication of the asking price plays a significant if not the most important role within the marketing mix, since only price determines the monetary amount a customer must give to obtain the vehicle. Thus, the decision about the ‘right’ pricing strategy is crucial for the success of the used-car retailer.

2.4.2 Analysis of Current Pricing Strategies

The initial price positioning and the adaptation of the price of a used vehicle is the most important decision made by a used car retailer. The value of the vehicle must be determined and factors like current demand, global and regional trends as well as seasonality and life cycle of the model must be considered. Finally, all these factors must be restricted to supply and demand functions.

A commonly used price management strategy is described here, but it should be noted that this is one of the better procedures within the used car sector. The majority of the dealers use less developed methods, so the optimization potential will be even greater there.

A common process of determining a pricing scheme for a used vehicle is separated into three steps: first, estimating an approximate market value; second, assessing the current market conditions and third, formulating a pricing policy, beginning with the initial price positioning and followed by price adjustments according to the current market condition.

Step one, estimating the current market value is done by using general market price lists provided by DAT and Schwacke[5], as well as monitoring real time sales at the site of the used car retailer. This step leaves room for uncertainties and intu-

[5] Both DAT and Schwacke are service providers of data collection and market analysis. Based on mathematical models, they determine reference lists of used vehicle valuations.

ition often plays an important role in estimating. In the second step, the process of evaluating the current market conditions, car dealers monitor proposals of similar vehicles offered by competing retailers and thus, get a good impression of the size and condition of the specific market for this vehicle.

The last step is the definition of the price strategy. First, an initial positioning is carried out in such a way that the asking price is situated in the upper 50 percent quantile of the vehicle's market identified in step 2. Subsequently, the asking price for the vehicle is adjusted every 30 days and decreases successively until, after approximately 60 days, it reaches the average market price. Should the vehicle not sell after 90 days, the price would be reduced into the lower 30 percent quantile (see Figure 2.4). Based on the findings from expert interviews, the average time on market of used vehicles is 60 days (cp. Wuppermann 2003, p. 71).

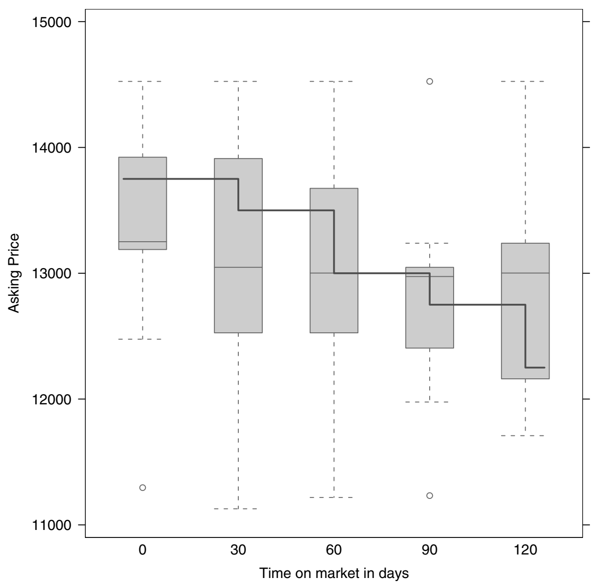


Fig. 2.4: Example of a commonly used pricing strategy for a used vehicle. The specific boxplots correspond to the population of similar vehicles offered at that time and the red line outlines an actual pricing strategy of a used car retailer, source: based on an expert interview with a used car retailer

Similar price policies are also used by other dealers. One alternative strategy is the usage of relative price steps (here depending on the market value as in the Schwacke list). A Ford dealer described his strategy as follows during an interview (cf. Wuppermann 2003, p. 237): up to the 30th day, the market price plus 15 percent is requested. For the next 30 days he asks for the market price plus 5 percent, up to 90th day the wholesale price according to the Schwacke list plus his preparation expenses. From the 91th on to the 120th day the wholesale price only and beyond that point in time, he will accept his own purchase price. Compared to the previous pricing scheme, the retailer here does not position the asking price in relation to the actual sub-market of the offered vehicle. Instead, he determines an average market value for the car and uses it as the basis for his pricing strategy by adding or subtracting a relative percental amount. Thus, he will achieve a decreasing pricing step function as the retailer in the previous example, however unconditioned on the current market environment for the specific vehicle.

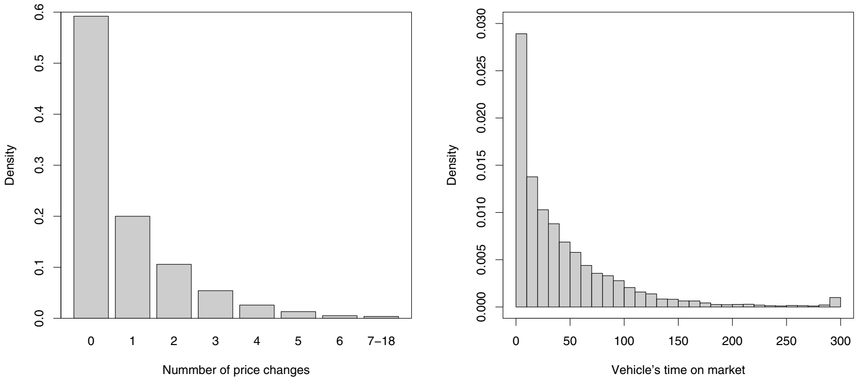
The above-mentioned pricing strategies already use in part information about supply and demand. They promise higher profits than normal waiting tactics and, therefore, can be considered as first attempts for the application of revenue management techniques in the used car sector. Nevertheless, weaknesses appear in the ability for reproduction of the procedure according to different market circumstances. Furthermore, the problem of traceability arises since the question of how and why exactly a certain price strategy is used cannot be justified clearly. In addition, operational calculations are missing, which support the target mark for the optimal average time on market (see for example Wuppermann 2003, p. 67). Also possible connections between asking prices and retail prices and the relation between asking prices and time on market must be considered.

The pricing scheme can be classified as a standard intuitive approach applied by a wide range of car dealers. As a general rule, these car dealers perform better than the industry's average and thus, provide a good benchmark for comparison with more sophisticated pricing techniques, which are developed in the forthcoming chapters.

2.4.3 Potential for Improvement

The past sections illustrated that the used car sector generates only marginal profits and shows the most potential for increased profit. To utilize the existing profit opportunities, especially in the pricing management, retailers require more sophisticated pricing strategies in automobile sales. In the following subsection, the potential for improvement is roughly demonstrated by the time on display and the number of price changes for a passenger sub-market.

In a sales study of medium-class passenger cars for the period of the year 2006, we analyzed the time on market, the respective price changes as well as the final price of sale for over 50,000 vehicles (please refer to section 5.2 for a detailed description of the mentioned market study). One of the findings was that the majority of all vehicles were sold without a price change during the whole offer phase, and only around 20 percent of all vehicles in this study experienced two or more price alignments, stated in Figure 2.5(a). Furthermore, analysis of the vehicles' time on



(a) Number of price changes until sale (b) Distribution of time on market

Fig. 2.5: Analysis of the potential for profit improvement regarding the number of price changes until a sale is made and the distribution of time on market of a used car

market revealed the use of non-optimal pricing strategies by numerous car retailers. Vehicles were offered below their market value, and thus, were sold very fast, reflected by the high fraction of vehicles with market times less than ten days in Figure 2.5(b). An optimal pricing strategy would have balanced price and expected time on market by suggesting a higher price at the beginning of the offer period and reducing the price successively over time. Another area for profit enhancement can be identified on the other extreme. Vehicles were offered for prices significantly higher than their actual market values and consequently, these cars were not sold at all, recognized by the higher bar at times of around 300 days at the right tail of Figure 2.5(b). Here, car dealers failed to recognize these shortcomings by not adjusting the price accordingly, and the cars were in the retailer's inventory for a long period (up to 300 days which corresponds to the total period of the market study), tying up capital and generating costs along with the depreciation of the vehicle. In the above mentioned study, more than six percent of vehicles offered at the beginning of the observation were not sold by the end of the study and more than half of this sample was still offered for the same price. By applying better pricing strategies, car dealers could reduce the time on market and realize higher selling prices at the same time.

These are only two examples to point out significant deficits in the pricing process for used passenger cars. Although these examples represent extremes, most of the profit potentials lie in between. To identify the full potential, a price-based revenue management strategy is developed in the subsequent chapters.

2.4.4 Outline of the Approach

The present chapter is concluded by outlining the concept of a price-based revenue management method. Focusing on used car retailers as the target group, the objective for this procedure is to maximize expected profit from the sale of one or several used vehicles under the influence of internal and external factors.

Following the description of the general revenue management framework in section 2.1.4, the present paper focuses on the two core components of the process, namely estimation of demand and optimization. Figure 2.6 shows the concept of the revenue management method. In light of the profit maximization problem of a used car retailer, the optimization module consists of several algorithms for determing optimal strategies under specific conditions. In chapter 3, deterministic and stochastic solutions are derived with regard to a continuous price region in addition to a discrete price set, enabling the retailer to adopt an appropriate revenue management system. Critical to the success of a revenue management system is the demand esti-

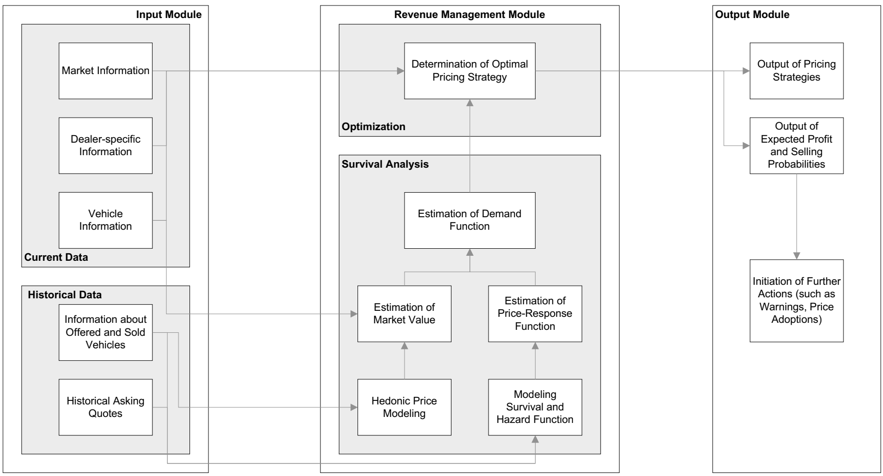


Fig. 2.6: Model of a price-based revenue management technique for a used car retailer

mation module. This work proposes the adoption of survival analysis for estimating individual demand functions, illustrated in chapter 4. There, the individual demand functions, determined on the basis of the estimate of the vehicles' market values, are modeled by parametric and semi-parametric survival functions to analyze the effects of variables such as the asking price and serve as input for the optimization module.

The proposed techniques can support retailers in establishing better and more profitable pricing strategies, including assigning an initial asking price and the adjustment of price over time. Furthermore, the program can be an early-warning system, alarming the retailer as to the point when a used vehicle has been displayed too long with the same asking price.

2.5 Summary

The German automobile industry is caught in an oligopolistic environment on account of the saturated markets of the Triad. Based on the stagnating markets, manufacturers are trying to generate growth at the expense of their competitors by the extension of the model palette and engine range. This and other components intensify price competition and lead to further shrinking profit margins. In summary, the industry is trapped in the predicament between consumption lull and growing Asian competition, rising expenses and overcapacity, and an greater price competition under shrinking profit margins.

The used car sector cannot free itself from these conditions. On the contrary, compared to the new vehicle, financial service, or spare parts and service sectors, almost no manufacturer profits from the used car sector. If losses from the used car business continue, many dealers and subsidiaries will be confronted with existential problems. None the less, in recent years, the used car market has been a release valve for automakers' overstock.

Based on these facts, used car retailers are obliged to reform the process of selling used cars to increase profits and reduce losses. Besides other process steps such as purchasing or bundling with other services, pricing offers the greatest window of opportunity for potential profit. One promising approach is to set prices of the used vehicles, incorporating internal as well as external factors to maximize expected profit. This way, a price-based revenue management method consists of two core components, namely the optimization and the demand estimation module. Whereas chapter 4 deals with the demand estimation module, the subsequent chapter concentrates on the optimization module and presents different profit-maximizing pricing algorithms.

Chapter 3
Modeling the Price-Based Revenue Management Problem

There are two fools in every market;
one asks too little, one asks too much.
Russian Proverb

The objective of this chapter is to address the optimization module of the price-based revenue management program. Optimal pricing strategies are derived for a used car retailer by applying stochastic optimal control theory and dynamic programming techniques to maximize expected profit.

3.1 Introduction

3.1.1 Statement of the Dynamic Pricing Problem

A used car retailer faces the challenge of determining a pricing strategy that maximizes the expected profit for a specific used vehicle. The dealer in particular is interested in determining the best initial asking price in addition to plotting it over the selling horizon. Internal and external factors influence this non-trivial task, in which the most important factors are supply and demand.

There are numerous pricing strategies for this price control problem. A simple but widespread approach is cost-based pricing, in which all cost variables add up to the wholesale price plus a profit margin. A temporary price adaptation occurs only unsystematically however, with the consequence that the time in stock can accumulate up to several months and even years. A more sophisticated strategy aims at reducing the average time on market by a dynamic price management. Here, a market price (or market value, respectively) is determined depending on the time elapsed. Such a pricing mechanism might propose a listing price equivalent to the market value plus ten percent for the first 30 days, then reducing the asking price to the market value plus five percent for the next 30 days and so forth. Although this strategy promises better results, it does not identify the optimum asking price for each time period, but applies rules of thumb regarding the pricing points and their duration. It does not take into account internal factors such as the remaining inventory of the retailer or the current market demand. Therefore, more sophisticated pricing strategies are needed to increase the profits of used car retailers.

In general, the dealer is faced with the dilemma of a low selling price versus a vehicle's long time in stock. If he asks for an unrealistically high price, the vehicle will remain unsold, generating high current expenses (for example, stand expenses and capital allocation expenses). On the other hand, if he asks for a relatively low retail price, the time on market will drastically decrease, but the dealer might lose profit by forfeiting a possibly higher attainable retail price. Thus, a pricing strategy has to determine the optimal set of prices that consider retail price and the cost associated with the time in stock by maximizing expected profit for a given time period.

The issue of optimal dynamic pricing is influenced by a number of factors. One of the most important criteria is capital allocation. A vehicle displayed by the dealer must be purchased, either self-financed with company capital or by credit. Instead of purchasing the car, the capital related to the vehicle could be invested in the capital market where it would gain a certain income return (the so-called *capital cost of opportunity*). Thus, cash flows generated by the sale of the vehicle have to be discounted by applying the car retailer's discount rate to account for the cost of capital. Another element are the expenses connected with the duration of the vehicle's display in the retailer's showroom. The vehicle must be regularly cleaned, maintained and repaired to keep it in good condition. All variable costs can be summarized as holding inventory costs associated with a specific vehicle.

Nevertheless, the determining criterion for the dynamic pricing function is the demand of the potential customers. The development of optimization algorithms in this chapter is undertaken assuming perfect knowledge about customer demand with regard to its functional form. The successive chapter presents a framework for estimating individual prices based on survival analysis.

3.1.2 Literature Review

In revenue management, techniques and strategies can be separated into two broad sections, namely quantity-based and price-based revenue management. The question of which approach to apply depends on the firm's ability to vary quantity or price in response to changes in market conditions. In quantity-based revenue management research, the main focus lies on the allocation of limited capacity to different demand classes under the assumption that demand can be segmented into multiple classes. In contrast, the term *dynamic pricing* as an integral part of price-based revenue management summarizes instruments of price control. If prices are viewed as variable and can be controlled on a continuous basis, then product prices can be set so as to maximize profit.

In the present chapter, the case of the single-product dynamic pricing problem with a finite stock of items and stochastic demand is considered. Due to the evolution of information technologies and the resulting availability of extensive data, academic interest into dynamic pricing has exploded over the last three decades, reflected by the increasing number of publications. A simple strategy to the problem of pricing a good is the single-price approach. Lazear (1986) considers a two-period model assuming a retailer with a single unit and a population of potential buyers whose valuation for the product is unknown. Feng and Gallego (1995) consider a two-price model with prices in both periods fixed and the only decision is when to switch from one to the other.

In contrast to periodic pricing policies, dynamic-price models are characterized by continuous price functions or, for a given set of allowable prices, by the time between two price changes. A landmark paper on dynamic pricing was published by Kincaid and Darling (1963), where the authors derived optimality conditions for a continuous-time model with demand assumed as a Poisson process and fixed intensity. Similarly, Gallego and van Ryzin (1994) consider the problem of pricing a given inventory in a continuous-time formulation. Demand is modeled as a controlled homogeneous, time-fixed Poisson process with an intensity rate. Bitran and Mondschein (1997) propose a periodic pricing review policy where prices are revised at a set of times, but generalize demand by modeling it as a non-homogeneous Poisson process.

Another approach to the single-product problem assumes a finite set of predetermined prices from which the retailer chooses. Feng and Xiao (2000) provide a systematic analysis of the pricing policy for the problem with a finite set of prices, and Zhao and Zheng (2000) study the case in which demand is modeled as a non-homogeneous, time-dependent Poisson process. For a more in-depth discussion on dynamic pricing, refer to the articles by Elmaghraby and Keskinocak (2003), Bitran and Caldentey (2003), and McGill and van Ryzin (1999).

3.1.3 Outline

The remainder of this chapter is organized as follows. In section 3.2, we define the basic deterministic, continuous-time model of a profit-maximizing car dealer by introducing the concepts of optimal control theory and dynamic programming. We finish this section by extending the deterministic model to the case where the state of the system is stochastic. Concentrating on the stochastic optimization problem in the remainder of the chapter, section 3.3 introduces the approach of discretizing the time horizon and states an algorithm for determining optimal prices, while section 3.4 describes the case when the price region is restricted to a finite set of discrete price points. In section 3.5 we extend the case to include discounting, inventory costs and values associated with the terminal state. Finally, the complete model for the single-product dynamic pricing problem is presented in section 3.6.

3.2 Basic Continuous-Time Model

In the subsequent section, the basic finite continuous-time pricing problem for a used car retailer is modeled by the *optimal control theory*. Optimal control theory was developed to find ways to control a dynamic system and can be seen as an extension of the classical calculus of variations. The control which maximizes a certain revenue function is called the *optimal control*. It can be derived by applying the maximum principle, first published by the Russian mathematicians Pontryagin, Boltyanskij, Gamkrelidze, and Mishchenko (1962). In this article, they analyzed a calculus of variations problem with constrained control variables and proved the maximum principle for these kind of problems. Pontryagin's maximum principle permits the decoupling of the dynamic problem over time using so-called *adjoint variables* into a series of problems, each of which holds at a single instant of time. The solution of the instantaneous problems is shown to give an optimal solution.

3.2.1 Statement of the Basic Control Problem

First, we describe the optimization problem as a *deterministic standard end constrained problem* within the context of optimal control theory. We begin by defining the optimal control problem in general to derive the optimal control problem of a profit-maximizing dealer afterward.

Consider a system whose state at time t is characterized by $x(t)$, also defined as the *state variable* of the system at time $t \in [0,T]$, where $T > 0$ is the terminal time for the system under consideration. Furthermore, assume that the state of the system can be controlled and let $u(t)$ be the *control variable* of the system at time t. In addition, assume that the rate of change of the state variable $x(t)$ depends on t and $u(t)$, with the initial inventory known as the initial state at time 0, $x(0) = x_0$. Thus, the evolution of $x(t)$ can be described by the controlled differential equation, known as the *state equation*,

$$\dot{x}(t) = g[x(t),u(t),t], \quad x(0) = x_0, \tag{3.1}$$

where $\dot{x}(t)$ is a commonly used notation for $\mathrm{d}x(t)/\mathrm{d}t$. Here, g is assumed to be a known and continuously differentiable function. The control variable $u(t)$ is piecewise continuous and constrained to the set of possible values $\Omega(t)$, hence

$$u(t) \in \Omega(t), \quad t \in [0,T]. \tag{3.2}$$

If the initial value x_0 of the state variable is known, that is the initial inventory, as well as the values of the control variable $u(t)$ over the whole time interval $[0,T]$, equation (3.1) can be integrated to obtain the values of $x(t)$ over the considered time interval (both are referred to as *control* and *state trajectory*, respectively).

Let $J(x,u,t)$ be the quantitative measure of the performance of the system over time, also known as the *objective function*. In other words, the objective function collects the revenue from selling the products over the available selling period $[t,T]$, and can be noted by

$$J(x,u,t) = \int_t^T f[x(s),u(s),s]\,\mathrm{d}s, \tag{3.3}$$

where $f(x(s),u(s),s)$ is assumed to be a known and continuously differentiable function. It can be considered as the instantaneous profit rate measured in currency units per time units.

Thus, the optimal control problem is formulated as follows. Among all the pairs $(x(t),u(t))$ that obey the state equation in (3.1) (so-called *admissible pair*) and satisfy the control constraints in (3.2), find that one that is the *optimal pair* (x,u), which maximizes the objective function in (3.3). More formally, the optimal control problem can be stated as

$$V(x,t) = \max_{u(s)\in\Omega(s)} \left\{ \int_t^T f[x(s),u(s),s]\,\mathrm{d}s \right\} \tag{3.4}$$

subject to

$$\dot{x}(t) = g[x(t),u(t),t], \quad x(0) = x_0,$$

where $V(x,t) = J(x^*,u^*,t)$ is the so-called *value function* or *optimal objective function*. After introducing the main concepts, we will define the continuous-time optimal control problem for a profit-maximizing car dealer.

Definition 3.1 (Optimal Control Problem)**.** Let $x(0) = x_0$ be the available initial inventory of a dealer, $p(t)$ be the price at time t, and the term $d(p(t))$ be the demand at a given time t. There are T periods and t indices the periods with the time index running forward. The dealer's objective is to determine a pricing strategy that maximizes the total revenue or the total profit under consideration of expenses, respectively. The control variable, price $p(t)$, is a piecewise continuous function of time and constrained to the set of possible values $\Omega_p(t)$. Then, the **optimal control problem (OCP)** of a profit-maximizing used car retailer can be stated as

$$V(x,0) = \max_{p(t)\in\Omega_p(t)} \left\{ \int_0^T p(t)d(p(t))\,\mathrm{d}t \right\} \tag{3.5}$$

subject to

$$\dot{x}(t) = -d(p(t)), \quad x(0) = x_0.$$

Following Gallego and van Ryzin (1994, p. 1003), several mild assumptions on the demand function $d(p,t)$ are considered. Let $r(p(t),d(t),t) = p(t)d(p(t))$ be the revenue rate for the optimal control problem of a profit maximizing used car retailer. Then, we assume that the revenue rate is continuous, bounded and concave, where the latter is based on the standard economic assumption that the marginal revenue is decreasing in output. Furthermore, we assume that the revenue rate has a unique maximizer defined by $p^* = p : r(p(t),d(t),t) = \max_{p \geq 0} r(p(t),d(t),t)$ and that there exists a unique correspondence between prices and demand so that a unique inverse $p(d(t))$ exists.

3.2.2 Dynamic Programming Approach

In this subsection, the finite-horizon dynamic program is formed corresponding to the preceding definition of the optimal control problem. First, the main concepts of dynamic programming are introduced and then the dynamic program for a profit-maximizing car dealer is defined.

Dynamic programming is a general method for solving discrete and continuous-time optimization problems and was first formalized by Bellman (1957). In his book, he stated the basic idea of dynamic programming, called the *principle of optimality*, as follows. An optimal policy has the property that whatever the initial conditions and control variables over some initial period, the control chosen over the remaining period must be optimal for the remaining problem, with the result based on the early decisions taken to be the initial condition.

In the following, sufficient conditions on the objective function are stated by applying the principle of optimality and considering what happens over a small increment of time Δt. The derived partial differential equation, also known as the *Hamilton-Jacobi-Bellman (HJB) equation*[1], is the fundamental element in optimal control theory and central to the development of the dynamic pricing models in this chapter.

Theorem 3.1. *Consider the dynamic optimization problem specified by equation* (3.4)*. The corresponding Hamilton-Jacobi-Bellman equation is given by*

$$-\frac{\partial V(x,t)}{\partial t} = \max_{u(t) \in \Omega(t)} \left\{ f[x(t),u(t),t] + \frac{\partial V(x,t)}{\partial x} g[x(t),u(t),t] \right\} \tag{3.6}$$

with the boundary conditions

$$V(x,T) = 0 \quad \forall x \quad \textit{and} \quad V(0,t) = 0 \quad \forall t. \tag{3.7}$$

[1] Named after the mathematicians William Rowan Hamilton, Carl Gustav Jacob Jacobi and Richard Bellman.

Proof. The HJB equation is derived following Sethi and Thompson (2000, p. 28). By the principle of optimality, the change of the objective function $J(x,u,t)$ consists of two parts. First, the incremental change in J from t to $t+\Delta t$, given by the integral of the instantaneous profit rate $f(x(t),u(t),t)$ from t to $t+\Delta t$, and second, the value function $V(x(t+\Delta t),t+\Delta t)$ at time $t+\Delta t$. Note that the objective function of the optimal control problem was given in (3.4) by

$$V(x,t) = \max_{u(s)\in\Omega(s)} \left\{ \int_t^T f[x(s),u(s),s]\,\mathrm{d}s \right\}.$$

Then, the principle of optimality is derived as follows. First, the integral of the objective function is divided into two subintervals by applying a small increment Δt and yielding

$$V(x,t) = \max_{u(s)\in\Omega(s)} \left\{ \int_t^{t+\Delta t} f[x(s),u(s),s]\,\mathrm{d}s + \int_{t+\Delta t}^T f[x(s),u(s),s]\,\mathrm{d}s \right\}.$$

Using the dynamic programming principle, it can be argued that the control function $u(t)$ has to be optimal for the problem beginning at time $t+\Delta t$ in state $x(t+\Delta t)$, which itself depends on state $x(t)$ and on the control function chosen over the period $s\in[t,t+\Delta t]$. Thus,

$$V(x,t) = \max_{\substack{u(s)\in\Omega(s)\\ s\in[t,t+\Delta t]}} \left[\int_t^{t+\Delta t} f[x(s),u(s),s]\,\mathrm{d}s + \max_{\substack{u(s)\in\Omega(s)\\ s\in[t+\Delta t,T]}} \left\{ \int_{t+\Delta t}^T f[x(s),u(s),s]\,\mathrm{d}s \right\} \right],$$

which can be restated by the recursive form

$$V(x,t) = \max_{\substack{u(s)\in\Omega(s)\\ s\in[t,t+\Delta t]}} \left\{ \int_t^{t+\Delta t} f[x(s),u(s),s]\,\mathrm{d}s + V(x[t+\Delta t],t+\Delta t) \right\} \tag{3.8}$$

with the boundary conditions

$$V(x,T)=0 \quad \forall x \quad \text{and} \quad V(0,t)=0 \quad \forall t,$$

since the scrap value at terminal time T is defined as zero. Similarly, the value function where the state variable equals zero is defined as zero, too. The fact that optimal strategies satisfy this equation is called Bellman's principle of optimality. Next, the integral in (3.8) can be estimated by $f[x(t),u(t),t]\Delta t$, since the instantaneous profit rate $f(x(t),u(t),t)$ is a continuous function and the control $u(t)$ can be considered as constant over the very small time increment Δt. Hence,

$$V(x,t) = \max_{u \in \Omega(t)} \{f[x(t),u(t),t]\Delta t + V(x[t+\Delta t], t+\Delta t)\}. \tag{3.9}$$

Further, assume that the value function V is a twice continuously differentiable function of its arguments and expand the second term on the right side by Taylor's theorem, obtaining

$$\begin{aligned} V(x,t) = \max_{u(t) \in \Omega(t)} &\left\{ f[x(t),u(t),t]\Delta t + V(x,t) + \frac{\partial V(x,t)}{\partial t}\Delta t + \frac{\partial V(x,t)}{\partial x}\dot{x}] \right\} \\ &+ \sigma(\Delta t), \end{aligned} \tag{3.10}$$

where $\sigma(\Delta t)$ denotes a collection of higher-order terms in Δt. Subtracting $V(x,t)$ on both sides, dividing by Δt and substituting $\dot{x} = g[x(t),u(t),t]$ gives

$$0 = \max_{u(t) \in \Omega(t)} \left\{ f[x(t),u(t),t] + \frac{\partial V(x,t)}{\partial t} + \frac{\partial V(x,t)}{\partial x} g[x(t),u(t),t] \right\} + \sigma(\Delta t). \tag{3.11}$$

Letting $\Delta t \to 0$ and dropping $V(x,t)$, since it does not depend on the control $u(t)$, finally yields

$$-\frac{\partial V(x,t)}{\partial t} = \max_{u(t) \in \Omega(t)} \left\{ f[x(t),u(t),t] + \frac{\partial V(x,t)}{\partial x} g[x(t),u(t),t] \right\} \tag{3.12}$$

with the boundary conditions

$$V(x,T) = 0 \quad \forall x \quad \text{and} \quad V(0,t) = 0 \quad \forall t.$$

The HJB equation in conjunction with the boundary conditions characterizes the optimal solution to the control problem.

Now, based on the introduction of Bellman's principle of optimality and the statement of the sufficient condition by the HJB equation as the central result in optimal control theory, the deterministic dynamic program for the optimal control problem in (3.5) is defined.

Definition 3.2 (Deterministic Dynamic Program)**.** Consider the dynamic optimization problem of a retailer specified by (3.5), where $V(x,t)$ denotes the optimal revenue to go. Then, the deterministic dynamic program for a used car dealer can be formed by the associated principle of optimality

$$V(x,t) = \max_{\substack{p(s) \in \Omega_p(s) \\ s \in [t, t+\Delta t]}} \left\{ \int_t^{t+\Delta t} p(s) d(p(s))\, \mathrm{d}s + V(x[t+\Delta t], t+\Delta t) \right\} \tag{3.13}$$

and by the corresponding HJB equation

$$-\frac{\partial V(x,t)}{\partial t} = \max_{p(t)\in\Omega_p(t)} \left\{ p(t)d(p(t)) - \frac{\partial V(x,t)}{\partial x} d(p(t)) \right\} \tag{3.14}$$

with the boundary conditions

$$V(x,T) = 0 \quad \forall x \quad \text{and} \quad V(0,t) = 0 \quad \forall t.$$

We refer to to this model as the **deterministic dynamic program (DDP)** of a profit-maximizing used car retailer.

3.2.3 Stochastic Dynamic Program for the Intensity Control Problem

In the preceding subsections, the optimal control problem and the corresponding dynamic program of a profit-maximizing car retailer are defined by assuming that the state variables of the system were known with certainty. Now, the model is extended by considering a problem of controlling a dynamic system where the state of the system over time is a stochastic process. More precisely, assume that the state of the system may be described as a *point process*. This is motivated by the fact that the actual demand process for selling a good is stochastic and may be modeled as a non-homogeneous Poisson process. Another stream of research assumes a so-called *Ito stochastic differential equation* as the state equation by introducing a stochastic process, also known as a Wiener process[2]. For further discussions on this approach, refer for example to Kamien and Schwartz (1981, p. 243).

Following the first mentioned approach, consider the univariate point process $N(t)$ defined on a measurable space. Let Ω be a set of admissible controls, and for each control $u \in \Omega$ associate a probability such that $N(t)$ admits the intensity $\lambda(u,t)$.

Now, consider the stochastic objective function

$$J(n,u,t) = E\int_t^T f[n,u(s),s]\,\mathrm{d}s, \tag{3.15}$$

where the function E takes the expected value and for each $n \in N_+$, $N(t)$ admits an intensity $\lambda(u,t)$. Then, the optimal solution of the maximizing problem

$$V(n,t) = \sup_{u(t)\in\Omega(t)} E\int_t^T f[n,u(s),s]\,\mathrm{d}s, \tag{3.16}$$

can be denoted by u^*, where $V(n,t) = J(n,u^*,t)$ is the value function. It is known that for this kind of model a non-randomized Markovian policy exists that is opti-

[2] The stochastic process obeys what is called Brownian motion or white noise.

mal with n as the state variable of the decision process. A sufficient condition for optimality can be stated by the stochastic Hamilton-Jacobi-Bellman equation.

Theorem 3.2. *Consider the stochastic dynamic optimization problem specified by* (3.16)*. Then, the stochastic Hamilton-Jacobi-Bellman equation for intensity-control problems is given by*

$$-\frac{\partial V(n,t)}{\partial t} = \sup_{u(t)\in\Omega(t)} \{f[n(t),u(t),t] + \lambda(u,t)\,[V(n-1,t) - V(n,t)]\} \tag{3.17}$$

with the boundary conditions

$$V(n,T) = 0 \quad \forall n \quad \textit{and} \quad V(0,t) = 0 \quad \forall t. \tag{3.18}$$

Proof. For the proof of the theorem, refer to Brémaud (1981, ch. VII).

There is a profound discussion and justification of (3.17) in journals about the *intensity control of point processes*, for instance in Gihman and Skorohod (1979).

In the following section, we apply the theory of stochastic intensity control and the stochastic Hamilton-Jacobi-Bellman equation to the problem of a profit-maximizing used car retailer. Assume that the customers arrive according to a non-homogeneous Poisson process. Then, the actual demand process given a certain pricing policy is a non-homogeneous Poisson process with the intensity $d(u,t)$. Assume that for any price p the intensity $d(u,t)$ is piecewise continuous in t. Furthermore, let the inventory n be the state variable. Then, the dynamic program is as follows.

Definition 3.3 (Stochastic Dynamic Program for Intensity Control)**.** Consider the dynamic optimization problem of a used car retailer specified by (3.5), where $V(x,t)$ denotes the optimal revenue to go. Then, the stochastic dynamic program for a profit-maximizing car dealer with the value function

$$V(n,t) = \sup_{p(t)\in\Omega_p(t)} E\left\{\int_0^T p(t)d(p(t))\,\mathrm{d}t\right\}$$

can be formulated by the corresponding stochastic Hamilton-Jacobi-Bellman equation

$$-\frac{\partial V(n,t)}{\partial t} = \sup_{p(t)\in\Omega_p(t)} \{d(p(t))\,[V(n-1,t) - V(n,t) + p(t)]\} \tag{3.19}$$

with the boundary conditions

$$V(n,T) = 0 \quad \forall n \quad \text{and} \quad V(0,t) = 0 \quad \forall t.$$

We refer to to this model as the **stochastic dynamic program (SDP)** of a retailer.

