Dynamic pricing under customer choice behavior for revenue management in passenger railway networks

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften der Universität Mannheim

vorgelegt von

Simon Hohberger
Frankfurt am Main

Dekan: Prof. Dr. Christian Becker

Referentin: Prof. Dr. Cornelia Schön-Peterson

Korreferent: Prof. Dr. Moritz Fleischmann

Tag der mündlichen Prüfung: 30.01.2020

Abstract

Revenue management (RM) for passenger railway is a small but active research field with an increasing attention during the past years. However, a detailed look into existing research shows that most of the current models in theory rely on traditional RM techniques and that advanced models are rare. This thesis aims to close the gap by proposing a state-of-the-art passenger railway pricing model that covers the most important properties from practice, with a special focus on the German railway network and long-distance rail company Deutsche Bahn Fernverkehr (DB). The new model has multiple advantages over DB's current RM system. Particularly, it uses a choice-based demand function rather than a traditional independent demand model, is formulated as a network model instead of the current leg-based approach and finally optimizes prices on a continuous level instead of controlling booking classes. Since each itinerary in the network is considered by multiple heterogeneous customer segments (e.g., differentiated by travel purpose, desired departure time) a discrete mixed multinomial logit model (MMNL) is applied to represent demand. Compared to alternative choice models such as the multinomial logit model (MNL) or the nested logit model (NL), the MMNL is significantly less considered in pricing research. Furthermore, since the resulting deterministic multi-product multi-resource dynamic pricing model under the MMNL turns out to be non-linear non-convex, an open question is still how to obtain a globally optimal solution. To narrow this gap, this thesis provides multiple approaches that make it able to derive a solution close to the global optimum. For medium-sized networks, a mixed-integer programming approach is proposed that determines an upper bound close to the global optimum of the original model (gap < 1.5%). For large-scale networks, a heuristic approach is presented that significantly decreases the solution time (by factor up to 56) and derives a good solution for an application in practice. Based on these findings, the model and heuristic are extended to fit further price constraints from railway practice and are tested in an extensive simulation study. The results show that the new pricing approach outperforms both benchmark RM policies (i.e., DB's existing model and EMSR-b) with a revenue improvement of approx. +13-15% over DB's existing approach under a realistic demand scenario. Finally, to prepare data for large-scale railway networks, an algorithm is presented that automatically derives a large proportion of necessary data to solve choice-based network RM models. This includes, e.g., the set of all meaningful itineraries (incl. transfers) and resources in a network, the corresponding resource consumption and product attribute values such as travel time or number of transfers. All taken together, the goal of this thesis is to give a broad picture about choice-based dynamic pricing for passenger railway networks.

Table of contents

List of figures .... **V**

List of tables .... **VI**

Chapter 1 Introduction **1**

Chapter 2 An overview on passenger railway revenue management **7**

2.1 Introduction 8

2.2 Passenger railway revenue management in practice 10

2.2.1 A survey with European passenger railway companies 10

2.2.2 Characteristics of passenger railway 12

2.2.2.1 Structural network characteristics 13

2.2.2.2 Demand related characteristics 14

2.2.2.3 Business characteristics 18

2.3 Passenger railway literature review 23

2.3.1 Quantity-based railway revenue management 23

2.3.2 Price-based railway revenue management 27

2.3.3 Railway revenue management studies with various focus 29

2.3.4 Final remarks on literature overview 30

2.4 Conclusion 33

Chapter 3 Continuous pricing in a capacitated network under mixed multinomial logit demand **35**

3.1 Introduction 36

3.1.1 Problem and motivation 36

3.1.2 Related work 38

3.1.3 Contribution and outline 48

3.2 Model development 49

3.2.1 Original model formulation (P0) 49

3.2.2 An equivalent formulation with inverse price-demand relationship (P1) 51

3.2.3 Convex approximations yielding upper bounds 53

3.2.3.1 Segment-based pricing relaxation (P_REL1) 53

3.2.3.2 Segment-based pricing relaxation with envelopes for the price consistency constraint (P_REL2) 54

3.2.3.2.1 Envelopes for the logarithm 54

3.2.3.2.2 McCormick envelopes for bilinear terms 56

3.2.3.3 Piecewise approximations through interval partitioning for product prices (P_REL3) 58

3.2.3.3.1 Piecewise McCormick approximations 58
3.2.3.3.2 Piecewise envelopes for the logarithm 61
3.3 Heuristics yielding lower bounds 63
3.3.1 Outer approximation heuristic 63
3.3.2 Price interval heuristic (3-step approach) 63
3.4 Numerical study 65
3.4.1 Systematic experiment 65
3.4.1.1 Experimental setup 65
3.4.1.2 Results 67
3.4.2 Railway network revenue management application 69
3.4.2.1 Experimental setup 69
3.4.2.2 Results 72
3.5 Conclusion and future research 78

Chapter 4 A simulation study for a choice-based dynamic pricing approach for large-scale railway networks 81
4.1 Introduction 82
4.1.1 Motivation and overview 82
4.1.2 Literature overview 83
4.2 The choice-based railway network dynamic pricing model 86
4.2.1 Model formulation 88
4.2.2 Solution approach 91
4.3 The simulation framework 95
4.3.1 Benchmark policies 95
4.3.2 Framework of the simulation 97
4.3.3 Test instances and data 101
4.3.4 Simulation of special O&D model versions 103
4.4 Simulation results 105
4.4.1 Qualitatively discussion of O&D model 105
4.4.2 Results of the main simulation 106
4.4.3 Results of special cases 110
4.5 Insights from practice 112
4.6 Conclusion 114

Chapter 5 An algorithm to create test data for large-scale railway network revenue management models with customer choice 115
5.1 Introduction 116
5.2 Theoretical background 118
5.3 The network data generation algorithm (NDGA) 123

5.3.1 Constructing the set of efficient connections....125
5.3.2 Deriving data for discrete choice models....128
5.3.3 Deriving operational data....129
5.4 Practical example....131
5.4.1 Comments on run time....133
5.5 Conclusion....135

Chapter 6 Conclusion and outlook....137

Appendix....VII

References....IX

List of figures

Figure 1: Thesis overview 3
Figure 2: Areas of improvement 11
Figure 3: Classification tree of railway literature 30
Figure 4: Example for f(p) with inflection point at p =73.26 51
Figure 5: Relation between #train connections & #legs in a railway network 66
Figure 6: Solution times of all instances as boxplot for railway network (1) 74
Figure 7: Solution times of all instances as boxplot for railway network (2) 75
Figure 8: Segmentation by departure time preference 90
Figure 9: Overview of adapted 3-step approach for the railway model 91
Figure 10: Example for itinerary price consistency 93
Figure 11: Overview on processed steps in each period 98
Figure 12: Overview on different settings in the simulation 104
Figure 13: Price decision and its dependency 113
Figure 14: Process flow of NDGA 125
Figure 15: Evolution of efficient connections. 127
Figure 16: Network with three trains 131

List of tables

Table 1: Summary of properties for railway companies in the context of revenue management 22
Table 2: Literature in context of practical properties 31
Table 3: Literature overview on passenger railway RM 32
Table 4: Overview on pricing & product assortment under MMNL choice 47
Table 5: Results of the experimental study...68
Table 6: Results for solving the first instance in the low demand scenario 69
Table 7: Statistics for tested railway networks 71
Table 8: Experimental setup for railway study 72
Table 9: Results for P0 without warm start 73
Table 10: Results for price interval heuristic (3-step approach)..76
Table 11: Improvement by applying the price interval heuristic 77
Table 12: Statistics for test networks 102
Table 13: Average prices from tickets sold 108
Table 14: Statistics for solution times in seconds and objective value gap between RAIL_ORG and upper bound model RAIL_INV...109
Table 15: Revenue improvement of the O&D model in the different demand cases and corresponding statistics about network utilization. 109
Table 16: Results for both special cases...111
Table 17: Exemplary schedule for two trains 123
Table 18: Exemplary data for one connection 128
Table 19: Statistics regarding the generated data..131
Table 20: Set of efficient connections 131
Table 21: Choice data for connection Berlin - Cologne..132
Table 22: Overview of the amount of generated data and run time (in sec.) for different network sizes 133

Chapter 1 Introduction

The passenger railway business experienced an increasing interest in the last years, not only in emerging markets like Asia where new high-speed routes are build, but also in countries that are typically known for a long history of rail transport like Germany (e.g., Chen and Haynes (2015), DB Netz (2017)). Reasons for the positive development are manifold. In addition to a general increasing mobility and wealth that makes travelling affordable for more people, rail transport is typically a convenient means of transport for limited distances (e.g., no security check or train stations in the city center). Further, if powered by green electricity, it also fulfils the demand for sustainable mobility, an issue that is increasing in demand at the present time of discussions about global warming. From a business perspective, this is accompanied by the challenge of meeting additional demand and, above all, of operating profitably in order to be able to operate the cost-intensive material such as rail tracks or trains. Motivated by the success of revenue management (RM) for airlines also railway companies started to use RM methods, e.g., the German long-distance passenger train provider DB Fernverkehr in 2002 (Link 2004). While passenger railway and passenger airline RM share some fundamental aspects (both sell tickets from A to B that can be purchased in advance and have a zero salvage value after departure) they differ in some substantial aspects. For example, an itinerary in the railway case consists of multiple legs compared to one or two legs at airlines, and typically a higher frequency of itineraries between two cities is offered. Further, the lack of a check-in process or the possibility to have standing passengers on the train are additional differences with impact on RM. For this reason, passenger railway revenue management (RRM) represents its own field of research. An analysis of existing research shows, however, that most of the RRM approaches to date apply very basic techniques known from revenue management theory whereas advanced approaches that include state-of-the-art demand models are still rare (e.g., Jiang et al. (2015), Zhai et al. (2016)).

For RM research in general, there has been a trend in recent years towards choice-based pricing models. The aim of using choice-based optimization models is to account for demand shifts in the optimization (e.g., to alternative products, the competitor or not to buy at all) as a result of the pricing/availability decision. In the context of RRM, a choice model represents the customer choice of a train connection from a set of several alternative train connections (e.g., train connections departing earlier or later). With typically multiple daily train connections between large cities, a choice model can help to significantly increase revenues by a better steering of demand. While price optimization based on multinomial logit (MNL) (e.g., Dong et al. (2009))

or nested logit (NL) (e.g., Gallego and Wang (2014)) choice models has already been intensively considered, research on pricing under the discrete mixed multinomial logit model (MMNL) with capacity constraints is still rare (e.g., Keller et al. (2014)).

Research questions

This thesis is dedicated to the development of a state-of-the-art RRM model that covers the most important properties from practice, with a focus on the project partner DB Fernverkehr AG. In order to achieve this goal, the following research questions are considered:

1. What is the current status in RRM in theory and practice?
2. What are the main features a RRM model should reflect and what limitations need to be considered?
3. How can these properties be represented in an optimization model?
    a. How to solve the model in terms of solution quality? How can structural properties be exploited?
    b. How to solve the model in terms of solution time for large-scale railway networks? How to find a good balance between covered effects and keeping the model tractable?
    c. How to derive the necessary data for large-scale networks automatically?
4. How does the new model perform compared to alternative benchmark RRM approaches?

The proposed model is a multi-product multi-resource dynamic pricing model under the MMNL. The price for each product (the combination of a train connection and comfort class) is considered to be a continuous variable. Furthermore, the model covers the tradition network effect by controlling prices on a product level. Therefore, it is able to evaluate the optimal revenue contribution of different products that share the same critical resource. Finally, by integrating the MMNL, the model can represent the customer choice of a product as a choice out of a set of alternative products. I.e., the optimization model anticipates the switching behavior of customers between products as a result of the price decision. The MMNL makes it able to account for multiple heterogeneous customer segments that differ in the evaluation of product attributes. On the one hand, multiple segments make it able to better represent actual demand, on the other hand this comes along with the challenge of a non-linear non-convex optimization problem (Li et al. 2018). Therefore, efficient solution techniques are required to find a good solution in reasonable time.

Thesis structure

The thesis consists of four essays that address the mentioned research questions (see Figure 1).