The sufficient condition expressed in (3.19) can be intuitively explained. Consider the case where the stochastic Hamilton-Jacobi-Bellman equation is satisfied at $p(t) = p^*(t)$ and restate the equation as follows:

$$\frac{1}{d(p^*(t))} \frac{\partial V(n,t)}{\partial t} + p^*(t) = V(n,t) - V(n-1,t). \tag{3.20}$$

Here, the term $V(n,t) - V(n-1,t)$ denotes the marginal expected revenue for the nth item at time t. The term $\frac{\partial V(n,t)}{\partial t}$ measures the marginal loss in revenue due to the elapse of time multiplied by the expected amount of time to reduce the inventory by one unit, whereas $p^*(t)$ states the gross revenue of selling one item at time t. Thus, the left side of the equation can be interpreted as the net revenue of selling one item for price $p^*(t)$, which, at the optimum, is equal to the expected marginal revenue of the nth item.

3.2.4 Closed Form Solutions

Considering the stochastic dynamic program for a profit-maximizing retailer, it is quite difficult to obtain an exact solution for arbitrary demand functions $d(p(t))$. Gallego and van Ryzin (1994, p. 1005) state a closed form solution for the demand function $d(p) = ae^{-\alpha p}$, where the arbitrary parameters $a > 0$ and $\alpha > 0$ are given.

Following Gallego and van Ryzin (1994), assume without loss of generality that $\alpha = 1$ holds for the exponential demand function by simply changing price units to $p' \leftarrow \alpha p$. Then, it can be shown that the maximum of the revenue function $r(p(t), d(t), t) = p(t)d(p(t))$ can be found at $d^* = \frac{a}{e}$ and corresponds to the optimum price $p^* = p(d^*) = 1$. Stadje (1990, p. 178), among others, derived the solution for this optimization problem with the following closed expression for the value function

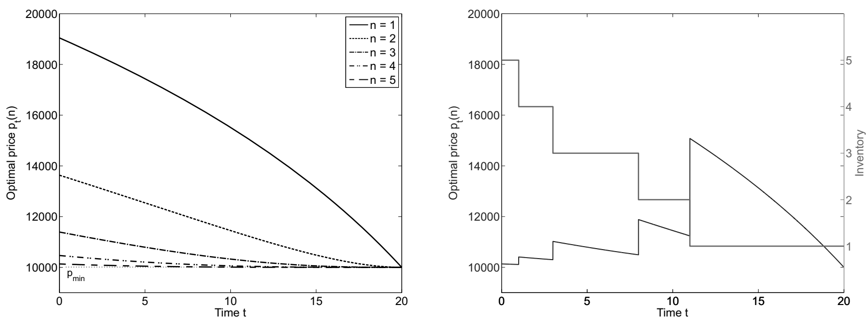
$$V(n,t) = \log\left(\sum_{i=0}^{n} (d^* t)^i \frac{1}{i!}\right),$$

whereas the optimal price path is given by $p_t^*(x) = V(x,t) - V(x-1,t) + 1$.

Example 3.1. Consider the case where demand follows a time-homogeneous Poisson process and is given by the exponential demand function $d(p) = 0.2e^{-0.0001p}$ ($a = 0.2$, $\alpha = 0.0001$). The selling horizon is assumed to be $T = 20$. Consequently, the value function can be stated as

$$V(n,t) = \log\left(\sum_{i=0}^{n} \left(ae^{-1}t\right)^i \frac{1}{i!}\right).$$

Figure 3.1(a) shows the optimal price paths for different inventory levels $n = 1,2,3,4,5$, where the dashed line corresponds to the minimum price p_{min} = 10,000. Note that the optimal prices are decreasing in time and in the level of inventory. These effects can be explained by the fact that with fewer periods to offer, there are fewer chances to sell the object. Thus, the asking price must be adjusted accordingly to increase the probability of a sale. Regarding the negative impact of the level of inventory on the optimal path of asking prices, a similar argument can be made. The probability of selling all objects by the end of the selling period decreases with the level of inventory and, therefore, the asking price must be reduced to increase the chance of selling a unit.



(a) Optimal prices for different inventories (b) Exemplary optimal pricing strategy

Fig. 3.1: Examples for optimal price paths of continuous stochastic dynamic program using a exponential demand function ($a = 0.2$, $\alpha = 0.0001$). The left figure plots optimal price paths for different inventory levels, whereas the right figure states the optimal strategy for a given demand process

In the case that the inventory contains $n = 5$ units, a sample path of the optimal price for the continuous-time model is given in Figure 3.1(b), including the corresponding graph of the time-dependent inventory. Note that the optimal price path features discrete positive jumps every time a sale occurs, whereas afterward, the optimal price diminishes until another sale can be observed.

3.3 Stochastic Discrete-Time Model

3.3.1 Statement of the Discrete-Time Control Problem

In the preceding section, we introduced the concepts of optimal control and dynamic programming and extended the model in which the state of the system is a stochastic point process. Closed form solutions for these optimal problems are found only for some specific demand functions $d(p(t))$.

Nevertheless, in terms of existing solutions for stochastic continuous-time dynamic programs, one thread of literature deals with discretizing the time horizon. Here, the time interval $[0,T]$ is divided into S subintervals of length Δt, and a finite difference equation is used to estimate the sufficient conditions.

Keeping in mind that the stochastic Hamilton-Jacobi-Bellman equation for intensity-control problems was given by

$$-\frac{\partial V(n,t)}{\partial t} = \sup_{u(t)\in\Omega(t)} \left\{ f[n(t),u(t),t] + \lambda(u,t)\left[V(n-1,t) - V(n,t)\right] \right\},$$

substitute the partial derivative $\partial V(n,t)/\partial t$ by the discrete expression $(V(n,t+\Delta t) - V(n,t))/\Delta t$. Then, the discrete Hamilton-Jacobi-Bellman equation can be stated as follows:

$$\begin{aligned} -\frac{V(n,t+\Delta t) - V(n,t)}{\Delta t} &= \\ &\sup_{u(t)\in\Omega(t)} \left\{ f[n(t),u(t),t] + \lambda(u,t)\left[V(n,t) - V(n-1,t)\right] \right\}. \end{aligned} \tag{3.21}$$

Note that the number of time intervals $S = T/\Delta t$ should be much larger that the expected demand $\int_0^T \lambda(u,t)\,\mathrm{d}s$ of the observed time interval $[0,T]$. In the following paragraph, the discrete-time stochastic dynamic program for a profit-maximizing retailer is defined.

Definition 3.4 (Discrete-Time Stochastic Dynamic Program for Intensity Control)**.** Consider the dynamic optimization problem of a retailer specified by equation (3.5), where $V(x,t)$ termed the optimal revenue to go. Note that for the stochastic optimal control problem, the state variable is denoted by n, the number of similar used vehicles in stock. Then, the stochastic dynamic program for a profit-maximizing used car retailer with a selling horizon of $S = T/\Delta t$ subintervals and the value function

$$V(n,t) = \sup_{p(t)\in\Omega_p(t)} E\left\{ \sum_0^S p(t)d(p(t)) \right\} \tag{3.22}$$

can be formulated by the corresponding discrete-time stochastic Hamilton-Jacobi-Bellman equation

$$-\frac{V(n,t+\Delta t)-V(n,t)}{\Delta t} = \sup_{p(t)\in\Omega_p(t)} \{d(p(t))\,[V(n-1,t)-V(n,t)+p(t)]\} \quad (3.23)$$

with the boundary conditions

$$V(n,S+1)=0 \quad \forall n \quad \text{and} \quad V(0,t)=0 \quad \forall t.$$

Refer to this model as the **discrete-time stochastic dynamic program (DSDP)** of a profit-maximizing used car retailer. Note that the boundary conditions for the discrete-time control problem differs from the continuous-time model, where the value function at terminal time T was defined as zero, leading to $V(n,T)=0$.

3.3.2 Solution Algorithm

In this subsection, a solution for the discrete-time stochastic dynamic program stated in definition 3.4 will be developed, assuming that demand is given by a differentiable distribution function.

Let $V(n,t)$ denote the optimal expected profit for a dealer with an inventory of n items, which obeys the sufficient condition stated by the stochastic Hamilton-Jacobi-Bellman equation in (3.19). Normalizing the small increment of time to $\Delta t = 1$, S can be replaced by the terminal time T and the corresponding Bellman equation can be formulated by

$$V(n,t) = V(n,t+1) + \max_{p(t)\in\Omega_p(t)} \{d(p(t))\,[p(t)-\Delta V(n,t+1)]\} \quad (3.24)$$

with the boundary conditions $V(n,T+1)=0$ for all n and $V(0,t)=0$ for all t, and where $\Delta V(n,t+1) = V(n,t+1)-V(n-1,t+1)$ is the marginal expected revenue of the nth item. Note that both the inventory level n and the time period t are integer variables. Assuming a differentiable demand function $d(p(t))$, the optimal pricing policy must obey the necessary and sufficient conditions stated in the following theorem.

Theorem 3.3. *Let the demand function $d(p(t))$ be continuously differentiable and strictly decreasing on $\Omega_p(t)$ and consider the discrete-time stochastic dynamic optimization problem specified by definition 3.4. Then, the optimal pricing policy at a time t for an inventory n has to obey the necessary condition given by*

$$p = -\frac{d(p(t))}{d'(p(t))} + \Delta V(n,t+1). \quad (3.25)$$

The necessary condition has a unique solution corresponding to the optimal price when the term $-d(p(t))^2/d'(p(t))$ *is decreasing in p.*

Proof. See Bitran and Mondschein (1993, p. 26).

At the optimum, it is necessary that marginal revenue equals marginal opportunity cost. Here, term $p + d(p(t))/d'(p(t))$ described the marginal revenue, and $\Delta V(n,t+1)$ the expected opportunity cost of selling the nth item in the future conditioned on not selling it at time t.

There are a number of demand functions which fulfill the necessary conditions and for which an optimum solution is found. For instance, the exponential and the Weibull distribution belong to this class of functions (cp. Bitran and Mondschein 1997, p. 69).

The solution of the discrete-time algorithm follows recursively. At first, the considered time horizon is divided into T subintervals of the length Δt. The algorithm begins at time $t = T$ (the end of the considered time period) and with a capacity of $n = 1$. Applying the boundary conditions $V(0,T) = 0$ and $V(n,T+1) = 0$, the equation in (3.25) can be simplified to

$$p = \frac{d(p(T))}{d'(p(T))}. \tag{3.26}$$

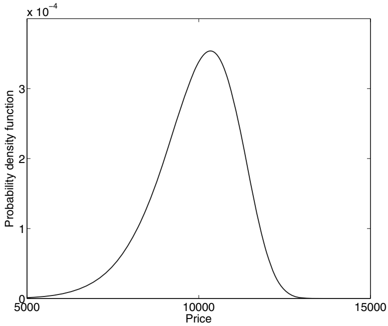
After solving the differential equation and calculating the corresponding expected optimal revenue $V(1,T)$, the algorithm can be continued by determining the optimal price for previous subintervals up to the time $t = 0$. Based on these results for $n = 1$, optimum price strategies for capacities of $n = 2, 3, \ldots, N$ can be determined using the same procedure. In the following, an example demonstrates the approach using the Weibull distribution as the demand function.

Example 3.2. Assume that demand for an object such as a used vehicle follows a two-parameter Weibull distribution, a continuous probability distribution with the probability density function

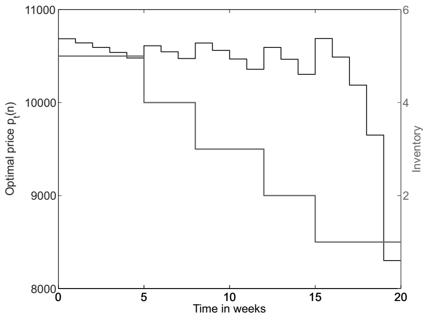
$$f(x;\beta,\eta) = \frac{\beta}{\eta}\left(\frac{x}{\eta}\right)^{\beta-1} e^{-(x/\eta)^{\beta}},$$

where $\beta > 0$ is the shape parameter and $\eta > 0$ is the scale parameter. Let the shape parameter be $\beta = 10$ and the scale parameter be $\eta = 10,450$. The Weibull distribution is a versatile distribution that obtains a variety of characteristics of other types of distributions based on the value of its parameters. For example, with a shape parameter of 3.25, the distribution appears similar to the standard normal distribution, and with $\beta = 1$ the Weibull distribution reduces to the exponential distribution.

The left graph in Figure 3.2 displays the shape of the probability density function for the stated example. Furthermore, consider an inventory of $n = 5$ identical used vehicles and a selling horizon of 20 weeks. Here, the choice of weeks instead of days as the time unit is motivated by the fact that multiple price changes within a day are rarely observed for used vehicles. In fact, prices are often changed only once during the course of a week. The right graph in figure 3.2 shows the path for an optimal pricing strategy and its corresponding inventory level for a particular outcome of the demand process. Note that similar to the example for the continuous-time model, the price path decreases until a sale occurs, where the path performs an upward jump and diminishes further until another sale is made.



(a) Weibull probability density function



(b) Price path and inventory level

Fig. 3.2: Example of an optimal price path for a discrete-time stochastic dynamic program utilizing a Weibull distribution ($\beta = 10$, $\eta = 10,450$) and an initial inventory of $n = 5$. The left figure shows the corresponding probability density function, whereas the right figure states the optimal strategy for an arbitrary demand process

3.4 Finite Price Sets

The derivation and analysis of solution procedures for a continuous set of prices is of great interest in the research community. Nevertheless, in most cases it misses the needs of the practitioners in real applications. Instead, companies often restrict prices to a small finite set due to strategic reasons, such as pricing according to the market competition. Furthermore, psychological reasons might explain a discrete pricing strategy as well. In practice, prices close to convenient whole amounts of monetary units are often used since they are familiar to customers, easy to understand and often form psychological thresholds for potential customers (Talluri and van Ryzin 2004, p. 193).

Besides, it might be difficult to implement models based on continuous price sets, as stated by Zhao and Zheng (2000, p. 383). When prices are chosen from a continuous set, the optimal pricing strategy changes continuously over time as well. Such a policy is not only difficult to implement, it is even harder to convince operators and customers of their usefulness. Mathematically, it might be hard, if not impossible, to calculate the optimal pricing strategy in closed form when prices are chosen from a continuous set. Consequently, in the following section the price region is discretized and restricted to a finite set of discrete price points.

Formally, let the control variable $p(t)$ be constrained to a finite number of M discrete prices

$$p(t) \in \Omega_p(t), \quad \Omega_p(t) = \{p_1, \ldots, p_M\}, \tag{3.27}$$

where we assume that $p_1 > \ldots > p_M$. Equivalently, the demand function $d(p(t))$ can be discretized, where $d(p_i,t)$ denotes the demand at time t when applying the price p_i.

3.4.1 *Solution for Continuous-Time Models*

In the following subsection, we will describe a procedure proposed by Zhao and Zheng (2000, p. 384) to obtain the optimal pricing policy for a continuous-time model. Assume that the optimal pricing strategy $p(n,t)$ decreases in t, that is the more time left, the higher the optimal price. Then, for fixed inventory n, the optimal pricing strategy $p(n,t)$ is a decreasing step function in t with at most $k-1$ steps, where k is the number of allowed prices. Furthermore, the time horizon can be divided into k subintervals, where the optimal price $p(n,t) = p_i$ is applied. Some of the subintervals might lead to an empty set, and thus, the price p_i will not be part of the optimal pricing strategy.

The algorithm to calculate the optimal pricing strategy follows a recursive setup starting from terminal time $t = T$ and an inventory of $n = 0$, where $V(0,t) = 0$ and $V(n,T) = 0$ hold. Due to the boundary conditions, for an inventory level of

$n = 1$ the value function $V(n,t)$ can be computed as well as the different time points $t_j(n)$ up to which each price p_j is applied as the optimal price. Now, based on the known $V(n-1,t)$, the value function $V(n,t)$ for the preceding inventory levels can be obtained. Given an inventory of $n = N$ items and a selling horizon of $t = T$, the algorithm of Zhao and Zheng (2000) can be characterized as follows.

Step 1: Initialization Set the temporary time variable to $s = T$, the temporary variable for the marginal revenue to $\Delta V = 0$, the temporary inventory level to $n = 1$ and let the number of allowed prices be given by the variable M, assuming that $p_1 > \ldots > p_M$. Furthermore, the index variable k equals the number of possible prices M increased by one, hence, $k = M + 1$. Note that the inventory level n is kept constant as long as the temporary time variable s is greater than zero. Then, the procedure is repeated again for the next inventory level until the given level N is achieved.

Step 2: Determination of optimal price at time s^- Find the highest price p_i, that is, find the smallest index i, that satisfies the following expression

$$\max_{i<k} \{d(p_i)\,[V(n-1,s) - V(n,s) + p_i(s)]\}\,.$$

Then, set $k = i$. Note that the applied expression is the right side of the corresponding Hamilton-Jacobi-Bellman equation.

Step 3: Determinating value function $V(n,t)$ **at** s^- Find $V(n,t)$, the solution of the Hamilton-Jacobi-Bellman equation

$$-\frac{\partial V(n,t)}{\partial t} = d(p_k)\,[V(n-1,t) + V(n,t) + p_k] \tag{3.28}$$

for the optimal price p_k, under consideration of the boundary conditions $V(0,t) = 0$, $V(n,T) = 0$ and $V(n,s) = V(n-1,s) + \Delta V$.

Step 4: Determination of optimal subinterval $t_k(n)$ Find the smallest time $t = s$ that satisfies the following expression

$$V(n,t) - V(n-1,t) = \min_{j<k} \frac{d(p_j,t)p_j - d(p_k,t)p_k}{d(p_j,t) - d(p_k,t)}\,.$$

Then, set $t_k(n) = s$ and $\Delta V = V(n,s) - V(n-1,s)$.

Step 5: Recursion If $s > 0$, then go to Step 2, else repeat the procedure while $n < N$.

In the first step, the procedure is initialized by setting the temporary variables for time to the terminal time, the marginal revenue to zero and the number of allowed prices to the given length of the price set increased by one. In the second step, the optimal price at the current time s is determined and the price index k is adjusted, accordingly. In agreement with the proposition that the optimal pricing strategy is a decreasing step function in t, the optimal current price p_k corresponds to the smallest price index that satisfies the HJB equation since the price set is given in descending order. In the next step of the algorithm, the value function at the current time s is determined by solving the corresponding differential equation for the HJB equation. However, the differential equation can be solved analytically for a few demand functions only, such as the exponential or the Weibull distribution. Otherwise, a numerical solution of the differential equation has to be found, for example by applying the Runge-Kutta method. Then, it must be determined for how long the current price remains optimal; hence, the time t_k at which the optimal price strategy has to be adjusted. The optimal time t_k for price p_k is found by a line search algorithm, where the left hand side of the equation in the fourth step increases in t, whereas the right hand side decreases in t. If the determined optimal time t_k is greater than zero, the algorithm is continued at the second step and the current optimal price is chosen from the remaining prices. Otherwise, the algorithm is calculated for the next inventory level until the given level N is reached.

Independently from this work, Feng and Xiao (2000) derived optimality conditions for the problem with a finite set of predetermined prices from which the retailer can choose. In addition, they showed that there is a maximum subset of prices, such that the revenue rate increases with the price and that the optimal price at any time belongs to this subset (referred to as the *maximum concave envelope*).

In the following example, we apply the algorithm of Zhao and Zheng on a profit-maximizing problem of a used car retailer.

Example 3.3. Consider a used car retailer with an inventory of $n = 1$. Demand is assumed to be a time-homogeneous Poisson process with an intensity rate following a Weibull distribution with a shape parameter of $\beta = 10$ and a scale parameter of η=10,450. Furthermore, let the selling horizon be $T = 20$ units. Two different price sets are applied, the first one consisting of prices as multiples of 1,000 ($\Omega_p^{1000}(t) = \{0, 1000, 2000, \ldots\}$) and the second one consisting of prices as multiples of 500 ($\Omega_p^{500}(t) = \{0, 500, 1000, 1500, \ldots\}$). Then, the objective of a profit-maximizing retailer is to determine the optimal price and its duration as the asking price until the next optimal price is chosen.

In Figure 3.3, the optimal price paths for both price sets are given. Regarding the first price set, the optimal pricing strategy would start with asking a price of 12,000 and, in case of no sale, would reduce the price after 0.81 time units to a price of 11,000. Since the second price set offers more prices to choose from, an optimal pricing strategy based on this price set would start asking for 11,500 before reducing the price to 11,000 after 6.95 time units. Both graphs roughly follow the same course, with the second pricing strategy at a more detailed path. Interestingly, the expected revenue for the second pricing strategy based on the wider price set generates only 0.3 percent more than the first pricing strategy, with $V_2(1,0)$=11,007.59 compared to $V_1(1,0)$=10,969.82.

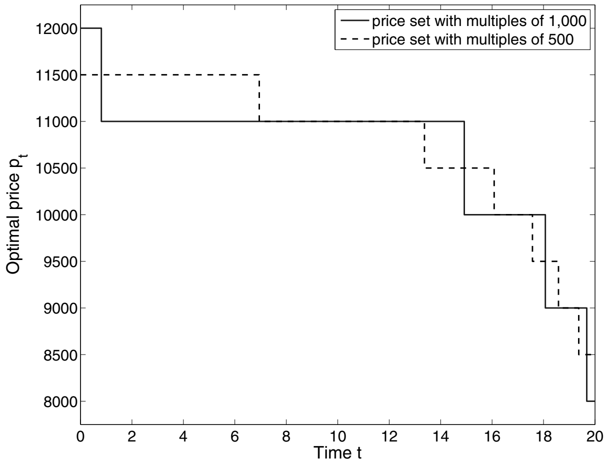


Fig. 3.3: Optimal price paths of continuous-time stochastic dynamic program for two finite price sets utilizing a Weibull distribution (β = 10, η=10,450) and an inventory of $n = 1$

3.4.2 Solution for Discrete-Time Models

The previous subsection presented an algorithm to obtain an optimal pricing policy for a continuous-time model under certain assumptions. However, if one of these assumptions does not hold, the standard approach of discretizing the time horizon can be followed. The basic principles were already stated in subsection 3.3.1 and a solution method for a continuously differentiable demand function with a continuous set of prices was proposed.

Often, the assumptions on the demand function cannot be fulfilled, but the price region can be restricted to a small discrete set. Then, a simplified algorithm based on the one proposed by Zhao and Zheng (2000) can be applied. There, the partial derivative $\frac{\partial V(n,t)}{\partial t}$ in step three for the determination of the value function $V(n,t)$ can be substituted by the discrete expression $(V(n,t+\Delta t) - V(n,t))/\Delta t$, yielding the following differential equation for calculating the value function

$$-\frac{V(n,t+\Delta t) - V(n,t)}{\Delta t} = d(p(t))\,[V(n-1,t) - V(n,t) + p(t)]. \tag{3.29}$$

Although the algorithm provides the potential to adjust prices on a small time interval, in practice, multiple price changes within a day are rarely observed by used car retailers. Instead, used car dealers often adjust prices only once during the course of a week. Therefore, the following example of a stochastic discrete-time control problem utilizes a finite price set and uses weeks as the time unit instead of shorter intervals such as minutes or days.

Example 3.4. Assume that demand for a specific used vehicle follows a two-parameter Weibull distribution with a shape parameter of $\beta = 10$ and a scale parameter of η=15,767. Furthermore, consider an inventory of $n = 1$ vehicle of the aforementioned type and let the selling horizon be 20 weeks. Given these assumptions, four different pricing strategies with different price sets are compared in the following example. More precisely, consider four different finite price sets containing multiples of 100, 250, 500, and 1,000 euro, thereby defining the possible price region to $\Omega_p^{100}(t) = \{0, 100, 200, \ldots\}$, $\Omega_p^{250}(t) = \{0, 250, 500, \ldots\}$, $\Omega_p^{500}(t) = \{0, 500, 1000, \ldots\}$ and $\Omega_p^{1000}(t) = \{0, 1000, 2000, \ldots\}$.

Figure 3.4 displays the progression of the different optimal pricing strategies over the course of 20 weeks. In general, the graphs follow roughly the same path, starting

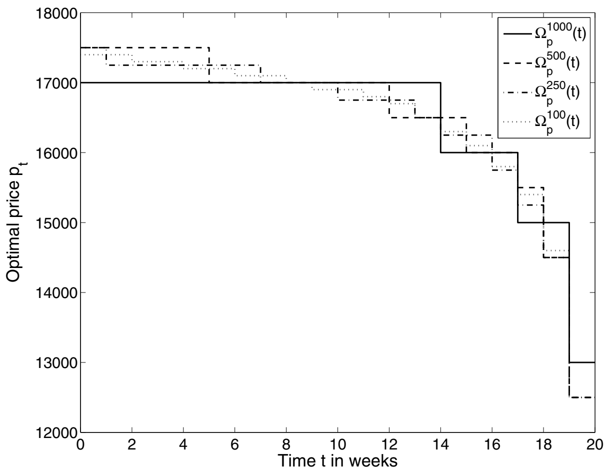


Fig. 3.4: Optimal price paths of discrete-time stochastic dynamic program for different finite price sets with multiples of 100, 250, 500, and 1,000 euro utilizing a Weibull distribution (β = 10, η=15,767) and an inventory of $n = 1$

with an initial asking price between 17,000 and 17,500 euro and closing the selling horizon with asking prices of 13,000 and 12,500 euro. Note that due to the limited number of possible asking prices, the optimal pricing strategy based on the price set containing multiples of 1,000 euro recommends its first price adjustment not until after 14 weeks and experiences only three price changes in total.

However, a comparison of the different expected profits associated with the respective optimal pricing strategies reveals only marginal differences between each price set. As expected and given in Table 3.1, the strategy using the price set of multiples of 100 euro achieves the highest expected profit. However, it generates only 0.4 percent more profit than the worst pricing strategy with multiples of 1,000 euro.

Price Set	Expected Value $V(1,0)$
$\Omega_p^{1000}(t)$	16735.6
$\Omega_p^{500}(t)$	16782.1
$\Omega_p^{250}(t)$	16797.0
$\Omega_p^{100}(t)$	16801.6

Table 3.1: Comparison of expected revenues generated by the optimal price paths for different finite price sets with multiples of 100, 250, 500, and 1000 euro utilizing a discrete-time stochastic dynamic program with a Weibull distribution ($\beta = 10$, η=15,767) and an inventory of $n = 1$

3.5 Extensions of the Basic Problem

In the previous sections, the basic concepts of optimal control theory were introduced together with the definitions of the corresponding problems for a profit-maximizing car retailer. In the succeeding part, several extensions of the basic problem are examined. First, the current value formulation is stated to discount cash flows over the time horizon. The second extension allows the revenue function to include a time-dependent cost term. Third, an additional function is included within the optimality criterion representing the value associated with the terminal state. Successively, the complete model is defined including all three extensions.

3.5.1 Current Value Formulation

In the used car sector, one of the most important cost drivers is summarized by the term *opportunity cost of capital*, since a displayed vehicle can be easily worth several tens of thousands of euro and the corresponding costs can amount to a remarkable value. In this setup, the concept of the time value of money can be applied by calculating the net present value of the expected cash flow, since cash flows in different time periods cannot be directly compared.

To obtain the net present value, expected future cash flows are discounted by the rate of return determined by comparable investment alternatives presented to the retailer. The rate of return is often referred to as the discount rate or opportunity cost of capital. In case of the car retailer, assume that a displayed vehicle is financed by debts, thereby equaling the discount rate to the interest of the debt.

Mathematically, the cost of capital can be modeled either in discrete-time or in continuous-time formulation. Let r denote the annual rate of interest. Then, the discrete discount factor is given by

$$\beta = \frac{1}{1+r},$$

and defines the number by which a future cash flow to be received at the next period has to be multiplied to obtain the current value. In continuous-time, the discount factor can be derived by taking the limit

$$\lim_{n\to\infty}\left(1+\frac{rt}{n}\right)^{n} = e^{rt}.$$

Considering the discounted cash flows, the stochastic optimal control problem defined in 3.3 can be extended as follows.

Definition 3.5 (Stochastic Dynamic Program with Discount Rate)**.** The value function for a profit-maximizing used car retailer with discounted cash flows can be defined as

$$V(n,t) = \sup_{p(t)\in\Omega_p(t)} E\left\{\int_0^T e^{-rt} p(t) d(p(t))\,\mathrm{d}t\right\} \tag{3.30}$$

and the corresponding sufficient condition for optimality given by the stochastic Hamilton-Jacobi-Bellman equation is

$$-\frac{\partial V(n,t)}{\partial t} = \sup_{p(t)\in\Omega_p(t)} \left\{d(p(t))\left[V(n,t) - V(n-1,t) + e^{-rt}p(t)\right]\right\} \tag{3.31}$$

with the boundary conditions

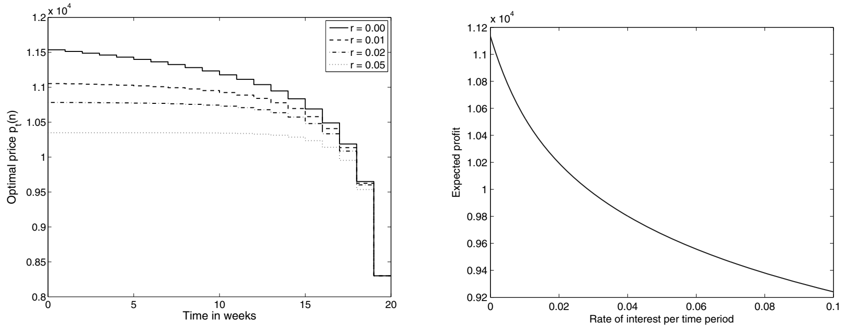
$$V(n,T) = 0 \quad \forall n \quad \text{and} \quad V(0,t) = 0 \quad \forall t.$$

To analyze the effect of the cost of capital by discounting the future expected cash flows, the example in 3.2 is modified by using different rates of interest.

Example 3.5. Consider the example in 3.2, where a used car retailer had an inventory of $n = 1$, the selling horizon was $T = 20$ and demand was assumed to be time-homogeneous Poisson process with an intensity rate following a Weibull distribution with a shape parameter of $\beta = 10$ and a scale parameter of η=10,450. Optimal prices are not restricted by a given price set, but can be chosen from a continuous price region. In Figure 3.5(a), optimal price paths are compared for different rates of interest, namely 0, 1, 2 and 5 percent. Note that these are rates per time period and are chosen for demonstration purposes only. In a real-world example, rates of interest would be much lower when applied on time units of ‘day’ or ‘week’. The graphs in Figure 3.5(a) suggest that higher interest rates result in lower optimal asking prices considering a specific time t. A higher interest rate penalizes an expected cash flow in the more distant future and supports a cash flow generated early in the course of time. Thus, the probability of a sale in the early periods is increased by lowering the asking price in comparison to a case with lower interest rates. This assessment is supported by a plot of the expected profit as a function of the interest rate in Figure 3.5(b). Here, higher interest rates correspond to lower expected profit values since cash flow generated in later time periods is penalized.

3.5.2 Inventory Costs

In all previous models, strictly speaking we maximized total expected revenue instead of expected profit since we did not include any cost terms. Therefore, we will introduce a linear holding cost term $c(n)$ that is charged on existing inventory n. At



(a) Optimal price paths for different interest rates (b) Expected profit as a function of interest rate

Fig. 3.5: Optimal price paths of continuous-time stochastic dynamic program for different interest rates utilizing a Weibull distribution ($\beta = 10$, $\eta = 10{,}450$) and an inventory of $n = 1$

time t, the existing level inventory is denoted by $n = n(t)$. Let $c(n,t)$ be the inventory cost associated with an existing capacity of n items at time t. Then, the stochastic optimal control problem with cost term can be stated as follows.

Definition 3.6 (Stochastic Dynamic Program with Cost Term)**.** The value function for a profit-maximizing car retailer with linear inventory costs $c(n,t)$ charged on existing inventories can be defined as

$$V(n,t) = \sup_{p(t)\in\Omega_p(t)} E\left\{\int_0^T \left[p(t)d(p(t)) - c(n,t)\right] \mathrm{d}t\right\} \tag{3.32}$$

and the corresponding sufficient condition for optimality given by the stochastic Hamilton-Jacobi-Bellman equation is

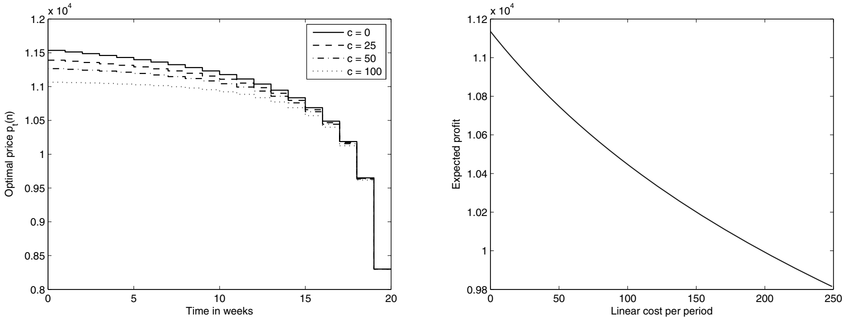
$$-\frac{\partial V(n,t)}{\partial t} = \sup_{p(t)\in\Omega_p(t)} \left\{d(p(t))\left[V(n,t) - V(n-1,t) + p(t) - c(n,t)\right]\right\} \tag{3.33}$$

with the boundary conditions

$$V(n,T) = 0 \quad \forall n \quad \text{and} \quad V(0,t) = 0 \quad \forall t.$$

In the following section, the effect of different levels of cost on the optimal price path and expected profit are analyzed.

Example 3.6. Consider the assumptions from the previous example regarding the basis model. Here, the setup is extended by a linear cost term with cost per time period of 0, 25, 50 and 100 monetary units. Similarly to the analysis of the effect of the interest rate, higher cost terms lead to flatter optimal price paths since cash flows generated in later time periods are penalized by an associated cost term. Therefore, it is more advantageous to increase the probability of a sale in earlier time periods by decreasing the corresponding asking price. Consequently, the expected profit decreases with higher cost terms, as stated in the Figure 3.6(b).



(a) Optimal price paths for different cost terms (b) Expected profit as a function of the cost term

Fig. 3.6: Optimal price paths of continuous-time stochastic dynamic program for different cost terms utilizing a Weibull distribution ($\beta = 10$, $\eta = 10,450$) and an inventory of $n = 1$

3.5.3 Values Associated with Terminal State

In economic optimization problems, often an additional function will be included within the optimality criterion representing the value associated with the terminal state. This function is called the scrape value function or salvage value function, and is needed so that the solution will make 'good sense' at the end of the time horizon. For example, a used car that is not sold at the end of the optimization period has a certain value that is not equal to zero, and hence, the usage of a salvage value is in order.

Let $S(n(T),T)$ be the salvage value of a terminal state $n(T)$ at terminal time T. Then, the stochastic optimal control problem from 3.3 can be restated as follows.

Definition 3.7 (Stochastic Dynamic Program with Salvage Value)**.** The value function for a profit-maximizing car retailer with a salvage value function $S(n(T),T)$ can be defined as

$$V(n,t) = \sup_{p(t)\in\Omega_p(t)} E\left\{\int_0^T [p(t)d(p(t))]\,\mathrm{d}t + S(n(T),T)\right\} \tag{3.34}$$

and the corresponding sufficient condition for optimality given by the stochastic Hamilton-Jacobi-Bellman equation is

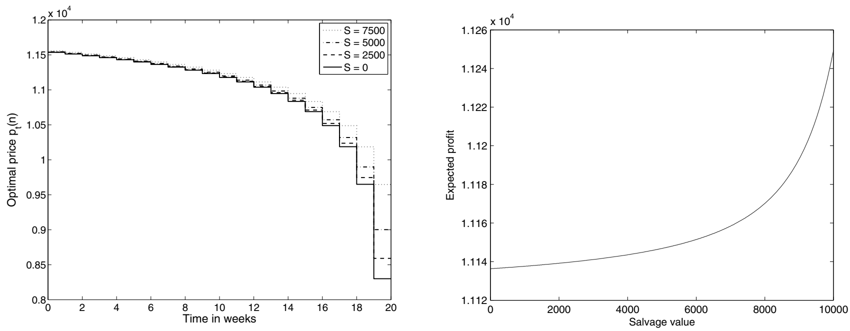
$$-\frac{\partial V(n,t)}{\partial t} = \sup_{p(t)\in\Omega_p(t)} \{d(p(t))\,[V(n,t) - V(n-1,t) + p(t)]\} \tag{3.35}$$

with the boundary conditions

$$V(n,T) = S(n,T) \quad \forall n \quad \text{and} \quad V(0,t) = 0 \quad \forall t.$$

An analysis of different levels for the salvage value illustrates the effect on the optimal price path and the corresponding expected profit.

Example 3.7. Carrying on with the assumptions from the previous example, in addition consider four different values at the terminal state $T = 20$, namely 0, 2500, 5000 and 7500. The determined optimal price paths in the left graph of Figure 3.7 show that the level of the salvage value influences the price path, especially at the end of the selling horizon. With few increments of time left to sell the object, the chance of not selling increases and, thus, the probability of drawing on the salvage value increases as well. However, the retailer can request a higher price for the vehicle, since not selling the object would result in a higher price (i.e. the salvage value). This effect diminishes as more selling time is left, noted by the left tail of the optimal price paths. Consequently, higher salvage values also result in higher expected profits, since they influence the last cash flow for the terminal state.



(a) Optimal price paths for different salvage values (b) Expected profit as a function of the salvage value

Fig. 3.7: Optimal price paths of continuous-time stochastic dynamic program for different salvage values utilizing a Weibull distribution ($\beta = 10$, $\eta = 10{,}450$) and an inventory of $n = 1$

3.6 The Complete Model

The chapter is concluded by defining the complete dynamic pricing model for the profit-maximizing car retailer, incorporating the cost of capital, inventory holding costs and the value associated with the terminal state.

Definition 3.8 (Complete Stochastic Dynamic Program)**.** The value function for a profit-maximizing used car retailer with discounted cash flows, linear inventory costs $c(n,t)$ and a salvage value of $S(n(T),T)$ can be defined as

$$V(n,t) = \sup_{p(t)\in\Omega_p(t)} E\left\{\int_0^T e^{-rt}\left[p(t)d(p(t)) - c(n,t)\right]\mathrm{d}t + e^{-rT}S(n(T),T)\right\} \tag{3.36}$$

and the corresponding sufficient condition for optimality given by the stochastic Hamilton-Jacobi-Bellman equation is

$$-\frac{\partial V(n,t)}{\partial t} = \sup_{p(t)\in\Omega_p(t)} \left\{d(p(t))\left[V(n,t) - V(n-1,t) + e^{-rt}\left(p(t) - c(n,t)\right)\right]\right\} \tag{3.37}$$

with the boundary conditions

$$V(n,T) = 0 \quad \forall n \quad \text{and} \quad V(0,t) = 0 \quad \forall t.$$

An optimal pricing strategy for a used car retailer can be derived by one of the several approaches presented in this chapter. Applied in an actual market situation, a discrete-time procedure will most likely prove to be the most realistic and most appropriate choice. This assessment is based on several reasons. First, a continuous-time formulation implicates continuously changing prices, as already stated in this chapter, and, thereby, generates costs associated with each price change. Furthermore, customers might not accept continuously changing prices. And lastly, discrete-time models are easier to formulate, to implement and to execute. For example, a retailer could determine optimal prices based on a weekly basis, thereby only adjusting prices once a week at most.

Another realistic and recommended constraint is the adoption of a finite price set, thereby reducing the number of possible asking prices. This approach is advantageous in several ways. First, the application of a finite price set decreases the complexity of the mathematical models, especially for procedures that do not provide closed form solutions. And second, the decision for a finite price is also based on findings in marketing studies, which conclude that some prices are more appealing to customers than others. Thus, prices close to convenient, whole amounts or marginally below are more familiar to customers and provide a higher incentive to buy the product.

In sum, all these procedures may determine optimal pricing strategies for a profit-maximizing used car retailer, given that the specific assumptions are met. In consideration of the above stated arguments however, a discrete-time finite price set approach is proposed for the application in practice and will be applied in the subsequent chapters.

3.7 Summary

In this chapter, the price-based revenue management problem for a profit-maximizing used car retailer was defined with a focus on selling individual durable goods. After introducing the deterministic, continuous-time model, the stochastic version was derived describing the state of the system by a point process and modeling the state equation as an Ito stochastic differential equation. Continuing with this approach, a closed form solution was stated. However, this is only possible for a limited number of specific demand functions, such as the exponential distribution. To overcome these obstacles, the considered time interval was divided into subintervals and algorithms were presented dealing with discrete-time problems.

Up to that point, the stated models determined optimal price paths by using a continuous price region, thereby allowing the optimal price to change with each single subinterval of time. Another thread of model approaches was associated with the set of possible prices. Here by contrast, the models restricted the choice of possible prices to a finite set and determined an optimal policy based on this subset of prices, considering both continuous-time and discrete-time models. The chapter concluded by extending the price-based revenue management problem to factors important in practice, such as discounting future cash flows, incorporating cost associated with existing inventory, and considering a salvage value of an object at the end of the selling horizon.

Chapter 4
Survival Analysis: Estimation of the Price-Response Function

The height of ability consists in knowing the price of things.
FRANCOIS DE LA ROCHEFOUCAULD
French Writer (1613–1680)

In the previous chapter, dynamic pricing models were developed to determine optimal pricing strategies for a profit-maximizing used car retailer. The analysis of these models clearly revealed the importance of the demand function and its functional form to the application of the optimization module. However, the identification of an adequate functional form and the estimation of its parameters poses great challenges for both researchers and practitioners, especially for products traded infrequently such as used vehicles. Therefore, the following chapter presents a conceptual framework for estimating individual demand functions by applying survival analysis with special regard to the used car sector and its particular characteristics.

Beginning with an introduction into the fundamentals of reservation prices and a discussion on different approaches retrieving them, section 4.1 further identifies the potential for the application of survival analysis in estimating customers' responses to price variations. Section 4.2 focuses on the basic concepts of survival analysis, on the different forms of regression models and finally on determining influential variables for estimating price response functions.

4.1 Reservation Price and Price Response Function

Economists and marketing researchers rely on measures of consumers' willingness-to-pay in estimating demand and in designing optimal pricing strategies (cp. Wertenbroch and Skiera 2002, p. 228). Developing pricing algorithms, researchers often ignore the complicated task of estimating customers' response to the offered pricing strategies. Mostly, the functional form of the demand function is assumed to be given or known, respectively, and sometimes even chosen in a mathematically tractable way (in terms of the discussed algorithm). But especially for practitioners, the success of the application of a theoretical pricing algorithm depends heavily on data processing and collecting, on data pre-processing and cleaning, and ultimately on an adequate functional form for demand.

4.1.1 The Reservation Price Concept

Price plays a significant role as one element of the marketing mix, since it determines the amount a customer must sacrifice to acquire a product or service. [1] Furthermore, price is distinguished from other marketing devices by the force and speed it has on sales, and by the shortness of time it takes to change it.

In this context, knowledge about customers' *reservation prices* is the key information used to develop rational and optimal pricing strategies. Following Jedidi and Zhang (2002, p. 1352), a consumer's reservation price for a specific product or service is the price at which the consumer is indifferent toward either buying or not buying a given product considering the alternatives available to the consumer. Balderjahn (2003, p. 389) defines the reservation price as the maximum price a consumer is willing to pay for a certain product. Since the reservation price is an upper limit of the acceptable price range, it corresponds directly with the perceived value of the product. Figure 4.1 illustrates the conceptual framework for the reservation price approach. A customer assesses a product's perceived value (measured in monetary units) and, in a particular purchase situation, compares this value with the current price asked by the supplier[2]. Assuming rational behavior, the customer selects from several competing products the offer with the highest *customer surplus*, that is, the greatest difference between perceived value and price. This is also known as the principle of maximization of customer surplus.

Consumer research literature provides two main approaches for explaining the impact of prices on consumer decisions (see Bettman, Luce, and Payne 1998, p. 187). *Rational choice theory* assumes a rational decision maker with well-defined

[1] Simon (1989, p. 1) defines the price of a product or service as the number of monetary units a customer has to pay to receive one unit of that product or service.

[2] For a discussion of the asking price and its different roles please refer to section 4.4.1.

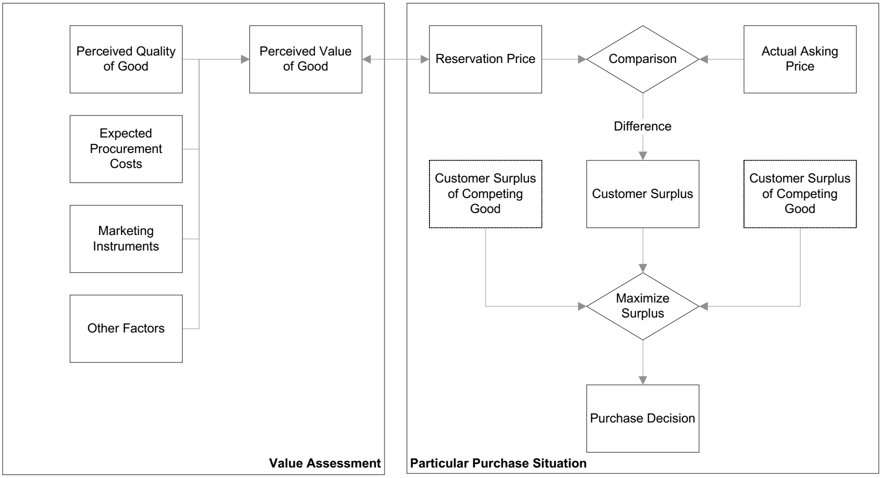


Fig. 4.1: Framework of the reservation price concept within a particular purchase situation

stable preferences. The rational decision maker assigns a utility to each option in a known choice set and has the ability to determine which option will maximize the received value, and thus selects accordingly. In contrast, the *information-processing approach* assumes that the decision maker has limitations on his capacities for processing information and limited rationality. He has no well-defined preferences and may construct them on the spot when making a choice.

Having information about customers' reservation prices (or their willingness to pay), one can derive the price response functions necessary for the development of optimal pricing strategies. Simon (1989, p. 14) differentiates between an individual and an aggregated level to explain the concept of price response functions. At an individual level, a customer faces two possible situations: the 'binary case', where he buys one unit at most (e.g. durable goods) and the 'variable quantity-case', where the customer buys several units depending on the price (e.g. many nondurable goods).

In the following section, we concentrate on the binary case, where the customer buys the good if the price is less than the product's perceived value, and the reservation price (or the maximum price) is equal to the product's value.

4.1.2 Classification of Estimation Methods

There are numerous techniques and methods for measuring and estimating price responses, which can be classified using different criteria. Balderjahn (2003) categorized methods based on the degree of aggregation, distinguishing between individual and aggregated willingness-to-pay. Nagle and Hogan (2006) apply a matrix consisting of two dimensions, one being the conditions of measurement and the other being the variable measured, whereas Breidert (2006) distinguishes between observations and surveys as the main classes for estimation methods.

Based on the classification of Sattler and Nitschke (2003), we organize the various techniques for estimating price sensitivity into three classes: market data (revealed preferences), stated preference data, and purchase experiments, as stated in Figure 4.2. One way to estimate the willingness-to-pay is to analyze *market data*, generated by past sales data or by experimentally-controlled experiments. Given the ability to track sales transactions under realistic marketing-mix conditions, past sales data enable marketers to analyze trends and predict future demand. However, sales data are frequently available only at an aggregated level and thus, conceal possible individual price differences. Furthermore, data about demand that is not satisfied is often not collected at all. Consequently, the reduction in price variation allows only limited explanation regarding the willingness-to-pay estimates. By applying laboratory purchase experiments to vary prices systematically, these problems can be prevented, but these experiments are quite time and financially cost intensive (cp. Nagle and Hogan 2006, p. 282). Another class of instruments used to measure price sensitivity is known as *stated preference data*, which can be distinguished between direct and indirect surveys. In direct surveys, test persons are asked to state how much they are willing to pay for a certain product whereas in indirect surveys, some sort of rating or ranking procedure of different products is applied (compare Breidert 2006, p. 40). The most successful methodology in market research is choice-based conjoint analysis, designed to determine trade-offs among product features or attributes including price (Wertenbroch and Skiera 2002, p. 229).

In contrast, *purchase experiments* attempt to simulate the realism of in-store experiments avoiding high financial and time allocations. One popular type of purchase experiment is represented by the Vickrey auction, where bidders submit written bids without knowing the bid of the other people in the auction. The highest bidder wins, but pays only the second highest bid price. Thus, in a Vickrey auction each bidder maximizes her expected utility by revealing her true valuation. Another subclass of purchase experiments is lotteries, with the Becker, DeGroot, and Marschak's (1964) procedure as the most established mechanism.

In the next subsection, we elaborate on these methods by focusing the interest on market data approaches since the most effective way to conduct a price-response analysis is by using non-experimental data.

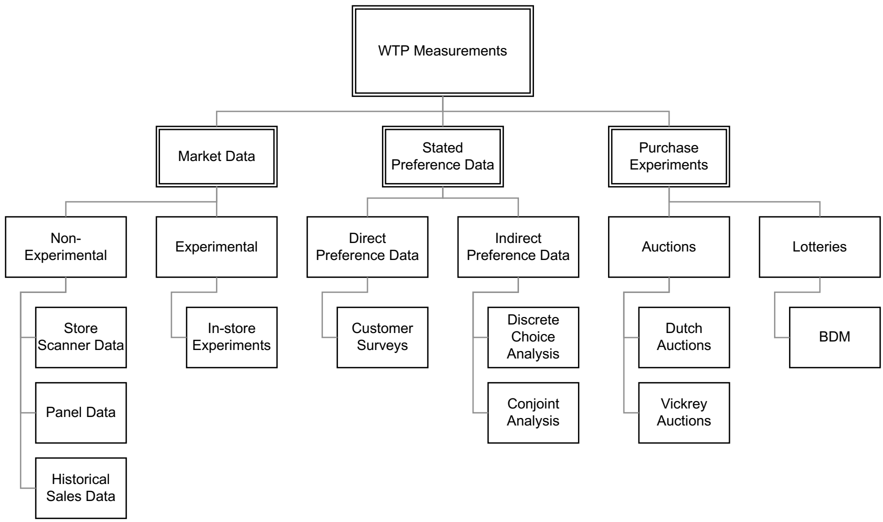


Fig. 4.2: Techniques for measuring price response functions, following Sattler and Nitschke (2003)

4.1.3 Further Discussion of Non-Experimental Market Data Methods

Often companies' marketers face the situation where market data on past sales and prices are readily available and the process of collection and preparation is already integrated within the existing data warehouse module of an enterprise resource planning system. Generally, marketing literature distinguishes between three types of past sales data to estimate price response functions: panel data, store scanner data, and historical sales data. Subsequently, these estimating techniques will be described with the scope of extending to the needs of the used car sector.

Panel Data Panel data consist of individual purchase information from different households, where each household keeps a daily record of products purchased and prices paid. Thus, price variation is observed at an individual level. The setup of panel data however is expensive and may not adequately represent the market (cp. Nagle and Hogan 2006, p. 284).

Store Scanner Data An alternative source of sales data may be derived from auditing price transactions and sales at individual stores. Using scanners to generate the data, one can observe responses to short-term price variations. Thus, scanner data have become a major source of information on price response, especially within the retail industry (cp. Nagle and Hogan 2006, p. 287).

Historical Sales Data Due to the increasing use of database systems, historical sales and price data are often available at a relatively low cost. Historical sales data however are frequently available only at an aggregate level, and thus conceal individual price differences. Consequently, the reduction in price variation allows only limited explanation regarding the estimate of willingness-to-pay. Therefore, Sattler and Nitschke (2003) assess historical data as not feasible for estimating price response function due to the missing price variation.

Market data therefore offer a cost-conservative and easy method for estimating price responses at an aggregate level.

4.1.4 Time Duration Market Data

Many industry sectors are characterized by few and infrequent transactions, which makes it difficult to develop feasible data sets based only on historical sales data. Furthermore, sales data only provide information about a specific point in time, disregarding data at the time of offer.

One approach would be to extend the method of market data by incorporating the quote history of a product. When price offer history is combined with sales success data and price adjustment information, it is possible to estimate the probability at which a sale is likely. It can also determine the decline in the probability in a sale as prices increase, and vice versa (cp. Nagle and Hogan 2006, p. 287). Thus, this approach is a specific form of estimating the price response function and in the following section will be named *time duration market data* according to the application within time duration models.

Definition 4.1 (Time Duration Market Data)**.** Time duration market data can be classified as a non-experimental market data method, as it includes information about past sales and corresponding sales prices. In addition to the information of the product and its characteristics, data is collected regarding the chronological course of the supplier's asking price. Thus, not only the time duration of a sale, but also the information that a particular good has not been sold for the offered price is utilized.

In consequence, with the application of time duration market data, an offered good does not only generate one data set with its disposal, but also a number of different data sets every time a price is adjusted. This increases the data base for estimating price response functions significantly. This fact is all the more important for sectors that are characterized by infrequent transactions, such as the used car market.

Real estate economics is one of the few areas where time duration market and listing price data are explicitly incorporated in estimating demand. There, search theory is used to explain the matching process of the time required to bring together a seller of a heterogeneous product with an appropriate buyer. The seller tries to maximize the discounted present value of the realized profits from sale by choosing an appropriate listing price, whereas the buyer searches among available products for the good that maximizes the utility provided by the good (cp. Anglin, Rutherford, and Springer 2003, p. 96).

Although several papers in real estate economics incorporate time duration market data into the search theoretic context, information about price changes during the product's marketing period is typically missing from data. Instead, only the most recent asking price and the corresponding time duration on market are considered. To our knowledge, only Knight (2002) presents a study where price revision information is incorporated in the data. Thus, one of the main results in this chapter is the development of a framework to estimate price response functions by applying information about price adjustments and their associated time on market, particularly with regard to the used car sector.

4.1.5 Application to the Used Car Sector

A used car retailer faces the trade-off between a too much time for selling the vehicle versus a too low of a price eventually received. Setting the asking price too high, the retailer discourages potential buyers and decreases the probability of encountering a buyer, thus risking having the vehicle on the market for an overly long time. Conversely, by setting the asking price too low, a quick sale might occur, but at the expense of lost potential profit with a better pricing strategy.

Standard search theory postulates a direct relationship between asking price and time-on-market, since the listing price may influence the rate at which buyers inspect a vehicle. Furthermore, the asking price acts as a signal to potential buyers as an upper bound of the seller's reservation price (Horowitz 1992).

To analyze the relationship between the asking price of a vehicle and the customers' response in the form of probability of sale, the collection of time duration market data is necessary. This information should include internal data such as the original list price, the asking price, the status of sale, the final selling price, and external data (such as market conditions, market size of comparable vehicles and the rank of the asking price in comparison to vehicles of the same market).

In the next section, the concept of survival analysis is introduced to estimate the probability of a vehicle's sale depending on a given set of internal and external variables.

4.2 Survival Analysis

4.2.1 Introduction

Individual market data on durable goods such as vehicles or assets indicate that these commodities spend a comparatively long time on the market before they are sold. The data are often limited, where standard regression methods cannot estimate the probability of a sale. Survival analysis is an excellent approach for analyzing these issues. The following section introduces the basic concepts of survival analysis and presents several models to determine explanatory variables. These techniques and models are applied to estimate price response functions for the used car sector on the basis of data provided by the sales study further described in section 5.2.

Survival analysis examines and models the time it takes for events to occur (in this context, the term ‘failure’ is often used). Such events might include the failure of components, the death of patients, or workers securing jobs after unemployment. The distribution of such survival times is intrinsically interesting, but attention more commonly focuses on the relationship between survival and one or more predictors, usually termed as *explanatory variables* or *covariates* in the survival analysis literature (Oakes 2001, p. 99). The central concept of statistical methods in survival analysis is based on the conditional probability of an event taking place (e.g., the probability of an individual leaving unemployment in the tenth week given that he or she has been unemployed for nine weeks, (see Kiefer 1988, p. 648)).

In this chapter, the used car sector is considered as an environment for the application of survival analysis. There, the vehicles are displayed with an associated asking price by the retailer. In this setting, the time until the disposal occurs, including possible price alignments, constitutes this particular vehicle’s survival time on the market. Furthermore, not only is the time observed until the car is sold, but also the factors affecting the survival time. That is, they are explanatory variables that help explain the observed survival times. These factors might include the price offered, the market value, the attributes of the vehicle, and external market conditions.

4.2.1.1 History

The origins of survival analysis can be traced back to the early work on mortality by John Graunt, who introduced the concept of ‘life tables’ in his book *Natural and Political Observations on the Bills of Mortality* in 1662 (Graunt 1662; Sutherland 1963). The modern era of survival analysis started at the beginning of the twentieth century with studies on durability of industrial devices. During World War II, the reliability of military equipment became a critical issue and the term ‘lifetime analysis’ came to be used by industrial reliability engineers. In the post-war period, methods of reliability were applied to the study of survival time for cancer patients and cancer researchers coined the term 'survival analysis'. Around this time, two landmark papers for the development of modern survival analysis were published, in which Kaplan and Meier (1958) formalized the product-limit estimator, and Cox (1972) introduced the proportional hazards model.

In the past five decades, survival analysis has become one of the most frequently used methods for analyzing data in various disciplines. Examples include survival times of patients in clinical trials (biomedical sciences), the lifetime of machine components (industrial engineering), and the duration of unemployment or durations of strikes in economics.

4.2.1.2 Censoring and Truncation

A special feature related to survival analysis arises from the fact that only some individuals experience the event, and subsequently, survival times are unknown for a subset of the sample. Such incomplete observation of the failure time is called *censoring* (Cox and Oakes 1985, p. 5). Possible censoring schemes include right censoring, where all that is known is that the individual is still alive at a given time, or left censoring, when all that is known is that the individual has experienced the event of interest prior to the start of the study.[3] Figure 4.3 illustrates the occurrence of right censoring. Here, individual 1 and 4 are examples for non-censored subjects since the observation starts after the beginning of the study and the subjects experience the event (such as a sale of a used vehicle) before the end of the study. In contrast, individual 2 is withdrawn from the study and, thus, subject to censoring (i.e. the survival time is set to the elapsed time up to the end of the study). Regarding individual 3, it experiences the event (assuming that it occurs) beyond the end of the market study, but only the duration up to the end of the study is recorded. Again, the survival time is underestimated compared to the actual time to event. Lastly, although individual 5 experiences its event within the study period, the complete duration is unknown since its starting time is located before the beginning of the study. The situations for individual 2, 3 and 5 are known as right censoring.

A second feature of survival data is *truncation*, which is due to a sampling bias that dictates that only individuals whose lifetimes lie within a certain interval $[Y_L, Y_R]$ can be observed. Again, cases can be subdivided into left truncation, where an event is observed if its lifetime is greater than a truncation variable Y_L, and into right truncation, where the measurement is less than a truncation variable Y_R.

Note that the appearance of censoring and truncation results in the need for special methods of analysis, because standard statistical methods are not applicable to survival times.

[3] For a comprehensive discussion on censoring and truncation, see Klein and Moeschberger (2003, p. 55).

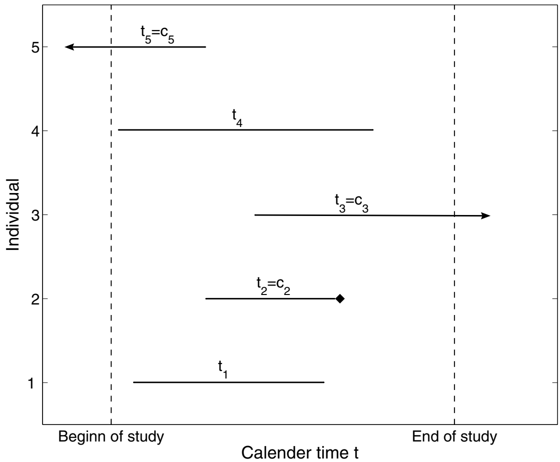


Fig. 4.3: Examples for right-censored data. The dashed lines correspond to the beginning and the end of the market study. The diamond signals a withdrawal of the subject from the study, whereas the arrow illustrates the fact that the survival time is right censored.

4.2.1.3 Requirements and Goals of Survival Analysis

Cox and Oakes (1985, p. 1) define three basic requirements for the application of survival analysis. A time origin must be unambiguously defined, a scale for measuring the passage of time must be agreed upon, and the meaning of failure must be entirely clear. Kleinbaum (1996) states three basic goals of survival analysis: 1) the estimation and interpretation of survival and hazard functions from survival data, 2) the comparison of survival and hazard functions, and 3) the assessment of the relationship of explanatory variables to survival time.

Each of these three statements plays an important role in the forthcoming analysis. The first two are needed to cluster the survival data into subgroups via comparison of individual survivor functions. However, the main focus is on the assessment of the explanatory variables, using different mathematical modeling approaches, such as the semi-parametric Cox model or parametric regression models.

4.2.1.4 Assessing the Used Car Sector

To our knowledge, the automobile industry in general and the used car sector in particular have not yet been scrutinized with survival analysis methodology, at least not in academic papers published before 2006. This is somewhat astonishing, because this industry sector provides excellent conditions in terms of information systems to generate, store, and process market data.

According to the requirements defined by Cox and Oakes (1985, p. 1), the used car sector is an ideal field for survival analysis. Consider the case where a used vehicle is displayed by a car retailer for sale. The time origin for the analysis starts with the first day of offering the vehicle for a certain price. At that point, one of three circumstances can occur: 1) the car is sold (at the asking price), and the case of failure (failure to survive on the market) appears; 2) the vehicle is not sold, but still offered for the current asking price; or 3) the vehicle is not sold for the asking price, and the retailer adjusts the price according to a pricing strategy. In this case, the generated data set is censored and a new data set starts with a new time origin. The scale for measuring the passage of time is measured in days for all data sets, because price changes rarely occur during the daytime.

The objective of this chapter is to provide a framework for determining individual price response functions. Within this scope, the main interest in survival analysis can be found in the assessment of the relationship of explanatory variables to survival time. More specifically, the influence of the asking price on the time on market of a specific vehicle is analyzed. Apart from this, other internal and external variables are considered as well as their relationship to survival time. The used car sector is selected as the backdrop, since it features excellent conditions for the application of survival analysis.

4.2.2 *Basic Concepts and Notation*

In this subsection, we consider a homogeneous population of individuals and introduce the basic mathematical terminology and notation for survival analysis.

4.2.2.1 Distributions of Failure Time

The survival time is a non-negative random variable measuring the time until some specific event. This event might be a failure, a response to treatment with medicines or even death. Survival data can be generally described in terms of two related functions, namely the *survival function*, which is the probability that an individual survives from the time origin to a specific time t, and the *hazard function* (which is also called the hazard rate), which is the chance that an individual during observation at time t experiences the event in the next instant.

Let T be a random non-negative variable from a homogeneous population. The relationship between the probability distribution function $f(t)$ and the cumulative distribution function $F(t)$ is then defined as

$$F(t) = P(T < t) = \int_0^t f(u)\,du, \tag{4.1}$$

where $F(t)$ can be interpreted as the probability that an individual does not survive beyond the time t. For a continuous random variable T, the survival function $S(t)$ is a non-increasing function in t with the value of 1 at the origin and 0 at infinity. It is defined as

$$S(t) = P(T > t) = \int_t^\infty f(u)\,du = 1 - F(t). \tag{4.2}$$

Thus, the probability density function $f(t)$ can be calculated as

$$f(t) = -\frac{d}{dt}S(t). \tag{4.3}$$

Another representation of the distribution of survival time is the hazard function (also known as the conditional failure rate, the inverse of the Mill's ratio and the hazard rate)

$$\begin{aligned} h(t) &= \lim_{\Delta \to 0} \frac{Pr[(t \leq T < t + \Delta t)|T \geq t]}{\Delta t} \\ &= \frac{f(t)}{S(t)}, \end{aligned} \tag{4.4}$$

which assesses the instantaneous risk of failure at time t, conditional on survival to that time. Blossfeld, Hamerle, and Mayer (1986) describe the hazard function as the central concept for the analysis of survival data. A related term is the *cumulative hazard function* $H(t)$, defined as

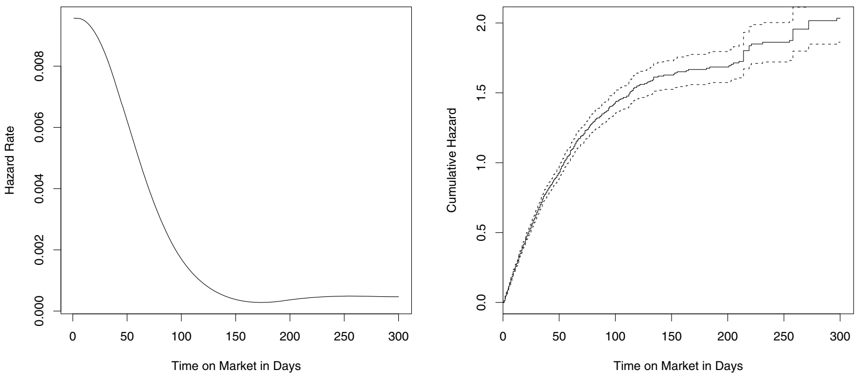
$$H(t) = \int_0^t h(u)\,du = -\ln S(t). \tag{4.5}$$

When T is a continuous random variable, the survival function $S(t)$ can be rearranged to

$$S(t) = \exp[-H(t)] = \exp\left[-\int_0^t h(u)\,du\right]. \tag{4.6}$$

The cumulative hazard function is interpreted as the number of events that would be expected for each individual by time t if the event were a repeatable process and is used as an intermediate measure for estimating the hazard function $h(t)$ (see Clark, Bradburn, Love, and Altman 2003, p. 234). The functions representing the survival data are mathematically equivalent and, consequently, either of them can be the basis of statistical analysis.

Example 4.1 (Hazard and cumulative hazard function). In a study examining the used car market in Germany from December 2005 to September 2006, asking prices were collected for specific model types. Each new asking price represented a new data set and was censored in the case when no sale occurred or the asking price was changed. Figure 4.4 shows the estimated hazard and cumulative hazard function for model ‘J-1’. Of specific interest is the hazard function, where the decreasing plot



(a) Hazard function (b) Cumulative hazard function

Fig. 4.4: Hazard and cumulative hazard function for time-on-market data for vehicle model ‘J-1’ in the period of January to September 2006 along with pointwise 95 percent confidence intervals for the cumulative hazard function

can be interpreted as follows. In the beginning of the offer phase, the likelihood of a sale (that is, of a failure to survive on the market) is higher and decreases as time passes. After being on the market for approximately 150 days, the hazard rate is constant over the remaining period. The estimate of the cumulative hazard function is extended by 95 percent pointwise confidence intervals.

4.2.3 Estimation of the Survivor Function

Dealing with real data, often the issue arises that data has to be grouped or clustered to provide more extensive samples for successive modeling and, thus, feature higher significance for the findings. More precisely, in the setup of the used car sector, it is of interest to find model types that follow the same characteristics in terms of survival and hazard rate. Then, types with similar survival functions can be analyzed as one sample providing higher explanatory power. The approach is to identify these correlations by comparing estimated survival functions before any parametric modeling is conducted.

Therefore, in the first subsection basic estimators for the survival function and the cumulative hazard function are introduced. In the second subsection, test statistics are described that make a weighted comparison between estimated survival functions of different samples. These test statistics represent the basis for clustering given data into sub-samples.

4.2.3.1 Nonparametric Methods

Nonparametric estimation of survival or distribution functions provide a useful way of analyzing unordered univariate survival data (see Lawless 1982, p. 71). Thus, it is possible to analyze survival data without any parametric assumption about the form of the underlying distribution.

Kaplan-Meier Estimator

The standard non-parametric estimator of the survival function from observed survival times was proposed by Kaplan and Meier (1958). An important advantage of the *Kaplan-Meier* (KM) method (also known as the Product Limit Estimator) is the extension of the estimate to censored data.

Let the observed times until an event from a sample of a population of size N be $t_{(1)} \leq t_{(2)} \leq \ldots \leq t_{(N)}$. Then, the probability of not experiencing the event at time $\hat{S}(t_{(i)})$ is given by the probability of surviving past the previous failure time, $\hat{S}(t_{(i-1)})$, multiplied by the conditional probability of surviving past time $t_{(i)}$, given survival to at least time $t_{(i)}$. The latter can be represented by the term $(1 - \frac{d_i}{n_i})$, with d_i as the number of events (e.g. death) at time $t_{(i)}$ and r_i the total number of individuals exposed to risk at time $t_{(i)}$. Referring to the used car sector, the risk set r_i contains all vehicles that are offered and not sold, yet. Then, the KM estimate of the survival function is defined as

$$\begin{aligned} \hat{S}(t_{(i)}) &= \hat{S}(t_{(i-1)}) \left(\frac{r_i - d_i}{r_i} \right) \\ &= \prod_{i:t_{(i)}<t} \left(1 - \frac{d_i}{r_i} \right). \end{aligned} \tag{4.7}$$

Since the probability of failure between two successive events is assumed to be zero, the estimated survival is a step function with jumps at times of observed events.

Nelson-Aalen Estimator

While the KM estimator can also be used to estimate the cumulative hazard function, an alternative approach is to estimate the cumulative hazard directly, first proposed by Nelson (1972) and further derived by Aalen (1978). The *Nelson-Aalen estimator* is defined by the term

$$\hat{H}(t_{(i)}) = \sum_{j=1}^{i} \frac{d_j}{r_j}. \tag{4.8}$$

The estimated cumulative hazard is the sum of the hazards at all failure times up to time $t_{(i)}$ and can be interpreted as the expected number of failures per unit at risk.