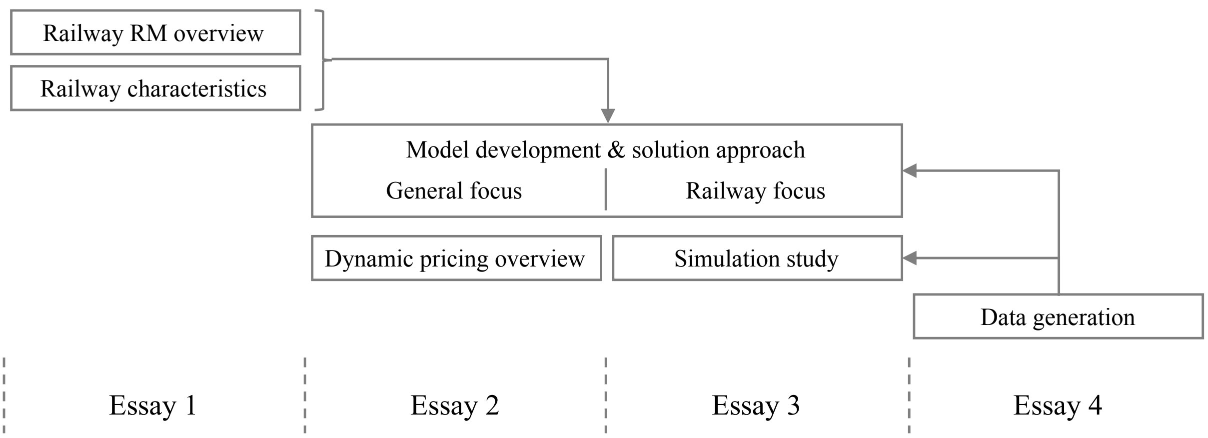

Figure 1: Thesis overview

The first essay (Chapter 2) gives a current overview on passenger railway revenue management from a theoretical and practical perspective. It presents the results from a survey with European passenger rail companies about RM in practice and shows the potential for advanced RM approaches. Furthermore, it discusses the main business model aspects of a passenger rail company that have an impact on RM based on the experiences at DB Fernverkehr. Finally, a literature review on RRM analyzes, discusses and classifies relevant papers published on the topic. It shows the increasing interest in research over the past years, but also reveals the gap of models that cover the most important conditions from practice. The first essay is the starting point and demonstrates the motivation for the development of the new pricing model in the following chapters.

The second essay (together with Cornelia Schön, Chapter 3) lays the theoretical foundation for the development of the network dynamic pricing model under the MMNL. It provides a literature review about choice-based pricing and shows the research gap of solving the MMNL-based pricing model to global optimality. The essay introduces multiple convex approximate reformulations of the original non-linear non-convex model that serve as upper bounds incl. a convex, piecewise approximation formulation that decreases the upper bound close to the globally optimal objective value. Furthermore, a heuristic approach is presented that makes it able to derive a good solution in reasonable time for large-scale networks. The presented models and heuristics are tested in two numerical experiments. The first experiment focuses on the upper bound models. It shows that the proposed convex models (especially the model that uses the piecewise approximation) help to decrease the objective value gap between the upper bound and the original non-convex formulation to <1.5%. The gap can be further decreased by choosing sufficiently small price intervals to make the piecewise approximation more precise. For the second experiment, data from DB Fernverkehr are used to test the heuristic for large-scale railway networks. The results show that the proposed heuristic makes it able to significantly decrease the solution time and to evaluate the (potentially local) solution by comparing the result with the solution from a convex upper bound model.

The third essay (Chapter 4) is based on the model and the theoretical results from essay two and extends it by integrating railway specific constraints to ensure price consistency between longer and shorter itineraries. This is necessary due to the missing check-in and check-out process so that customers could buy a ticket for a longer train connection and embark later and/or disembark earlier. Therefore, prices for longer train connections should not be lower than for shorter ones to avoid strategic acting customers. Additionally, from DB's product-strategic perspective, first- and second-class tickets for the same train connection should fulfill a minimum price-gap that is ensured by a second type of price consistency constraints.

An extensive simulation study is conducted to evaluate the performance of the new model in terms of revenue and network utilization compared to two leg-based benchmark policies that are based on DB's current RM policy. The strong collaboration with DB Fernverkehr made it possible to use real data from practice (e.g., train schedules, demand data, booking history etc.). Furthermore, technical statistics (e.g., solution times and variances, maximum gaps to the upper bound model) are presented and discussed to give insights into the quality of the solution from a theoretical perspective. The results show a significant revenue improvement by the new choice-based network dynamic pricing model.

Finally, the technical essay number four (together with Cornelia Schön, Chapter 5) lays the foundation for all tests that are done for railway networks in this thesis. Since one challenge with models for networks of realistic size is the availability of data, an algorithm is presented that automatically derives most of the operational data based on raw train schedules. This includes the set of train connections in the network (incl. transfers), the corresponding resource consumption for the network capacity constraints and product attribute values such as travel time or the number of transfers for the choice model. To derive the set of train connections, a number of efficiency rules are defined that ensure that only meaningful train connections are considered (e.g., without long detours).

All taken together, the thesis provides a broad picture of passenger railway revenue management with a special focus on network dynamic pricing models under the MMNL.

Project partner DB Fernverkehr AG

This thesis is not only motivated by its scientific questions as presented above, but also by the close cooperation with the project partner DB Fernverkehr AG in a direct practical context. DB Fernverkehr AG is a 100% subsidiary of Deutsche Bahn AG and provides long-distance travel within Germany and adjacent countries, mainly with its ICE (Intercity-Express) and IC/EC (Intercity/Eurocity) trains. With 147.9m travelers in 2018 and a revenue of 4.7bn EUR (Deutsche Bahn 2019), DB Fernverkehr is by far the largest provider of long-distance rail traffic in Germany. The revenue management department is part of the marketing division and consists of two teams, one team for the operational, short-term control (e.g., classical revenue managers) of prices and for the monitoring of train utilizations and a second team for data analytics and the development of the RM system. The joint research project contributes to this topic by answering the most important questions from a theoretical and practical perspective.

Chapter 2 An overview on passenger railway revenue management

Single author:
Simon Hohberger
Chair of Service Operations Management, Business School, University of Mannheim

Abstract

Revenue management (RM) is a discipline in Operations Research that is widely applied in practice, especially at airlines or hotels. The passenger railway industry can further be mentioned as an example with an increasing attention in research and practice especially in the last two decades driven by global investments in the extension of rail networks. This paper provides an overview of passenger railway in the context of revenue management. This includes an extensive study of the business model of a railway company to analyze the properties that have an influence on revenue management. Further, detailed insights into RM methods and challenges in practice based on a survey with European railway companies are provided. Finally, a literature overview analyzes existing research on railway revenue management from both a theoretical and practical perspective. This overview can serve as a starting point and source of information for future research in this still underdeveloped research area.

2.1 Introduction

Revenue management (RM) is an already well-established discipline in Operations Research, with wide applications in practice. Starting in the early 1970s, British Airways (back then called BOAC) started selling discounted flight tickets for customers willing to book early (McGill and van Ryzin 1999). The following deregulation of the airline market in 1978 is often seen as the starting point of an ongoing success story of revenue management in the airline sector (Talluri and van Ryzin 2005). To date, other industries have followed, including car rental companies or hotels (Chiang et al. 2007). Another sector to be mentioned is railway passenger transport, especially the long-distance rail market. Compared to airlines, hotels or car rentals this field is underrepresented both in theory and practice. One reason may be that revenue management research in the early years primarily took place in the USA where the focus is strongly on air traffic. At the same time, railway passenger transport is well suited for applying RM methods. E.g., following the prerequisite criteria for RM defined by Kimes (1989), a railway company has a fixed seat capacity, the product (i.e., a train ride from A to B) has a fixed expiry date and can be booked in advance, low variable costs arise with each additional passenger, demand is fluctuating and demand segments can be priced individually (although the latter one highly depends on the government's regulations). Thus, the right use of RM methods can potentially generate additional revenue and profit for railway companies.

The rail market in general shows a strong, positive development in many countries of the world. Networks in Asia (e.g., high-speed rail line Beijing-Shanghai opened in 2011 (Chen and Haynes 2015)) or Europe (e.g., the high-speed connection Berlin-Munich opened in 2017 (DB Netz 2017)) are expanding and increase the attractiveness of the product by shortening travel times. The demand for sustainable, green means of transport additionally favors the demand for railway. However, rising passenger numbers not only lead to rising revenues, but also to capacity bottlenecks with overcrowded trains. Thus, RM has not only a purely profit-driven role but also a steering component, as passengers can be deterred from taking full trains by setting appropriate prices. The interest in railway revenue management (RRM) is also reflected in the agenda of well-known rail conferences, with specific streams of presentations on revenue management (e.g., "Rail World Festival 2018" in Amsterdam, or "Asia Pacific Rail 2019" in Hong Kong). It will be shown in this paper that many rail companies are already applying some kind of RM technique. However, a detailed look into the results of the conducted survey also shows that the methods used are mainly basic and so potential for improved RM approaches exists. The reasons for the limited application in practice are manifold: railway companies are often state-owned and focus on the domestic market. Thus, in addition to the influence of the government that limits the flexibility in terms of pricing strategy in some countries, the strong geographical separation of markets leads to limited intramodal competition. Accordingly, a lack of competition can lead to a low level of innovation. In addition, decisions concerning public transport are often strongly and critically discussed by society. As a result, changes to the pricing system often take place in small steps. When DB Fernverkehr (the long-distance division of the German railway company Deutsche Bahn) applied RM for the first time in 2002 by introducing controlled, discounted tickets, the headwind of the population was immense as customers had to get used to the train-bounded tickets (until then they were used to purely flexible tickets) and the image damage was considerable (Link 2004). However, the inhibitory reasons mentioned above seem to dissolve more and more. The discounted, controlled tickets at DB Fernverkehr have enjoyed great popularity and are a reason for the sharp rise in passenger numbers and revenue.

In summary, it can be stated that the conditions for RM in railway passenger transport are met and the interest in practice is given, although advanced methods (e.g., network approaches) are missing. At the same time, a glance at research on RM for passenger railway shows a very limited number of publications. You (2008) and Ciancimino et al. (1999) can be described as the most prominent contributions. The working paper of Armstrong and Meissner (2010) gives the so far only overview on railway RM. Based on their results, this paper presents a broad and up-to-date overview of RM in passenger transport with a focus on relevant properties in practice. The scientific contribution is divided into three parts: First, the results of a survey conducted with European railway companies about RM in practice are presented, showing the already popular role of RM in practice as well as a high potential for advanced railway specific demand and optimization models. Second, particularities of rail transport that should be considered when developing a RM system are explained. Finally, an overview on existing literature that deals with passenger railway revenue management analyzes current approaches from both a theoretical and practical perspective.

2.2 Passenger railway revenue management in practice

Revenue management as a discipline is already widely applied and well-studied for various industries (see, e.g., Chiang et al. (2007) for an industry overview). An open question is still, if and how revenue management is applied at railway companies. Some practical insights are provided by, e.g., Sibdari et al. (2008) for the auto train at Amtrak (US rail company), Bharill and Rangaraj (2008) for Indian railway and Bao et al. (2014) for Chinese railway. Abe (2007) gives some insights into European and American railway companies but to the author's knowledge, there is no up-to-date analysis available that gives a structured overview about RM in practice.

2.2.1 A survey with European passenger railway companies

To make a first step for closing that gap a survey with passenger railway companies was conducted in spring 2018 by Feng (2018) as part of a master thesis at the University of Mannheim. In total, we contacted 28 companies (mainly in Europe) and received nine responses, from eight European and one Asian company. The online questionnaire covered different topics that are related with revenue management. To make the comparison and interpretation of the results easier for different companies in the same geographical/political region, only the results from the eight European companies are presented (i.e., n=8 sample size). Although the sample size is obviously not sufficient to derive statistically significant implications, it can give valuable insights and help to show the gap between theory and practice. The survey was aimed at three insights. First, to understand whether RM is applied at all and if so, which models and approaches are used. Second, to receive information about factors that influence the price decision and third, to understand the reasons that make an implementation difficult.

From the eight European rail companies that took part, seven are offering controlled tickets (i.e., tickets that, e.g., vary in price or availability) so that they can be considered to use RM techniques to some extent. Therefore, RM seems to be already well established in European railway practice. Only one company stated that they are not applying RM. From the seven companies, two are using a pure capacity control mechanism, four are having a mix of capacity- and price-based approach and one uses a fixed quota control depending on the day to departure so that it is considered to be a (very basic) pure price-based control. This indicates that capacity allocation models play a very important role in practice and that pure and advanced price-based approaches are not applied so far.

Furthermore, answers show that most of the rail companies (75%) offer both flexible and train-bounded tickets and the remaining 25% offer only flexible tickets. No company mentioned to offer only train-bounded tickets which shows the popularity and importance of flexible tickets in railway practice. Furthermore, seven companies mention to offer discount cards and most of them also offer discounts for groups, families or for different periods of validity (e.g., monthly passes). Asking for the most important factors that influence the pricing decision, the participants mentioned that the departure time is the most important one, closely followed by the number of seats available and the day of the week.

From a technical perspective, all seven companies mention to use a leg-based rather than a network-based approach. This is a surprising result since RRM represents a classical network problem with multiple products that share the same resource. In addition, three companies are controlling the product availabilities/prices completely manually, one uses an algorithmic approach and the remaining three others use both an algorithmic and manual control. This shows a high potential for automated RM methods since manual control is still dominant in practice.