Klein and Moeschberger (2003) identified two primary uses of the application of the Nelson-Aalen estimator in survival analysis. First, it can be used to assess a parametric model by plotting the transformed estimator in a way that the graph should be approximately linear if the underlying parametric model is correct. And secondly, the estimator provides a simple estimate of the hazard rate $h(t)$. In addition, Breslow (1972) suggested estimating the survival function based on the Nelson-Aalen estimator as $\hat{S}(t_{(i)}) = e^{\hat{H}(t_{(i)})}$.

Example 4.2 (Nonparametric estimation of survival function). Continuing with the market study introduced in example 4.1 and considering the data examining the used car market in Germany, two estimators of the survival function are compared for model 'J-1'. Figure 4.5 compares the two estimators on the survival data along with 95 percent confidence intervals for the Kaplan-Meier, where a difference between the Kaplan-Meier and the Nelson-Aalen estimator is barely noticeable. Note that the survival functions are defined only up to 300 days, the largest of the observation times. Furthermore, the plot shows that the estimated survival functions, although decreasing, do not go down to zero at the end of the study. This appearance is based on the fact that right censoring has occurred within the data since not all vehicles were sold at the end of the study and prices were changed during the observation. This leads to this specific data set and beginning a new one with the new asking price.

4.2.3.2 Hypothesis Testing Comparing Survival Estimators

Hypothesis testing for the equality of the survival function of two or more groups is a common requirement in survival analysis. Within the scope of this work, these test statistics are applied to make statements about possible clustering of populations within a given data set. Therefore, methods are presented testing the null hypothesis that there is no statistically significant difference between the analyzed populations in the probability of an event at any time point. Following Klein and Moeschberger (2003, p. 191) the set of hypotheses for K populations and their associated hazard

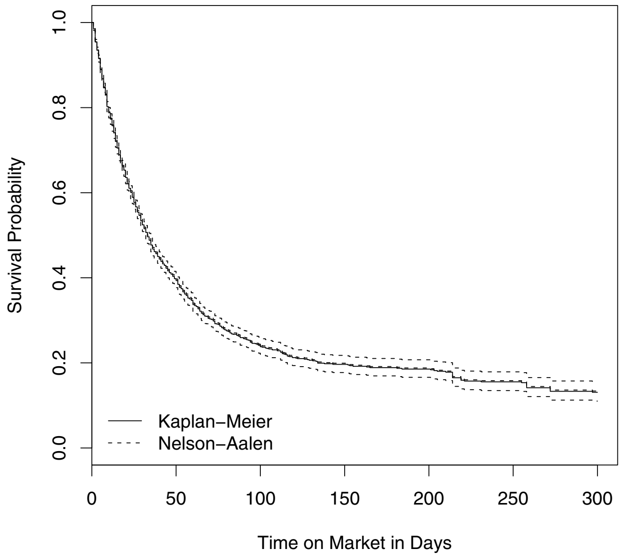


Fig. 4.5: Kaplan-Meier (solid) and Nelson-Aalen (dashed) estimate for the time-on-market data set of model 'J-1', along with pointwise 95 percent confidence intervals for Kaplan-Meier. Note that the difference of both estimates is barely observable

rates is formally given by

$$\begin{aligned} &H_0 : h_1(t) = h_2(t) = \ldots = h_K(t), \text{ for all } t \leq \tau, \text{ versus} \\ &H_1 : \text{at least one of the hazard rates is different for some } t \leq \tau \end{aligned} \tag{4.9}$$

with τ as the longest time in the study. The alternative hypothesis rejects the null hypothesis if at least one of the populations differs from the others at some time. Let d_{ij} be the observed events at time t_i in the jth population out of r_{ij} at risk, and $d_i = \sum_{j=1}^{K} d_{ij}$ as well as $r_i = \sum_{j=1}^{K} r_{ij}$ be the total number of events and the total number at risk at time t_i. The test statistic is based on the Nelson-Aalen estimator and basically calculates for each group, at each event time, the number of events one would expect since the previous event if there were no difference between the groups. The difference to the observed number of events is then multiplied by a weight function $W_j(t_i)$ and summed up over all event times. Defining a class of weight functions where the common weight is shared by each group, thus $W_j(t_i) = W(t_i)$, the test statistic is given by

$$Z_j(\tau) = \sum_i^D W(t_i)\left[d_{ij} - \frac{r_{ij}d_i}{r_i}\right], \quad j = 1, \ldots, K. \tag{4.10}$$

The value of the test statistic is compared to a χ^2 distribution with $(D-1)$ degrees of freedom, where D is the number of groups.

A variety of different weight functions can be found in the literature, with the log-rank test (Peto, Pike, Armitage, Breslow, Cox, Howard, Mantel, McPherson, Peto, and Smith 1977) as the most widely used method for comparing survival functions with a common weight function of $W(t) = 1$. Another choice of the weight function is $W(t_i) = r_i$, based on the generalization of the Kruskal-Whitney-Wilcoxon test (Gehan 1965) and the generalization of the Kruskal-Wallis test (Breslow 1970). Therefore, the test statistic is also known as the Breslow test. Finally, Tarone and Ware (1977) proposed a weight function of $W(t_i) = \sqrt{r_i}$, which gives more weight at points with the most data.

Example 4.3 (Test statistic comparing two vehicle models). In example 4.1, a market study examining the used car market in Germany was introduced. Now, the question whether there is a difference in the probability of a sale for two populations arises, namely vehicles from the new model line (solid graph) and vehicles from the previous model line. Table 4.1 summarizes different test statistics applying different weight functions.

Test Statistic	Weight Function $W(t_i)$	χ^2	p-Value
Log-rank test	1.0	335.16	0.000
Breslow test	r_i	260.95	0.000
Tarone-Ware test	$\sqrt{r_i}$	298.99	0.000

Table 4.1: Test statistics for two-sample tests

Figure 4.6 shows the estimated survival functions for the two populations. Here the curves appear to be significantly different, with the new model's probability of survival on the market always higher than for the previous model, a fact supported by the test statistics given in Table 4.1 with p-values of zero to three decimal places for all tests. Keeping in mind that the survival on the market corresponds to the circumstance that the vehicle is not sold, the comparison of the model's survival estimators allows the conclusion that used vehicles belonging to the previous model are sold with a higher probability than used vehicles from the new model. Although counterintuitive on the first glance, the assessment might be explained by the fact that asking prices for used vehicles from the previous model are lower with regards to the vehicles' estimated market value than for vehicles belonging to the new model.

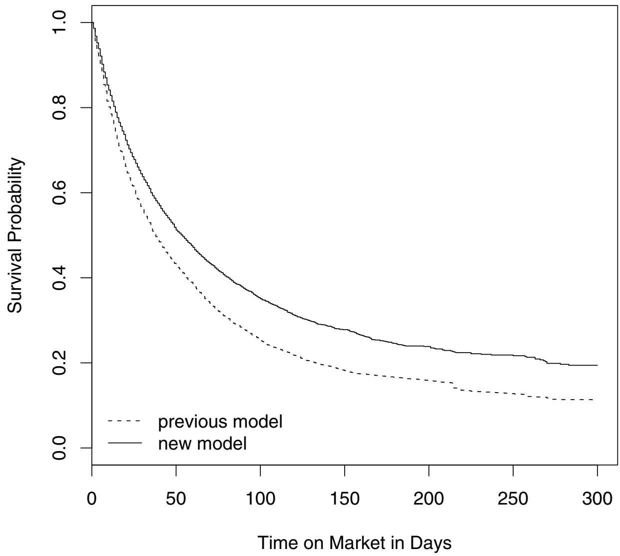


Fig. 4.6: Estimated survival function for the new model line (solid) versus the previous model line (dashed)

However, the used Nelson-Aalen estimator is a non-parametic estimation method and does not include the asking price as an explanatory variable.

4.3 Parametric Regression Modeling

The previous sections introduced the basic idea of survival analysis, including the definition of the hazard and survival function, the construction of nonparametric estimators and tests for differences between groups. This section reviews parametric regression models which are used to analyze the effect on failure time by explanatory variables.

Klein and Moeschberger (2003, p. 45) stated that one challenge in dealing with survival data is that of adjusting the survival function to account for concomitant information (also referred to as covariates or explanatory variables). Furthermore, it is of interest to ascertain the relationship between failure time and explanatory variables. Therefore, allowing survival to be assessed with respect to several factors simultaneously might improve the overall accuracy and offers an explanation of the effect for individual factors.

Basically, approaches to modeling covariate effects are divided into two classes, namely *proportional-hazard models* (including the semi-parametric Cox model) and fully parametric regression models and secondly, *accelerated failure time models* (also known as location-scale models for $\log T$). Both approaches are discussed in the forthcoming section.

4.3.1 The Cox Proportional Hazards Model

4.3.1.1 Introduction and Notation

The Cox proportional-hazards regression model was introduced in an seminal paper by Cox (1972), and is one of the most commonly used approaches for analyzing survival data. The model examines the relationship between survival, expressed by the hazard function and a set of covariates. The Cox model is a sub-class of the proportional hazards family, which is characterized by the property that different individuals have hazard functions proportional to each other. Thus, the hazard function of T, given x, can be written in the form

$$h(t|x) = h_0(t)g(x'\beta), \tag{4.11}$$

where T is a nonnegative variable representing survival time and $x = (x_1, x_2, \ldots, x_q)'$ a $q \times 1$ vector of covariates (also known as explanatory variables) associated with the survival time. The term $h_0(t)$ can be interpreted as the baseline hazard function, and both, h_0 and $g(\cdot)$ may involve unknown parameters. Treating the baseline hazard function as nonparametrically, Cox (1972) proposed a particular proportional hazards model of the form

$$h(t|x) = h_0(t)e^{x'\beta}, \tag{4.12}$$

where $\beta = (\beta_1, \ldots, \beta_q)'$ represents a vector of regression coefficients. It is called a semi-parametric model due to the fact that a parametric form is assumed for the covariates while the baseline hazard is treated non-parametrically. The model is also considered a proportional hazard since the ratio of two individuals with fixed covariate vectors x_i and x_j is constant over time:

$$\frac{h(t|x_i)}{h(t|x_j)} = \frac{h_0(t)e^{x_i'\beta}}{h_0(t)e^{x_j'\beta}} = \frac{e^{x_i'\beta}}{e^{x_j'\beta}}.$$

The Cox model is essentially a multiple linear regression of the logarithm hazard on the covariates x with the baseline being the intercept term that varies over time. The covariates are multiplied on the hazard with regard to the key assumption of the proportional hazards model (cp. Bradburn, Clark, Love, and Altman 2003, p. 432).

Equivalently, one can formulate the Cox model in terms of the regression coefficients and the survivor function $S(t|x)$. Then, the survivor function arising from (4.6) equals

$$S(t|x) = \exp\left[-\int_0^t h_0(u)e^{x'\beta}\,du\right] = [S_0(t)]^{e^{x'\beta}}, \tag{4.13}$$

with the baseline survivor function $S_0(t) = \exp[-H_0(t)]$.

4.3.1.2 Estimation of Regression Parameters

One goal of survival analysis is to make an inference about the influence of the covariates by estimating the relating coefficients. Taking censoring into consideration, one approach would be to maximize the likelihood function for the observed data simultaneously with respect to the regression coefficients and the baseline survivor function (cp. Lawless 1982, p. 345).

A more attractive approach of estimation represents the *method of partial likelihood* due to Cox (1972), where the likelihood function does not depend upon $h_0(t)$. Following the elaboration in Klein and Moeschberger (2003, p. 232), the partial likelihood for the case of no ties between the event times can be motivated from the probability Pr_i that a subject experiences an event at time t_i with covariates $x_{(i)}$, given one of the subjects in the risk set $r(t_i)$ experiences an event at the same time t_i:

$$\begin{aligned} Pr_i &= \frac{Pr[\text{individual dies at } t_i \mid \text{survival to } t_i]}{Pr[\text{one death at } t_i \mid \text{survival to } t_i]} \\ &= \frac{h_0(t_i)e^{x_{(i)}'\beta}}{\sum_{j\in r(t_i)} h_0(t_i)e^{x_j'\beta}}. \end{aligned}$$

The partial likelihood is formed by multiplying these probabilities over all distinct failure times d:

$$PL(\beta) = \prod_{i=1}^{d} \frac{h_0(t_i)e^{x_{(i)}'\beta}}{\sum_{j\in r(t_i)} h_0(t_i)e^{x_j'\beta}}$$
$$= \prod_{i=1}^{d} \frac{e^{x_{(i)}'\beta}}{\sum_{j\in r(t_i)} e^{x_j'\beta}}. \tag{4.14}$$

The natural log of Cox's partial likelihood $pl(\beta) = \ln PL(\beta)$ can then be derived to

$$pl(\beta) = \sum_{i=1}^{d} \left\{ x_{(i)}'\beta - \ln \sum_{j\in r(t_i)} e^{x_j'\beta} \right\}. \tag{4.15}$$

Using the score vector $s(\beta) = \left(\frac{\partial}{\partial\beta_1}PL(\beta), \ldots, \frac{\partial}{\partial\beta_q}PL(\beta)\right)'$ by taking the derivatives with respect to β, the partial maximum likelihood estimates are found by solving the set of q equations $s(\beta) = 0$. This can be done numerically by using iterative methods such as the Newton-Raphson technique (see Therneau and Grambsch 2000, p. 40).

A comprehensive discussion for the extension of partial likelihood formulation when ties are present is found in Klein and Moeschberger (2003, p. 237). There, three ways of constructing partial likelihoods are suggested, including the Breslow approximation (Breslow 1974), the Efron approximation (Efron 1977) and the exact partial likelihood function (Cox 1972). The exact partial likelihood function involves permutations that may be time-consuming to compute. Harrell (2001, p. 467) presents three commonly used algorithms for computing the partial likelihood in consideration of the ties. The Breslow approximation derives a satisfactory approximate log likelihood function when the number of ties is relatively small, whereas the Efron estimate is significantly more accurate than the Breslow one. It is very close to the exact partial likelihood, unless the proportion of ties is extremely large. In section 5.3.1, different methods of dealing with tied data are applied estimating the regression coefficients for a Cox model.

4.3.1.3 Example for Adopting the Cox Model to The Used Car Problem

This section concludes with the adoption of the Cox model to the profit-maximizing problem of a used car retailer. More precisely, in the following subsection there is an example determining the explanatory variables for the survival of a used vehicle on the market assuming the semi-parametric functional form of a Cox model. Based on the determined covariates and the corresponding coefficients, section 4.4 illustrates the process of estimating individual price response functions on the basis of a survival model as the input variable for calculating optimal pricing strategies developed in chapter 3.

This example uses data from a market study where the course of the asking prices and final sales were monitored over a period of ten months. The study and its findings are discussed in more detail in the next chapter. Here, the data are only used to illustrate the estimate of a Cox proportional hazards model.

Example 4.4 (Adoption of a Cox Proportional Hazards Model). In section 5.2, a market study is presented of 4,564 vehicles within the German used car market. The primary outcome variable is represented by time-on-market (TOM) of a used vehicle and, thus, characterizes its survival. The factors considered in the following survival analysis are degree-of-overpricing ('DOP'), size of the relative market ('market size'), the position within this market based on the relative asking price ('quantile'), the number of previous owners ('po'), and the age of the vehicle in weeks ('age') (refer to section 5.2.1 for a more in-depth discussion of the variables).

Assume that the relationship of the asking price relative to the market value of the vehicle (denoted as 'DOP') plays a significant role in explaining the vehicle's time on market. The results of the semi-parametric proportional hazards regression is stated in Table 4.2 along with an 'analysis of variance' table describing the estimated standard error, the relative risk of the effects and the univariate Wald test with its corresponding p-value. The relative risk, given by the exponent $\exp(b)$ of the coefficient b, represents the difference of one unit in the covariate values, whereas the univariate Wald test infers about single coefficients.

Variable	Coefficient	Std. Error	Wald χ^2	Degrees of Freedom	p-Value	Relative Risk
DOP	-5.1258	0.284	325.05	1	$< .0001$	0.006
Market size	-0.0045	0.001	38.35	1	$< .0001$	0.996
Quantile	-0.4829	0.119	16.36	1	0.0001	0.617
PO	0.0194	0.035	0.31	1	0.5790	1.020
Age	-0.001	0.001	16.82	1	$< .0001$	0.999

Table 4.2: Analysis of variance table for a Cox model fitting the used car example including the variables DOP, market size, quantile, po, and age, utilizing the Efron method for handling ties

The fit of the Cox proportional hazards model in Table 4.2 suggests that all regression coefficients are statistically significant with the exception of the variable 'number of previous owners' (po), where the corresponding p-value of the Wald test is $p = 0.5790$. The degree-of-overpricing (DOP) is the most important variable, as indicated by the large Wald χ^2-value. Each one unit change in DOP is associated with a change of 0.006 in the relative risk of leaving the market. Therefore, based on the regression we can conclude that higher asking prices and thus higher degrees-of-overpricing lead to a decreased risk of leaving the market. Consequently, the expected time-on-market will increase for a used vehicle. On the contrary, lower asking prices and, consequently, a lower DOP will cause a reduction in the expected time-on-market of a used vehicle.

4.3.2 Accelerated Failure Time Models

4.3.2.1 Introduction and Notation

Referring to Cox and Oakes (1985, p. 51), the accelerated failure time (AFT) model represents the primary competitor to the proportional hazards model. It relates covariate linearly to the logarithm of the failure time (Kalbfleisch and Prentice 1980, p. 32).

As before, let T be the nonnegative survival time and $x = (x_1, x_2, \ldots, x_q)'$ be a $q \times 1$ vector of covariates (also known as explanatory variables) associated with the survival time. In this setting, the components of x may represent continuous variables, dichotomous or categorial characteristics as well as time-dependent variables.

Assuming constant covariates, the AFT model is written mathematically as an ordinary regression model with the natural logarithm of survival time T of the form

$$Y = \ln T = -x'\beta + \sigma W, \quad W \overset{iid}{\sim} S_0(\cdot), \tag{4.16}$$

with β as the vector of regression coefficients, σ as a scale parameter, $S_0(\cdot)$ as a known baseline survivor function, and W as an error term with a suitable distribution independent of x. The name of the AFT model is justified by the fact that the effects of identified covariates accelerate or decelerate the time scale.

Assume that all covariates are equal to zero. Then, the error term σW can be viewed as a reference distribution for $x = 0$, and will be translated to the time scale by defining $T_0 = e^{\sigma W}$. Let $S_0(t)$ denote the survival function when the covariate vector $x = 0$, that is

$$S_0(t) = P(T_0 > t) = P(e^{\sigma W} > t). \tag{4.17}$$

Including now the covariates x and keeping in mind that they have a multiplicative relationship on survival time, the survival function of the accelerated failure time model is defined as

$$
\begin{aligned}
S(t|x) &= P(T > t|x) = P(Y > \ln t|x) \\
&= P(-x'\beta + \sigma W > \ln t|x) \\
&= P(e^{\sigma W} > te^{x'\beta}|x) \\
&= S_0[te^{x'\beta}].
\end{aligned} \tag{4.18}
$$

The last expression is interpreted as the probability that an individual characterized by the covariates x will be alive at time t is equal to the probability that a reference object will be alive at time $te^{x'\beta}$. Thus, the effect of the explanatory variables in the original time scale is to accelerate or decelerate the time scale by $e^{x'\beta}$. Based on the defined survivor function, one can derive the hazard function in terms of the baseline hazard h_0:

$$
h(t|x) = h_0[te^{x'\beta}]e^{x'\beta}. \tag{4.19}
$$

Mostly, the AFT model is applied as a parametric model, where the baseline survivor function S_0 (or the distribution of the error term W, respectively) is taken from some parametric class of distribution and the parameters to estimate, besides the regression parameters β, are those of the selected distribution. For example, if the baseline function is assumed to follow a standard normal distribution, then the survival time will follow the log-normal distribution.

In the parametric case, the AFT model is named after the distribution of the survival time. Table 4.3 presents often applied parametric AFT models and their corresponding error term distributions.

Distribution of Error Term W	Distribution of Survival Time T
Extreme value (1 parameter)	Exponential
Extreme value (2 parameters)	Weibull
Normal	Log-normal
Logistic	Log-logistic
Log-gamma	Gamma

Table 4.3: Typical distributions and corresponding AFT models

4.3.2.2 Estimation of Regression Parameters

The standard approach for the estimate of the influence of explanatory variables is to maximize the likelihood function for the observed data simultaneously with respect to the regression coefficients and the baseline survivor function. Assuming that data is only subject to right-censoring and that censoring times are independent of survival time, data can be represented by the tuple (T_i, δ_i), where δ represents the censoring indicator. The term δ indicates whether an event occurs at time T ($\delta = 1$) or not. Thus, if C_i is the time an individual left the study, the observed time equals $\min(T_i, C_i)$.

For a parametric model with x as the vector of covariates and β as the corresponding vector of regression coefficients, the probability function is defined as

$$f(t|x,\beta) = h(t|x,\beta)S(t|x,\beta),$$

and the survival function as

$$S(t|x,\beta) = exp\left[-\int_0^t h(u|x,\beta)\,du\right],$$

respectively. Then, the likelihood function can be derived by

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{n} f(t_i|x_i,\beta)^{\delta_i} S(t_i|x_i,\beta)^{1-\delta_i} \\ &= \prod_{i=1}^{n} h(t_i|x_i,\beta)^{\delta_i} S(t_i|x_i,\beta) \\ &= \prod_{i=1}^{n} h(t_i|x_i,\beta)^{\delta_i} exp\left[-\int_0^{t_i} h(u|x_i,\beta)\,du\right]. \end{aligned} \tag{4.20}$$

The natural logarithm of the likelihood function $l(\beta) = \ln L(\beta)$ can then be written as

$$l(\beta) = \sum_{i=1}^{n}\left(\delta_i \ln h(t_i|x_i,\beta) - \int_0^{t_i} h(u|x_i,\beta)\,du\right). \tag{4.21}$$

Other accelerated failure time models can be found in the literature, including a method proposed by Buckley and James (1979) and semi-parametric AFT models. In the case of the semi-parametric estimation, the baseline survivor function is unknown and estimated non-parametrically. This class of models has not been widely used, due to the complicated estimate procedures, but an overview can be found in Wei (1992). Bagdonavicius and Nikulin (2004) assess the AFT model as a good choice when the survival time distribution class is to be known. Nevertheless, the AFT models have some shortcomings, such as the restricting assumption of the error term. Furthermore, the effect of the regression variables is assumed to be constant and multiplicative on the time scale, whereas in the proportional hazards model it acts multiplicatively on the hazard function for T.

4.3.2.3 Example for Adopting the AFT Model to the Used Car Problem

The illustration of the accelerated failure time model is concluded by its adoption to a used car dataset. To fit the parametric model, in the following example the log-logistic distribution is assumed as a model for survival data. The hazard rate for this distribution is hump-shaped; it initially increases and then decreases.

Example 4.5 (Adoption of a Accelerated Failure Time Model). Here, the dataset from the Cox model in example 4.4 will be re-examined employing the log-logistic AFT model using the main effects of degree-of-overpricing ('DOP'), size of the relative market ('market size'), the position within this market based on the relative asking price ('quantile'), the number of previous owners ('po'), and the age of the vehicle in weeks ('age').

Assume that the survival time t follows a log-logistic distribution. Then, the survival function, where time is modeled directly, is given by

$$S(t) = [1 + \lambda t^{\alpha}]^{-1} \tag{4.22}$$

with the corresponding hazard rate

$$h(t) = \alpha\lambda t^{\alpha-1}[1 + \lambda t^{\alpha}]^{-1}. \tag{4.23}$$

Note that the hazard rate of the log-logistic AFT model increases initially and, then, decreases with time. Transforming survival time by taking the natural logarithm $Y = ln(T)$ and considering a vector of covariates where λ is replaced by $\exp(x'\beta)$, the survival function is given by

$$S(t|x) = \left[1 + \exp\left(\frac{\log(t) - x'\beta}{\sigma})\right)\right]^{-1} \tag{4.24}$$

with $\alpha = 1/\sigma$.

In Table 4.4, the coefficient estimates, corresponding standard errors, Wald χ^2 and p-values for the estimated AFT model utilizing a log-logistic distribution for survival time are given. The results confirm the findings from fitting the Cox model; all variables are statistically significant, with the exception of the 'number of previous owners'. The regression further acknowledges that 'degree-of-overpricing' is the most important variable in estimating the survival time.

4.3.3 Summary

This section reviewed parametric regression models to analyze the effects on failure time by explanatory variables. The Cox proportional hazards model and the accel-

Variable	Coefficient	Std. Error	Wald χ^2	p-Value
(Intercept)	−3.4104	0.353		
DOP	5.9168	0.333	315.80	$< .0001$
Market size	0.0044	0.001	37.97	$< .0001$
Quantile	0.4384	0.134	10.68	0.0011
PO	−0.0434	0.039	1.22	0.2686
Age	0.0013	0.001	20.46	$< .0001$
Log(Scale)	−0.2701	0.016		

Table 4.4: Analysis of variance table for an AFT model utilizing a log-logistic distribution

erated failure time model were presented and adapted to a used car dataset. Based on this foundation, the question of how an individual price response function can be estimated utilizing survival analysis models arises. The following section answers this question, whereas the forthcoming chapters apply the stated theory to the data from a market study including the processes of model selection and model building.

4.4 Estimation of the Price Response Function

The objective of the present chapter is to develop methods for estimating the price response function and to determine optimal pricing strategies in the used car market. Since this industry is characterized by few and infrequent transactions, the method of time duration market data was proposed to benefit from information about the asking price history of the used vehicle. Thus, not only the knowledge about a sale but especially the information about the decline of a sale by a potential customer is incorporated into the modeling of demand for a certain durable good such as a used vehicle. Next in this chapter, survival analysis is introduced as an approach to model price response functions by estimating a used car's duration on the market until it has been sold. Parametric regression models are described in order to make inferences about the explanatory variables and their effect on the survival of used vehicles.

In this section, the actual process of estimating an individual price response function of a used vehicle is illustrated. The price response function represents the basis for the optimization algorithm and is therefore one of the most critical parts of the price-based revenue management application.

To be able to estimate customers' responses to prices and price changes by analyzing survival times, there has to be a significant relationship between asking prices and time-on-market. This section starts by motivating and describing the relationship among the asking price of a durable good, its time-on-market, and the probability of a sale. However, the asking price cannot be directly incorporated within survival regression as a significant covariate due to the problem of comparing and clustering heterogeneous goods such as used vehicles.

Therefore, the variable 'degree-of-overpricing' is introduced as an abstract measure for a vehicle's asking price relative to its market value and the estimation of the market value of used vehicles is illustrated by applying hedonic price modeling. Based on these calculations, a survival model can be regressed on given data and the expected survival for an individual used vehicle can be estimated. Assuming that the degree-of-overpricing ratio is identified as an explanatory variable and thereby the asking price as part of the ratio, a function can be derived capturing the dependence between asking prices and probability of a sale, known as a price response function.

4.4.1 Relationship between Asking Price and Time on Market

The primary goal of a profit-maximizing used car retailer is to sell the vehicle for as high a price as possible and as quickly as possible. While these are two separate objectives, they are closely related through the asking price of the retailer, since the asking price affects the time period required to find a buyer (cp. Yavas and

Yang 1995, p. 347). Despite the importance of asking prices in the used car sector, academic literature does not focus much on the function of asking prices in the automobile market. However, real estate economics face similar issues regarding the role of listing prices, and the relationship between the asking price and duration on the market has been an objective of several academic papers.

In a heterogeneous market such as the used vehicle sector, potential buyers have different preferences for specific features in a vehicle and, therefore, valuate an offered car differently. In such a market, the asking price occupies several roles: first, it provides a signal of the seller's reservation price. Second, it serves as an upper limit for the seller's reservation price since it is perceived as the price the seller is willing to accept. Third, it influences the rate at which offers arrive. Standard search theory theorizes that sellers willing to wait longer increase the probability of encountering a buyer with a high reservation price. However, goods that remain unsold on the market for an overly long time period may become stigmatized, because potential buyers regard the long duration as evidence for some defect. This effect is known as 'negative herding' (see Taylor 1999). And lastly, the asking price can be interpreted as a signal for the quality of the offered product, since in some situations customers tend to prefer higher-priced goods when price is the only information available. For a more in-depth discussion of the price-quality relationship, please refer for example to (Monroe 1979, p. 41), (Simon 2003, p. 604) or (Völckner 2006, p. 473).

Therefore, the choice of the asking price influences the arrival rate of buyers as well as the distribution of bids received from potential buyers. Setting a high price relative to its value reduces the buyer arrival rate and thus extends duration on the market of a durable good (cf. Knight 2002, p. 215). On the other hand, if the asking price is substantially lower than the expected asking price (serving as a proxy for the vehicle's market value), then a prospective buyer is more likely to visit the retailer and to bargain over the vehicle. Hence, a critical element in modeling individual price response functions (by estimating the duration on the market) represents the term of the actual asking price related to the expected market value. For the forthcoming analysis, the ratio of both values is defined as the 'degree-of-overpricing' (DOP), and the hypothesis is supported that vehicles with a smaller DOP will sell faster than vehicles with a higher DOP. To analyze the influence of the degree-of-overpricing on the probability of a sale, the expected market value of a used vehicle has to be determined.

4.4.2 Hedonic Pricing: Estimation of Expected Market Value

In the subsequent section it was hypothesized that the relationship between asking price and market value of a specific car in part explains its survival time on the market. A critical part in the corresponding survival analysis is estimating the market value of used vehicles before they are traded.

In contrast to used vehicles, pricing of new vehicles is relatively straight-forward. Each model vehicle can be associated with a basic model and a corresponding original list price. For each additional piece of equipment and configuration, such as metallic paint, leather or air conditioning, a surcharge is added, determined by a price list. Accumulating all extras, the total price based on the price list is fixed for each individual car, and the only difference between actual sales prices for identical vehicles can occur in the process of negotiation between the car dealer and the customer. Although theoretically countless possibilities of combinations for a car and its configuration exist, finally each vehicle is associated with a total list price. In comparison to new vehicles, determining the market value for used cars is significantly more complex, since there are a number of new dimensions to be added. As a durable good, a used vehicle is characterized by its age, its mileage and its overall condition, just to name a few examples. In addition to the presence of different characteristics, each potential customer might possess individual utility functions and therefore value each of the characteristics of a used vehicle differently. Consequently, a used car with a given set of basal factors (such as mileage, age, and number of previous owners) and an additional set of options (such as air conditioning or hi-fi systems) might be valued differently by different potential customers.

A major stream in research literature suggests valuing a durable good's individual components to determine its total value. Within this market study, hedonic price modeling is applied as one of the most promising approaches [4]. Hedonic pricing hypothesizes that each good can be looked upon as a bundle of attributes and that a function relationship exists between these attributes and the price of a good. The development of hedonic pricing theory is closely connected to the estimation of assets, with Haas (1922) and Wallace (1926) as the first researchers developing hedonic models for estimating the value of farmland. Court (1939) developed the first hedonic price index for automobiles, but in the following decades studies in asset economics influenced the development of hedonic pricing theory with the papers of Lancaster (1966) and Rosen (1974) as the most important ones. However, only little attention was paid to the automobile industry by academics in terms of hedonic pricing theory, with Griliches (1961) and Berry, Levinsohn, and Pakes (1995) as two of the few examples. For a more in-depth discussion of hedonic price modeling, refer to Follain and Jimenez (1985), Sheppard (1999), or Malpezzi (2002).

Generally, the hedonic model can be stated as:

$$\text{Value of good} = f(\text{basal characteristics, external factors}),$$

where the estimated value of a durable good such as a used vehicle is a function of its basal characteristics and external factors such as the state of the economy.

[4] Other pricing models are *repeat sales price indices* and *hybrid indices* (cp. Malpezzi 2002, p. 6).

For estimating the value of a used vehicle, several different functional relationships can be assumed with the linear and the semi-logarithmic form as two basic models, whereas the translog functional form and the Box-Cox form represent the foundation for more complex models.

Within the context of this study, market value for a used vehicle is estimated assuming the semi-logarithmic form

$$\ln V = x'\beta,$$

with $\ln V$ as the natural log of the vehicle's value, β as the vector of coefficients and x as the vector of the vehicle's characteristics.

4.4.3 Estimation of the Expected Survival Functions

The previous subsections established and justified the relationship between the asking price of a used vehicle, its duration on the market and the probability of its sale, where the 'degree-of-overpricing' was identified as a critical element relating these factors. Indispensable for regressing a survival model is the estimate of the expected market value for each subject, carried out by hedonic price modeling introduced in the last subsection.

Based on these elements, a survival model can be adapted to a given dataset by identifying explanatory variables and estimating the corresponding coefficients. Keeping in mind that the objective of this chapter is to determine individual price response functions for used vehicles, this subsection describes the estimation of individual expected survival for a vehicle with a given set of characteristics. Under certain assumptions, a functional form can be derived capturing the dependence between covariates and the expected survival of a vehicle for a specific survival time. In this case, it is possible to establish a functional connection between the asking price and the probability of a sale as a key element for an optimization module of a price-based revenue management system.

Consider the Cox proportional hazards model stated in section 4.3.1. Here, the estimation of the expected survival function is based on Breslow's estimator for the baseline cumulative hazard rate $H_0(t)$. Let $t_1 < t_2 < \cdots < t_D$ denote distinct event times and d_i be the number of events at time t_i. Then, Breslow's estimator is given by

$$\hat{H}_0(t) = \sum_{t_i \leq t} \frac{1}{\sum_{j \in r(t_i)} e^{x_j{}'\beta}}$$

for the case where at most one event at any time occurs. Its extension to the case of tied data with d_i events at a given time can then be stated as

$$\hat{H}_0(t) = \sum_{t_i \le t} \frac{d_i}{\sum_{j \in r(t_i)} e^{x_j{}'\beta}}$$

with $r(t_i)$ as the set of all individuals who are still under study at a time just prior to t_i. Using the estimate of the baseline survival function, $\hat{S}_0(t) = \exp[-\hat{H}_0(t)]$, the estimate of the survival function for an individual with a covariate vector $x = x_0$ can be derived to

$$\hat{S}(t|x = x_0) = \left[\hat{S}_0(t)\right]^{e^{x_0{}'\beta}} . \tag{4.25}$$

Example 4.6 (Estimation of an Individual Price Response Function). Based on the dataset from the market study mentioned in example 4.4, for the purpose of simplicity a Cox proportional hazards model is analyzed. It is assumed that the variable 'degree-of-overpricing' is the only explanatory variable affecting the survival on the market, or in other words, the probability of a vehicle's sale. Recall that the variable 'degree-of-overpricing' represents the quotient of the asking price of a specific used vehicle and its market value estimated by hedonic pricing.