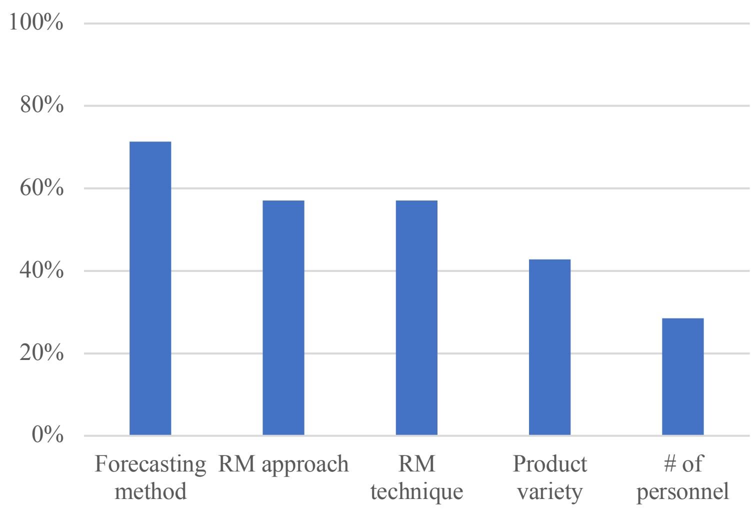

Figure 2: Areas of improvement

When asked about possibilities for improvement (see Figure 2), advanced forecasting methods were named most, followed by RM approach (e.g., going from leg-based to network model) and RM technique (e.g., improved heuristics, dynamic programming).

From an implementation perspective, challenges in the IT-setup are the most critical obstacles (100% approval) next to problems in customer acceptance (71% approval). The results about the organizational structure show that the RM teams (excluding IT support) are rather small (57% with 1-5 employees, 43% with 6-10 employees) and the teams are on a low (43%) or medium level (53%) in the hierarchy of the organization.

Taken all together one can say that RM is widely known and mostly applied in European railway practice but with significant potential for further improvements in the applied RM and forecasting techniques, especially with network-based models. Based on the desire for advanced RM approaches and techniques the question arises which (special) properties should be considered when developing a state-of-the-art railway RM model.

2.2.2 Characteristics of passenger railway

Armstrong and Meissner (2010) give the first overview of research for rail freight and rail passenger revenue management and a brief look into practice. They conclude that passenger railway RM is closely related to airline RM with some industry specific features. These include on average less utilized capacities, high number of walk-in customers or the possibility to stand on the train (i.e., no fixed maximum capacity). Furthermore, they point out that rail journeys typically consist of multiple legs so that it is necessary to control capacities/prices on an origin and destination (O&D) based level, which stands in contrast with the insights from the survey. They present a good overview of railway RM literature, the corresponding models and at the same time highlight that railway RM is still very unpopular in theory and (in some countries) also in practice and provides opportunities for future research in theory and practice.

Their contribution is used as a starting point to further extend the findings and present evidence from practice. The objective is to give a broad picture about properties that should be considered when building a railway RM system. Some of them can be determined as unique for the passenger rail business. The presented findings are mainly based on data and experiences from a joint project with DB Fernverkehr, the long-distance passenger rail unit of the main German railway company Deutsche Bahn. The attributes are split into three main areas, (1) structural (2) demand and (3) business model related characteristics. Important to mention is that the business model of a railway company can vary from country to country due to restrictions by, e.g., the government, geographical characteristics and historical influences. I.e., the characteristics presented in the following must not be universally valid for every country or company, but give a good overview of possible limitations and challenges when applying RM. In the following, statistics are based on internal data of DB (if not stated otherwise). Throughout the paper, the expression *O&D* is used to denote a pairing of an origin and destination (e.g., two cities/train stations) and the expression *itinerary* to denote a specific train connection. Example: There exist multiple daily itineraries for the O&D Frankfurt-Berlin. Finally, a *fare class* (or *fare product*) represents the combination of a price and corresponding fare rules (e.g., cancellation or advance booking policies), in the airline sector typically defined by capital letters (e.g., Y, M, C).

2.2.2.1 Structural network characteristics

The rail network in some countries shows similarities to a hub-and-spoke network (e.g., France or Great Britain), where train lines are meeting in a hub-train-station (e.g., Paris in France). In contrast, other countries show a highly-connected network with little/medium (Italy) or large (Germany) number of train lines. The structure of the railway network can have a significant influence on the challenges that come along with RM.

Challenge of defining the set of itineraries

Especially for large and highly-connected rail networks without a hub-and-spoke or point-to-point structure, there are typically multiple options to travel from city A to city B, e.g., some fast itineraries and some with (inefficient) detours. Since the set of itineraries is a crucial input in a network RM model, it is important to derive it in a meaningful way. At Deutsche Bahn, the algorithm that determines the offered train connections is not under direct control of the RM department but can be influenced by collaborating with the department in charge. Especially in situations of unexpected high demand, this can lead to suboptimal decisions. Imagine a high demand situation (e.g., soccer game Frankfurt vs. Berlin, taking place in Berlin on Friday evening) where many soccer fans want to travel from Frankfurt to Berlin. Since a normal Friday is already a day with strong demand, the additional passengers would lead to overcrowded trains. Therefore, it could be beneficial to offer alternative train connections that use less demanded trains (e.g., with a slight detour) so that not all soccer fans choose the same trains on the standard route from Frankfurt to Berlin. An approach that combines the decision of the offered product set (assortment problem) and the pricing decision (pricing models) would be a promising way. Since such an optimization model would be hard (if not impossible) to solve for large networks, a sequential step, where the set of itineraries is determined before the price optimization, is the more practicable way. In practice, there exists an additional challenge. On the website of DB, customers see by default only comfortable/fast train connections, but they have the option to deactivate an option called “prefer fast connection” so that alternative train connections are displayed. From an RM perspective that comes along with the question “What is the set of itineraries to optimize?”. While in practice, one would use the set of itineraries that is displayed by default as a product set, it is still necessary to have a heuristic that derives prices for the non-optimized/alternative train connections.

Multiple transfers when including regional train traffic

An itinerary consists of all used trains on the journey from your starting point (origin) to your final destination, i.e., it can include the transfer from one train to another. For the long-distance part of a journey, most of the booked itineraries have no transfer but still 18% include one or more transfers. Thus, a RM model should be able to deal with train connections including transfers.

When taking regional train services into consideration (i.e., the transport to/from the long-distance train station) the number of transfers further increases (50% of long-distance travelers have a regional train as part of their ticket). In practice, customers generally buy their regional and long-distance ticket as a bundle, i.e., not separately. On the one hand, especially in an O&D based model, this can create additional difficulties, e.g., for the forecasting (see demand related characteristics below) as well as the optimization since the number of itineraries to optimize would increase significantly.

An itinerary consists of many legs

In the airline business, an itinerary typically consists of only one or two (in minor cases three or four) legs and each additional leg typically causes a transfer to another aircraft. In contrast, trains generally travel along many intermediate stations. Further, every section between two stations represents a leg so that a trip from city A to city D via two intermediate stops (B and C) already results in an itinerary of three legs. When analyzing big networks like the DB long-distance network, you can see that an itinerary consists on average of eight legs which is by far larger than in any airline case (Hohberger and Schön 2019). At the same time, this means that a leg is used by multiple itineraries and, therefore, a network approach seems to be promising.

2.2.2.2 Demand related characteristics

Not only the network structure in the rail business can differ from other industries, also the demand for the products is special.

Many intermodal alternatives

Train services typically cover limited distances. This results in high competition with alternative means of transport such as intercity bus, private car, flight and carpooling. With increasing transparency based on price comparison websites, competition will probably be even stronger in future since customers can easily access and evaluate different alternatives and prices and finally switch, e.g., from train to car. If applicable, also other train service providers should obviously be considered as direct competitors. As rail companies are often government-owned and the only large rail provider in the operating market, competition between different train companies is weak/non-existing but this varies from country to country and sometimes also depends on the O&D. An extensive analysis of the competitive situation is necessary and demand forecasting techniques that allow for incorporating competition could be beneficial.

Many rail alternatives to go from A to B

There exist typically many train connections to go from city A to B within a day (e.g., 48 train connections between Berlin and Frankfurt each day). On the one hand, a tight schedule is good for the customers, on the other hand it makes forecasting difficult. Especially for train connections with similar departure times, an independent demand forecasting would ignore the choice effects of the customers, since they have many alternative train connections within a given desired departure time frame to choose from. Therefore, a demand model that accounts for (neighboring) alternative products is beneficial. In RM this challenge can be solved by incorporating some kind of customer choice model that becomes relevant especially in the context of RRM. Using a choice model helps the company to anticipate the itinerary choice of customers based on the pricing decision which is not possible with a traditional independent demand model.

Many changes in the schedule

Rail schedules are typically changed twice a year, in June and December. Major changes take place in December and the changes in June are typically smaller. During the year, additional, more spontaneous changes often occur, e.g., construction work that is so far not considered in the schedule. The problem with forecasting algorithms based on historical, realized passenger numbers is that any change in the timetable (i.e., planned or unplanned) will degrade the forecasting quality. For example, omitted stops (e.g., due to detours through construction sites) or supply reductions/expansions (i.e., fewer/more trains on the same O&D) have a strong effect on future demand that is not reflected in the historical data. From this point of view, forecasting methods that can account for timetable changes should be preferred. E.g., a discrete choice model can be considered. The idea is to use an O&D market demand forecast that is independent of one's own supply/timetable and to allocate the demand (by applying the choice model) to all possible itineraries on the basis of the respective attractiveness. Changes in the timetable are reflected directly in the utility values of the set of itineraries and thus lead to a forecast which takes the new timetable into account.

High percentage of tickets are purchased close to departure

As already mentioned by Hetrakul and Cirillo (2014), many passengers buy their ticket close to the departure day or even on the day of departure (at DB more than 50% of bookings occur in the week before departure). This leads to two challenges: The time to react to unforeseen demand situations (lower or higher than predicted) is small. Therefore, the optimization technique should be able to adjust prices/quantities in reasonable time. In addition, forecasting techniques that include information of accumulated bookings on hand have only a small database during the first months of the booking horizon and, therefore, deviations from the forecasted demand pattern can only be observed at a late point in time.

Manual counting of passengers

It seems to be surprising but the evaluation of the realized demand (i.e., the actual number of passengers in a comfort class on a train) is a tough task in practice. Due to the missing check-in process or electronic counting on the train (e.g., at the doors) there is no automatic record of passengers at DB. Statistics of sold tickets can help to determine the number but especially passengers with flexible tickets or monthly/yearly passes have no (fixed) booking record and also no-shows cannot be observed. At DB, the train conductors carry out a manual count on predetermined legs on a train run. Obviously, manual processes are always a source of errors and, e.g., in overcrowded trains it is even impossible to count the passengers. Furthermore, not each leg is counted separately so that precise numbers for each leg are missing. A second problem of the manual counting lies in the manual postprocessing steps. Until early 2019, passenger counts at DB were documented manually on a piece of paper and entered manually in the IT-system. Therefore, it took some days until the data were available for the RM team and forecasting system. Today, train conductors can record the passenger counts electronically, but problems of different kind still occur in daily practice. This is a very crucial point in practice since a forecasting system is only as good as the data are.

Segmentation of customers

As in the case of airlines, customers of a railway company can be divided into different customer segments. Typical criteria for a segmentation are the travel purpose (e.g., business or leisure travelers) or socio-demographic factors (e.g., age). Since the booking behavior and the willingness to pay (WTP) of these customer groups differ, several price points/booking classes can help to skim off the WTP. Additionally, the possibility of adjusting the price or booking class availability decision during the booking horizon would make segment-specific pricing to some extent possible. From a modelling perspective, this means that a multi-fare multi-period model could cover both properties by taking varying price levels over the booking horizon into account.

Customers have preferred departure time

In the context of demand models (e.g., a discrete choice model) that incorporate external (other transportation modes) and internal alternatives (other itineraries, e.g., an itinerary one hour later or earlier), the fact that rail passengers typically have a preferred departure time becomes relevant. Therefore, it is advisable to divide the day into several time windows to better control the consideration set of a customer group. E.g., customers willing to travel in the morning only consider train connections departing in the morning. Additionally, one can increase the accuracy of the predicted choice behavior of a segment by considering that train connections with a larger deviation from the desired departure time have a lower utility than similar train connections that depart exactly within the preferred time frame.

Little demand history for unpopular O&Ds

Standard forecasting models are based on historical data to predict the future (e.g., time series regression, machine learning techniques). As with every statistical analysis, the quality of the forecast increases with the amount and quality of available data. Depending on the optimization model, leg based, O&D-market based (how many people travel between an O&D) or itinerary-based (how many people travel between two stations using train no. 123), different kind of data are necessary. The problem especially in the two latter cases is that data for unpopular O&Ds (e.g., between two small cities) is very limited and therefore robust forecasting difficult. One way to tackle that problem in practice would be using an O&D based approach only for popular O&Ds and an alternative approach (e.g., leg-based approach) for the remaining part of the network. This challenge is not unique for railway companies. Lufthansa Airlines announced in 2019 to go back to a leg-based forecasting system in some part of its system to avoid the problem of too small O&D demand values (and therefore high chance of forecasting errors) (Pölt 2019).