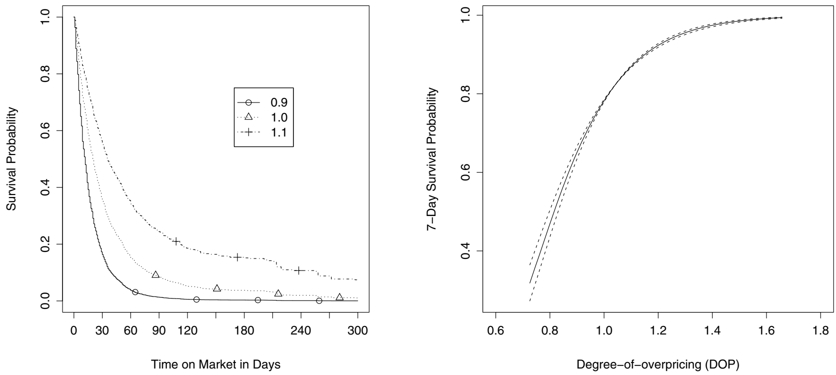
The results of the Cox regression are given in Table 4.5 along with the analysis of variance. Based on the regression results, Breslow's estimator extended for handling ties in the dataset can be utilized to determine the expected survival function.

Variable	Coefficient	Std. Error	Wald χ^2	Degrees of Freedom	p-Value	Relative Risk
DOP	-5.6151	0.207	734.5	1	$< .0001$	0.004

Table 4.5: Analysis of variance table for a simple Cox model fitting the used car example including the variables DOP utilizing the Breslow method for handling ties

Note that for each survival time t, a distinct survival function with a corresponding baseline survival function exists. For example, the estimates for a survival time of $t = 7$ days are plotted in Figure 4.7. The left graph displays the expected survival for different values of 'degree-of-overpricing' over the course of the offer period. Clearly, increasing the offering time of a vehicle results in the increase of the probability of its sale. The right graph analyzes the survival probability for the survival time at $t = 7$, examining the relationship between asking price and the probability of a sale in more detail. Here, the question of how the probability of survival up to 7 days is influenced by the asking price is answered, expressed by the variable 'DOP'. Higher values of 'DOP' result in higher probabilities of survival, which in reverse implies that higher asking prices result in lower probabilities of a vehicle's sale.

In contrast, the estimation of expected survival for accelerated failure time models follows the standard approach of regression by estimating the influence of the



(a) Sensitivity of DOP versus time-on-market (b) Sensitivity versus DOP

Fig. 4.7: Plots of estimated survival probability with regard to different pre-defined values of DOP (left figure) and at time $t = 7$ (right figure) for a Cox proportional hazards model

explanatory variables. For example, when survival time follows a log-logistic distribution, the estimate of the survival function based on maximum likelihood estimates of the covariate vector β can be derived to

$$\hat{S}(t|x) = \left[1 + \exp\left(\frac{\log(t) - x'\beta}{\sigma})\right)\right]^{-1} \tag{4.26}$$

with $\alpha = 1/\sigma$ (compare (4.24)).

4.5 Summary

In this chapter, survival analysis is applied to estimate individual price response functions with regard to the used car sector. After proposing time duration market data as a technique to extract information from price quote histories, the concept of survival analysis is introduced. To identify and model the effects of explanatory variables on the survival of used cars and thereby to model the probability of a sale, two approaches for parametric regression models are discussed, namely the Cox proportional hazards model and the accelerated failure time model. The chapter concludes with the presentation of a procedure to estimate expected survival based on a fitted Cox model and, thereby, estimating the individual price response function since the asking price was included in the Cox model as the only explanatory variable.

Having introduced the concept of survival analysis for estimating individual demand functions, the next chapter analyzes market data from the German used car sector and validates selected survival models as a basis for the application of the complete optimization module in chapter 6.

Chapter 5
Validation of the Survival Analysis Approach: a Case Study within the Used Car Market

There is hardly anything in the world
that someone cannot make a little worse and sell a little cheaper,
and the people who consider price alone
are that person's lawful prey.
JOHN RUSKIN
(1819–1919)

The objective of the present chapter involves applying survival analysis for estimating individual price response functions. Based on an extensive market study concerning the German used car sector, several models introduced in the previous chapter are fitted to the datasets and their applicability on predicting survival is determined, thereby estimating individual sales probabilities of used vehicles.

5.1 Introduction

In the previous chapter, survival analysis was used to estimate individual price response functions for used vehicles. Hypothesizing that the sale of a used vehicle is influenced by internal factors such as the asking price as well as external factors like the market conditions, one way to determine these relationships is to assume a vehicle's time on the market until a sale occurs as the dependent variable. Then, the influence of explanatory variables such as the asking price can be assessed using survival analysis regression. Different models were introduced, including Cox proportional hazards models and accelerated failure time models as two categories of parametric regression modeling.

In this chapter, methods and models stated in the previous chapter are applied on a dataset from a study conducted in the German used car market with the scope of determining explanatory variables and their effect on the failure time of used vehicles. Here, the failure of a used car on the market translates to a sale of the vehicle. Then, based on a fitted model, survival of a specific used vehicle, or the probability of its sale, can be estimated with regard to its characteristics and the given market conditions. Furthermore, the survival regression model can be used to determine optimal pricing strategies of a used car retailer, if the asking price has a significant influence on the survival of a used car.

In section 5.2, the market study regarding the German used car market is described and the process of model building is illustrated. Section 5.3 selects and fits a Cox proportional hazards model on the dataset including an assessment of its adequacy. Afterward, an accelerated failure time model is identified, and the fitted model assessed and validated. In section 5.5, the model that best fits the data is extended by restricted regression splines to incorporate possible non-linearity in the covariates. The chapter is concluded in section 5.6 by a presentation and discussion of the final extended survival model that best fits the data from the market study.

The calculations and plots in this chapter were performed with the R software (R Development Core Team 2006), using the `Design` and `Hmisc` library, written by Frank E. Harrell (Harrell 2005, 2007) as well as the `Survival` library, written by Terry Therneau (Therneau and Lumley 2006).

5.2 Used Car Study

The automobile industry is characterized by sophisticated data and information systems. Although innumerable amounts of data are generated and stored, information about the development of individual asking prices in the used car sector is not available and neither are academic studies regarding the analysis of these quote histories. This study fills this ap and investigates the relationship between asking prices, vehicles characteristics, external market conditions and the probability of a vehicle's sale.

In a market study from December 2005 to September 2006, the German used car market was analyzed with a focus on authorized dealerships of a specific major German car manufacturer. Daily data were collected to obtain information regarding vehicles' individual characteristics and listing prices and to determine whether a car was sold or not. Over the course of the market study, more then 300,000 vehicles were observed, with the majority positioned in the medium and premium sector. In this period, the German economy appeared to be in a significant economic upswing, benefiting from the strong growth of the global economy. Similarly, market conditions in the German used car sector have been affected by increased revenue and a stable number of vehicles selling after a five-year period of negative compound growth rates (cp. Schönleber 2006, p. 51).

In the following section, we concentrate our analysis on a specific volume model in the compact class segment, referred to as model 'J-1', where to date more than one million units have been sold worldwide. The first generation of model 'J-1' was introduced in the late 1990s and a second generation appeared in 2004. The current model is available as a three- or five-door hatchback, but the predecessor was only offered as a five-door hatchback. Within the used car market study, 59,549 vehicles of type 'J-1' are analyzed, generating 94,828 data sets, since each price change is treated as a separate data point. The subsequent analysis consists of a subgroup of the complete data set, reflecting vehicles from the first generation with a gasoline engine of 1600 cubic centimeter capacity.

5.2.1 Dataset and Descriptive Statistics

The dataset of used cars consisting of type 'J-1' includes 4,564 vehicles, of which 2,544 have been sold until the end of the analysis period. The earliest observations are for vehicles offered in December 2005, whereas the market study was finished in September 2006. Vehicles not sold at that time are considered censored (right censoring). Table 5.1 presents a summary of descriptive statistics for the data. The variable 'time-on-market' (TOM) denotes the time from the beginning of the offering period until the asking price is changed or the vehicles are sold or withdrawn from the market. Thus, if the vehicle's asking price is changed, a new data is generated. Recognizing this fact, the average TOM for all observations is 34 days, although the average total time for selling a vehicle is 48 days.

				Quantiles of Variable Distribution						
Variable	*n*	unique	mean	.05	.10	.25	.50	.75	.90	.95
TOM	4564	187	33.84	2.0	3.0	8.0	21.0	41.0	69.0	102.8
DOP	4564	4555	1.091	0.95	0.98	1.03	1.09	1.15	1.21	1.25
Market size	4564	173	24.7	3	4	6	11	23	66	128
Quantile	4564	1110	0.39	0.071	0.100	0.182	0.344	0.571	0.750	0.833
Age (in weeks)	4564	397	209.4	95.0	113.3	131.0	181.0	267.0	361.0	393.0
Original price	4564	1298	20271	18025	18450	19262	20109	21134	22359	23153
Previous owners	4564	6	1.466							

Table 5.1: Descriptive statistics for observed vehicles of type 'J-1'

One of the critical elements for this analysis is the degree-of-overpricing (DOP), measured as the percentage ratio of the actual asking price and the expected market value given the observable characteristics of the vehicle. For a short discussion of determining the expected market value, refer to the subsequent section 4.4.2. In the upcoming analysis hypothesize that vehicles with smaller DOP sell faster than vehicles with higher DOP since prospective buyers are less likely motivated to inspect an offered vehicle featured with a higher relative asking price and bargain over the final selling price.

To caption magnifying effects by niche markets, the variable 'market size' gives the number of vehicles which are similar to the observed one. Here, a similar vehicle is defined as follows: the type and model of the car must be the same, the mileage has to be within plus or minus ten thousand kilometers, and the age has to be plus or minus two months with regard to the considered vehicle.

Since the expected market value determined by the hedonic model often only captures mid- to long-term trends, there is another compound variable to make statements about short-term changes. More precisely, the offered vehicles are ranked within the constructed markets by their relative asking prices. In this context, the relative asking price is defined as the quotient of the asking price with its original listing price, and, thus, provides the ability to compare heterogeneous vehicles by their relative asking prices. After ranking the vehicles, the vehicle-specific quantiles can be calculated (noted as 'quantile'). It is advantageous to apply this variable since it can explain situations in which a vehicle might be overpriced, but when compared to vehicles in the same market segment, it is one of the lowest priced due to the fact that the majority of similar cars are offered at a higher price. This would be characterized by a small quantile value such as 0.1 for the vehicle; consequently, the probability to sell this vehicle should be higher than for vehicles with higher asking prices.

Note that all variables are considered time-invariable, determined at the beginning of the period under observation. A specific used vehicle, its market size, degree-of-overpricing, its ranking within the market and its age as calculated for the first day of its appearance in the dataset are considered. Once the asking price of the vehicle is updated, its corresponding variables are updated as well.

5.2.2 Model Building

5.2.2.1 Model Building and Selection Process

One objective of this chapter is to identify variables and how they effect the probability of a used vehicle's sale. The scope is to be able to predict the distribution of time to a vehicle's sale from a list of explanatory variables. In this context, the question is how to incorporate these variables in a model and compare different modeling approaches with the objective to find the one best fitting the dataset.

The model building and selection process used in this paper is based on an approach suggested by Box, Jenkins, and Reinsel (1994, p. 16). The Box-Jenkins model building strategy is primarily an iterative method of identifying a possible useful model from a general class of models and consists of four main stages: selection of a general class of models, model specification, parameter estimation, and diagnostic checking. The model building and selection approach based on the Box-Jenkins procedure is summarized in Figure 5.1. In the first stage of the process, a

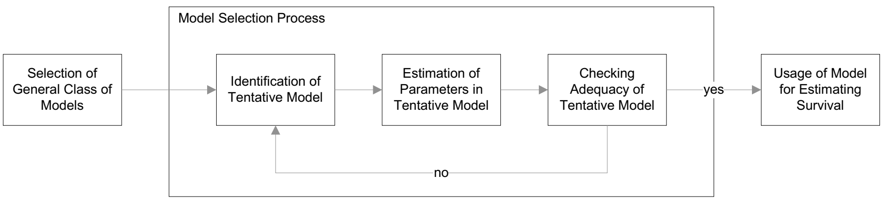


Fig. 5.1: Stages of the model building process adapted from the Box-Jenkins approach

useful class of models is selected for determining explanatory variables in predicting the survival of a used vehicle. As general classes the following ones are considered in this chapter: the Cox proportional hazards models, the accelerated failure time models and the extended models for both classes. In the second stage, subclasses of the general models are identified for fitting the given dataset from the market study. The exponential and log-logistic distribution are examples for different sub-classes in the general class of accelerated failure time models. Criteria for selecting a tentative model are discussed in the following sub-section. In the third stage, the tentative model selected in the previous stage is fitted to the observed data with estimated parameters. Finally, diagnostic checks assess the model's adequacy regarding functional form and predictive ability. The iterative process is repeated until an adequate model is found, which then can be applied on the objective of estimating the survival of used vehicles for a given set of characteristics.

5.2.2.2 Selection Criteria

In published articles, there are two main categories of selection strategies. The first category consists of all-subset strategies with different optimization criteria such as the Akaike information criterion (AIC) or the Bayesian information criterion (BIC). These strategies use the error sum of squares to measure of the goodness of fit and a penalty term for model complexity. The second category is represented by sequential procedures such as forward and stepwise selection or backward elimination. Starting from a null model or from a full model, at each step a new variable is added or deleted until a pre-specified selection level (so called *p-value*) is reached (see for example Sauerbrei 1999). The p-value as the selection level can be derived using different techniques to make an inference about a subset of the coefficients. These techniques are also called local tests, in contrast to global tests, which make an inference about the coefficient vector β in a global sense. There are three main test procedures. The Wald test is based on the estimators standardized by use of the information matrix, the likelihood ratio test uses the maximized likelihood itself, and the score test applies the first derivatives of the log likelihood (for a detailed description, refer to Klein and Moeschberger 2003, pg. 429).

The application of standard significance testing for selecting variables for a model is controversial, as it only indirectly addresses the issue of whether a variable is sufficiently important to be included in the model (see for example Kuha 2004). Therefore, the results determined by the sequential procedures based on hypothesis testing are verified applying all-subset strategies with penalized criteria. Here, the task is to find a model that would best reflect reality given the data recorded and, thus, to minimize the loss of information. In addition, the selection is penalized by the number of parameters included in the model.

We use the AIC as an information-theoretic selection strategy as well as the BIC, also known as Schwarz's information criterion, as a Bayesian model selection strategy. These criteria are defined as

$$AIC = -2\log PL(\hat{\beta}) + 2k, \tag{5.1}$$

$$BIC = -2\log PL(\hat{\beta}) + 2\log n, \tag{5.2}$$

where $PL(\hat{\beta})$ is the maximized partial likelihood function, k the number of free parameters in the model, and n the number of observations. For further discussions on these topics, please refer to Burnham and Anderson (2002, 2004), among others.

5.3 Cox Model

In this section, the Cox proportional hazards model estimates explanatory variables with the coefficients influencing the probability of a vehicle's sale for the 'J-1' used car dataset. Different variable selection strategies are applied with different selection criteria. Also, different computing algorithms determine the partial log likelihood since the used car data set contains tied data. After evaluating assumptions of the proportional hazards model and the functional form of the covariates, the final model is assessed regarding replication stability, selection bias, and an overoptimistic estimate of the predictive value of the model.

5.3.1 Model Identification

In section 4.3.1, the Cox proportional hazards model was introduced by its characteristic survival function

$$S(t|x) = [S_0(t)]^{e^{x'\beta}}, \tag{5.3}$$

with the baseline survivor function $S_0(t) = \exp[-H_0(t)]$ and its corresponding hazard function

$$h(t|x) = h_0(t)e^{x'\beta}. \tag{5.4}$$

The coefficients of explanatory variables can be determined by deriving the maximum estimates for Cox's log partial likelihood:

$$pl(\beta) = \sum_{i=1}^{d} \left\{ x_{(i)}{}'\beta - \ln \sum_{j \in r(t_i)} e^{x_j{}'\beta} \right\}. \tag{5.5}$$

Note that the partial likelihood for the Cox model is developed under the assumption of continuous data, thereby assuming distinct failure times d. However, the actual used car data set contains multiple tied events, given that the time is recorded daily. In subsection 4.3.1.2, three different approaches incorporating tied data within the partial log likelihood function were addressed, namely the Breslow approximation, the Efron approximation and the exact partial log likelihood function. In the following, all three different methods are applied to determine the coefficients of explanatory variables for the regression of the 'J-1' used car dataset. The primary outcome variable that is used to assess prognostic quality is the 'time-on-market' (TOM). The factors considered are 'degree-of-overpricing' (DOP), size of the relative market ('market size'), the position within this market based on the relative asking price ('quantile'), the number of previous owners ('po'), and the age of the vehicle in weeks ('age').

The following Table 5.2 states the comparison of the three different approaches to handle ties using the 'J-1' used car dataset. The estimated results are fairly typical for a dataset with a significant amount of tied data, especially at the left tail of the survival time. The results for the different approaches are fairly similar with regard

	Regression Coefficients				
Algorithm	DOP	Market size	Quantile	po	Age
Breslow	−5.0474	−0.0044	−0.4816	0.0196	−0.0011
Efron	−5.1258	−0.0045	−0.4828	0.0194	−0.0011
Exact	−5.2715	−0.0045	−0.4675	0.0203	−0.0011

Table 5.2: Comparison of tied data methods for vehicles of type 'J-1'

to the coefficients of the variables market size, po and age, but differ significantly for the predictors 'degree-of-overpricing' and quantile. The Efron approximation is chosen for the forthcoming modeling since its estimate is feasible even with large tied datasets and close to the exact log likelihood function. However, the application of the exact partial log-likelihood function should be considered for further modeling approaches.

Based on the Efron approximation of the log likelihood function, the selection process of explanatory variables for the 'J-1' used car dataset continues. Two different selection strategies are carried out to confirm the results, namely sequential procedures with pre-defined selection levels and all-subset strategies with different optimization criteria. First, sequential procedures are applied with a pre-specified selection level for the explanatory variables. For the application of sequential procedures on the used car dataset, a selection level of $p = 0.05$ is defined for the elimination and the re-inclusion of explanatory variables. The calculated results were the same for executing forward and stepwise selection as for backward elimination. Table 5.3 summarizes the parameter estimates including the p-values for the Wald test. The variable 'po' (number of previous owners) was not included at all stages since its p-value of $p = 0.576$ (for the model in step four, similar results in the other steps) provides strong indication that this term can be dropped from the model and that the probability of a sale does not depend on the number of previous owners.

A drawback of standard significance testing for selecting variables for a model is that it only indirectly addresses the issue of whether a variable is sufficiently important to be included in the model. Therefore, we apply another category of selection strategy on the dataset, namely all-subset strategies with different optimization criteria. Table 5.4 summarizes the model selection process using the AIC as the optimization criterion. Confirming the selection strategy based on significance testing, the variable 'po' (number of previous owners) is excluded from the final model.

Step	Variable	Coefficient	Standard Error	Wald χ^2	p-Value	Exp(B)
1	DOP	−5.540	0.207	717.387	0.000	0.004
2	Market size	−0.003	0.001	20.637	0.000	0.997
	DOP	−5.500	0.205	722.638	0.000	0.004
3	Market size	−0.004	0.001	38.274	0.000	0.996
	DOP	−5.784	0.212	740.996	0.000	0.003
	Age	−0.001	0.000	26.498	0.000	0.999
4	Market size	−0.004	0.001	38.160	0.000	0.996
	DOP	−5.059	0.283	318.506	0.000	0.006
	Quantile	−0.482	0.119	16.267	0.000	0.618
	Age	−0.001	0.000	17.308	0.000	0.999

Table 5.3: Forward selection procedure for 'J-1' used car dataset, utilizing likelihood ratio test for significance testing and Efron approximation for handling ties

Note that the model with the lowest AIC is to be chosen as the 'best' among all models specified for the data. Similar results were achieved applying the BIC as

	Model	Excluded	AIC
Start	S ∼ DOP + market size + quantile + po + age		38049
Step 1	S ∼ DOP + market size + quantile + age	po	38047
	S ∼ DOP + market size + po + age	quantile	38063
	S ∼ DOP + market size + quantile + po	age	38064
	S ∼ DOP + quantile + po + age	market size	38090
	S ∼ market size + quantile + po + age	DOP	38351
Step 2	S ∼ DOP + market size + age	quantile	38061
	S ∼ DOP + market size + quantile	age	38063
	S ∼ DOP + quantile + age	market size	38089
	S ∼ market size + quantile + age	DOP	38352
End	S ∼ DOP + market size + quantile + age		38047

Table 5.4: Fitted model based on Akaike information criterion for 'J-1' used car dataset

the selection criterion, supporting the decision to exclude the variable 'number of previous owner' from the model.

Summarizing the model selection process, the final fitted Cox proportional hazards model consists of the explanatory variables degree-of-overpricing, market size, quantile, and age. These results were confirmed applying both all-subset strategies and sequential procedures with a pre-specified selection level. In the next section, the adequacy of assuming a Cox proportional hazards model will be assessed with regard to regression assumptions and proportional hazards assumptions.

5.3.2 Adequacy Checking of the Cox Model

The Cox proportional hazards model makes assumptions on the functional form of the covariates and on the independence of time of the relative hazard. Both categories of assumptions are tested in the forthcoming subsection for the fitted Cox model based on the 'J-1' used car dataset.

5.3.2.1 Evaluating Regression Assumptions

In its basic setup, the Cox model applied to the used car dataset assumes that the relationship between predictors and the log hazard function should be linear, and in absence of interaction terms the predictors should operate additively (cp. Harrell 2001, p. 418). Hence, the adequacy of these assumptions is checked and the correct functional form for the covariates is explored. Therneau and Grambsch (2000, p. 87) suggest a simple approach of plotting the martingale residuals from a null model against each covariate separately and superimposing a smoothed fit of the scatter diagram. In this context, a null model is one with a coefficients vector of $\hat{\beta} = 0$. A martingale residual is a slight modification of the Cox-Snell residual and can be interpreted as the difference over time of the observed number of events minus the expected number of events under the assumed Cox model. Thus, it represents an estimate of the excess number of events seen in the data but not predicted in the model (cf. Klein and Moeschberger 2003, p. 334). A smoothed-fitted plot for each covariate gives an indication of the correct functional form to explain the effect of this covariate on survival. If the plot is linear, the assumption of linearity is confirmed and no transformation of the variable is needed. Figure 5.2 illustrates the application of the martingale residual plot for the 'J-1' used car dataset. Here, a Cox model with a null vector of coefficients is fitted and the martingale residuals are plotted along with a LOWESS smooth[1]. A LOWESS smooth that is approximately linear implies that the regression slope is a good estimate of change throughout the time period used for trend analysis, whereas curves indicate short-term fluctuations within the time period. The smoothed curve for the covariate 'degree-of-overpricing' decreases only slightly up to values of about 0.9, but the decrease is sharper up to values of about 1.3, where the slope changes again. This analysis suggests that customers are

[1] A smooth curve plot where each smoothed value is given by a weighted linear least squares regression and is intended to show the natural trend of the center of mass of the data.

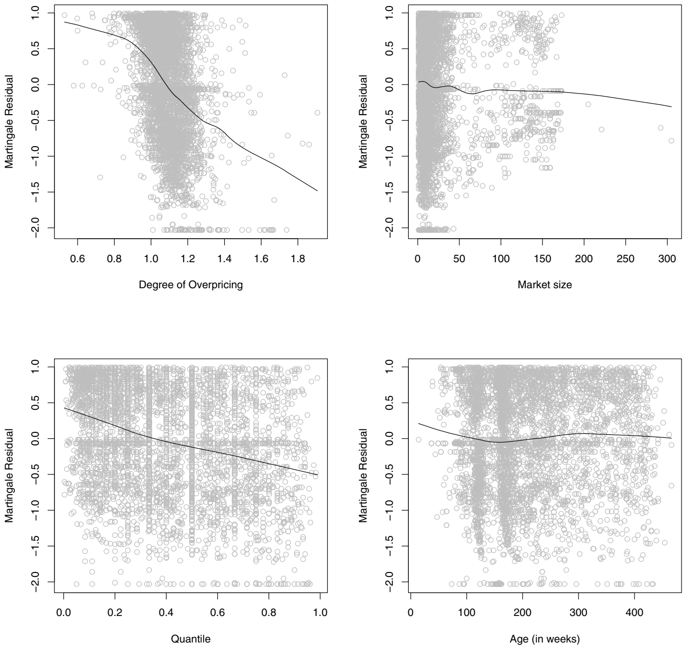


Fig. 5.2: Plot of martingale residuals and LOWESS smooth for degree-of-overpricing, market size, quantile, and age for 'J-1' used car dataset

more price-sensitive when the asking price is roughly similar to the vehicle's market value. Here, a change in asking price has a stronger impact on the chance of selling a vehicle. However, where the asking price differs significantly from the market values, a change in the asking price does not change the probability of a sale with the same attitude as in the region around the market value.

A more advanced approach is stated by Grambsch, Therneau, and Fleming (1995), which extended the basic martingale method to address both linear and nonlinear relationships, applying a Poisson regression approach. Sophisticated modeling tools are already developed for Poisson residuals and can be used to determine appropriate functional forms for the effect of the covariates on survival. Figure 5.3 illustrates the main-effects plots for each covariate, including upper and lower twice-standard-error curves. The martingale residuals are examined by non-parametric smooths fitting the scatterplots of residuals against the covariates of interest, where the smooths are generated using a Poisson regression implementation of the generalized additive model. Table 5.5 summarizes the significance tests of non-linearity of the smoothed curves applying a score test for each of the non-parametric terms. Note

Term	Degrees of Freedom for Linear Component	Degrees of Freedom for Nonparametric Component	χ^2	p-Value
s(DOP)	1	3	33.992	< 0.001
s(market size)	1	3	7.405	0.060
s(quantile)	1	3	12.160	0.007
s(age)	1	3	24.158	< 0.001

Table 5.5: Significance test for non-linearity in the Cox model using the Poisson approach for ‘J-1’ used car dataset

that all covariates show a significant curvature based on the test of non-linearity. The plots support this assessment graphically for all covariates. For instance, the shape of the smoothed curve for ‘quantile’ decreases up to values of approximately 0.4, and then slightly rises again. Furthermore, the smooth curve for the covariate ‘degree-of-overpricing’ confirms the evaluation based on martingale residuals, in which the dependence of the covariate on survival can be separated into three subintervals. Summarizing, further analysis must be carried out regarding all explanatory variables in order to identify parametric families of functions that approximate the curves.

5.3.2.2 Evaluating Proportional Hazards Assumptions

The regression analysis for the used car Cox model relies on the proportionality of hazard rates of individuals with distinct values of a covariate (cf. Klein and Moeschberger 2003, p. 337). More precisely, the relative hazard for any two subjects should be constant and independent of time and should hold individually for each covariate in the model. Therefore, graphical and analytical techniques verify the proportional hazards assumption in this section.

In other academic papers, numerous methods are found for the evaluation. For example, a simple graphical test is given by looking at the survival curves for each level of a variable. The curves should steadily drift apart if the proportional hazard assumption holds (cp. Therneau and Grambsch 2000, p. 127). A comprehensive discussion on techniques to verify the proportional hazards assumption is found in Harrell (2001, p. 483), among others.

In this analysis, we follow an approach proposed by Therneau and Grambsch (2000, p. 130) and extend the Cox model by a time-dependent coefficient. Then, a method is applied for graphing the proportional hazards assumption based on scaled

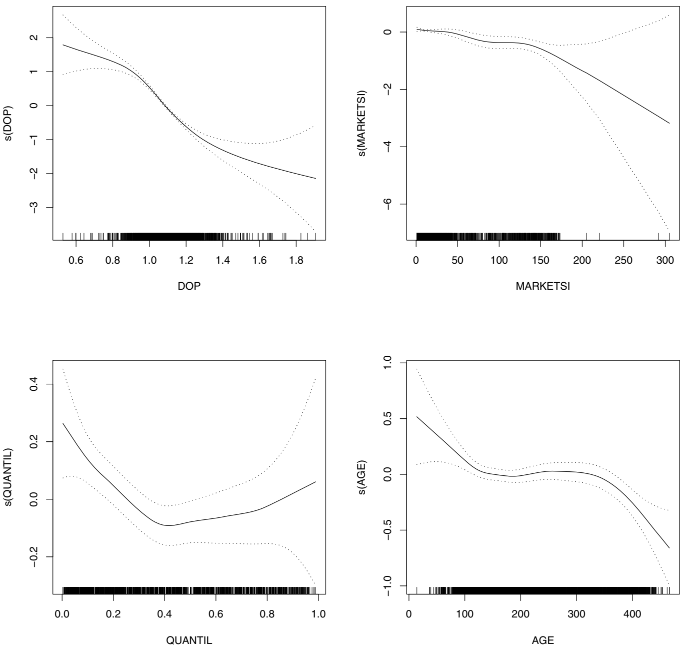


Fig. 5.3: Assessing functional form utilizing Poisson approach for degree-of-overpricing, market size, quantile, and age for 'J-1' used car dataset

Schoenfeld residuals[2] and according to each predictor in the fitted model. Plotting the scaled Schoenfeld residuals over time, a nonzero slope of a fitted line for the plot is evidence against the proportional hazards assumption. Furthermore, a formal test for significance is derived assessing the correlation with time of the time-dependent coefficient. Table 5.6 summarizes the 'correlation with time' test statistics for the 'J-1' used car dataset. Here, column 'ρ' represents the Pearson product-moment correlation between the scaled Schoenfeld residuals and the time scale. The other two columns give the test statistics with the corresponding *p*-value (refer to Therneau and Grambsch (2000, p. 134) for the mathematical equations). We used a logarithmic time scale since the data set is characterized by long-tailed survival distributions. The largest test statistics for non-proportionality are observed for the covariates 'age' and 'degree-of-overpricing'. The corresponding residual plots including

[2] The Schoenfeld residual is defined as the covariate value for the individual that failed minus its expected value. Instead of a single residual for each individual, there is a separate residual for each individual for each covariate (Schoenfeld 1982).

Covariate	ρ	χ^2	p-Value
DOP	0.0546	7.1627	$7.444 \cdot 10^{-3}$
Market size	0.0092	0.2088	$6.477 \cdot 10^{-1}$
Quantile	0.0112	0.3204	$5.714 \cdot 10^{-1}$
Age	0.0819	16.7367	$4.294 \cdot 10^{-5}$
GLOBAL		29.5421	$6.066 \cdot 10^{-6}$

Table 5.6: Significance test regarding proportional hazards assumption for ‘J-1’ used car dataset

a spline smooth and 90 percent confidence intervals (dotted lines) are shown in Figure 5.4. There, the slightly positive correlation with time can be seen graphically for both covariates by examining the smoothed fits of the residual plots. In the remaining part of this section, we ignore possible effects based on non-proportionality and revert to this problem in section 5.5 by considering spline regression for modeling the influence of covariates.

5.3.3 Internal Validation of the Cox Model

Earlier subsections introduced the ‘J-1’ used car dataset and developed a fitted Cox proportional hazards model including the selection of adequate predictor variables and the evaluation of the model assumptions. In the following part, the determined Cox model is validated with regard to predictive accuracy to assess the likely performance of the model on a new series of objects and, thus, testing the applicability with real data.

In general, two major aspects of predictive accuracy are often considered, namely calibration and discrimination (cf. Steyerberg, Harrell, Borsboom, Eijkemans, Vergouwe, and Habbema 2001, p. 775). Discrimination is the model’s ability to separate subjects’ outcomes and, thus, distinguish low-risk subjects from high-risk subjects. Calibration, or reliability, refers to the ability of the model to make unbiased estimates of the outcome by assessing whether predicted probabilities agree with observed probabilities.

Several different types of validation can be identified, namely apparent validation, internal validation and external validation. For testing the predictive accuracy, apparent validation applies to a dataset that was used for model development. It can only provide a first impression of the model’s performance since testing is performed on the same dataset and, thus, apparent validity will be optimistic. More reliable types of validity are methods derived from internal data. Here, the model is assessed by using only one set of subjects. A subset of the original dataset is sam-

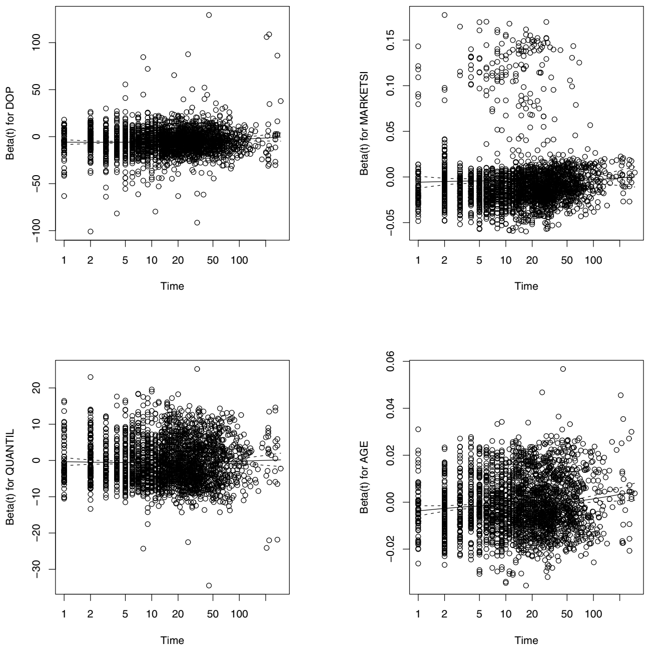


Fig. 5.4: Scaled Schoenfeld residuals testing the Proportional hazards assumption of Cox model for 'J-1' used car dataset, where a nonzero slope of the fitted line is evidence against the proportional hazards assumption

pled for model development and another subset (or the rest of the original dataset, respectively) is used for model testing. Finally, external validation refers to testing a final model developed with one set of subjects on another set of subjects.

In this section, internal validation methods are applied for assessing the predictive accuracy of the 'J-1' used car dataset. Many different validation techniques can be found in other studies, with data-splitting and resampling methods being the most cited. Data-splitting separates the dataset into a training sample for model development and test sample for model validation. Although easy to implement, the method requires larger samples since it reduces the complete sample size and does not validate the final model, but only a model built on the training set (cf. Harrell 2001, p. 92). To overcome these obstacles, resampling methods validate the model in the subsequent sections. Resampling methods refer to model validation by using random subsets of the original data with permutation, cross-validation and bootstrapping as the most noticeable techniques. In this analysis, validation is based on the bootstrap approach, a simulation technique for studying properties of statistics without the need to have the infinite population available. The use of the bootstrap involves taking random samples with replacement from the original dataset, where each random sample has the same number of observations as the original dataset. Here, some of the original subjects may be omitted from the random sample and some may be sampled more than once. Then, the model is repeatedly fitted in the bootstrap sample and the performance of each sub-model is evaluated on the original sample. The estimate of the likely performance of the final model on future data is then estimated by the average of all the indices computed on the original sample (for further discussions on bootstrapping, confer Efron 1981; Efron and Tibshirani 1993).