Low average load factor

Armstrong and Meissner (2010) mention that trains have typically a load factor <100% so that overbooking or cancellations do not have to be considered. On the one hand, data from DB confirm that finding with an average load factor of 56.1% in 2018 (Deutsche Bahn 2019). On the other hand, an average load factor for the whole network does not give any information about the number of single, highly utilized legs which is an important consideration in RM. Internal numbers from DB show that 4.1% of the legs (for the whole network on a standard Wednesday in late summer 2018) had a load factor of >95%. I.e., the chance that an itinerary uses a highly-utilized leg is given because an itinerary typically consists of many legs. Following, capacity consideration becomes (despite the low average load factor) relevant.

Fencing of fares

To avoid buy up or buy down behavior, airlines use fare rules to fence lower from higher fare classes like the popular Saturday night stay. I.e., cheaper booking classes are only available if the outward and return flight include a Saturday night stay at the destination to avoid bookings from price insensitive business travelers that are typically at home on weekends (Talluri and van Ryzin 2005). Since railway tickets are typically one-way tickets (approx. 76% of all travelers at DB) and the different booking classes at DB do not have different fare rules, this self-selecting customer segmentation is less given and the buy down problematic becomes highly relevant. Therefore, a stronger fencing of fares can also be useful for railway companies, but the rules must be adapted to the fact that the majority of sales are one-way tickets.

2.2.2.3 Business characteristics

As the last category of characteristics, properties that are closely linked to the business model of the railway company are presented. Some of them are self-imposed rules that can be defined from a strategic point of view.

Price consistency within the network

As no check-in (check-out) is required when boarding (leaving) trains at DB, strategically acting customers could buy a ticket for a longer than desired itinerary (i.e., with earlier and/or later departure/arrival station) if it is cheaper than the original itinerary. To avoid that, DB wants to ensure that price consistency holds between the itineraries so that a longer itinerary is always more expensive than a shorter one when the same train is used. This policy obviously reduces the flexibility in terms of pricing and can, therefore, lead to less revenue. An example where it could be beneficial to charge higher prices for a short itinerary compared to a longer itinerary is, e.g., if the short O&D-market has a high demand of business travelers with high WTP and the longer O&D has mostly leisure demand (with low WTP). It is therefore advisable to analyze the effect of adding the price consistency constraint vs. the potential loss due to strategic customers and derive a business decision based on the results.

Itinerary as a product bundle

An itinerary could be interpreted as a product bundle of multiple leg-products. For example, a long, direct itinerary from Munich (southern Germany) to Hamburg (northern Germany) can be split into multiple, partial itineraries without having the risk of missing a connecting train, since the passenger can just stay seated. This is also a reason why this consideration is more important in the railway than airline case, since in the latter one, splitting an itinerary into single tickets would result into a risk for the customer of missing the connecting flight without having any right for compensation. To avoid customers that buy multiple partial itineraries to rebuild their journey (with the aim to save money), one can add a constraint that makes sure that the sum of partial itineraries is always more expensive than the original itinerary. Since (especially long) itineraries can be represented by multiple different combinations of partial itineraries, this would lead to many additional constraints and probably strong limitations in terms of pricing flexibility so that it is again a business decision whether to consider this type of price constraint.

Flexibility in defined capacity limits

In contrast to airlines where the number of seats in an aircraft exactly determines the maximum number of passengers on the airplane, the capacity of a train is higher than the number of seats as it is also allowed to have standing passengers. Therefore, the capacity used in optimization models is determined rather by a service level than by a fixed, physical seating capacity (e.g., 80-120% of the number of seats).

Open system: Flexible tickets allow for entering any train

Some rail companies (such as DB) offer flexible tickets that allow for taking any train within a given timespan without doing any reservation, e.g., a flexible ticket for a single trip or even monthly or yearly passes for commuter traffic. At DB it is even possible to buy a ticket from the conductor on board of a long-distance train for an additional service fee. This has two implications. First, this is a big issue for the forecasting since passengers with flexible tickets have no recordings in the database about the train that is finally chosen and thus harder to predict than passengers with controlled tickets. Second, a large number of passengers with flexible tickets may choose one particular train, leading to overcrowding.

Discount loyalty card reduces realized revenue

Some rail companies (e.g., DB in Germany, SBB in Switzerland) sell discount loyalty cards that reduce every ticket by a certain percentage (e.g., 25% or 50% off each ticket). From a revenue perspective this means that the (optimized) price is reduced by certain percent so that the realized ticket revenue is smaller than the expected revenue in the optimization.

Fixed prices determined by government

While DB in Germany already has the possibility to adjust the price of a train connection during the booking horizon, this is not given in every country. In some countries (e.g., China) prices for tickets are restricted by the government and different price levels for a train connection are not possible. Therefore, revenue management models in that environment typically control the amount of tickets that are available for sale for different O&Ds that share the same resource (e.g., Bao et al. (2014), Jiang et al. (2015)).

High public awareness

In many countries, the railway companies are still government-owned. Especially in Germany, every change of the ticket pricing policy is critically discussed in public. The introduction of a new pricing scheme in 2002 (Link 2004, Brenck et al. 2003) or new fare rules in 2018 (Bild 2018) have experienced significant nationwide attention. The flexibility in pricing is therefore limited since not only technical/modelling hurdles have to be taken, but also (critical and sensitive) public reactions have to be considered.

Preliminary remarks

The analysis shows that the railway business has many characteristics that should be considered when discussing new RM methods and techniques. As mentioned above, most of the presented findings are based on practical experiences at DB Fernverkehr. Since the railway business in a country has typically historic roots and rail networks are partly separated from neighboring countries, the application and intensity of the above-mentioned characteristics varies from country to country. Table 1 summarizes the presented characteristics and gives an overview on the impact on forecasting and optimization. Although the overview shows a large number of properties that have an influence on railway RM, six very important aspects can be identified that a state-of-the-art model should focus on. These are: (1) network consideration, (2) integration of train connections with transfers, (3) multiple price points for each itinerary, (4) considering multiple booking periods, (5) covering the customers' switching behavior between train connections caused by the (ticket price) decision and (6) fulfilling price consistency between long and short itineraries. Based on these findings the question arises whether there already exist models in literature that cover the mentioned aspects and are applicable in railway business. Therefore, the literature overview in the following analyzes current railway RM models from both a practical and theoretical perspective.

Category	Property	Influence on... demand model	Influence on... optimization model
Structural characteristics	Challenge of defining set of itineraries	Network-based demand model requires information about forecasted itineraries	Preceding algorithm necessary that determines fixed set of itineraries to optimize
	Multiple transfers	For itineraries incl. regional trains, demand forecasting is difficult due to limited demand history	(1) Capacity constraint should be able to work with indirect train connections, (2) number of itineraries explodes if also regional trains are considered
	Many legs per itinerary		Strong network effects in case of bottleneck situations and thus a network model is favorable
Demand related characteristics	Many intermodal alternatives	Demand function should automatically react on price and quality changes of competitors	Quality and price levels of competitors influence own demand and optimal decisions
	Many rail alternatives	Demand for itinerary i is a function of multiple, alternative itineraries	Significant switching effects between itineraries caused by price decision can be anticipated by, e.g., integrating a discrete choice model
	Many changes in schedule	Forecasting should be able to immediately react on changes in the schedule	
	High #bookings close to departure	Forecasting has limited data base to learn from bookings on hand	Optimization should be fast to make frequent reoptimization possible
	Manual counting of passengers	Data base of historic, realized demand is error-prone	
	Segmentation of customers	Demand and WTP varies for different segments	Multiple price levels and decision based on booking time should be possible
	Preferred departure time	Desired departure time is a criterion for customer segmentation and thus segment-specific demand forecasting reasonable	Multiple segments increase complexity for optimization model
	Little demand history for unpopular O&Ds	Future demand is hard to predict accurately	
	Low average load factor		Most of the capacity constraints are not binding
	Fencing of fares		Strong buy-down behavior without appropriate fencing
Business characteristics	Price consistency		Can be ensured by, e.g., additional constraints
	Itinerary as a product bundle		Can be ensured by, e.g., additional constraints
	Flexibility in defined capacity limit		Overbooking is less relevant and violations of capacity constraints can be accepted within certain limits
	Open system with flexible tickets	Difficult to forecast since realized travel routes of passengers with flexible tickets are not recorded in database	Reduces available capacity to sell for controlled tickets
	Discount cards		Realized revenue is smaller than assumed in optimization
	Government influence		Additional constraints on RM-control can apply and less flexible for changes
	Public awareness		Limited flexibility in terms of price strategy adjustments

Table 1: Summary of properties for railway companies in the context of revenue management

2.3 Passenger railway literature review

Passenger railway RM is a small but active research field. The following literature overview is split into three parts: (1) quantity-based models (i.e., fares are fixed and decisions are made about which product/booking class to offer), (2) price-based models (i.e., price is the decision variable) and (3) railway RM studies with various focus.

At the end of this chapter, the literature is presented from two perspectives: First, an analysis with a theoretical focus and second, a practical perspective to check which of the practical requirements mentioned in Chapter 2.2.2 are already covered by the various articles.

2.3.1 Quantity-based railway revenue management

As a starting point, consider the standard (non-rail) specific deterministic linear programming model (*DLP*) for the quantity-based network case following, e.g., Williamson (1992, Ch. 4.1) or Talluri and van Ryzin (2005, Ch. 3.3.1),

$$max \sum_{i \in J} p_i y_i \tag{2.1}$$

$$s.t. \sum_{i \in J} \alpha_{il} \, y_i \leq c_l \quad \forall \, l \in L \tag{2.2}$$

$$0 \leq y_i \leq \mu_i \quad \forall \, i \in J \tag{2.3}$$

with $i \in J$ representing the set of products, p_i the price and μ_i the expected demand for product i, $\alpha_{il} \in \{0,1\}$ defining whether product i uses leg l ($\alpha_{il} = 1$) or not ($\alpha_{il} = 0$). Finally, the decision variable y_i defines the number of seats to assign to product i. Hence, the objective is to maximize total revenue under the consideration of capacity and demand limitations. Most of the following articles strongly rely on this basic network model and extend or adjust it to the considered setting.

Ciancimino et al. (1999) are one of the first researchers that analyze RM for rail companies from a scientific perspective. They present a multi-leg single-fare network model for RRM to find the optimal booking limit for each itinerary. I.e., they consider a fixed price for each itinerary and try to find the optimal availabilities for all (long and short) itineraries in the network. They present two approaches, a deterministic linear program and a probabilistic non-linear formulation. The deterministic version follows the standard *DLP* approach with the slight difference of stricter bounds on the number of assigned seats to the itineraries due to, e.g., a minimum number of tickets to assign for an O&D due to public interests. In the non-linear formulation, they adapt the objective function to integrate the probabilistic demand behavior by assuming normal distributed demand. In both formulations demand is assumed to be independent. Their numerical study shows that both models achieve revenue gains compared to a first-come-first-serve (FCFS) policy when being re-optimized during the booking horizon, while the probabilistic non-linear model outperforms the deterministic linear model. They also show that the advantage of an optimization model (over a FCFS policy) increases with the size of the rail network. From a computation time perspective, it can be shown that the non-linear model is (expectedly) slower than the linear model. While the linear model can be easily solved using standard software (e.g., CPLEX), they propose an own solution algorithm for the non-linear approach and show that it outperforms both tested solvers (minos and lancelot). Since the model considers only one fare class for each itinerary, i.e., decides how many tickets to assign (=booking limit) to each possible itinerary on a train running from A to B via many intermediate stations, it cannot account for multiple price levels for a specific train connection. You (2008) extends the probabilistic non-linear model of Ciancimino et al. (1999) to the two-fare class case (a full fare vs. discounted fare) and presents a solution approach to solve the probabilistic non-linear multi-leg two-fare problem for railway RM. The goal is to find the optimal booking limit for each fare-class-itinerary combination in a sequential rail network. As a further extension to Ciancimino et al. (1999), You (2008) takes overbooking into account by considering the tradeoff between additional revenue through overbooking and expected bumping costs. Further, he does not allow the number of full fare tickets to exceed the capacity, so that overbooking becomes only relevant for discount fare demand. Because of the probabilistic formulation, the model is a non-linear integer problem and the focus of the paper is on the development of a heuristic approach to solve the model efficiently (i.e., a good solution in reasonable time). The idea is to first solve the relaxed integer problem and use the optimal solution as a starting point for a particle swarm optimization algorithm. In terms of solution quality and time the approach is superior to the results of both tested non-linear solvers (Lingo and DICOPT).

Dutta and Ghosh (2012) provide insights into the business of an Asian railway company und present a deterministic network approach for optimizing seat allocation where the decision variable represents the number of tickets assigned to each itinerary in a network of multiple O&Ds. The model formulation is based on the above mentioned *DLP* and is extended in two ways: First, they consider an additional type of ticket (called *Tatkal*) that can be purchased shortly before departure by an extra fee (but the number of assigned seats to this ticket type is determined upfront). Second, they take revenues from cancellation fees into account. Since

fares are fixed for each itinerary in each cabin class, the model considers only a single fare. They test their model based on real data for one specific train and run a simulation study where various stochastic demand realizations are used as inputs to solve the deterministic model. Furthermore, they emphasize the importance of accurate demand forecasts and describe briefly how an EMSR (Expected Marginal Seat Revenue) based system could work for a rail company.
Bao et al. (2014) present a single-fare network seat allocation model for a Chinese railway provider that determines the optimal quantity of tickets to sell for each itinerary. They mention that in China ticket prices for each O&D are fixed and controlled by the government, so that the most effective control mechanism to increase revenue is the optimal assignment of tickets to different itineraries. They give detailed insights into the rail business in China and discuss different techniques to forecast demand. To optimize the seat allocation, they present three different methods. A non-nested booking limit strategy, a nested booking limit strategy and a bid-price strategy. The non-nested booking limit strategy again follows the classical deterministic linear problem as in *DLP*. For the nested booking limit setting, they propose a stepwise method that takes the dual variable of the capacity constraint of the standard *DLP* approach into account. Finally, the third proposed approach is a traditional bid price control, where the shadow prices of the capacity constraint of the *DLP* are used as bid price approximations. Hence, a ticket for an itinerary is offered if its fare is larger or equals the sum of the bid prices of all used resources (see, e.g., Talluri and van Ryzin (2005, Ch. 3.1.2.3)). Although they mention that the model is formulated for a single train, the problem formulation should also work for more complex networks including train connections with transfers. The conducted experiment shows that the non-nested booking control with daily updates performs best among all approaches.
Jiang et al. (2015) present a seat allocation model for a Chinese high-speed rail. Since fares are fixed for each O&D, they underline that they focus on how to assign tickets to different itineraries in peak seasons to different stations and thus follow the idea of Bao et al. (2014). This practical example shows that due to national regulations, RRM models often need to be tailored to the specific conditions in the operating country. Their deterministic model optimizes revenue by finding the optimal number of seats assigned to an itinerary in a comfort class (again based on the basic *DLP* formulation). In the case of limited capacity, they propose to first determine the so-called restricted number of stations. The limited set of stations makes sure that long-distance demand is prioritized over short distances. They propose to resolve the model within the booking horizon with updated demand forecasting data. The model is solved by

applying a genetic algorithm. In their numerical test they show that their approach increases revenue by 13.5% compared to the existing ticket assignment policy.

Zhai et al. (2016) present a deterministic single-fare network model that maximizes the total travel distance of passengers or number of passengers. Since they assume the prices to be a linear function of the travel distance, travel distance maximization equals revenue maximization. Compared to the capacity constraint as in *DLP* they formulate the resource consumption as a passenger flow, where for each station the remaining, boarding and leaving passengers are determined so that the utilization of the leg between two stations can be determined and restricted to the available capacity. The motivation for this complex representation is not mentioned. The model is tested with a medium sized network (solved using a branch-and-bound method provided in the LINGO software) with various crossing train lines. The results show that under passenger maximization, the number of short-haul ticket sales increases compared to the travel distance maximization case.

Compared to all aforementioned studies, Wang et al. (2016) reject the assumption of independent demand and formulate a choice-based stochastic seat-allocation model that applies a discrete random demand function. More precisely they present (1) a stochastic single-stage model where seat allocation is not changed during the selling horizon (e.g., optimized at the beginning of the selling horizon) and (2) a stochastic multi-stage model, that allows for an updated seat allocation during the selling horizon. Both stochastic formulations are then transformed to its deterministic representation and solved using CPLEX. Similar to Ciancimino et al. (1999) they consider the single-fare setting. Although they test the model only for a sequential rail network (i.e., only direct itineraries without transfers), the definition of a product and formulation of the capacity constraint should also work for indirect train connections. The applied multinomial logit model (MNL) includes the basic attributes for each product like travel time and comfort and, therefore, allows to represent the choice of customers based on product attributes. As only fixed product attributes are part of the customer choice model, it represents passengers' choice independently from the RM decision (i.e., the seat allocation). They propose to relax this assumption in future research. To test the proposed approach, they run a numerical test based on a Chinese rail line with two trains serving the same O&D and compare the results for two different demand scenarios (low and high demand). They show that the revenue benefit of a multi-stage compared to a single-stage model is small especially when the single-stage model is re-solved at the beginning of each booking period.

As a preliminary conclusion it can be stated that the presented quantity-based approaches are limited to the single-fare setting, with the exception of You (2008) and to a limited extent Dutta and Ghosh (2012). This can be explained by the practical orientation to the Chinese railway market which currently does not permit a multi-fare RM model (Bao et al. 2014, Jiang et al. 2015, Wang et al. 2016). Further, deterministic network approaches play a dominant role, that are mostly based on a basic capacity allocation model as in *DLP* and adjusted to local conditions.

2.3.2 Price-based railway revenue management

Compared to quantity-based RM with only a few publications for passenger rail transport, there is even less literature available for price-based railway RM approaches.

Sibdari et al. (2008) study the railway RM problem for the specific application of auto trains at the US company Amtrak. In this application, the RM problem becomes a product bundling problem instead of a (traditional) network problem since customers buy a ticket for both, themselves (coach seat or sleeper) and their car. Thus, two capacity buckets are required to make a product available. Their proposed approach is a single-leg multi-product multi-period problem that determines the optimal price for each of the three products in each booking period. They use stochastic dynamic programming to determine the optimal prices for the three products (coach seat, sleeper and car accommodation), where prices can be chosen from a discrete set. Following, the state space is defined as (m, n, s, t) where m, n and s represent the remaining capacity of car accommodation, coach seat and sleeper respectively and t the current planning period. Numerical results show that their approach yields higher revenues than the current (manual) strategy at Amtrak. Their approach also outperforms three other benchmark pricing policies. Since auto trains are a niche in the passenger railway business, the proposed model is hardly transferrable to the standard passenger rail setting.

Sato and Sawaki (2012) present a choice-based single-leg multi-period dynamic programming approach that uses price as decision variable. The passengers' choice is described by the multinomial logit model (MNL) and represents the product choice over a set of multiple intramodal (other train connections) and intermodal alternatives (e.g., air traffic, bus services). The approach also takes possible no-shows and cancellations into account. The focus is on modelling the choice of customers under competition and the analysis of optimal prices under different demand situations. Since the original formulation of the objective function with price as decision variable is non-concave, they use the common approach for MNL-based price optimization models of choosing choice probabilities as decision variable. Thus, they can solve a convex optimization problem. They show in their numerical study that price changes of the competitors influence the own price and that prices without competition (monopoly market) are higher than in the oligopoly case. They only consider single-leg itineraries, i.e., trains run from A to B without any intermediate stops so that the authors point out that the model should be extended to a more advanced network model.

Hetrakul and Cirillo (2014) use a MNL and latent class demand model to describe the customers' booking time decision and implement it into a deterministic multi-product multi-period network railway RM model that jointly considers the pricing and seat allocation decision (i.e., the price for each itinerary and the accepted fraction of demand). Although they jointly optimize price and accepted demand, the article is assigned to pricing models since the main focus of the contribution is on the pricing part. They use real data of a railway company to estimate the choice parameters for the booking time decision and demand data and run an experiment to evaluate the performance of the proposed approaches. They can show significant revenue improvements compared to the current approach used in practice. The considered network consists of multiple trains running from north to south during a day. Since the multiple trains are not connected in the model, they do not consider switching behavior between train connections. That is why they propose to incorporate more product attributes into the choice model and to allow alternative departures to represent demand shifts between connections. Furthermore, they propose to extend the model and analysis for bigger (hub-and-spoke) networks of realistic size.

Finally, Hohberger (2019a) presents a deterministic choice-based multi-period multi-product network dynamic pricing model for passenger railway RM. The model covers multiple effects from practice by integrating a discrete mixed logit demand function into the optimization model that reflects the product choice of heterogeneous customer segments over multiple alternative train connections. The network formulation makes it also able to account for network effects (e.g., between short- and long-haul travelers) and prices can be chosen from a continuous set. Additional rail specific constraints ensure two types of price consistency. First, with respect to longer or shorter itineraries and second, with respect to other comfort classes. Finally, a solution algorithm is proposed that makes it possible to solve even large-scale instances and it is shown that the new O&D-based model outperforms two benchmark models in terms of revenue (the current approach by DB Fernverkehr and an EMSR-b based method).

2.3.3 Railway revenue management studies with various focus

In addition to the presented studies that focus strongly on the model development for the standard passenger railway RM case, there are some interesting articles that provide insights into practice or consider different objectives.

Terabe and Ongprasert (2006) study the question whether revenue maximization is the right objective when talking about optimizing the seat allocation. They highlight that revenue maximization is a short-term decision that could potentially lead to lower customer satisfaction since revenue maximization could lead to high numbers of rejected customers. Further, from a political perspective it could be desirable to maximize the number of served customers (instead of revenue). Therefore, they propose and test three different approaches (maximizing revenue, maximizing the average passenger load factor, minimizing the number of rejected customers) against a FCFS policy that is still used by some rail companies. Their empirical study shows that the three objectives are not in contrast to each other. Rather, when optimizing one objective also the other two are optimized simultaneously in 70% of the cases and the other two objectives are at least improved in 92% of the cases (compared to the FCFS policy). They also show that revenue management techniques do not have any impact on revenue when trains are not fully occupied and prices are fixed.

Bharill and Rangaraj (2008) give practical insights into the pricing strategy of the Indian Rail Company (IR) and analyze internal booking data. At the time of the study, IR had a very simple fare structure and the analysis shows a high potential for intensified RM techniques combined with more advanced data analytics to detect possible sources for revenue gains.

Gopalakrishnan and Rangaraj (2010) also analyze IR and consider the problem of optimally assigning seats to different travel segments on a train serving multiple intermediate stations. As the authors mention, the proposed model is again based on the *DLP*, but the objective is different. Instead of revenue maximization, it minimizes the number of required seats so that the demand is fulfilled. Their numerical study shows that under the new linear model total revenue and load factor increase significantly, i.e., it creates a positive financial impact to IR (in the range of +2.6% to +29.3% revenue increase). It also increases the amount of confirmed bookings that will probably result in higher customer satisfaction.

Finally, Xiaoqiang et al. (2017) consider the case of group bookings which is a special case of railway RM with an increasing relevance due to the popularity of traveling in groups. Their approach (formulated using a Bellman equation) determines the discount rate for group tickets that is applied on the fixed price for the single ticket. Based on the real-world example of the itinerary Beijing-Shanghai (a single-leg network) they conduct a numerical study to illustrate their approach.

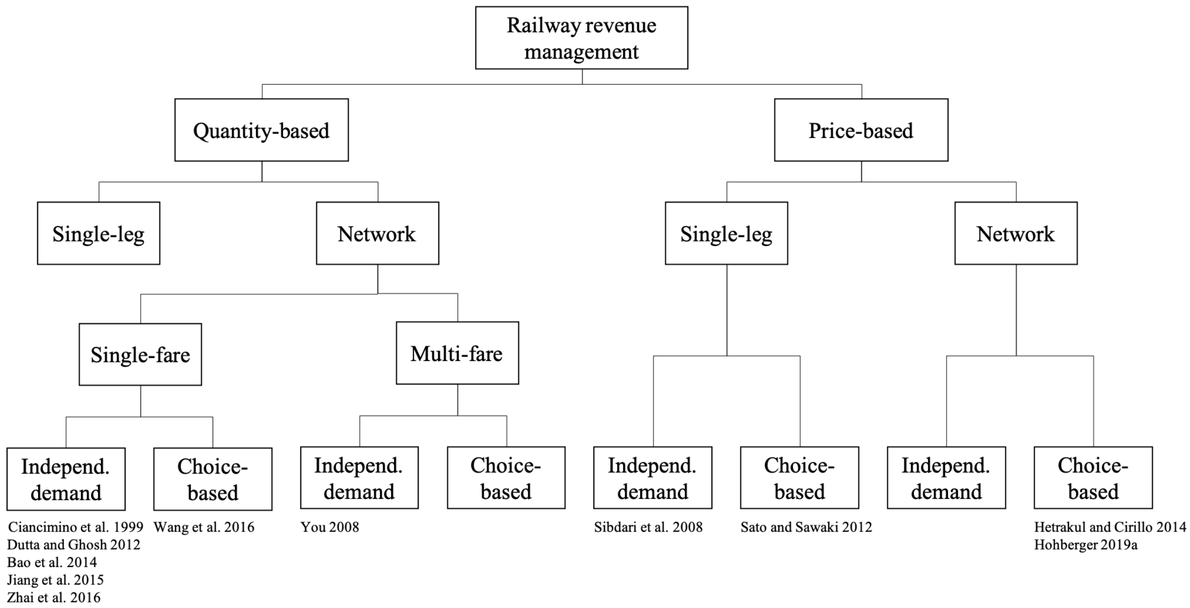

Figure 3: Classification tree of railway literature

2.3.4 Final remarks on literature overview

The overview in Figure 3 emphasizes once again the strong focus on network approaches which makes sense for the structure of a rail network. In addition, it can be seen that quantity-based approaches are in the majority, partly due to the lack of flexibility in pricing in the Chinese market where some of the research is coming from. There is also a slight trend towards models that use a choice-based demand model, especially more recent contributions.