In the following, the fitted Cox model for the 'J-1' used car dataset will be evaluated regarding model discrimination and model calibration.

5.3.3.1 Validation of Model Discrimination

As stated above, discrimination assesses the model's ability to separate subjects having low responses from subjects having high responses. In terms of survival analysis, it can be quantified by the measure of the index c, the probability of concordance between predicted and observed survival and, thus, can be interpreted as the proportion of all pairs of subjects whose survival time can be ordered such that the subject with the higher predicted survival is the one who in reality survived longer (cp. Harrell 2001, p. 493). A value of $c = 0.5$ stands for random prediction, a value of $c = 1$ for a perfectly discriminating model and a value of $c = 0$ shows that the 'opposite' predictor has perfect discriminatory power. Equivalently, one can apply Somers' D_{xy} rank correlation using the relationship of $D_{xy} = 2(c - 0.5)$ and, thus, rescale the index to 0 for random prediction.

Another index assessing the discriminatory power is provided by Nagelkerke's R_N^2 (Nagelkerke 1991). This index can assess how well a model compares to a 'perfect' model and ranges from 0 to 1. It is useful for quantifying the predictive strength of a model and is defined as

$$R_N^2 = \frac{1 - \exp(-LR/n)}{1 - \exp(-L_0/n)},$$

where L_0 is the $-2log$ likelihood for a model that has no predictive information (that is, the worst model), and LR is the difference between L_0 and the -2 log likelihood of the fitted model. The latter can be interpreted as the global log likelihood ratio statistic for testing the importance of all predictors in the model. Attention should be paid to the fact that often maximum values of 1 cannot be achieved. Generally, values of 0.2 to 0.4 indicate very strong discrimination power of the model.

Before assessing the 'J-1' used car dataset for discriminatory power, the bootstrap technique is used to penalize the Cox model for possible overfitting. The bootstrap estimates for 200 repetitions are stated in Table 5.7. The column 'Original Sample' describes the estimate for the original dataset, the 'Training Sample' column refers to the average estimate from the training samples of the bootstrap method and the 'Test Sample' column describes the estimate when applying the bootstrap model to the original sample. For the fitted Cox model, the optimisms added to the estimates are relatively small, keeping the biased-corrected estimates almost unchanged compared to the biased estimates. Thus, only a moderate amount of overfitting can be detected from the analysis. The apparent Somer's rank correlation index D_{xy} is -0.3353 and the bias-corrected index is $D_{xy} = -0.334$, suggesting that the model has some predicting power in comparison with the theoretical value of $D_{xy} = 0$, with which a model would make only random predictions. Assessing the Nagelkerke's R_N^2, the value of $R_N^2 = 0.1570$ is comparable to values of other survival analysis studies with similar setups and suggests that some predictive ability is provided by the fitted Cox model. Note that the last index, the calibration slope, was added to the table due to convenience; however, the index is discussed in the forthcoming paragraph of model calibration.

Index	Original Sample	Training Sample	Test Sample	Optimism	Corrected Index
D_{xy}	-0.3353	-0.3359	-0.3346	-0.0013	-0.3340
R_N^2	0.1590	0.1603	0.1583	0.0020	0.1570
Slope	1.0000	1.0000	0.9880	0.0120	0.9880

Table 5.7: Bootstrap estimates of discrimination accuracy of Cox model utilizing $B = 200$ repetitions for the 'J-1' used car dataset

5.3.3.2 Validation of Model Calibration

In the following section, model calibration for the 'J-1' used car dataset is validated. As stated before, calibration refers to the reliability of predicted values; that is, the question of to what extent predicted values agree with observed values. The basis for this assessment are bias-corrected estimates derived from bootstrap sampling, as explained in the previous section. Verweij and Van Houwelingen (1993) proposed the estimate and application of a global shrinkage factor, which reduces the bias of parameter estimates caused by model building through shrinking each regression coefficient. Similarly, Miller, Langefeld, Tierney, Hui, and McDonald (1993) motivated the usage of a 'calibration slope', i.e. the slope of the linear predictor, which is identical to the shrinkage factor as noted by Steyerberg, Harrell, Borsboom, Eijkemans, Vergouwe, and Habbema (2001, p. 775). The calibration slope can be interpreted as the only regression coefficient in a logistic model with the linear predictor as the only covariate. Thus, values close to 1 may indicate that hardly any overfitting occurred. Applied on the 'J-1' used car dataset, the bias-corrected calibration slope has a value of 0.988, indicating that almost no overfitting can be observed in the dataset.

Next, the fitted and bias-corrected model is validated for calibration accuracy by analyzing the difference between Cox predicted survival probability and observed values represented by Kaplan-Meier survival estimates (cf. Harrell 2001, p. 493). This analysis can only be carried out at fixed survival times t and is done for 7-day, 14-day, 30-day, and 60-day survival. To construct these calibration curves, the original sample is stratified on n intervals of predicted survival at time t. Then, for each interval the average survival probability is estimated based on the fitted Cox model, and subsequently compared with the corresponding Kaplan-Meier estimate at time t, stratified by intervals of estimated survival probability. For the 'J-1' used car dataset, bootstrapping with $B = 200$ repetitions is used to obtain bias-corrected estimates. Furthermore, the original sample is stratified with subsets of approximately 250 subjects per interval of predicted survival. Thus, 18 subsets are created. The estimated calibration curves for 7-day, 14-day, 30-day, and 60-day survival, including their 95 percent confidence intervals for the Kaplan-Meier estimates, are shown in Figure 5.5. There, dots represent apparent calibrations for an interval and crosses are the bias-corrected calibrations. Note that the difference between both is negligible for almost all estimates. The calibration plots have to be read as follows. Ideally, if predicted values and observed values agree over the whole range of probabilities, the plot would follow the dotted line. Only bootstrap calibration for 30-day survival shows some evidence of calibration power, where the estimated plot follows roughly the predicted dotted line. Bootstrap calibrations for earlier survival times (here, 7-day and 14-day survival) identify differences between predicted and observed values represented by Kaplan-Meier estimates. Here, survival is significantly better than predicted, whereas bootstrap calibration for 60-day suggests that observed values are worse than predicted survival.

In summary, the Cox proportional hazards model fitted on the 'J-1' used car dataset reveals some serious drawbacks with regard to its calibration power as well as to its ability to predict the survival of used vehicles. Reasons for these issues might be the non-linearity of the explanatory variables and the presence of non-proportionality of the hazard rates, both suggesting that the adequacy of the assumed Cox model cannot be verified and should be further analyzed to improve the model estimating survival of used vehicles.

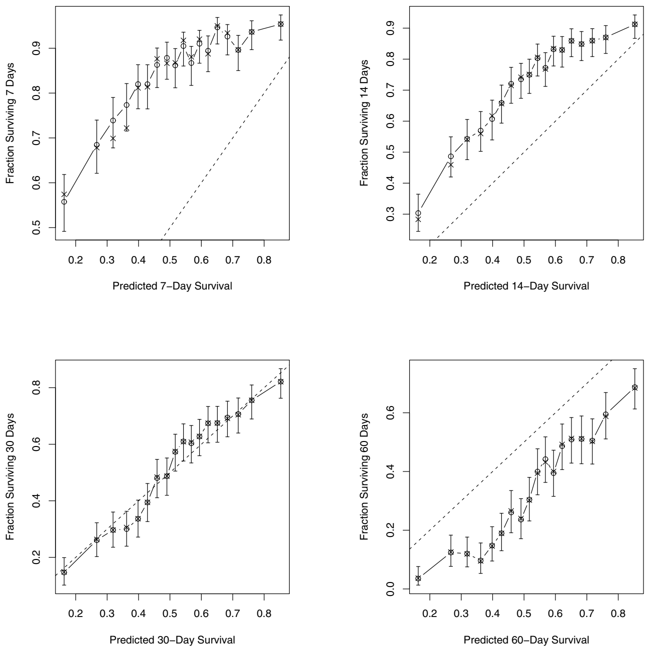


Fig. 5.5: Bootstrap validation of calibration curves for fitted Cox model based on the 'J-1' used car dataset

5.4 Accelerated Failure Time Model

In the previous section, we derived results from the multivariate analysis of the Cox proportional hazards model in order to capture the relationship between the distribution of survival time of a vehicle on the market and the values of explanatory variables for the 'J-1' used car dataset. Evaluating the proportional hazards assumption identified two covariates, namely 'DOP' and 'Age', where the impact of the variable changed significantly over time. An alternative to the Cox proportional hazards model was presented in section 4.3.2 involving parametric modeling of the hazard function in form of the accelerated failure time models. Therefore, in this section parametric accelerated failure time models are applied on the 'J-1' used car dataset.

5.4.1 Model Identification

Recall that the survival function of the accelerated failure time model was defined generally as

$$S(t|x) = S_0[te^{x'\beta}], \tag{5.6}$$

assuming that the baseline survivor function S_0 adequately describes the form of the distribution of survival time, and that the factor $e^{x'\beta}$ models the variation in the scale of the distribution. Furthermore, the natural logarithm of survival time $Y = \ln(T)$ is modeled linearly for Y

$$Y = -x'\beta + \sigma W, \quad W \overset{iid}{\sim} S_0(\cdot). \tag{5.7}$$

In the following section, a number of different theoretical distributions for the baseline survival function S_0 are applied, namely the exponential, Weibull, normal, log-normal, and log-logistic distribution. Table 5.8 summarizes the statistical distributions including their probability density function $f(t)$, their survivor function $S(t)$, and the corresponding hazard function $h(t)$.

Distribution	Density Function $f(t)$	Survivor Function $S(t)$	Hazard Function $h(t)$
Exponential	$\lambda \exp(-\lambda t)$	$\exp(-\lambda t)$	λ
Weibull	$\alpha\lambda t^{\alpha-1}\exp(-\lambda t^{\alpha})$	$\exp(-\lambda t^{\alpha})$	$\alpha\lambda t^{\alpha-1}$
Normal	$\dfrac{\exp\left[-\frac{1}{2}\left(\frac{t-\mu}{\sigma}\right)^2\right]}{(2\pi)^{0.5}\sigma}$	$1-\Phi\left[\dfrac{t-\mu}{\sigma}\right]$	$\dfrac{f(t)}{S(t)}$
Log normal	$\dfrac{\exp\left[-\frac{1}{2}\left(\frac{\ln t-\mu}{\sigma}\right)^2\right]}{t\,(2\pi)^{0.5}\sigma}$	$1-\Phi\left[\dfrac{\ln t-\mu}{\sigma}\right]$	$\dfrac{f(t)}{S(t)}$
Log-logistic	$\alpha\lambda t^{\alpha-1}[1+\lambda t^{\alpha}]^{-2}$	$[1+\lambda t^{\alpha}]^{-1}$	$\alpha\lambda t^{\alpha-1}[1+\lambda t^{\alpha}]^{-1}$

Table 5.8: Parametric distributions and their corresponding characteristics employed in fitting the 'J-1' used car dataset

Corresponding to these underlying distributions, the accelerated failure time (AFT) models determine important prognostic factors associated with the survival of a used vehicle on the market, thereby estimating the probability of a sale. Identical to the Cox model, possible factors to consider for model selection are degree-of-overpricing, market size, quantile, number of previous owners, and age in weeks. Furthermore, both categories of selection strategies are applied on the AFT models, namely sequential procedures based on a pre-specified selection level and all-subset strategies with penalized criteria. The results of the selection process are stated in

Table 5.9. Both strategies lead to similar results and include ‘DOP’, ‘market size’, ‘quantile’, and ‘age’ as explanatory variables for all AFT models except for the normal model. In addition to these four variables, the normal-distributed model include ‘po’ as a fifth covariate.

Model	Log Likelihood	AIC
Exponential	-12413	24836
Weibull	-12405	24821
Normal	-14801	29616
Log normal	-12318	24647
Log-logistic	-12309	24630

Table 5.9: Comparison of parametric AFT models derived on the basis of the ‘J-1’ used car dataset

To select an appropriate parametric accelerated failure time model for further analysis, the decision is based on the Akaike information criterion, where the model with the smallest AIC is selected. Table 5.9 suggests that the log-logistic distribution provides the best fit to the ‘J-1’ used car dataset, although the AIC of the log normal model is almost as small as the log-logistic AIC. Based on these results, in the subsequent paragraphs the log-logistic AFT model is considered and further analyzed with regard to its appropriateness and its discrimination and calibration power to compare it with the Cox proportional model developed in the previous section.

5.4.2 Assessment of Model Fit for the Log-Logistic Distribution

In the last subsection, the log-logistic AFT model was selected as the distribution providing the best fit to the ‘J-1’ used car dataset. In this subsection, the appropriateness of this model is verified, including general adequacy of the log-logistic distribution and the functional form of the covariates.

5.4.2.1 Adequacy Checking of Log-Logistic AFT Model

The considered AFT model assumes that the baseline survival function S_0 follows a log-logistic distribution. Graphical checks are applied rather than formal statistical tests since these tests tend always to reject a given model for large samples. These graphical checks do not prove that the parametric model is correct, but rather serve as a means for rejecting clearly inappropriate models. Harrell (2001, p. 434) proposes a graphical verification of the distributional assumption by plotting the estimated survival function of standardized residuals. To obtain more stringent assessments, residuals can be stratified by important variables as well as quantiles of these variables along with the theoretical log-logistic distribution. The assumption regarding the underlying distribution should be rejected when the plots do not follow the theoretical distribution plot. In Figure 5.6 the thick plot represents the theoretical log-logistic distribution and the residuals are stratified by 'DOP', 'market size', 'quantile', and 'age' as well as by two quantiles of the continuous variables. The

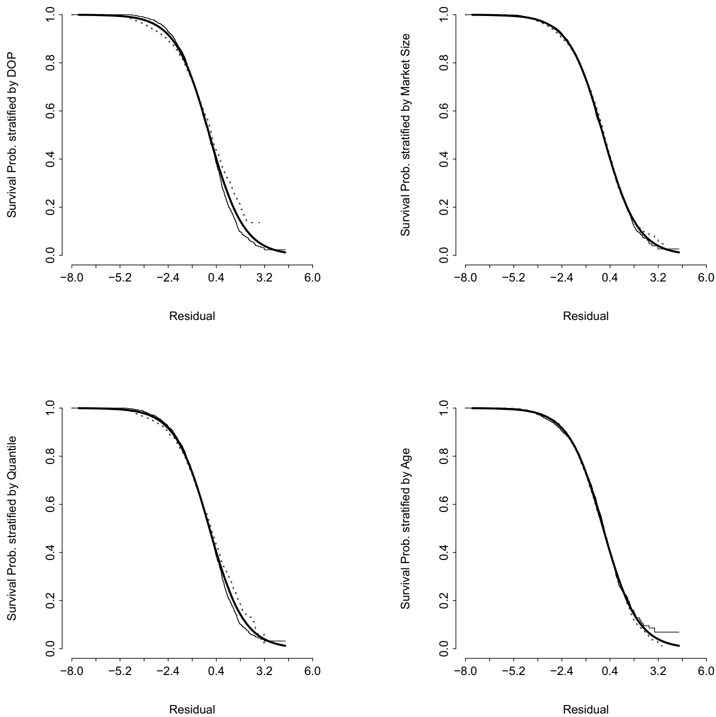


Fig. 5.6: Kaplan-Meier estimates of distributions of normalized, right-censored residuals from fitted log-logistic AFT model for 'J-1' used car dataset in comparison with the theoretical log-logistic distribution

fit for all covariates follows roughly the theoretical distribution plot, suggesting that the log-logistic distribution is indeed appropriate and should not be rejected.

5.4.2.2 Analyzing the Functional Form of Incorporated Covariates

The current formulation of the fitted log-logistic AFT model assumes the linearity and additivity of incorporated predictors. The adequacy of these assumptions is checked and the correct function form of predictors' influence is assessed in this subsection. Comparable to the assessment of the Cox model, a plot of the residuals from a null model against each covariate is applied and a smooth fit of the scatter diagram is included for displaying the correct functional form of the covariate. In contrast to the Cox model, where martingale residuals were used for the plot, here deviance residuals are plotted to obtain the correct functional form of the covariates. Martingale residuals[3] give a measure of the excess number of events seen in the data, but not predicted by the model. The usage of deviance residuals applied in the following plot is an attempt to make the martingale residuals more symmetric about 0. Figure 5.7 shows the plot of the deviance residuals from a fitted log-logistic null model against each covariate. The smoothed fits of the scatter plots indicate that all covariates act non-linearly on the survival time of a used vehicle. Note that similar results were obtained assessing the Cox model, although the slopes of the smoothed fits of the Cox model had reverse effects due to the functional form of the Cox proportional hazards model. The analysis of the 'degree-of-overpricing' shows the non-linearity of a covariate's effect. Here, the smoothed fit demonstrates linearity for values smaller than approximately 0.8, greater than 1.15 and in between this interval. Further analysis regarding the non-linearity of covariates and incorporating regression splines are in section 5.5.

5.4.3 Internal Validation of the Log-Logistic AFT Model

After fitting the log-logistic accelerated failure time model to the 'J-1' used car dataset and checking the appropriateness of the underlying distribution, the log-logistic model is validated with regard to its predictive accuracy.

5.4.3.1 Validating Discrimination Accuracy of Log-Logistic AFT Model

Before assessing the discrimination accuracy of the fitted log-logistic model, the bootstrap resampling approach is applied with $B = 200$ repetitions to penalize the model for possible overfitting. Judging from the results contained in Table 5.10, only a small amount of overfitting can be detected (identified by the 'Optimism'

[3] Klein and Moeschberger (2003, p. 397) point out that the derivation of the martingale residual does not hold for parametric models, but the name carries through since they are similar in form to those of the Cox model.

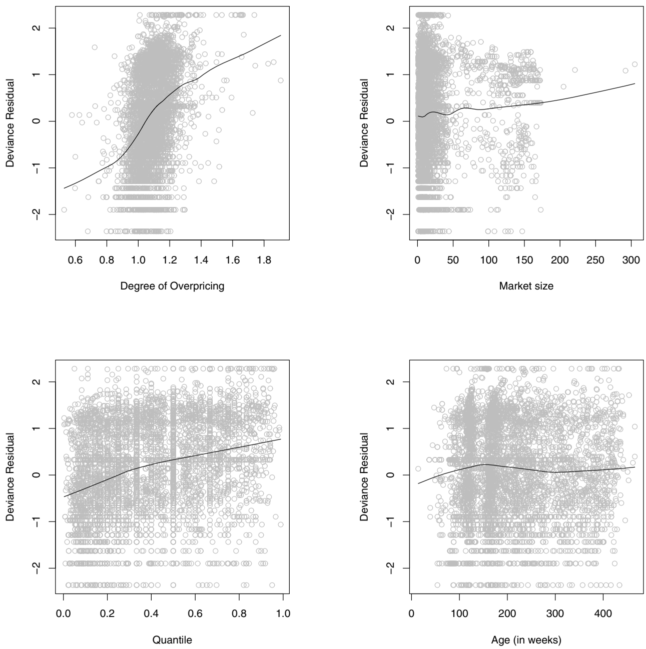


Fig. 5.7: Plot of deviance residuals and LOWESS smooth against each covariate of log-logistic AFT model for 'J-1' used car dataset

column). An estimate of future predictive discrimination on similar used vehicles is

Index	Original Sample	Training Sample	Test Sample	Optimism	Correction
D_{xy}	0.3359	0.3370	0.3352	0.0018	0.3341
R^2_N	0.1641	0.1648	0.1634	0.0014	0.1627
Intercept	0.0000	0.0000	0.0074	−0.0074	−0.0074
Slope	1.0000	1.0000	0.9981	0.0019	0.9981

Table 5.10: Bootstrap estimates of discrimination accuracy utilizing $B = 200$ repetitions for fitted log-logistic AFT model based on 'J-1' used car dataset

given by Somer's D_{xy} rank correlation. The apparent correlation index D_{xy} is 0.3359, only slightly better than the bias-corrected value of $D_{xy} = 0.3341$. Both reveal the discrimination power of the model. Assessing the further discrimination indices of the log-logistic AFT model, Nagelkerke's bias-corrected index $R^2_N = 0.1627$ suggests that the model possesses some predictive ability, similar to the Cox model and comparable to index values of other survival analysis studies with the same amount of censoring and tied data.

5.4.3.2 Validating Calibration Accuracy of Log-Logistic AFT Model

The calibration accuracy of the bias-corrected log-logistic model is validated by analyzing the difference between predicted survival probability by the AFT model and the observed values estimated by Kaplan-Meier survival estimates. The calibration slope in Table 5.10 with a value of 0.991 indicates that almost no overfitting can be observed after estimating the log-logistic AFT model. To get bias-corrected calibration curves, the bootstrap resampling method is applied with $B = 200$ repetitions, where the subjects of the dataset are stratified by their survival estimates into 18 groups of approximately 250 subjects per interval. This analysis can only be carried out at fixed survival times t and is done for 7-day, 14-day, 30-day, and 60-day survival. The estimated calibration curves are presented in Figure 5.8, where dots represent apparent calibrations for an interval and crosses display bias-corrected calibrations. Furthermore, 95 percent confidence intervals for Kaplan-Meier estimates are added. The bootstrap calibration appears to validate well for all four predicted survival times except for the smallest interval groups for predicted 7-day and 14-day survival as well as several intervals at predicted 60-day survival. These plots suggest that although the calibration validates well for most intervals, the analysis of non-linear covariates might better capture their effects on survival. This is further discussed in the next section.

Recapitulating, parametric accelerated failure time models provide a good fit for estimating the survival of the 'J-1' used car dataset, with the log-logistic distribution considered as the best fitted model. Internal validation suggested the applicability of the log-logistic model with regard to predicting survival of used vehicles on the market. However, analysis of the adequacy of the log-logistic model revealed some non-linearity in the explanatory variables. Therefore, the next section introduces more complex functional forms to overcome these obstacles.

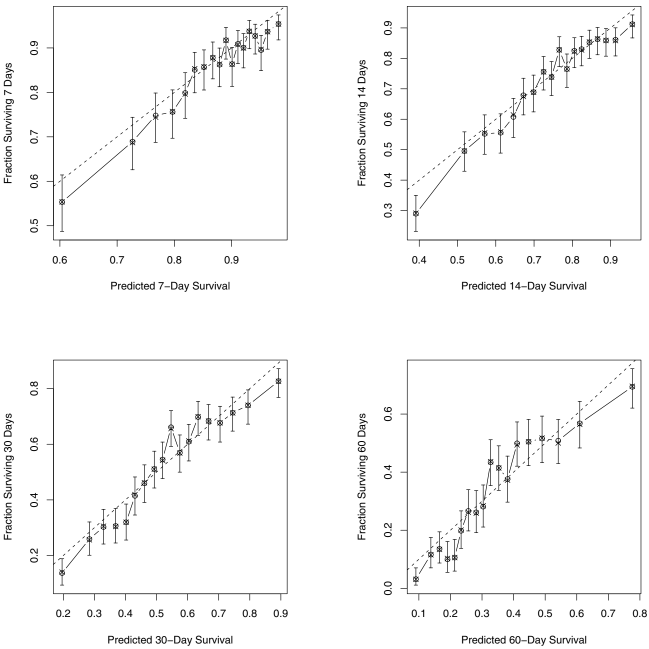


Fig. 5.8: Bootstrap validation of calibration curves for fitted log-logistic AFT model based on the 'J-1' used car dataset

5.5 Spline Regression Extended Model

In the previous sections, two different survival time models were developed based on the 'J-1' used car dataset. First, the Cox proportional hazards model and its explanatory variables were determined, followed by the log-logistic model as the accelerated failure time model that best fitted the data. The functional form analysis of the covariates showed significant non-linearity relationships for both approaches. In this section, the relationship between predictors and survival is estimated without assuming linearity by applying regression splines. After choosing between the Cox and the log-logistic model as the best suitable for estimating survival of the 'J-1' used car dataset, the selected model is extended by regression splines as a flexible fitting function for exploring non-linear relationships. Finally, the extended model is validated regarding discrimination and calibration accuracy.

5.5.1 Model Selection for Spline Regression Extension

Fitting the used car dataset, in the previous sections two different approaches were proposed with the derivation of a Cox proportional hazards model and a log-logistic accelerated failure time model. For estimating the best survival model of used vehicles on the market, in this section results from both models are first compared and then, the appropriate model is selected on the basis of minimum Akaike information criterion.

Model	Log Likelihood	AIC	Bias-Corrected R_N^2
Cox PH	−19019	38047	0.1570
Log-logistic	−12309	24630	0.1627

Table 5.11: Comparison of Cox PH model and log-logistic AFT model derived from 'J-1' used car dataset

The results summarized in Table 5.11 suggest that it might be more appropriate to use the log-logistic accelerated failure time model. The Akaike information criterion for the log-logistic AFT model is smaller than for the Cox model, confirmed by the log likelihood of both models. In addition, Nagelkerke's R_N^2 provides further evidence for choosing the AFT model. But even more important are the validations of the models' calibration accuracy. The plots were shown in the corresponding subsections and demonstrate a significant difference at 7-day, 14-day, and 60-day survival in favor of the log-logistic model. Therefore, the log-logistic AFT model is chosen and extended by spline regression in the subsequent section capturing non-linear effects of explanatory variables.

5.5.2 Regression Splines

The concept of splines originated from drafting techniques of using a thin, flexible strip called a spline to draw smooth curves through a set of points. Splines are piece-wise polynomial functions that join points called knots and allow the regression line to change direction abruptly. They are fitted essentially by adding restricted dummy variables to the regression equation. The simplest setup is a linear spline function, where the x-axis is divided into different intervals, for example with endpoints a, b, and c. Then, the piecewise linear spline function is given by

$$f(x) = \beta_0 + \beta_1 x + \beta_2 (x-a)_+ + \beta_3 (x-b)_+ + \beta_4 (x-c)_+, \tag{5.8}$$

where

$$(u)_+ = \begin{cases} u, & u > 0, \\ 0, & u \leq 1, \end{cases} \tag{5.9}$$

and a, b, and c are referred to as knots (cp. Harrell 2001, p. 18). In most settings, linear spline functions do not approximate relationships very well, and piecewise polynomials of higher order are necessary. Common polynomials used in practice are cubic splines, which are made to be smooth at the knots by forcing the first and second derivatives of the function to agree there. A cubic spline function with k knots is given by

$$f(x) = \sum_{i=0}^{3} \beta_i x^i + \sum_{j=1}^{k} \beta_{(i+j)} (x - t_j)_+^3, \tag{5.10}$$

where the k knots are given as t_j, $j = 1, \ldots, k$. Both presented forms of spline fit are called *regression splines*. Often, a disadvantage of these spline functions is the behavior before the first and after the last knot. Therefore, *restricted cubic splines*, also called *natural splines*, are applied, which obey the constraint of linearity beyond the range of the knots.

Usually, the knots are placed at the quantiles of a predictor's marginal distribution, although the placement of the knots can be part of the estimate. However, in restricted cubic spline regression the goodness of fit does not depend on the location of the control points, but much more on the choice of the number of knots. Therefore, in the subsequent development of the log-logistic AFT model, restricted cubic splines with fixed knots are applied to model the functional form of explanatory variables for the 'J-1' used car dataset.

5.5.3 Model Development

Estimates of the basic log-logistic accelerated failure time model in section 5.4 suggest non-linearity in several explanatory variables. To overcome these issues, restricted cubic splines are applied for fitting the model to the 'J-1' used car dataset.

The fitting of restricted cubic splines in covariates to the survival data is illustrated exemplarily by the 'DOP' variable. Here, restricted cubic splines with $k = 3$, 4, 5 and 6 equally spaced knots in terms of quantiles between 0.05 and 0.95 are compared to the linear model. Based on the Akaike information criterion (AIC), the results stated in Table 5.12 suggest that a model with $k = 4$ knots represents the best fit to the used car dataset. The other continuous explanatory variables are expanded similarly by restricted cubic splines to describe non-linear relationships. Table 5.13 summarizes different combinations of spline expanded covariates with their corresponding log-likelihood and AIC indices. Note that each spline regression model stated in the table is fitted by using the optimal number of knots based on the AIC

Knots k	Log Likelihood	AIC
0	−12309	24630
3	−12305	24624
4	−12299	24613
5	−12298	24615
6	−12298	24617

Table 5.12: Spline regression extended log-logistic AFT model with different number of knots for 'DOP'. The model with 4 knots represents the best fit among the proposed ones.

index for each model (similar to the procedure described for variable 'DOP'). An analysis of log likelihood and AIC reveals that the expansion with restricted cubic splines results in a better fit, with the exception of the variable 'market size'. Thus, the best log-logistic accelerated failure time model incorporates the covariates 'DOP', 'quantile' and 'age' with restricted cubic splines, whereas the 'market size' covariate is added linearly to the regression.

Model	Log Likelihood	AIC
DOP, market size, quantile, age (all linear)	−12309	24630
market size, quantile, age, spline in DOP	−12299	24613
market size, age, spline in (DOP and quantile)	−12293	24605
age, spline in (DOP, market size and quantile)	−12293	24607
market size, spline in (DOP, quantile and age)	−12238	24511

Table 5.13: Analysis of different spline regression extended log-logistic AFT models. The model with restricted cubic splines in 'DOP', 'quantile' and 'age' represents the best fit.

5.5.4 Validation of the Extended Log-Logistic Model

After expanding the explanatory variables with restricted cubic splines, the fitted log-logistic AFT model has to be validated regarding to discrimination and calibration accuracy. In addition, comparisons are made between the linear log-logistic and the extended model in terms of improvements in important indices.

5.5.4.1 Discrimination Accuracy of Extended Log-Logistic AFT Model

Consistent with the validation of former models, the bootstrap resampling approach is conducted with $B = 200$ repetitions to penalize for possible overfitting. The optimism indices displayed in Table 5.14 suggest that only a small amount of overfitting can be detected, although the values increased compared to the results of the linear log-logistic AFT model in Table 5.10. For instance, the optimism subtracted from Somer's D_{xy} rank correlation index increased from 0.0018 to 0.0055 and similar results can be assessed for the other indices.

Somer's D_{xy} rank correlation index provides an estimate for future predictive discrimination on similar used vehicles. A bias-corrected value of $D_{xy} = 0.3542$ gives evidence for discrimination power of the model and is greater than the linear log-logistic model's index with $D_{xy} = 0.3341$. This conclusion is confirmed by Nagelkerke's R_N^2 indices for the different models. Here, the Cox model has a value of $R_N^2 = 0.1570$ and the linear log-logistic AFT model a value of $R_N^2 = 0.1627$, whereas the regression spline extended log-logistic AFT model provides a Nagelkerke index of $R_N^2 = 0.1844$.

Index	Original Sample	Training Sample	Test Sample	Optimism	Correction
D_{xy}	0.3597	0.3630	0.3575	0.0055	0.3542
R_N^2	0.1899	0.1926	0.1871	0.0055	0.1844
Intercept	0.0000	0.0000	0.0469	−0.0469	0.0469
Slope	1.0000	1.0000	0.9866	0.0134	0.9866

Table 5.14: Bootstrap estimates of discrimination accuracy using $B = 200$ repetitions for regression splines extended log-logistic AFT model based on the 'J-1' used car dataset

It can therefore be ascertained that although overfitting increased compared to the fitted Cox model and to the linear log-logistic AFT model, discrimination accuracy increased also significantly.

5.5.4.2 Calibration Accuracy of Extended Log-Logistic AFT Model

The calibration curves for the linear log-logistic AFT model suggest potential for enhancing model building due to some misfits at the tails of the calibration curves (compare Figure 5.8 for log-logistic AFT model). Even more potential for improvement was revealed by the calibration validation of the Cox model, where only the predicted 30-day survival plot suggested good calibration accuracy.

In conformation with the former analysis, validation of calibration accuracy for the regression spline expanded log-logistic AFT model is based on bias-corrected calibration curves by applying bootstrapping with $B = 200$ repetitions and stratification of the dataset's subjects by their survival estimates into 18 groups of approximately 250 subjects per interval. Conducted at the fixed 7-day, 14-day, 30-day, and 60-day survival, the analysis compares the difference between survival probabilities predicted by the extended log-logistic AFT model and the observed values estimated by Kaplan-Meier survival values for the 'J-1' used car dataset. The calibration curves for the extended log-logistic AFT model in Figure 5.9 clearly expose its improved calibration accuracy compared to the Cox model and the linear log-logistic AFT model.

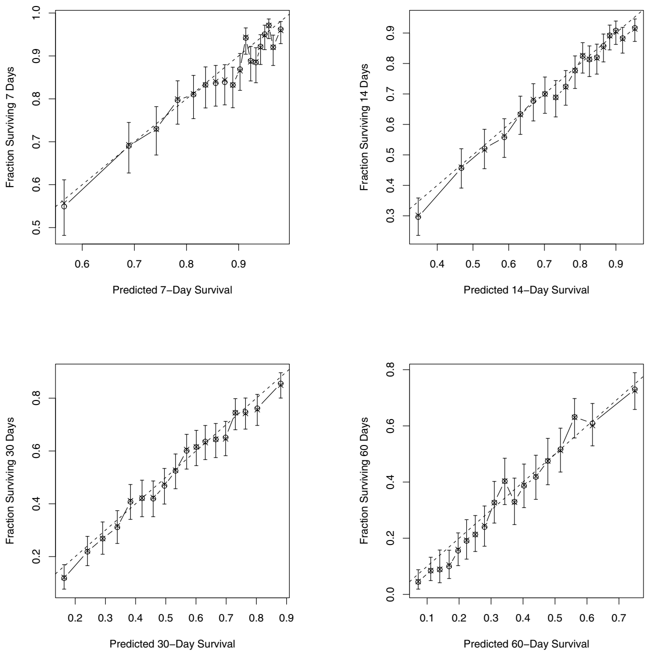


Fig. 5.9: Bootstrap validation of calibration curves for regression splines extended log-logistic AFT model based on 'J-1' used car dataset

The previous sections applied the Cox proportional hazards model and the log-logistic model as the accelerated failure time model best fitting the 'J-1' used car dataset. In this section, both approaches were compared based on Akaike information criteria and the bias-corrected Nagelkerke's index R_N^2, whereas the log-logistic model was chosen to be extended by spline regression. Here, the final fitted model incorporated the explanatory variables ‘DOP’, ‘quantile’, and ‘age’ with the help of restricted cubic splines and the covariate ‘market size’ added linearly to the model. In comparison with the previous models, overall discrimination and calibration accuracy increased significantly. Finally, the predictive ability of estimating survival improved by comparing predicted survival with observed survival at different times.