Table 2 analyzes the articles from 2.3.1 and 2.3.2 from a practical perspective. The most important properties mentioned in Chapter 2.2.2 are used to check if the model formulations fulfill the practical requirements. The fields marked with n/a in the "price consistency" row indicate that price consistency does not play a role for these models by restricting them to a single-leg or single-fare. The numbers in brackets correspond to the listed and ordered articles in Table 3. Table 2 shows that, with the exception of Hohberger (2019a), none of the models are covering all six practical requirements.

Table 3 analyzes and summarizes the literature with a theoretical focus. For the classification of *Solution quality*, a heuristic solution is defined as an approach that finds a locally optimal solution and an exact approach to find a globally optimal solution of a proposed optimization model. This means, a deterministic optimization model with a globally optimal solution is defined to be an exact approach. This is important to highlight since *DLP* is often defined as an approximation for the original stochastic dynamic programming formulation (e.g., Talluri and van Ryzin (2005, Ch. 3.3)). For the categorization in *Formulation*, Dynamic Programming refers to models using a Bellman equation and static to standard mathematical optimization problems, e.g., as in *DLP*.

Finally, when comparing the survey results (in 2.2.1), where leg-based approaches are stated to be dominant in practice, with existing research, where network models are in the majority, a large gap between theory and practice can be observed. A possible explanation for this could be the lack of models that consider the most important attributes as seen in Table 2. A stronger cooperation between research and practice for railway RM would be a possibility to narrow the gap in future.

Property	Quantity-based railway RM							Price-based railway RM			
	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]	[9]	[10]	[11]
Network consideration	✓	✓	✓	✓	✓	✓	✓			✓	✓
Transfers possible	✓			✓			✓				✓
Multi-fare (>2 price points)								✓	✓	✓	✓
Multiple periods			✓	✓			✓	✓	✓	✓	✓
Switching between train connections caused by (ticket price) decision									✓		✓
Price consistency between long and short itineraries	n/a		n/a	n/a	n/a	n/a	n/a	n/a	n/a	✓	✓

Table 2: Literature in context of practical properties

				Optimization model												Test instances					
				Network structure		Fare structure		Demand model		Uncertainty		Formulation		Solution quality		Network size			Rail network		
		Author(s)	Year	Single-leg	Multi-leg	Single-fare	Multi-fare	Independent	Choice-based	Determi-nistic	Probabi-listic	Static	Dynamic Programming	Heuristic	Exact	Small	Medium	Large	Single train	Multiple (parallel) trains	Highly-connected network
Quantity-based	[1]	Ciancimino et al.	1999		✓	✓		✓		✓	✓	✓			✓	✓			✓		
	[2]	You	2008		✓		✓	✓			✓	✓		✓		✓			✓		
	[3]	Dutta and Ghosh	2012		✓	✓		✓		✓		✓			✓	✓			✓		
	[4]	Bao et al.	2014		✓	✓		✓		✓		✓		✓	✓	✓			✓		
	[5]	Jiang et al.	2015		✓	✓		✓		✓		✓		✓		✓			✓		
	[6]	Zhai et al.	2016		✓	✓		✓		✓		✓			✓		✓			✓	
	[7]	Wang et al.	2016		✓	✓			✓	✓		✓			✓	✓				✓	
Price-based	[8]	Sibdari et al.	2008	✓			✓	✓			✓		✓		✓	✓			✓		
	[9]	Sato and Sawaki	2012	✓			✓		✓		✓		✓		✓	✓				✓	
	[10]	Hetrakul and Cirillo	2014		✓		✓		✓	✓		✓		✓		✓				✓	
	[11]	Hohberger	2019a		✓		✓		✓	✓		✓		✓				✓			✓

Table 3: Literature overview on passenger railway RM

2.4 Conclusion

Railway revenue management is already widely recognized and applied in practice. The survey results have shown that leg-based approaches are still dominant at European rail companies and that managers see high potential for improved demand forecasting and optimization algorithms. To satisfy the market request for advanced railway RM models, an extensive overview on properties that should be considered when developing a state-of-the-art RM model was presented. The analysis showed that railway RM in general shares some aspects with airlines, which is in fact the most prominent industry for RM, but also has some special characteristics that make railway RM unique (e.g., price consistency, multiple legs per itinerary, typically one-way bookings). The information given is the result of a joint project with DB Fernverkehr. The following literature overview analyzed current research from two perspectives. First, a theoretical and mathematical focus with the result that quantity-based approaches are in the majority compared to price-based models and that deterministic network models are dominant. Bringing this together with the insights from the survey, a large gap between theory and practice can be noticed. Further, it can be observed that current research does not fulfill the stated requirements from practice. This is a possible reason for still less advanced RM models in practice and shows a high potential for collaborations between research and practice in future. Finally, especially recent contributions are mainly dealing with an application for an Asian passenger railway provider, showing the significant interest in this emerging market. All taken together, a high potential for the research area of railway RM can be observed since the demand for improved models from industry is rising and existing research is still underdeveloped.

Chapter 3 Continuous pricing in a capacitated network under mixed multinomial logit demand

First author:
Cornelia Schön
Chair of Service Operations Management, Business School, University of Mannheim

Second author:
Simon Hohberger
Chair of Service Operations Management, Business School, University of Mannheim

Abstract

The deterministic multi-product multi-resource pricing problem (DMMP) arises in many applications in practice where pricing decisions for multiple products should be optimized jointly and capacities are constrained. We study the DMMP under the discrete mixed multinomial logit model, i.e., a product is considered by multiple heterogeneous customer segments. Since the objective function and constraints of DMMP are non-linear and non-convex, we present various convex reformulations that serve as upper bounds incl. a piecewise approximation formulation that decreases the upper bound close to the globally optimal objective value of DMMP. Further, a heuristic approach with a focus on large-scale applications is presented and applied to a real-world revenue management case study of the German long-distance railway network with several thousand train connections per day.

3.1 Introduction

The deterministic multi-product multi-resource pricing (DMMP) problem arises in many applications in practice where pricing decisions for multiple products should be optimized jointly, in particular when products have cross-effects on demands of other products and/or different products share the same resources (Talluri and van Ryzin 2005). Direct applications are manifold, such as revenue management for airlines, railways, and hotels, assortment pricing in retailing or product line pricing in consumer goods industries. The objective is to find the profit-maximizing set of prices under the consideration of limited capacities. A recent research direction are pricing models that integrate a choice-based demand function to overcome the limitations of independent demand models used in previous studies. While the basic multinomial logit model (MNL) is already extensively analyzed, more advanced choice models such as the nested logit (NL) or the discrete mixed multinomial logit model (MMNL) that relax the IIA-assumption in the MNL have experienced increasing attention in the pricing literature.

3.1.1 Problem and motivation

In the following, we consider the deterministic multi-product multi-resource pricing (DMMP) problem with continuous prices under the MMNL choice model. More precisely, assume a seller wants to sell a given set J of products over a finite planning horizon with sales periods $t \in T$ and determine the profit-maximizing product prices p_{jt} $(j \in J, t \in T)$ within some predefined lower and upper bounds p_j^L and p_j^U, respectively. There is a set of resources R with capacities CAP_r $(r \in R)$ which are non-replenishable during the planning horizon and each resource $r \in R$ is required by a (sub)set $\dot{J}_r \subset J$ of products. Each unit sale of product j incurs a variable cost of c_j. The total market potential in period t is λ_t and the share that product j attracts in period t if offered at price $\boldsymbol{p_t} = \left(p_{jt}\right)_{j \in J}$ is given by a function $\pi_{jt}(\boldsymbol{p_t})$ which will be specified shortly for the MMNL model.

Then, the DMMP problem can be formulated as follows:

$$Max \sum_{t \in T} \sum_{j \in J} \lambda_t \left(p_{jt} - c_j\right) \pi_{jt}(\boldsymbol{p_t}) \tag{3.1}$$

$$s.t. \sum_{t \in T} \sum_{j \in \dot{J}_r} \lambda_t \pi_{jt}(\boldsymbol{p_t}) \leq CAP_r \quad \forall\, r \in R \tag{3.2}$$

$$p_j^L \leq p_{jt} \leq p_j^U \quad \forall\, j \in J, t \in T \tag{3.3}$$

The deterministic problem formulation plays an important role as an approximation of the stochastic version of the problem and may be used for developing powerful heuristics (Gallego and van Ryzin 1997, Maglaras and Meissner 2006). The corresponding problem instances of the DMMP in practice are typically of large-scale size. For example, a railway service provider sells different fare products, i.e., scheduled itineraries from the customer's origin to the desired destination at defined fare class conditions; each sale of a fare product requires capacity of one or multiple scheduled train legs also utilized as resources by other fare products; and the price of a second class ticket for the train from Frankfurt to Berlin at 10 a.m. on a selected day will probably impact first class demand on the same train as well as second class ticket demand for the next train from Frankfurt to Berlin two hours later – a cross-effect which should be properly captured by the demand model. To get an estimate of a possible problem size, assume a medium railway network with 500 origin-destination (O&D) pairs, 5 connections per day per O&D pair on average and two comfort classes per connection. Then, there would be 5,000 fare products for each day of departure that need to be dynamically priced during the reservation period prior to departure. With hundreds of scheduled trains in the national long-distance network per day, Deutsche Bahn (DB) serves around 410,000 customers daily (Deutsche Bahn 2019). Based on publicly available schedule data of DB, Hohberger and Schön (2019) estimate algorithmically that DB offers tens of thousands of fare products on each day in the future in the national long-distance railway network that must be priced accordingly.

In face of the large size of problem instances in practice, efficient solution techniques are required. However, the tractability of the DMMP problem largely depends on the assumed choice model $\pi_{jt}(\boldsymbol{p_t})$ that is incorporated into the optimization problem. Most commonly used is the standard MNL choice model. While the optimization problem under the logit function is tractable, the MNL model suffers from well-known limitations such as the assumption of a) homogeneous choice behavior across the potential customer base and b) independence of irrelevant alternatives (IIA) (see below for a more in-depth discussion). To overcome the limitations of the standard MNL model, recent approaches have also incorporated more advanced choice models, such as the nested logit (addressing IIA) and the MMNL model (allowing for heterogeneity and thereby also resolving IIA, at least partially).

The MMNL model is considered to be one of the most powerful discrete choice models (Hensher and Greene 2003). However, the DMMP problem under MMNL choice is hard to solve, with non-convex nonlinearities included in both objective function and capacity constraints. Despite its theoretical and practical relevance, the MMNL model and its incorporation into the DMMP problem has only received scant attention in the dynamic pricing literature, probably due to the resulting complexity of the optimization. How to efficiently solve the continuous pricing problem under MMNL choice to optimality is still an open problem (Keller et al. 2014) and we want to contribute to narrow this gap. In particular, we develop a convex mixed-integer programming approach that approximates the original problem and allows to determine near-optimal solutions and an upper bound close-to-optimum of the original problem. Before we position our contribution in more detail, we review the related work and the current state-of-the-art in multi-product multi-resource pricing under choice behavior.

3.1.2 Related work

Excellent reviews on dynamic pricing and capacity allocation for revenue management are provided, e.g., in Bitran and Caldentey (2003) and Strauss et al. (2018). In their seminal work on multi-product multi-resource dynamic pricing, Gallego and van Ryzin (1997) develop and analyze a deterministic and a stochastic version of the problem under some general model for the demand rate. The multi-product demand rate function is required to satisfy some regularity conditions such as the existence of a so-called null price that turns off demand, invertibility in the price vector, and a resulting revenue function that is continuous, bounded and concave in demand. This is the case, e.g., for the log-linear demand function used exemplary in Gallego and van Ryzin (1997). Considering demand rather than price as the decision variable has also played an important role in multi-product multi-resource dynamic pricing problems with choice behavior. We will elaborate this in the following, and to maintain brevity and focus, concentrate our review to multi-product pricing under the MNL, the NL, and the MMNL choice model. Furthermore, we assume cross-effects of demand between periods are neglectable and, for the sake of simplicity, introduce choice model specifications $\pi_j(\boldsymbol{p})$ for a single period, suppressing the time index t for price variables and choice parameters.