5.6 Presentation of the Extended Log-Logistic Model

In this section, the final restricted cubic spline extended log-logistic accelerated failure time model for the 'J-1' used car dataset is presented and interpreted by analyzing the effect of each covariate. The final model consists of four predictors, namely the degree-of-overpricing (DOP), the size of the market for similar vehicles (market size), the ranking of the subject within this market (quantile), and the age of the vehicle measured in weeks, whereas all variables are transformed by restricted spline regression with the exception of 'market size'.

First, the effect of each predictor is analyzed and interpreted, plotting a graph of the relationship between each predictor and the log survival time. In Figure 5.10, the y-axis states the log survival time for each predictor when the other predictors are set to reference values as stated below each single graph. Furthermore, 95 percent confidence intervals are included (dotted lines) and rug plots are added representing the data density of each covariate. Note that predicted values have been centered so that predictions at predictor reference values are zero.

Comparing the effects of all four predictors, clearly the 'degree-of-overpricing' exerts the highest influence on the log survival time due to the steepest slope. Previous analysis of the functional form already revealed the non-linearity of the variable's effects on survival, which is confirmed by the plot in Figure 5.10. Generally, higher values of 'DOP' result in higher log survival times, whereas a relatively long survival time of a used vehicle can be interpreted as a low probability of its sale. Consequently, higher probabilities of a vehicle's sale are demonstrated by lower survival times and vice versa. Recall that the 'degree-of-overpricing' represents a quotient of the vehicle's asking price and its estimated market value. Since the latter is retained unchanged, 'DOP' is varied only by modifying the asking price and, thus, 'DOP' serves as a proxy for the relative asking price of the used vehicle. Coming back to the effect of 'DOP' on log survival time, the plot can be roughly separated into three subintervals characterized by different slopes. Up to values for 'DOP' of around 1.05, the model extended by restricted cubic spline regression suggests a linear relation between the variable and the log survival time. In the region between 1.05 and 1.2, the slope steepens, whereas for values greater than 1.2, the slope is similar to the first subinterval. The analysis of this curve progression suggests that customers are more price sensitive in price regions where the asking price is slightly higher than the vehicle's market value. Here, a change in the asking price has a stronger impact on the probability of selling the vehicle than in the other two subintervals. However, at the tails where the asking price differs significantly from the market values, a change in the asking price does not change the probability of a sale with the same attitude as in the region around the market value.

The plot of the 'market size' as a predictor on log survival time reveals that a greater number of similar vehicles offered at the same time results in higher log survival times, although the graph's slope is relatively flat indicating that the influence

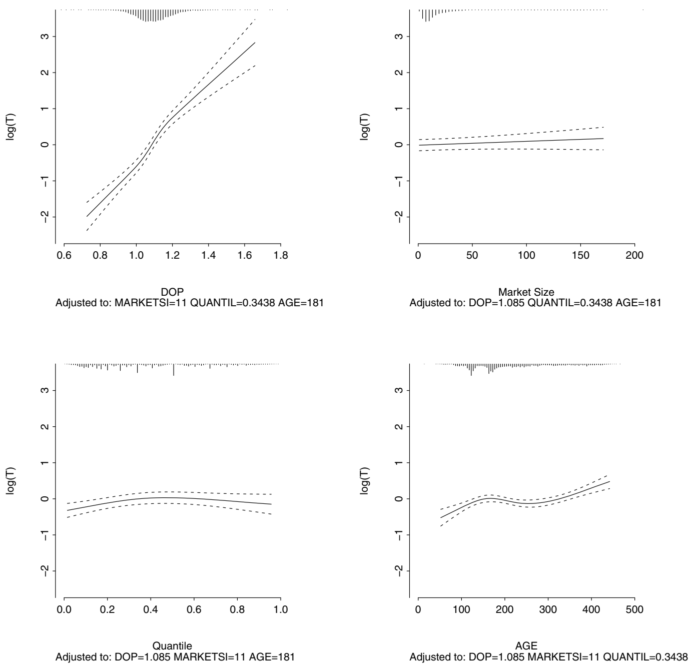


Fig. 5.10: Effect of each predictor on log survival time for 'J-1' used car dataset

of the predictor is low in comparison to the remaining covariates. Next, analyzing the effect of the variable 'age' on log survival uncovers a general positive interrelation between both parameters. However, non-linearity of the variable's effect on log survival can be observed from the plot with an increase up to approximately 150 weeks, an slight decrease up to values of around 250 weeks and again an increase from there on. One explanation for this behavior might be the fact that the variable 'age' has a significant impact on the transaction price when the vehicle is still perceived as good as new, but for vehicles with an age between three and five years, it plays only a secondary role in the prediction of its time on market.

Another interesting curve progression is given by the effects plot for the 'quantile' variable. Here, the position of the vehicle's relative asking price compared to similar vehicles is ranked within this set and expressed by its quantile. Observing the shape of the smoothed curve for 'quantile', the plot shows an increase up to values of approximately 0.4, and then a slight decrease. Intuitively, the progression suggests that a vehicle with an asking price relatively low in comparison to similar vehicles offered at the market should experience a higher probability of a sale.

However, this is only true for the bottom region of this market. In the medium range of offered vehicles, the rank plays no significant role. Here, the probability of a sale does not depend on the vehicle's pricing position within the according market. Contrarily, an asking price positioned at the upper end of the ranking results in smaller log survival times and, consequently, in higher sales probabilities. This behavior might be explained by the fact that customers associate a higher asking price with higher overall quality and better general condition of the vehicle. Thus, the asking price serves as a proxy for the unobserved characteristics of the vehicle and justifies the relatively high asking price compared to similar vehicles on the market.

5.7 Summary

In this chapter, a market study concerning the German used car sector was presented and analyzed. Based on a selected dataset, two general classes of parametric survival models, namely Cox proportional hazards and accelerated failure time model, were fitted and analyzed with regard to its applicability on predicting survival functions and thereby estimating individual sales probability. The log-logistic failure time model was selected to be extended by restricted cubic spline regressions since the functional form analysis of the covariates showed significant non-linearity relationships. Validating the discrimination and calibration power of the extended survival model, the following analysis exposed the variable ‘degree-of-overpricing’ as the most influential predictor on the survival probability of a used vehicle. Since the ‘DOP’ acts as a proxy for the asking price based on the definition of ‘DOP’ as the quotient of the asking price of a used vehicle and its market value, a functional relationship between the probability of a vehicle’s sale and its asking price is developed, considering other significant factors such as the market size or the position of the asking price within the market.

Due to the estimate of individual demand functions by applying survival analysis, the most important input parameter for the price-based revenue management program is extracted. Thus, the next chapter demonstrates the complete idea of the revenue management approach for a profit-maximizing used car retailer and provides a ‘proof-of-concept’ including the identification of profit potential through the revenue management program.

Chapter 6
Computational Analysis: Proof of Concept

If it is a miracle, any sort of evidence will answer,
but if it is a fact, proof is necessary.

MARK TWAIN
(1835–1910)

The objective of the present chapter is to verify the concept of the price-based revenue management program with regard to profit-maximizing used car retailers. After describing the complete revenue management module in section 6.1, the framework is applied to a chosen used vehicle in section 6.2. In section 6.3, the general potential for profit enhancement is estimated by analyzing an extensive independent dataset and comparing the discounted profit of observed sales with the corresponding expected profits determined by the revenue management program.

6.1 General Description of the Revenue Management Program

In the previous chapters, the application of a price-based revenue management program was developed with a focus on the used car sector. Before the revenue management module is applied on the dataset of a used vehicle, the revenue management program should be outlined. Figure 6.1 illustrates the general setup of the price-based revenue management program. Whereas the revenue management module is the core element of the program and consists of two major components, namely the optimization component and demand forecasting component, the program also includes corresponding input and output modules. Since both core components, optimization and demand forecasting, rely on current and historical data, a crucial part represents data analysis including the collection, cleaning, processing and analysis of information, which is carried out in the input module. For example, to model individual survival functions as a basis for predicting individual demand, historical

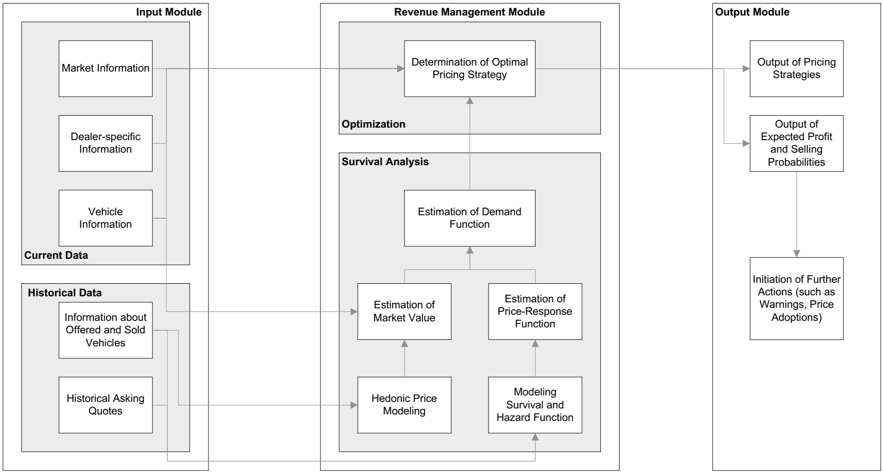


Fig. 6.1: Model of a price-based revenue management program for a used car retailer

data of used vehicles offered in the past is needed with regard to the vehicles' characteristics and their asking price paths. Additionally, the estimate of optimal pricing strategies for a current used vehicle depends on the vehicle's characteristics, internal market information regarding for example the retailer and external market data such as the current market conditions. Consequently, the design and implementation of a professional data analysis structure is crucial for the success of the price-based revenue management program.

The output module should be adapted to the needs of its user, the used car retailer. Based on the determination of optimal pricing strategies, the program should calculate these price paths and carry out price adjustments after confirmation by the retailer. Furthermore, information about expected revenue for different scenarios can be used to confirm a chosen pricing strategy, supported by additional sensitivity analysis. In a more passive way, the program might be applied to support the decision process of a used car retailer by alerting him at the point when the current pricing strategy is no longer optimal and an adjustment should be considered. In summary, both the input and output module are important for the success of the price-based revenue management program, but the core component is represented by the revenue management module.

6.1.1 The Optimization Module

In chapter 3, dynamic pricing models were developed to determine optimal pricing strategies for a profit-maximizing used car retailer. Several approaches were followed, distinguishing between deterministic and stochastic treatment, between continuous-time and discrete-time, and between finite and continuous price region. Analyzing the requirements of car retailers in practice, a stochastic discrete-time model with a finite price set might be the best model to select. Consequently, the example conducted in the following section will use this model for determining an optimal pricing strategy.

6.1.2 The Demand Forecasting Module

In practice, the estimating and forecasting of demand represents the most important though also the most critical and complex component of any revenue management method. In the presented revenue management module, survival analysis was proposed as a way to estimate the probability of a vehicle's sale and to analyze its predictors. Confirming the hypothesis that the quotient of asking price and market value of a specific vehicle (denoted as 'degree-of-overpricing') plays a significant role in explaining the probability of a sale, an interrelationship was established between the asking price of a used vehicle and the chance to sell it on the market, given the internal and external factors affecting the sale. Concurring with the vehicle's market value by the application of hedonic price modeling, individual demand functions can be derived as a prerequisite for the optimization module.

6.2 Case Study for a Selected Used Vehicle

In this section, the application of the revenue management module is demonstrated on a specifically selected used vehicle. First, the individual price response function is estimated by survival analysis and hedonic price modeling. Second, an optimal pricing strategy for a profit-maximizing used car retailer is determined on the basis of the estimated demand function. Then, this is compared with the observed outcome of the vehicle. Furthermore, this demonstration serves as a case for the procedure applied in the subsequent section, where a comprehensive dataset from a market study is analyzed according to its profit enhancement by applying dynamic pricing through the revenue management module. Note that the estimation of the individual price response functions is based on a database generated by the market study presented in section 5.2 with a total of 94,828 datasets from 59,549 vehicles covering a time period from December 2005 to September 2006. In the subsequent section however, an independent test is conducted by applying an external dataset of 1,816 vehicles from a time interval of October 2006 to December 2006. In the optimization module, a stochastic discrete-time model with a finite price set is applied, where optimal prices should be updated on a weekly basis. Consequently, expected survival functions are estimated for a survival time of $t = 7$ days.

6.2.1 Description of a Selected Example

To demonstrate the revenue management approach, an specific used vehicle is selected from the external dataset gathered by the mentioned market study in chapter 5. The object belongs to model 'J-1' and features the following basal characteristics: at the time of its first offering, the used car was 275 weeks old with a reading of 52,000 kilometers. The specific market for similar vehicles consisted of a total of 14 vehicles. These used cars were the same model with an age in the range of plus or minus two months, with an odometer reading of plus or minus 10,000 kilometers. Furthermore, the original price of the vehicle was 19,209.60 euros.

6.2.2 Estimation of the Individual Price Response Function

The first step in the revenue management process is determining the individual demand function, serving as a prerequisite for the optimization module. As already discussed in great detail previously, survival analysis is applied to reveal the relationship between the asking price of the used vehicle and its sale probability. In the present case, a log-logistic accelerated failure time model extended by regression splines is considered for the survival analysis. This is consistent with the findings in chapter 5, where the spline regression log-logistic AFT time model was revealed to be the best fit for the given dataset. Furthermore, the variable 'degree-of-overpricing' is selected as the only explanatory variable, focusing the analysis exclusively on the relationship between the asking price and the probability of a sale. Besides this, the decision is supported by the comparison of the variables' contribution in predicting survival time for the final spline regression extended log-logistic AFT model of chapter 5. Although the final model consisted of the four predictors 'DOP', 'market size', 'quantile' and 'age', Figure 6.2 reveals that 'DOP' is by far the strongest contributor in predicting survival time utilizing the Wald χ^2 statistics. Consequently, a spline regression extended log-logistic AFT model with 'degree-of-overpricing' as the only covariate is considered in the subsequent analysis.

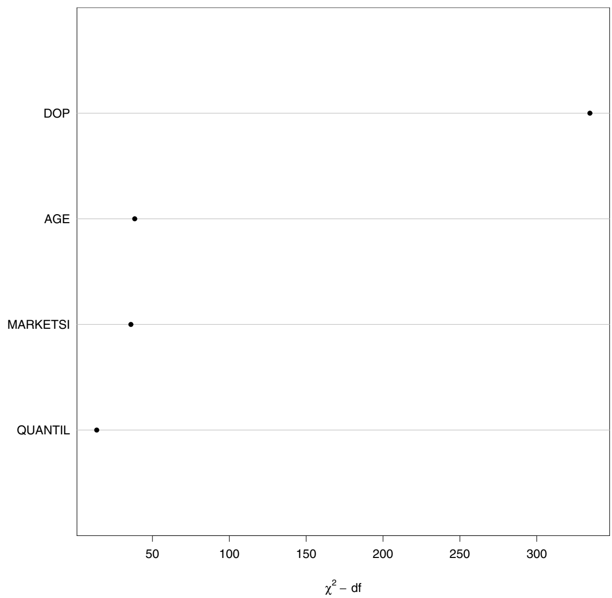


Fig. 6.2: Contribution of variables in predicting survival time of the final spline regression extended log-logistic AFT model from chapter 5 utilizing the Wald χ^2 statistics

To estimate the individual demand function for the selected used vehicle, the log-logistic AFT model is fitted to the corresponding subset of the market study. This dataset contains 4,564 records of the same model as the selected example vehicle. As mentioned before, the predictor 'DOP' for the log-logistic AFT model is extended by regression splines. Thus, the optimal number of knots must be determined. Based on the Akaike information criterion, the results in table 6.1 suggest that a model extended by restricted cubic splines with $k = 4$ knots for the covariate 'DOP' represents the best fit to the given dataset.

Knots k	Log Likelihood	AIC
0	−12309	24630
3	−12305	24624
4	−12299	24613
5	−12298	24615
6	−12298	24617

Table 6.1: Spline regression extended log-logistic AFT model for assessing the optimal number of knots for 'DOP'

Next, the log-logistic AFT model is estimated and the corresponding results of the regression are given by table 6.2 along with an 'analysis of variance' table describing the estimated standard error, the relative risk of the effects and the univariate Wald test with its corresponding p-value. As expected, the analysis suggests that the covariate 'degree-of-overpricing' is statistically significant in explaining the survival of a used vehicle. Note that for each knot a coefficient is estimated. However,

Variable	Coefficient	Standard Error	Z Statistics	p-Value
(Intercept)	−2.456	0.6390	−3.84	< .0001
rcs(DOP, 4)DOP	5.312	0.6467	8.21	< .0001
rcs(DOP, 4)DOP'	8.075	2.3247	3.47	< .0001
rcs(DOP, 4)DOP"	−33.656	8.2708	−4.07	< .0001
Log(scale)	−0.268	0.0163	−16.39	< .0001

Table 6.2: Analysis of the variance table for the regression of the log-logistic AFT model fitting the used car dataset ('J-1' using restricted cubic splines with 4 knots on the variable 'DOP')

the fitted coefficients are not particularly meaningful, since they do not correspond to the y-coordinates of the control points. Therefore, Figure 6.3 visualizes the effect of the predictor 'DOP' on the log survival time. Supporting the conclusions in section 5.6, the plot suggests that higher values of 'DOP' result in higher log survival

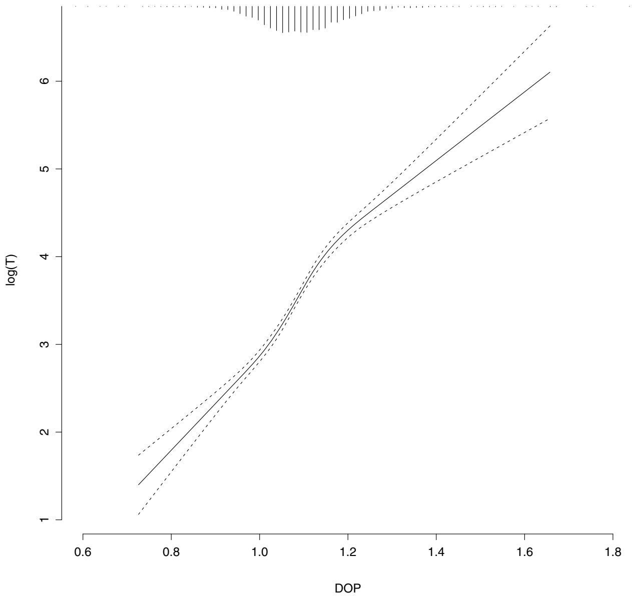


Fig. 6.3: Effect of the predictor 'DOP' on log survival time for the fitted log-logistic AFT model

times and, thus, in higher probabilities of survival on the market. Consequently, higher 'DOP' values lower the chance of selling a particular vehicle. In addition, the plot can be roughly separated into three subintervals characterized by different slopes. This indicates that customers are more price sensitive where the asking price is slightly higher than the vehicle's market value.

The fitted log-logistic AFT model for estimating the individual expected survival function is given by

$$\begin{aligned} S(t|x) &= P(T > t|x) \\ &= \left[1 + \exp\left(\frac{\log(t) - x'\beta}{0.765})\right)\right]^{-1} \end{aligned}$$

where

$$\begin{aligned} x'\beta \; = \; & -2.456 + 5.312\text{DOP} + 85.338(\text{DOP} - 0.946)^3_+ \\ & -355.675(\text{DOP} - 1.051)^3_+ + 344.565(\text{DOP} - 1.120)^3_+ \\ & -74.228(\text{DOP} - 1.254)^3_+ \end{aligned}$$

and $(a)_+ = a$ if $a > 0,\ 0$ otherwise. In the subsequent analysis, expected survival is estimated for the 7-day survival, since the optimization module assumes that prices are only updated on a weekly basis. Bearing in mind that the predictor ‘degree-of-overpricing’ represented a quotient of the vehicle’s asking price and its estimated market value, ‘DOP’ is varied only by modifying the asking price. The estimated market value remains unchanged. Thus, the expected market value has to be estimated for the fitted individual price response function within the optimization module.

6.2.3 *Estimation of Market Value Applying Hedonic Price Modeling*

In section 4.4.2, hedonic price modeling was presented to estimate the market value of a used vehicle based on the idea that its value can be determined by looking upon it as a bundle of attributes and valuing each attribute separately. Regarding the underlying functional form, a semi-logarithmic model of the form $\ln V = x'\beta$ is applied to estimate the market value of the used vehicle. For identifying the hedonic pricing model, sequential selection procedures with pre-specified selection levels are chosen. For elimination and re-inclusion of an explanatory variable, a selection level of $p = 0.05$ is defined. Table 6.3 summarizes the final semi-logarithmic hedonic pricing model, which shows similar results for forward and stepwise selection as well as for backward elimination. Applying the fitted semi-logarithmic hedonic pricing

	Non-Standardized		Standardized		
Variable	Coefficient b	Standard Error	Coefficient B	T	p-Value
Constant	8.878	0.014		641.332	¡0.001
Age (in days)	−0.001	0.000	−0.477	−58.188	< 0.001
List price	0.001	0.000	0.371	59.125	< 0.001
Mileage	−0.001	0.000	−0.318	−50.652	< 0.001
Date of sale	0.001	0.000	0.070	11.200	< 0.001

Table 6.3: Estimated coefficients for determining a vehicle’s market values using semi-logarithmic hedonic price modeling

model on the given used car example, a market value of V =8,506 euros is predicted for the given set of characteristics.

Based on these results, the individual demand function can be determined on the basis of both the fitted regression spline extended log-logistic AFT model and the hedonic pricing model for the market value estimation. A visualization of the corresponding demand function in dependence of the asking price is shown in Figure 6.4. Here, the plotted function can be divided into subintervals with different character-

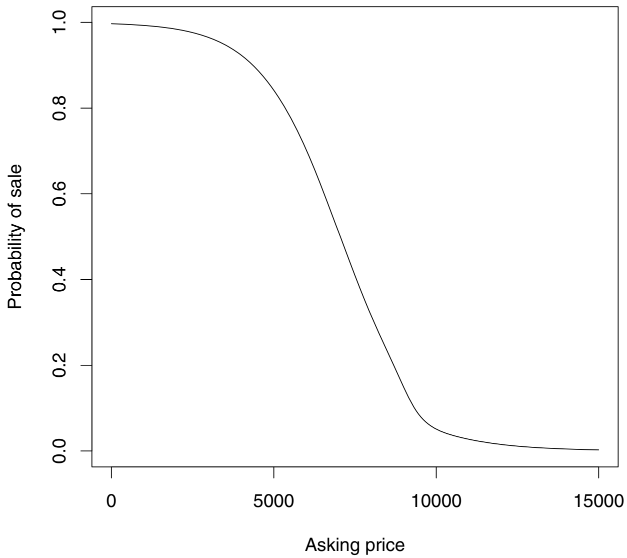


Fig. 6.4: Plots of selected vehicle's estimated demand function at time $t = 7$ for the regression spline extended log-logistic AFT model

istics. Up to an asking price of about 5,000 euros, the demand is constantly marginal below one. Then, the sale's probability decreases with the asking price, following a linear relationship up to values of about 8,000 euros, where any further progression results in a reduction of the probability. For asking prices higher than 15,000 euros, demand approaches zero.

6.2.4 Determination of the Optimal Pricing Strategy

After determining the price response function for the used car example, the optimal pricing strategy is derived for a profit-maximizing used car retailer. In previous sections, it was concluded that a stochastic discrete-time model with a finite price set would best address the requirements of a retailer. Thus, the present problem is a discrete-time stochastic dynamic program under consideration of a finite price set.

Assume that two different finite price sets contain multiples of 250 euros and 100 euros respectively, thereby defining the possible price region to $\Omega_p^{250}(t) = \{0, 250, 500, 750, 1000, \ldots\}$ and $\Omega_p^{100}(t) = \{0, 100, 200, 300, \ldots\}$. Furthermore, consider a time scale observed in units of weeks, weekly costs of 10 euros per existing vehicle and a rate of interest of $r = 0.2$ percent per week, corresponding to a annual rate of 10.95 percent. The selling horizon is defined as 20 weeks, based on assessments from used car retailers and professionals within the automobile sector. The salvage value associated with the terminal state is set to zero, thereby choosing a conservative approach for the optimization module. Given the set of characteristics described in subsection 6.2.1, the predicted market value of the used vehicle is 8,506 euros.

The objective of the dynamic pricing algorithm is to maximize the expected profit by determining optimal prices in each period, thereby constructing an optimal price path. Applying the stochastic discrete-time model to the first finite price set with multiples of 250, the optimal price path is shown in Figure 6.5. Here, the retailer should start offering the selected car for 10,250 euross. If the vehicle is not sold during the first two weeks, the asking price should be reduced to 9,750 euros, remaining there for another week in case of no sale. Over the course of the offering period, the vehicle's price decreases successively, thereby increasing the probability of a sale. If the vehicle is not disposed at the end of the offer phase, then listed at 5,500 euros, it is assumed to gain no salvage value, for instance by selling it to secondary markets overseas. Therefore, it supports an optimal pricing strategy that will sell the used vehicle with a high probability within the selling horizon.

Additionally, a second pricing path is included in the figure for comparison with the second finite price set containing only multiples of 100 euros. Here, the second finite-based strategy follows roughly the first strategy, although prices are adjusted in shorter time intervals. Even more interesting is a comparison of expected profit of the associated price paths. Remarkably, the first proposed strategy with multiples of 250 euros generates a negligible 0.03 percent less profit than the second price strategy with price set $\Omega_p^{100}(t)$, with an expected profit of 8,768 euros for the first, and 8,771 euros for the latter mentioned. Considering the outcome of the two approaches, we conclude that the optimal price within a single time period is not important, but the identification of an appropriate pricing strategy over the course of the complete selling horizon.

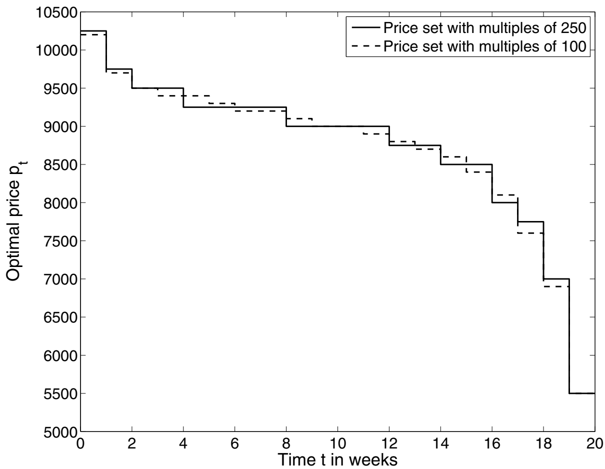


Fig. 6.5: Optimal pricing strategy for a finite price set (solid line) in comparison to the continuous-price strategy (dashed line) for the considered used vehicle

6.2.5 Comparison of Expected versus Observed Revenue

Completing the demonstration of the price-based revenue management approach applied on the selected example, the results of the optimization are compared with the actual sale. In reality, the retailer started at a price of 10,449 euros for the selected vehicle. During the next 27 weeks the car was not sold, but the asking price was kept at 10,449 euros. In week 28, the list price was reduced to 7,950 euros, and one week later, a customer bought the vehicle. The actual price path is shown in Figure 6.6. To compare the actual profit obtained by the retailer with the expected profit determined by dynamic pricing, all cash flows have to be discounted back to its present value at the beginning of the offer period. Assuming the same rate of interest (0.2 percent per week) and weekly cost of existing inventory (10 euros per week) as in the previous section, the net present value of the actual sale equals 7,221 euros.

However, the optimal price strategy determined by the stochastic discrete-time model with a finite price set of multiples of 100 euros generated an expected profit of 8,771 euros, thereby increasing profit by 21.5 percent. Clearly, this enhancement is based on the fact that the actual pricing strategy overestimated the vehicle's market value and that the retailer refused to adjust the asking price accordingly after observing the prevailing market circumstances.

For a more reliable assessment of the potential in profit enhancement in association with the price-based revenue management program, the next section analyzes

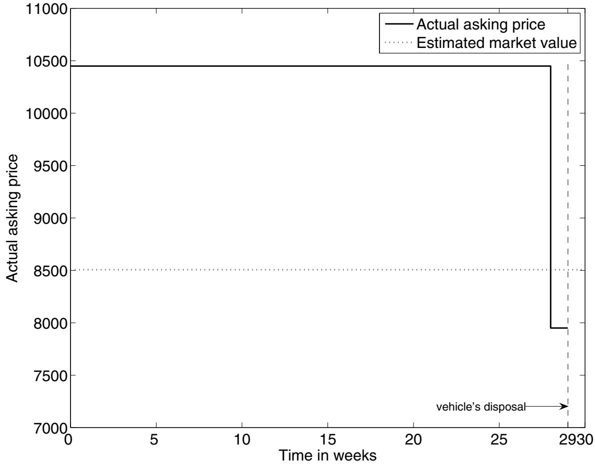


Fig. 6.6: Actual price path chosen by the retailer for the considered used vehicle. The dotted line represents the market value estimated by hedonic price modeling

an extensive additional sample of the dataset generated by the German used car market study.

6.3 Assessment of Potential for Profit Enhancement

In chapter 2, the used car sector was identified as the sector with the most potential for profit improvement within the German automobile industry, as scarcely any manufacturer gains a positive contribution to the operating result from the used car sector. Besides other approaches to the problem, dynamic pricing offers the greatest opportunity for retrieving potential profit. The objective of this section is to quantify the potential in profit enhancement by applying the price-based revenue management program developed in this thesis to an extensive dataset of used vehicles. First, the setup of the experiment is explained. Then, similar to the example in section 6.2.5, the net present values of observed sales are calculated and the corresponding expected profit under the application of the revenue management program is determined. Finally, both discounted profits and expected profits are compared for the complete dataset and conclusions about the potential are established.

6.3.1 Calculation of Discounted Profit for Observed Sales

A market study regarding the German used car sector was described in chapter 5 and data from December 2005 to September 2006 were used to estimate individual demand functions. In this section an external set of data is gathered from the study with used vehicles offered by used car retailers between October and December 2006. All selected vehicles belong to the model 'J-1' and experienced a sale to a customer up to the end of the observation in July 2007. The complete dataset consists of 1,816 entries and serves as the foundation for the forthcoming analysis.

Similar to the calculation for the example vehicle in section 6.2, the concept of net present value is applied to compare cash flows occurring at different time periods with each other. Concluding on the basis of several experts interviews, the weekly cost of a retailer is assumed to amount to 10 euros per existing vehicle and the rate of interest is supposed to be $r = 0.2$ percent per week, corresponding to a annual rate of 10.95 percent. Note that since the sample contains only objects that have been sold, the respective analysis is somewhat distorted. More precisely, objects that have not been sold are not included in the assessment of profit enhancement. In most cases, an unsold can be examined due to an inappropriate pricing strategy by the retailer, where the asking price does not correspond to its market value. There is however potential for increased profit, given by the fact that these vehicles were not sold at current market conditions and therefore, are not part of this analysis. Consequently, the achieved results serve only as a conservative estimate and actually leave room for even higher profit margins in practice.

Table 6.4 summarizes the analysis of the discounted cash flows associated with actual sales of vehicles from the considered dataset. The average net present value amounts to 14,641.45 euros, with an estimated standard deviation of 3,623.88 euros.

	N	Min.	Max.	Mean	Mean Std. Error	Std. Dev.	Var.
NPV	1,816	4,072.22	28,260.90	14,641.45	85.039	3,623.88	13,132,488

Table 6.4: Descriptive statistics for net present value associated with profit from observed sales

In the next step, the corresponding values of expected profit applying the price-based revenue management program are calculated.

6.3.2 Determining Expected Profit Applying Optimal Pricing Strategies

To compare the price-based revenue management program with the outcome of actual pricing schemes in practice, expected profit is calculated applying the discrete-time stochastic dynamic model under consideration of a finite price set, similar to the selected example previously stated. More precisely, similar to the analysis for the selected example, the price region is restricted to multiples of 100 euros with $\Omega_p^{100}(t) = \{0, 100, 200, 300, 400, \ldots\}$, the price is determined on a weekly basis and the selling horizon is considered to be twenty weeks, that is, after this time period the used vehicle has to be sold for the salvage value, only. Furthermore, assume that the salvage value associated with the terminal state is set to zero. Note that the first assumption was chosen in accordance with findings from expert interviews, whereas the latter reflects the conservative approach of the present analysis, since section 3.5.3 demonstrated that higher salvage values would result in higher expected profits.

On the basis of a separate analysis, the data was clustered into subgroups, differentiating between specific model types and engine sizes. Applying hypothesis tests for the equality of the survival function of two or more subgroups introduced in subsection 4.2.3.2, these subgroups are joined together for further calculations in case there is no significant difference in the probability of a sale for two considered populations. Eventually, fourteen subgroups are composed of the initial dataset and a demand function is estimated separately for each subset. Then, for each subset a log-logistic AFT model extended by regression splines for ‘degree-of-overpricing’ as the only explanatory variable is fitted the datasets. Then, the corresponding expected survival function is estimated as a basis for the individual price response functions. Note that since optimal prices are determined on a weekly basis, the probability of survival is of interest at the time $t = 7$ days and, thus, the baseline survival function has to be estimated accordingly.

For each object within the selected dataset, the expected profit is calculated on the basis of the determined optimal pricing strategy, applying the discrete-time stochastic dynamic model under consideration of a finite price set. Respectively, the required individual demand function for a used vehicle is constructed corresponding to the log-logistic AFT survival model of the subgroup with which the considered object can be associated. The results of these calculations are given in Table 6.5. Here,

	N	Min.	Max.	Mean	Mean Std. Error	Std. Dev.	Var.
Exp. Profit	1,816	6,107	30,236	15,312.03	84.391	3,596.29	12,933,307

Table 6.5: Descriptive statistics for expected profit from the vehicles' sales utilizing discrete-time stochastic dynamic model under consideration of a finite price set

the average expected profit determined by the optimal pricing strategy is 15,312.03 euros with a standard deviation of 3,596.29. Both the variance and the standard deviation are smaller than the values for the discounted profit associated with the actual sale of the considered vehicle. One reason for this observation might be that negotiating partners sometimes behave irrationally. Customers might buy a vehicle for a price in excess of its market value and retailers might sell a car below value. Therefore, the range of prices for the same product is much wider when determined by the optimal pricing algorithm.