Pricing under the MNL choice model

The most common approach to capture cross effects in demand has been to assume homogeneous choice behavior and incorporate the standard MNL choice model into a (deterministic or stochastic) multi-product multi-resource pricing problem. In case of the MNL model, the probability that product $j \in J$ is chosen is given by (see, e.g., McFadden (1974), Anderson et al. (1992)):

$$\pi_j(\boldsymbol{p}) = \frac{e^{\alpha_j+\beta_j p_j}}{A_0 + \sum_{n \in J} e^{\alpha_n+\beta_n p_n}} \quad (j \in J) \tag{3.4}$$

with parameters $\alpha_j \in \mathbb{R}$ and A_0, $\beta_j < 0$ to be estimated. The parameter A_0 specifies the attraction of consumers to the outside option to buy a competitor product or not to buy at all.

Recent references discussing pricing under the MNL include, among many others, Song and Xue (2007), Dong et al. (2009), Akcay et al. (2010), Keller et al. (2014). The standard logit profit function is known to exhibit nonlinearities that are not pseudo-concave in price (Hanson and Martin 1996). However, if price coefficients are uniform for all products, then the logit profit function is unimodal with respect to the markups (prices minus costs) and the optimal markups are constant for the products. Therefore, the pricing problem in this case can be easily solved by the first-order optimality conditions (see, e.g., Akcay et al. (2010)).
Furthermore, also for the general case of product-specific parameters, the profit function is known to be concave when considering demand or market shares rather than prices as decision variables. In case of a homogeneous group of customers, there is a linearizable one-to-one mapping between product prices and MNL choice probabilities such that the resulting optimization problem is convex in demand and can be solved efficiently. Concavity of the (expected) revenue function in demand under the MNL model was first shown by Song and Xue (2007) and Dong et al. (2009) under the assumption that the MNL price sensitivities β_j $(j \in J)$ are identical and equal to one across products. Li and Huh (2011) extend this result and prove that the revenue function under MNL choice remains concave in demand as long as price sensitivity parameters are uniform, but not necessarily equal to one. Schön (2010a) and Keller et al. (2014) consider own product shares as decision variables and introduce an additional variable related to the probability of making no sale. With this auxiliary variable, they show that the revenue function is concave under the general setting where price sensitivities are product-specific.
Hausman and Chen (2000) analyze pricing under the aggregate MNL choice model in the context of optimal product line selection with the objective to maximize total profit. In their case, prices are considered to be discrete and can be chosen from a given discrete price set. The decision of the price level for each product is represented by binary decision variables. Capacity constraints are not considered. They show that the problem can be solved efficiently by considering the relaxed problem (i.e., dropping the binary constraint) and applying standard linear or fractional programming solution methods.

In summary, while the pricing problem under logit choice behavior exhibits structural properties that can be well exploited for its efficient solution, the aggregate MNL model fails to account for the empirically more realistic case where customers have heterogeneous preferences. Furthermore, the choice model suffers from the well-known IIA property, i.e., the share of any product which is removed from the offered set is allocated among the remaining products proportional to their original choice probabilities (McFadden 1974). The IIA assumption is quite strong since it is often violated in empirical data in practice, in particular if a consideration set contains a subset of alternatives that are closer substitutes than those outside the subset. To overcome the limitations of the standard MNL, recent approaches have also incorporated more advanced choice models, in particular, the NL and the MMNL model.

Pricing under the NL choice model

The general idea of a nested choice model is that a consumer's decision is a hierarchical process with two or more stages, first selecting the nest (e.g., the type of accommodation such as hotel vs. apartment) and in a second step choosing an alternative within the selected nest (e.g., hotel 1 vs. hotel 2). With a NL model, the IIA assumption can be relaxed by allowing the unobserved factors to be correlated over alternatives. As a result, the NL model exhibits the IIA property only for alternatives within the same nest but not for choices across nests, and more flexible substitution patterns can be captured. The standard MNL model is included as a special case.

Research on pricing under the NL choice model mainly differs in the number of possible stages/levels, considered constraints and the flexibility of nest coefficients and price betas. Most of the models so far assume homogeneous preferences and thus focus on a single segment NL model, such as Li and Huh (2011) and Gallego and Wang (2014).

Li and Huh (2011) consider a multi-product pricing problem with a single-segment NL model. They show that the objective function is jointly concave in market shares if the price betas are equal for all products within each nest (but can vary over nests) and the nest coefficients are between zero and one.

Gallego and Wang (2014) show that this property is lost for the general case when the price sensitivity parameters are allowed to be different for each product or the nest coefficients are >1. In particular, the authors provide an example with two products, product-differentiated price betas and a nest coefficient of 0.1 that shows that the profit function is not jointly concave in market shares. Nevertheless, the authors demonstrate how to still exploit the mathematical structure of the problem by introducing the notion of an adjusted markup (i.e., price minus cost minus the reciprocal of the product's price beta). They show for the uncapacitated problem that the optimal adjusted markup is constant for all products within a nest. Furthermore, the adjusted average markup for all products in the same nest is also constant for all nests. Accordingly, the optimization problem can be reduced to a single variable with an objective function that is shown to be unimodal under mild conditions and therefore efficiently solvable.

Li et al. (2015) consider both the product selection (assortment) problem and pricing problem under the NL model with multiple stages (d-nested logit) and present a novel algorithm for the pricing problem that solves it locally optimal and much faster than a traditional gradient ascent.

Rayfield et al. (2015) study the uncapacitated pricing problem under the nested logit model in two variants, i.e., when the set of products is fixed and when the set of offered products is jointly determined with the prices. Additionally, both models include the possibility to define lower and upper price bounds. The authors present an approximation method to solve the problems with a performance guarantee and show that it can solve the pricing problem in reasonable time for medium-sized instances.

More recently, Davis et al. (2017) present an algorithm to solve the uncapacitated pricing problem with two types of quality consistency constraints (within a nest or between nests), where prices are chosen from a discrete set.

We note that all of the presented studies for nested logit models ignore preference heterogeneity. As mentioned by, e.g., Gallego and Wang (2014), choice models with multiple heterogeneous customer groups (such as the MMNL) that differ in the evaluation of product attributes are a promising research area.

Pricing under the MMNL choice model

Mixed multinomial logit models are an alternative extension to account for heterogeneous cross-effects in demand and are considered to be among the most powerful state-of-the-art discrete choice models (Hensher and Greene 2003). Here, a consumer's choice probability is a mixture of logits, allowing the parameter vector of the logit specification to be randomly distributed across the population according to some mixing distribution. McFadden and Train (2000) have shown that a mixed logit model can approximate any random utility choice model arbitrarily close by an appropriate specification of the mixing distribution.

The mixing distribution can be continuous or discrete. For the continuous case, a parametric distributional form has to be assumed, e.g., a normal or a lognormal distribution. Then, the parameters are estimated for the assumed specification (Train 2008). A mixture model with a discrete mixing distribution, unobserved segment membership, and observed choices following a multinomial logit distribution within each segment is also called a latent class (LC) model (see Wedel and Kamakura (2012, p. 77)).

In the LC model, the observed choices in a sample are assumed to stem from an unknown number of discrete segments with uncertain segment size. Though the segments are not directly observable, they can be recovered from the data by "unmixing" the sample through the estimation of a finite mixture model (see Wedel and Kamakura (2012, Ch. 6)). Accordingly, the LC model is based on a post hoc segmentation where the number of segments, their weights, and their choice parameters are estimated jointly from the data (see Wedel and Kamakura (2012, Figure 3.1, p. 17)).

Denote the set of consumer segments by I and assume that each segment $i \in I$ has a relative weight $\widetilde{\omega}_i$ and a consideration set $\tilde{J}_i \subset J$ of products that are considered to be alternative options to buy from the seller. Further, let $\tilde{I}_j$ be the set of segments $i \in I$ that have included product $j \in J$ in their consideration set, i.e., $\tilde{I}_j \coloneqq \{i \in I: j \in \tilde{J}_i\}$. Then, the probability that product $j \in J$ is chosen according to the discrete MMNL choice model is given by

$$\pi_j(\mathbf{p}) = \sum_{i \in \tilde{I}_j} \widetilde{\omega}_i \frac{e^{\alpha_{ij} + \beta_{ij} p_j}}{A_{i0} + \sum_{n \in \tilde{J}_i} e^{\alpha_{in} + \beta_{in} p_n}} \quad (j \in \mathrm{J}) \tag{3.5}$$

with segment-specific choice parameters $\alpha_{ij} \in \mathbb{R}$, $\beta_{ij} < 0$ $(i \in I, j \in \tilde{J}_i)$ and outside option value $A_{i0} > 0$ $(i \in I)$ to be estimated. Thus, conditional on segment membership, the probability that a consumer from segment $i \in I$ chooses product $j \in J$ is the standard MNL formula. The unconditional likelihood that product $j \in J$ is chosen is defined as a weighted mixture of the segment-level probabilities, where the weights $\widetilde{\omega}_i$ correspond to the relative size of each segment i (Wedel and Kamakura 2012, Ch. 6).

While the IIA property is still present within each segment of a latent class model (Wedel and Kamakura 2012, p. 289), the aggregate LC probability that an alternative is chosen will usually not exhibit the IIA property, i.e., the LC model may resolve the IIA limitation of the standard aggregate MNL choice model to a large extent by accounting for preference heterogeneity. In fact, Gensch and Ghose (1997) show that in face of preference heterogeneity across individuals, it is virtually impossible to meet the IIA assumption at the aggregate market share level. Thus, the authors argue that observed violations of the IIA property at the aggregate could stem primarily from preference heterogeneity rather than IIA violations at the individual level. To

reduce this aggregation bias in face of preference heterogeneity, Gensch and Ghose (1997) recommend segmentation and LC models.
We note that some authors define the term mixed logit in a narrow sense where the mixing distribution is restricted to be continuous and discrete distributions are not considered as part of this class. Other authors (e.g., McFadden and Train (2000), Train (2008), and us) consider the distinction rather arbitrary and include any mixing distribution in the definition of mixed logit, in particular since any continuous distribution can be approximated by a discrete one with numerous support points. Likewise, the accuracy of the approximation of a latent class model to the "true" distribution may be increased by increasing the number of parameters (segments), as long as the computational effort of the maximum likelihood estimation is not prohibitive. In fact, the discrete mixed logit model is less restrictive with regard to the assumptions of the mixing distribution than the continuous one (Train 2008).
We follow this more conclusive definition of the term mixed logit. In particular, we include a discrete mixed logit model into the pricing problem in order to capture customers of multiple types that choose according to different logit models. Our definition of discrete mixed logit includes LC models based on a post-hoc segmentation with latent classes (unobserved heterogeneity) as well as multi-segment MNL models based on an a-priori segmentation (Wedel and Kamakura 2012, p. 17) where the type and number of segments are observable and determined in advance (observed heterogeneity). Furthermore, we may also include mixed logit models that are combinations of a-priori and post-hoc discrete segments, e.g., when the population is split into different observable markets and each market can be further split into latent classes with different preferences.

MMNL models have received increasing attention in the empirical literature for estimating travel choice behavior, e.g., for railways and airlines (see, e.g., Warburg et al. (2006), Carrier (2008), Hetrakul and Cirillo (2013), Hetrakul and Cirillo (2014)). But despite its empirical relevance in theory and practice, the MMNL model and its incorporation into the DMMP problem has only received scant attention in the prescriptive dynamic pricing literature (e.g., Keller et al. (2014), Li et al. (2018)), probably due to the resulting complexity of the optimization.

With a particular focus on railway revenue management, Hetrakul and Cirillo (2014) estimate multinomial logit and latent class choice models to describe the ticket purchase timing decision of railway passengers as a function of price and other attributes. The demand functions are then incorporated into a revenue management optimization model that jointly determines prices and seat allocations for a single train. The non-linear optimization problem is not further analyzed, but solved with a commercial solver so that only local solutions can be expected. For future research in railway RM, they propose to include the choice-based demand model into a more complex railway network structure and the consideration of trips with station transfers. Additionally, they suggest optimizing revenue over multiple departures to account for demand shifts between train connections.

Keller et al. (2014) consider the multi-product multi-resource problem with continuous prices under attraction choice models (e.g., the MNL and MCI demand model). They provide an example that the objective function is non-concave in price and only locally optimal solutions can be expected. To globally solve the optimization problem for a single customer segment, they use the convex reformulation based on inverse attraction functions where demand rather than price is the decision variable. For the more general case of multiple overlapping customer segments, the authors suggest replacing the multi-segment demand model (e.g., the MMNL model) by an approximate attraction model that renders the resulting optimization problem convex and thus tractable. Theoretical bounds on the maximum error of the approximate demand model with respect to the true multi-segment demand model are provided. With regard to the optimization, a numerical experiment shows that a solution of the approximate problem is computed quickly and yields objective function values close to the (in general local) optimum of the original non-convex problem formulation. However, since the global optimum of the original problem is not known, there is neither a theoretical performance guarantee nor an experimental assessment of the solution quality.