6.3.3 Comparison and Analysis

After separately determining the observed profits expressed in net present values and the expected profits estimated by the discrete-time stochastic dynamic model, both calculations are compared individually and analyzed to identify the potential for profit enhancement by applying the dynamic pricing program. The descriptive statistics of the comparison are shown in Table 6.6. For a total of 1,816 objects in the dataset, the average difference between expected profit generated by applying the optimal pricing strategy and the net present value of the actual sale amounts to 670.59 euros. With an average net present value of 14,641.45 euros for all actual sales, the surplus by the price-based revenue management program is equivalent to an enhancement of 4.6 percent. A more in-depth analysis by subgroups reveals that the optimal pricing strategy is superior for all subsets. Two subgroups, namely

Type	N	Mean	Mean Std. Error	Std. Deviation	Variance
1	120	174.33	98.75	1,081.71	1,170,099
2	66	421.77	141.04	1,145.85	1,312,983
3	147	769.63	80.85	980.24	960,877
4	21	585.24	262.21	1,201.59	1,443,827
5	113	864.38	96.55	1,026.36	1,053,414
6	7	1,574.01	139.98	370.35	137,162
7	182	573.36	108.40	1,462.37	2,138,512
8	456	795.76	70.85	1,512.88	2,288,804
9	363	530.31	65.83	1,254.18	1,572,976
10	178	771.55	123.15	1,642.98	2,699,368
11	1	1975.10			
12	100	893.71	122.90	1,228.98	1,510,397
13	46	776.68	228.29	1,548.34	2,397,371
14	16	673.71	568.19	2,272.76	5,165,440
Total	1816	670.59	32.09	1,367.34	1,869,609

Table 6.6: Descriptive statistics of comparison between expected profit determined by the optimal pricing strategy and the net present value of actual sales. The analysis is subdivided according to types and engine sizes utilizing a cluster analysis.

group 6 and 11, feature an average expected profit enhancement of more than 1,000 euros per vehicle. This observation can be explained by the fact that the subgroups consist of only 7 and 1 vehicles respectively, which are of a more exclusive type in terms of engine, interior and exterior equipment. Furthermore, subgroup seven is characterized by a significantly lower standard deviation and variance due to the limited number of records within the respective samples.

In summary, the analysis of monetary impacts identified a significant potential for profit enhancement by applying optimal pricing strategies. The increase in profit compared to observed sales averages approximately 4.6 percent for the observed dataset, thereby increasing profit margins within the used car sector by a multiple of current profits. Even more remarkable is an analysis of the total potential for enhancement. Adding the expected profit over all vehicles within the analyzed dataset, a total amount of 1,217,785 euros for 1,816 vehicles could have been gained by the involved retailers if they had used optimal pricing strategies. Furthermore, translating the potential increase in profit to all vehicles examined during the year 2006 within the market study, the total of over 50,000 offered vehicles belonging to the same model would hold a medium double-digit million euro sum in potential profit.

Besides the positive monetary impacts, the price-based revenue management program implicates operational, tactical, and strategic benefits. On an operational level, the revenue management program can help retailers by monitoring the inventory and indicating the appropriate amount of time on the market. In a more passive approach, the program can only suggest price adjustments, but not perform them. Considering the tactical level, besides the proposition of optimal pricing strategies, executive vendors are freed by a price-based revenue management program from the task of determining prices. Especially the issue of information gathering and data preparation with regard to deriving asking prices without the help of a program is time consuming and distracts the retailer from his principal duty, consultation and guidance of customers. At the strategic level, the management of a used car retailing company does benefit from the revenue management program not only with the pricing strategy, but also by identifying such profit potential within the selling process and adjacent business areas. The program can be used to recommend types of vehicles that promise higher than average profit margins.

6.4 Summary

This chapter demonstrated the concept of the price-based revenue management program and identified significant potential for profit enhancement for its application within the used car sector. An extensive dataset was analyzed, comparing the discounted profit of observed sales and the corresponding expected profits determined by the revenue management program. The dynamic pricing approach achieved significantly higher profits than the actual sales strategy applied on the observed vehicles and, thus, promises an increase in profit margins. There are not only financial incentives with the application of the revenue management program, but also managerial benefits at an operational, tactical and strategic level.

Chapter 7
Conclusions and Further Directions

The final chapter identifies possible extensions within this work and areas for future research activities in section 7.1 and summarizes the main findings of the book in section 7.2.

7.1 Directions for Future Research

The work in the present book was related to a number of different research areas with the main focus put on the interaction of survival analysis, hedonic price modeling and optimal control theory. Hence, several interesting directions of future research can be identified. Concerning the development of the optimization module, three extensions are outlined in the following.

Variable final time In the present paper, optimization problems were restricted to the case where the selling horizon was assumed to be fixed. However, for some real-world problems it might be of interest to determine the final time optimally, along with the optimal price path. In comparison with the standard fixed final time problem, the variable final time problem is extended by an additional unknown variable.

Group demand models In practice, often the inventory of used car retailers contains types of vehicles which are not identical but only similar to each other, such as cars from the same class of models but different manufacturers. Then, instead of considering the disposal of a single item of a given number, an extended model is required to incorporate group demand.

Competition An important extension for any real-world implementation of a revenue management program is competition. In the existing models, the retailer is assumed to act as a monopolist, but the incorporation of competition would make a model much more realistic and, thus, more appealing to practitioners.

Regarding the demand forecasting module, various fields of research are affected by this work, including hedonic price modeling for estimating the market value of a object and survival analysis to predict demand by estimating the survival probability on the market. Hence, several extensions and directions for further research can be pointed out.

Time-dependent variables The presented demand models account for only a limited number of predictors to identify and select an appropriate survival model for estimating demand. In addition, selected predictors, such as the market value or the size of the specific market, were restricted to time-independent values. However, in practice, the explanatory variables might vary with time and therefore should be treated as time-dependent in the survival analysis. Also, future studies should incorporate additional internal and external factors to better represent actual market demand forces.

Multiple events per subject In the developed survival models, we assumed that an object within the dataset was either sold to a customer or the price was adjusted accordingly. In the latter case, a new data object was generated, thereby accepting that the underlying data set consisted of objects with only one event per subject. However, to identify depended variables based on the course of the price quoting history, it might be advantageous to treat these quoting histories as interrelated datasets with multiple events per subject.

Simultaneous estimation of market value and survival In the presented models, the market value of an object and its probability of sale were estimated separately. There might be circumstances in practice in which both the time on market as a proxy of the probability of sale and the market value influence each other. In this case, the execution of simultaneous estimates might be appropriate. Furthermore, in addition to the semi-logarithmic regression model, more sophisticated approaches can be applied to estimate the market value of a used vehicle. One such approach could be estimating using neural networks.

We believe that the proposed framework cannot only provide a basis for further research in dynamic pricing and individual demand estimation, but it can serve as a prototype for testing the revenue management program at used car retailers' sites.

7.2 Summary

In the present book, the application of revenue management techniques to the automobile industry was conducted by developing a dynamic pricing program for the used car sector. Assessing the automobile industry as suitable for revenue management, the used car sector was identified as a segment promising significant potential for increased profit since the segment generates only marginal contributions to the operating results of automobile companies. The analysis performed within the scope of this work revealed that used car retailers should reform their pricing processes by establishing professional pricing management procedures in order to increase profits or at least reduce losses.

In this work, a price-based revenue management framework was developed for retailers to maximize their profits by determining optimal pricing strategies for used vehicles. Using stochastic control theory, different models and corresponding algorithms were derived, incorporating both theoretical and practical requirements by considering continuous- and discrete-time models as well as restricting the price region to a finite set.

To estimate individual price response functions, a promising approach for the automobile industry was followed in applying survival analysis. A new category of market data was introduced with the time duration type to use information not only about the sale of a good but also about the price quoting history as well. The assumption of using survival analysis to estimate individual demand functions was based on the hypothesis that the asking price of an offered vehicle influences the probability of its sale. Hence, a functional relationship between the asking price and the probability of a sale was established, serving as a proxy for predicting demand for the vehicle. This assumption was confirmed within this work.

The model selection and validation process regarding the determination of suitable survival analysis models and, thus, price response functions, was conducted experimentally by evaluating data from a market study in the German used car market. The configuration of this market study comprised the collection and analysis of price quoting histories over the course of a ten-month period. On the basis of this data, the accelerated failure time model, extended by regression splines, was selected as the best parametric model for the estimation of used cars' sales probabilities.

A case study was undertaken to reveal the potential of the presented price-based revenue management approach with regards to profit enhancement. The general concept of the program was demonstrated by taking a used vehicle and estimating the corresponding demand function by survival analysis, the market value by hedonic price modeling and, finally, determining the optimal pricing strategy by stochastic control theory. An evaluation of an extensive sample of 1,816 used vehicles uncovered the potential for an increase in profitability by 4.6 percentage points on the basis of revenue by applying the presented price-based revenue management program, with a total expected profit amounting to 1,217,785 euros for all analyzed vehicles.

Even more remarkable is a translation of the potential increase in profit of all vehicles examined during the year 2006 within the market study, which would amount to a medium double-digit million euros sum in potential profit for a total of over 50,000 vehicles offered. Taking into consideration that the profit margins in the used car sector amount to less than one percent, the identified enhancement translates to an increase in profit by a multiple of actual profits. On basis of these findings, the developed price-based revenue management framework can support used car retailers to reform their selling process in the future, leading to reduced losses and increased profits.

References

AALEN, O. (1978): “Nonparametric Inference for a Family of Counting Processes,” *The Annals of Statistics*, 6(4), 701–726.

ACHTERHOLT, U., AND R. SCHMID (2007): *Momentum 2007 KPMG Global Auto Executive Survey*. KPMG International.

AL-SIBAI, J., O. MÖLLER, AND M. HOFER (2004a): “Pricing Prozesse in der Automobilindustrie,” in *Automotive Management. Strategie und Marketing in der Automobilwirtschaft*, ed. by B. Ebel, M. B. Hofer, and J. Al-Sibai, vol. 1, pp. 351–365. Springer, Berlin.

——— (2004b): “Verkaufsförderungsmaßnahmen und Customer Incentives im Spannungsfeld zwischen Marktanteil und Gewinn,” in *Automotive Management. Strategie und Marketing in der Automobilwirtschaft*, ed. by B. Ebel, M. B. Hofer, and J. Al-Sibai, vol. 1, pp. 475–486. Springer, Berlin.

ANGLIN, P. M., R. RUTHERFORD, AND T. M. SPRINGER (2003): “The Trade-off Between the Selling Price of Residential Properties and Time-on-the-Market: The Impact of Price Setting,” *Journal of Real Estate Finance & Economics*, 26(1), 95–111.

BAGDONAVICIUS, V., AND M. NIKULIN (2004): “Statistical Modeling in Survival Analysis and its Influence on the Duration Analysis,” in *Advances in Survival Analysis*, ed. by N. Balakrishnan, and C. R. Rao, vol. 23 of *Handbook of Statistics*, pp. 411–429. Elsevier, Amsterdam.

BALDERJAHN, I. (2003): “Erfassung der Preisbereitschaft,” in *Handbuch Preispolitik : Strategien, Planung, Organisation, Umsetzung*, ed. by H. Diller, vol. 1, pp. 387–404. Gabler, Wiesbaden.

BANHAM, R. (2003): “The Right Price: Ford has an ambitious Revenue-Management Strategy, but can it Save the Company?,” *CFO, The Magazine for Senior Financial Executives*, 19(13), 66–73.

BELLMAN, R. E. (1957): *Dynamic Programming*. Princeton University Press, Princeton, NJ.

BELOBABA, P. P. (1987): “Airline Yield Management. An Overview of Seat Inventory Control,” *Transportation Science*, 21(2), 63–73.

——— (1989): “Application of a Probabilistic Decision Model to Airline Seat Inventory Control,” *Operations Research*, 37(2), 183–197.

BERRY, S., J. LEVINSOHN, AND A. PAKES (1995): “Automobile Prices in Market Equilibrium,” *Econometrica*, 63(4), 841–890.

BETTMAN, J. R., M. F. LUCE, AND J. W. PAYNE (1998): “Constructive Consumer Choice Processes,” *Journal of Consumer Research*, 25(3), 187–217.

BITRAN, G., AND R. CALDENTEY (2003): “An Overview of Pricing Models for Revenue Management,” *Manufacturing & Service Operations Management*, 5(3), 203–229.

BITRAN, G., AND S. GILBERT (1996): “Managing Hotel Reservations with Uncertain Arrivals,” *Operations Research*, 44(1), 35–49.

BITRAN, G., AND S. MONDSCHEIN (1993): “Pricing Perishable Products: An Application to the Retail Industry,” Discussion paper, Alfred P. Sloan School of Management, Massachusetts Institute of Technology.

——— (1995): “An Application of Yield Management to the Hotel Industry considering Multiple Day Stays,” *Operations Research*, 43(3), 427–443.

BITRAN, G. R., AND S. V. MONDSCHEIN (1997): “Periodic Pricing of Seasonal Products in Retailing,” *Management Science*, 43(1), 64–79.

BLOSSFELD, H.-P., A. HAMERLE, AND K. U. MAYER (1986): *Ereignisanalyse: Statistische Theorie und Anwendung in den Wirtschafts- und Sozialwissenschaften*, vol. 569 of *Campus Studium*. Campus Verlag, Frankfurt/Main.

BOX, G., G. JENKINS, AND G. REINSEL (1994): *Time Series: Forecasting and Control*. Prentice-Hall, Englewood Cliffs, NJ.

BRADBURN, M. J., T. G. CLARK, S. B. LOVE, AND D. G. ALTMAN (2003): “Survival Analysis Part II: Multivariate Data Analysis–An Introduction to Concepts and Methods,” *British Journal of Cancer*, 89(3), 431–436.

BRANDSTAD, P., T. WILLIAMS, AND T. RODEWIG (1999): *Challenges Facing the Global Automotive Industry*Booz Allen & Hamilton, Chicago, Illinois.

BREIDERT, C. (2006): *Estimation of Willingness-To-Pay: Theory, Measurement, Application*, Gabler Edition Wissenschaft. Deutscher Universitäs-Verlag, Wiesbaden.

BRESLOW, N. (1970): “A Generalized Kruskal-Wallis Test for Comparing K Samples Subject to Unequal Patterns of Censorship,” *Biometrika*, 57(3), 579–594.

BRESLOW, N. (1972): “Contribution to the Discussion of the Paper by D.R. Cox,” *Journal of the Royal Statistical Society*, 34(2), 216–217.

——— (1974): “Covariance Analysis of Censored Data,” *Biometrics*, 30, 89–99.

BRÉMAUD, P. (1981): *Point Processes and Queues: Martingale Dynamics*, vol. 1 of *Springer Series in Statistics*. Springer, New York, NY.

BUCKLEY, J., AND I. JAMES (1979): “Linear Regression with Censored Data,” *Biometrika*, 66(3), 429–436.

BURNHAM, K., AND D. ANDERSON (2002): *Model Selection and Multi-Model Inference: A Practical Information-Theoretic Approach*, vol. 2. Springer, New York, NY.

——— (2004): "Multimodel Inference: Understanding AIC and BIC in Model Selection," *Sociological Methods & Research*, 33(2), 261–304.

CARROLL, W., AND R. GRIMES (1995): "Evolutionary Change in Product Management: Experiences in the Car Rental Industry," *Interfaces*, 25(5), 84–104.

CIANCIMINO, A., G. INZERILLO, S. LUCIDI, AND L. PALAGI (1999): "A Mathematical Programming Approach for the Solution of the Railway Yield Management Problem," *Transportation Science*, 33(2), 168–181.

CLARK, T. G., M. J. BRADBURN, S. B. LOVE, AND D. G. ALTMAN (2003): "Survival Analysis Part I: Basic Concepts and First Analyses," *British Journal of Cancer*, 89(2), 232–238.

COURT, A. (1939): "Hedonic Price Indexes with Automotive Examples," in *The Dynamics of Automobile Demand*, pp. 98–119. The General Motors Corporation.

COURTY, P. (2000): "An Economic Guide to Ticket Pricing in the Entertainment Industry," *Louvain Economic Review*, 66(2), 167–189.

COX, D. (1972): "Regression Models and Life Time Tables," *Journal of the Royal Statistical Society*, 34, 187–220.

COX, D. R., AND D. OAKES (1985): *Analysis of Survival Data*, Monographs on Statistics and Applied Probability. Chapman and Hall, London.

CROSS, R. (1995): "An Introduction to Revenue Management," in *The Handbook of Airline Economics*, ed. by D. Jenkins, vol. 1, pp. 443–458. McGraw Hill, New York, NY.

CROSS, R. (1997): *Revenue Management: Hard-Core Tactics for Profit-Making and Market Domination*. Orion Business, New York, NY and London.

DATAMONITOR (2005): *Used Cars in Germany: Industry Research Report*.

DEFREGGER, F., AND H. KUHN (2003): "Revenue Management in Manufacturing," in *Operations Research Proceedings*, ed. by D. Ahr, R. Fahrion, M. Oswald, and G. Reinelt, pp. 17–22.

DONAGHY, K., U. MCMAHON-BEATTIE, AND D. MCDOWELL (1997): "Implementing Yield Management: Lessons from the Hotel Sector," *International Journal of Contemporary Hospitality Management*, 9(2), 50–54.

EFRON, B. (1977): "The Efficiency of Cox's Likelihood Function for Censored Data," *Journal of the American Statistical Association*, 72(359), 557–565.

——— (1981): "Censored Data and the Bootstrap," *Journal of the American Statistical Association*, 76(374), 312–319.

EFRON, B., AND R. TIBSHIRANI (1993): *An Introduction to the Bootstrap*, vol. 59 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, New York, NY.

ELIMAM, A., AND B. DODIN (2001): "Incentives and Yield Management in Improving Productivity of Manufacturing Facilities," *IIE Transactions*, 33(6), 449–462.

ELMAGHRABY, W., AND P. KESKINOCAK (2003): "Dynamic Pricing in the Presence of Inventory Considerations: Research Overview, Current Practices, and Future Directions," *Management Science*, 49(10), 1287–1309.

ENGELKE, J. (2004): "Preisgestaltung von Sonderausstattungen," in *Automotive Management. Strategie und Marketing in der Automobilwirtschaft*, ed. by B. Ebel, M. B. Hofer, and J. Al-Sibai, vol. 1, pp. 377–390. Springer, Berlin.

FENG, Y., AND G. GALLEGO (1995): "Optimal Starting Times for End-of-Seasons Sales and Optimal Stopping Times for Promotional Fares," *Management Science*, 41(8), 1371–1391.

FENG, Y., AND B. XIAO (2000): "A Continuous-Time Yield Management Model with Multiple Prices and Reversible Price Changes," *Management Science*, 46(5), 644–657.

FOLLAIN, J., AND E. JIMENEZ (1985): "Estimating the Demand for Housing Characteristics," *Regional Science and Urban Economics*, 15, 77–107.

FRIEGE, C. (1996): "Yield-Management," *Wirtschaftswissenschaftliches Studium*, 25(12), 616–622.

GALLEGO, G., AND G. VAN RYZIN (1994): "Optimal Dynamic Pricing of Inventories with Stochastic Demand over Finite Horizons," *Management Science*, 40(8), 999–1020.

GEHAN, E. A. (1965): "A Generalized Wilcoxon Test for Comparing Arbitrarily Singly-Censored Samples," *Biometrika*, 52(1-2), 203–223.

GERAGHTY, M. K., AND E. JOHNSON (1997): "Revenue Management Saves National Car Rental," *Interfaces*, 27(1), 107–127.

GIHMAN, I. I., AND A. V. SKOROHOD (1979): *Controlled Stochastic Processes*. Springer, New York, NY.

GRAMBSCH, P., T. THERNEAU, AND T. FLEMING (1995): "Diagnostic Plots to Reveal Functional Form for Covariates in Multiplicative Intensity Models," *Biometrics*, 51(4), 1469–1482.

GRAUNT, J. (1662): *Natural and Political Observations on the Bills of Mortality*. John Martyn and James Allestry, London.

GRILICHES, Z. (1961): "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change," *The Price Statistics of the Federal Government*, 73, 173–196.

HAAS, G. (1922): "Sales Prices as a Basis for Farm Land Appraisal," Technical Bulletin 9, The University of Minnesota, Agricultural Experiment Station, St. Paul, MN.

HARRELL, F. (2001): *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*, vol. 1. Springer, New York, NY.

HARRELL, F. E. (2005): “Design: Design Package,” R package version 3.3-1.

——— (2007): “Hmisc: Harrell Miscellaneous,” R package version 3.3-1.

HARRIS, F. H., AND P. PEACOCK (1995): “Hold My Place, Please,” *Marketing Management*, 4(2), 34–46.

HARRIS, F. H., AND J. PINDER (1995): “A Revenue Management Approach to Demand Management and Order Booking in Assemble-to-Order Manufacturing,” *Journal of Operations Management*, 13(4), 299–309.

HECHING, A., G. GALLEGO, AND G. VAN RYZIN (2002): “Mark-Down Pricing: An Empirical Analysis of Policies and Revenue Potential at one Apparel Retailer,” *Journal of Revenue & Pricing Management*, 1(2), 139–160.

HIRAOKA, L. S. (2001): *Global Alliances in the Motor Vehicle Industry*. Quorum Books, Westport, CN.

HOFER, M. B., B. EBEL, AND J. AL-SIBAI (2004): “Erstpreispositionierung und Preisoptimierung von Neufahrzeugen,” in *Automotive Management. Strategie und Marketing in der Automobilwirtschaft*, ed. by B. Ebel, M. B. Hofer, and J. Al-Sibai, vol. 1, pp. 366–376. Springer, Berlin.

HOROWITZ, J. (1992): “The Role of the List Price in Housing Markets: Theory and an Econometric Model,” *Journal of Applied Econometrics*, 7(2), 115–129.

JEDIDI, K., AND Z. J. ZHANG (2002): “Augmenting Conjoint Analysis to Estimate Consumer Reservation Price,” *Management Science*, 48(10), 1350–1368.

JONES, P. (1999): “Yield Management in UK Hotels: A Systems Analysis,” *Journal of the Operational Research Society*, 50(11), 1111–1119.

KALBFLEISCH, J. D., AND R. L. PRENTICE (1980): *The Statistical Analysis of Failure Time Data*, Wiley Series in Probability and Mathematical Statistics. Wiley, New York, NY.

KAMIEN, M. I., AND N. L. SCHWARTZ (1981): *Dynamic optimization: The Calculus of Variations and Optimal Control in Economics and Management*. North-Holland, New York, NY.

KAPLAN, E. L., AND P. MEIER (1958): “Nonparametric Estimation from Incomplete Observations,” *Journal of American Statistical Association*, 53, 457–481.

KASILINGAM, R. (1996): “Air Cargo Revenue Management: Characteristics and Complexities,” *European Journal of Operations Research*, 96(1), 36–44.

KIEFER, N. (1988): “Economic Duration Data and Hazard Functions,” *Journal of Economic Literature*, 26(2), 646–679.

KIMES, S. (1999): “Implementing Restaurant Revenue Management: A Five-step Approach,” *Cornell Hotel and Restaurant Administration Quarterly*, 40(3), 16–21.

KIMES, S., D. BARRASH, AND J. ALEXANDER (1999): “Developing a Restaurant Revenue-Management Strategy,” *Cornell Hotel and Restaurant Administration Quarterly*, 40(5), 18–29.

KIMES, S. E. (1989): “The Basics of Yield Management,” *Cornell Hotel & Restaurant Administration Quarterly*, 30(3), 14–19.

KIMES, S. E., R. B. CHASE, S. CHOI, P. Y. LEE, AND E. N. NGONZI (1998): "Restaurant Revenue Management: Applying Yield Management to the Restaurant Industry," *Cornell Hotel and Restaurant Administration Quarterly*, 39(3), 32–39.

KIMMS, A., AND R. KLEIN (2005): "Revenue Management im Branchenvergleich," *Revenue Management.*, 1, 1–30.

KLEIN, J. P., AND M. L. MOESCHBERGER (2003): *Survival Analysis: Techniques for Censored and Truncated Data*, Statistics for Biology and Health. Springer, New York, NY.

KLEIN, R. (2001): "Revenue Management: Quantitative Methoden zur Erlösmaximierung in der Dienstleistungsproduktion," *Betriebswirtschaftliche Forschung und Praxis*, 53(3), 245–259.

KLEINBAUM, D. G. (1996): *Survival Analysis: A Self-Learning Text*, Statistics in the Health Sciences. Springer, New York, NY.

KNIGHT, J. R. (2002): "Listing Price, Time on Market, and Ultimate Selling Price: Causes and Effects of Listing Price Changes," *Real Estate Economics*, 30(2), 213–237.

KRÄMER, A., AND H. LUHM (2002): "Peak Pricing oder Yield-Management? Zur Anwendbarkeit eines Erlösmanagement-Systems bei der Deutschen Bahn," *Internationales Verkehrswesen*, 54(1+2), 19–23.

KUHA, J. (2004): "AIC and BIC: Comparisons of Assumptions and Performance," *Sociological Methods & Research*, 33(2), 188–229.

LADANY, S. P., AND A. ARBEL (1991): "Optimal Cruise-Liner Passenger Cabin Pricing Policy," *European Journal of Operational Research*, 55(2), 136–147.

LANCASTER, K. (1966): "A New Approach to Consumer Theory," *The Journal of Political Economy*, 74(2), 132–157.

LAWLESS, J. F. (1982): *Statistical Models and Methods for Lifetime Data*, Wiley Series in Probability and Mathematical Statistics. Wiley, New York, NY.

LAZEAR, E. P. (1986): "Retail Pricing and Clearance Sales," *American Economic Review*, 76, 14–32.

LIEBERMAN, W. (1993): "Debunking the Myths of Yield Management," *Cornell Hotel and Restaurant Administration Quarterly*, 34(1), 34.

——— (2004): "Revenue Management Trends and Opportunities," *Journal of Revenue and Pricing Management*, 3(1), 91–99.

LITTLEWOOD, K. (1972): "Forecasting and Control of Passenger Bookings," in *AGIFORS 12th Annual Symposium Proceedings*, pp. 95–117.

MALPEZZI, S. (2002): "Hedonic Pricing Models and House Price Indexes: A Selective and Applied Review," in *Housing Economics and Public Policy*, ed. by A. O'Sullivan, and K. Gibb. Blackwell, Oxford.

MANTRALA, M., AND S. RAO (2001): "A Decision-Support System that Helps Retailers Decide Order Quantities and Markdowns for Fashion Goods," *Interfaces*, 31(2), 146–165.

MCGILL, J. I., AND G. J. VAN RYZIN (1999): "Revenue Management: Research Overview and Prospects," *Transportation Science*, 33(2), 233–256.

MERCER (2005): *Systemprofit Automobilvertrieb 2015 – Die Agenda für profitables Wachstum der Markenkanäle*. Mercer Management Consulting, München.

MILLER, M., C. LANGEFELD, W. TIERNEY, S. HUI, AND C. MCDONALD (1993): "Validation of Probabilistic Predictions," *Medical Decision Making*, 13(1), 49–58.

MONROE, K. (1979): *Pricing: Making Profitable Decisions*. McGraw-Hill Book Company, New York NY.

NAGELKERKE, N. (1991): "A Note on a General Definition of the Coefficient of Determination," *Biometrika*, 78(3), 691–692.

NAGLE, T. T., AND J. E. HOGAN (2006): *The Strategy and Tactics of Pricing: A Guide to Growing More Profitably*. Pearson/Prentice Hall, Upper Saddle River, NJ.

NELSON, W. B. (1972): "Theory and Applications of Hazard Plotting for Censored Failure Data," *Technometrics*, 14, 945–965.

OAKES, D. (2001): "Biometrika Centenary: Survival Analysis," *Biometrika*, 88(1), 99–142.

ORKIN, E. (1988): "Boosting your Bottom Line with Yield Management," *Cornell Hotel and Restaurant Administration Quarterly*, 28(4), 52–56.

PETO, R., M. C. PIKE, P. ARMITAGE, N. E. BRESLOW, D. R. COX, S. V. HOWARD, N. MANTEL, K. MCPHERSON, J. PETO, AND P. G. SMITH (1977): "Design and Analysis of Randomized Clinical Trials Requiring Prolonged Observation of Each Patient. II. Analysis and Examples," *British Journal of Cancer*, 35(1), 1–39.

PHILLIPS, R. (2005): *Pricing and Revenue Optimization*. Stanford University Press, Stanford, CA.

PINCHUK, S. (2002): "Future of Revenue Management – Revenue Management Does far More than Manage Revenues," *Journal of Revenue and Pricing Management*, 1(3), 283–285.

PONTRYAGIN, L., V. BOLTYANSKIJ, R. GAMKRELIDZE, AND E. MISHCHENKO (1962): *The Mathematical Theory of Optimal Processes*. Interscience Publishers, New York and London.

R DEVELOPMENT CORE TEAM (2006): *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

ROSEN, S. (1974): "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition," *The Journal of Political Economy*, 82(1), 34–55.

ROTHSTEIN, M. (1971): "An Airline Overbooking Model," *Transportation Science*, 5, 180–192.

——— (1974): "Hotel Overbooking as a Markovian Sequential Decision Process," *Decision Sciences*, 5(3), 389–404.

SATTLER, H., AND T. NITSCHKE (2003): “Ein empirischer Vergleich von Instrumenten zur Erhebung von Zahlungsbereitschaften,” *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung*, 55, 364–381.

SAUERBREI, W. (1999): “The Use of Resampling Methods to Simplify Regression Models in Medical Statistics,” *Applied Statistics*, 48(3), 313–329.

SCHÖNLEBER, J. (2006): *DAT-Report 2006*, Autohaus. Springer Transport Media GmbH, München.

SCHOENFELD, D. (1982): “Partial Residuals for the Proportional Hazards Regression Model,” *Biometrika*, 69(1), 239.

SETHI, S. P., AND G. L. THOMPSON (2000): *Optimal Control Theory: Applications to Management Science and Economics*. Kluwer Academic Publishers, Boston, MA.

SHEPPARD, S. (1999): “Hedonic Analysis of Housing Markets,” *Handbook of Regional and Urban Economics*, 3(1), 595–1635.

SHOEMAKER, R., AND S. SUBRAHMANYAN (1996): “Developing Optimal Pricing and Inventory Policies for Retailers Who Face Uncertain Demand,” *Journal of Retailing*, 72(1), 7–30.

SIMON, H. (1989): *Price Management*. Elsevier, Amsterdam.

SIMON, H. (2003): *Preismanagement: Analyse, Strategie, Umsetzung*. Gabler, Wiesbaden, 3 edn.

SMITH, B., J. LEIMKUHLER, R. DARROW, ET AL. (1992): “Yield Management at American Airlines,” *Interfaces*, 22(1), 8–31.

SMITH, S. A., AND D. D. ACHABAL (1998): “Clearance Pricing and Inventory Policies for Retail Chains,” *Management Science*, 44(3), 285–300.

STADJE, W. (1990): “A Full Information Pricing Problem for the Sale of Several Identical Commodities,” *Zeitschrift für Operations Research*, 34(3), S. 161–181.

STEYERBERG, E., F. HARRELL, G. BORSBOOM, M. EIJKEMANS, Y. VERGOUWE, AND J. HABBEMA (2001): “Internal Validation of Predictive Models Efficiency of some Procedures for Logistic Regression Analysis,” *Journal of Clinical Epidemiology*, 54(8), 774–781.

STRASSER, S. (1996): “The Effect of Yield Management on Railroads,” *Transportation Quarterly*, 50(2), 47–55.

STUHLMANN, S. (2000): *Kapazitätsgestaltung in Dienstleistungsunternehmungen: Eine Analyse aus der Sicht des externen Faktors*. Deutscher Universitäts-Verlag, Wiesbaden.

SUTHERLAND, I. (1963): “John Graunt: A Tercentenary Tribute,” *Journal of the Royal Statistical Society. Series A (General)*, 126(4), 537–556.

TALLURI, K. T., AND G. J. VAN RYZIN (2004): *The Theory and Practice of Revenue Management*. Boston, Mass.: Kluwer Academic Publ.

TARONE, R. E., AND J. WARE (1977): “On Distribution-Free Tests for Equality of Survival Distributions,” *Biometrika*, 64(1), 156–160.

TAYLOR, C. R. (1999): "Time-on-the-Market as a Sign of Quality," *Review of Economic Studies*, 66(3), 555–578.

THERNEAU, T., AND P. GRAMBSCH (2000): *Modeling Survival Data: Extending the Cox Model*. Springer-Verlag, New York.

THERNEAU, T., AND T. LUMLEY (2006): "Survival: Survival Analysis, Including Penalised Likelihood," R package version 2.4.

TSCHEULIN, D. K., AND H. LINDENMEIER (2003): "Yield-Management – Ein State-of-the-Art," *ZfB – Zeitschrift für Betriebswirtschaft*, 73(6), 629–662.

VALKOV, T. V., AND N. SECOMANDI (2000): "Revenue Management for the Natural Gas Industry," *Energy Industry Management*, 1(1).

VDA (2006): *Auto Annual Report 2006*. VDA Presse- und Öffentlichkeitsarbeit.

VERWEIJ, P., AND H. VAN HOUWELINGEN (1993): "Cross-Validation in Survival Analysis.," *Statistics in Medicine*, 12(24), 2305–2314.

VÖLCKNER, F. (2006): "Determinanten der Informationsfunktion des Preises: Eine empirische Analyse," *ZfB – Zeitschrift für Betriebswirtschaft*, 76(5), 473–497.

WALLACE, H. (1926): "Comparative Farmland Values in Iowa," *The Journal of Land & Public Utility Economics*, 2(4), 385–392.

WEATHERFORD, L. R., AND S. E. BODILY (1992): "A Taxonomy and Research Overview of Perishable-Asset Revenue Management: Yield Management, Overbooking, and Pricing," *Operations Research*, 40(5), 831–844.

WEI, L. (1992): "The Accelerated Failure Time Model: A Useful Alternative to the Cox Regression Model in Survival Analysis," *Statistics in Medicine*, 11, 1871–1879.

WELCH, D. (2003): "Ford Tames the Rebate Monster," *Business Week*, 3831, 38.

WERTENBROCH, K., AND B. SKIERA (2002): "Measuring Consumers' Willingness to Pay at the Point of Purchase," *Journal of Marketing Research*, 39(2), 228–241.

WIRTZ, J., S. E. KIMES, J. H. P. THENG, AND P. PATTERSON (2003): "Revenue Management: Resolving Potential Customer Conflicts," *Journal of Revenue & Pricing Management*, 2(3), 216–226.

WUPPERMANN, M. (2003): *Schnelldreher statt Langsteher: Best-Practice-Lösungen für Ihr Gebrauchtfahrzeug-Geschäft*. Auto-Business-Verlag, Ottobrunn.

YAVAS, A., AND S. YANG (1995): "The Strategic Role of Listing Price in Marketing Real Estate: Theory and Evidence," *Real Estate Economics*, 23(3), 347–368.

ZHAO, W., AND Y.-S. ZHENG (2000): "Optimal Dynamic Pricing for Perishable Assets with Nonhomogeneous Demand," *Management Science*, 46(3), 375–388.

Index