Li et al. (2018) study the continuous pricing problem under the discrete MMNL model as we do. However, capacity constraints are not considered. The authors consider a transformed problem formulation where the total profit under multi-segment MMNL choice is expressed as a function of segment 1 demand only (rather than of prices as decision variables). They show that the transformed objective function is not necessarily concave (in segment 1 demand) under the MMNL model and that the optimal markup is in general not equal across products (as is in the basic MNL), even with symmetric price sensitivities across all products and all segments. Only if also the variation of the price-independent attraction parameter is sufficiently small across segments, the objective function is proven to be concave in segment 1 demand. For this case, the authors propose an efficient bisection search algorithm (including convex feasibility subproblems) for globally solving the problem under the MMNL model. For the more general case with asymmetric choice parameters and non-quasi-concave profit function, a gradientdescent method with multiple random starting points is proposed to search for stationary point solutions.

Zhang and Lu (2013) consider a dynamic stochastic network pricing problem under a multi-segment MNL choice model with uniform price betas (equal to one across products and segments) and disjoint consideration sets. To obtain approximate dynamic pricing policies, the authors suggest a resource decomposition approach that uses dual values from a deterministic approximation model. Similar to the DMMP problem under MNL choice, the problem with multiple segments but disjoint consideration sets can be transformed into a convex programming problem by considering demand rather than price as the control variable. Since every product is only considered by one single segment, the transformed formulation represents a one-to-one mapping between demand and prices and can therefore be used to derive a globally optimal solution to the problem. The authors propose an Augmented Lagrangian method to efficiently solve the convex programming problem. Finally, a numerical study shows that the dynamic pricing policies from their network decomposition approach can achieve substantial revenue improvements compared to dynamic availability controls, deterministic static pricing and deterministic pricing with re-optimization, although the gap to the latter is overall smaller.

Schön (2010a) shows that the convex problem structure of the multi-product multi-resource continuous pricing problem under a discrete mixed logit choice with heterogeneous price sensitivity parameters can still be maintained if it is feasible to simultaneously quote individual prices for the same product to each segment according to a first- or third-degree price discrimination. However, this requires the capability to identify a priori which customer segment an incoming sales request belongs to. In the more common case considered here, where the same product is offered at a uniform price to all customer requests occurring at the same time, the convexity property is lost, since non-convex constraints need to be introduced to ensure price consistency across segments with overlapping consideration sets. Accordingly, the DMMP problem under MMNL choice is non-linear non-convex and thus difficult to solve in general.

While the MMNL choice model has been less considered in the dynamic pricing literature so far, it has received increasing attention in the related field of product assortment (PA) optimization (see, e.g. Schön (2010b), Mendez-Diaz et al. (2014), Feldman and Topaloglu (2015), Kunnumkal (2015)). The PA problem is to determine which products should be included in an assortment and how to price them such that total profit is maximized. The PA problem is somewhat related to the DMMP problem considered here. However, prices are usually treated as discrete decision variables (including the null price) if not predetermined.

Capacity constraints may or may not be included. The non-linear PA problem has often been reformulated as a mixed-integer linear program through Big-M or McCormick linearizations of bilinear mixed-binary terms (e.g., Sen et al. (2018)). It has been shown to be NP-hard for the MMNL model even in the uncapacitated case (Rusmevichientong et al. 2014, Désir et al. 2014). Therefore, much work has been focused on deriving upper bounds and efficient approximations.
As an exception, Sen et al. (2018) consider the PA problem with shelf-space constraints and predetermined prices and propose a conic mixed-integer reformulation of the problem. Combining the conic formulation with the traditional McCormick constraints, the authors are able to exactly solve different problem instances with 500 products and 50 customer classes with CPLEX in less than 2 minutes on average.

In summary, multiple restrictive assumptions have been made to efficiently solve the DMMP problem under MMNL choice, either with restrictions for the choice parameters, with regard to pricing policies (allowed to be non-uniform) or the consideration set. Table 4 summarizes the few approaches that consider the general case of a discrete mixed logit model with heterogeneous price sensitivity and uniform pricing, and compares them with regard to solution quality and run time performance reported from their numerical experiments.

Reference	Constraints		Solution quality	Test instance
Continuous pricing under MMNL choice				
Li et al. (2018)	Cap. constraints	✗	▪ Locally optimal for general case ▪ Globally optimal for case with high symmetry across segments	|J|= 4,096 (products) |I|= 16 (segments) Sol. time = 54s
	Price bounds	✗		
Keller et al. (2014)	Cap. constraints	✓	▪ Locally near optimal for general case	|J|= 3 |I|= 7 Sol. time: n/a
	Price bounds	✓		
Schön and Hohberger (2019)	Cap. constraints	✓	▪ Globally near-optimal for general case ▪ Locally optimal heuristic for large-scale instances	1) |J|=50 and |I|=125 Sol. time: 350s, gap: <1.5%[1] 2) |J|=40,012 and |I|=72,240 Sol. time: 1,328s
	Price bounds	✓		
Product assortment under MMNL choice				
Sen et al. (2018)	Cap. constraints[2]	(✓)	▪ Globally optimal for general case	|J|= 500 |I|= 50 Sol. time: 119s
	Price bounds	n/a		

Table 4: Overview on pricing & product assortment under MMNL choice

[1] Numbers show base case results; refer to Table 5 for more detailed results.
[2] Sen et al. (2018) consider shelf-space as capacity constraint.

3.1.3 Contribution and outline

How to efficiently solve the continuous pricing problem with multiple segments to optimality is still an open problem and we want to contribute to narrow this gap. Our contributions are as follows:

a. First, we analyze the DMMP problem with continuous prices and price consistency constraints under the MMNL choice model in detail with regard to its mathematical structure.
b. We present an approximate optimization problem to derive an upper bound on the optimal profit and to determine heuristic solutions. The approximate problem is convex and can therefore be solved efficiently even for large problem instances. An experimental study shows that the approach is very promising with regard to run time performance and solution quality.
c. We present a convex mixed-integer programming approach that allows to tighten the upper bound close-to-optimum and to determine near-optimal solutions of the original problem. To our knowledge, this is the first approach to approximately tackle the problem under the MMNL choice model. In our experiments, we are able to approximately solve medium-sized problem instances in reasonable time.
d. The suggested dynamic pricing approach is applied to a real-world revenue management case study of the German long-distance railway network with several thousand train connections per day.

The paper is structured as follows: Chapter 3.2 discusses the original non-convex model formulation of DMMP and presents various approximate, convex reformulations and solution techniques with a focus on upper bound models. Subsequently, Chapter 3.3 presents two heuristic procedures that serve to find a good lower bound of the original non-convex model. A numerical study in Chapter 3.4 tests the performance and solution quality of the proposed models and solution approaches. Finally, Chapter 3.5 summarizes the findings and gives an outlook for future research directions.

3.2 Model development

In this section, we present and discuss the DMMP problem formulation. Chapter 3.2.1 is concerned with the original problem formulation with prices as decision variables, while Chapter 3.2.2 presents an equivalent problem formulation with inverse price demand relationship. In Chapter 3.2.3, different convex approximations are developed that provide both upper bounds on the optimal objective function value for the original problem and good starting points for solving the original problem at least locally. For convenience, we focus on a single period only to discuss the structural properties. The extension to the multi-period setting is straightforward. Further, as common for RM models we assume variable costs c_j to be zero and therefore drop the parameter in following.

3.2.1 Original model formulation (P0)

Under the discrete MMNL choice model (3.5), the DMMP problem looks as follows:

$$Max \sum_{i \in I} \frac{\omega_i \sum_{j \in \tilde{J}_i} p_j e^{\alpha_{ij} + \beta_{ij} p_j}}{A_{i0} + \sum_{n \in \tilde{J}_i} e^{\alpha_{in} + \beta_{in} p_n}} \tag{3.6}$$

$$\sum_{i \in I} \frac{\omega_i \sum_{j \in \ddot{J}_{ir}} e^{\alpha_{ij} + \beta_{ij} p_j}}{A_{i0} + \sum_{n \in \tilde{J}_i} e^{\alpha_{in} + \beta_{in} p_n}} \leq CAP_r \quad (r \in R) \tag{3.7}$$

$$p_j^L \leq p_j \leq p_j^U \quad (j \in J) \tag{3.8}$$

We denote this problem formulation, where prices are decision variables, as $\mathbf{P0}$. Let $j \in \ddot{J}_{ir}$ denote the set of products that are considered by segment $i \in I$ and use resource $r \in R$. Further, let ω_i be the absolute size of segment $i \in I$.

In general, both the objective function and the capacity constraints can exhibit nonlinearities that are not quasi-convex in the price vector. An example of a non-convex problem instance with one segment and two products is provided in Keller et al. (2014). To better understand the mathematical structure of the objective function, we note that it is the sum of $\sum_{i \in I} |\tilde{J}_i|$ non-linear ratios $f_{ij}(\mathbf{p})/g_i(\mathbf{p})$ $(i \in I, j \in \tilde{J}_i)$, which is one of the most difficult fractional programs (Schaible and Shi 2003). Here, the denominator $g_i(\mathbf{p}) :=\ A_{i0} + \sum_{n \in \tilde{J}_i} e^{\alpha_{in} + \beta_{in} p_n}$ is convex in $\mathbf{p} \geq \mathbf{0}$ and the numerator $f_{ij}(\mathbf{p}) := \omega_i p_j e^{\alpha_{ij} + \beta_{ij} p_j}$ is quasi-concave in $\mathbf{p} \geq \mathbf{0}$ (see Figure 4) with an inflection point in $\bar{\bar{p}}_{ij} := -2/\beta_{ij}$ where f_{ij} changes from being concave to convex.

Since the objective function is the sum of quasiconcave-convex ratios, we could exploit the structure by approximately guessing the value of each numerator and each denominator for an optimal solution and solving a parameterized convex feasibility problem for each guess (similar to, e.g., Désir et al. (2014)). However, with $m := \sum_{i \in I} |\tilde{J}_i|$ ratios to combine, the approach will be prohibitive unless the number of ratios is very small. Furthermore, how to handle the nonlinearities in the capacity constraints remains an open question.

The structure has some interesting implications if product prices can be restricted to not exceed $p_j^{max} := \min_{i \in \tilde{I}_j} \bar{\bar{p}}_{ij}$ such that f_{ij} and thus $\sum_{j \in \tilde{J}_i} f_{ij}$ is concave. It is well known that the ratio of a nonnegative concave and a positive convex function is known to be pseudo-concave in the differentiable case (Avriel et al. 1988). Thus, each segment-specific ratio

$$h_i(\mathbf{p}) = \frac{\sum_{j \in \tilde{J}_i} f_{ij}(\mathbf{p})}{g_i(\mathbf{p})} = \frac{\omega_i \sum_{j \in \tilde{J}_i} p_j e^{\alpha_{ij} + \beta_{ij} p_j}}{A_{i0} + \sum_{n \in \tilde{J}_i} e^{\alpha_{in} + \beta_{in} p_n}} \quad (i \in I) \tag{3.9}$$

of the objective function is pseudo-concave on the set $\tilde{S} := \{\mathbf{p} \geq \mathbf{0} : p_j \in [0, p_j^{max}]\}$. Accordingly, a problem instance with a single segment, no capacity constraints and prices restricted to $\tilde{S}$ corresponds to a fractional program with pseudo-concave objective function and linear (bound) constraints. In this case, the KKT conditions are sufficient for a global optimum such that the problem can be efficiently solved, e.g., by transforming it into a program that is convex in price (see, e.g., Schaible (1976)).

Unfortunately, with multiple segments, the pseudo-concavity of the objective function may be lost, as the properties of concave-convex single-ratio fractional programs do not generally extend to the sum of concave-convex ratios. Freund and Jarre (2001) show that the sum-of-concave-convex-ratios-problem is NP-complete. The difficulty of the problem mainly arises from the number of ratios, i.e., segments. For a small number of ratios, different solution methods such as branch-and-bound algorithms can be used (see, e.g., Benson (2002), Schaible and Shi (2003), Gao and Shi (2010)). Alternatively, if the number of segments is small, the uncapacitated objective function could be maximized on the restricted set $\tilde{S}$ by parameterizing ratios and solving convex feasibility problems at selected parameter values from a set of appropriately chosen grid points (Boyd and Vandenberghe 2004, Section 4.2.5).

However, approaches that assume a small number of segments do not seem promising for our purposes. First of all, with different O&D markets and heterogeneous preferences for travel time and purpose, the number of segments is usually rather large. Second, the restriction $\mathbf{p} \in \tilde{S}$ is not realistic in many cases as empirical studies show. For example, Carrier (2008, Table 6.2) estimates a two-class LC model of airline itinerary choice with beta parameters -0.0125 for segment 1 and -0.0273 for segment 2. Thus, the maximum allowable price for the optimization would be p_j^{max} = €73.26 (Figure 4), while real ticket prices were as high as €362 in the empirical study.

In summary, while there are sufficient conditions for the data that would allow exploiting generalized convexity substructures of the problem in price, they appear far too restrictive. In general, due to the non-convex nonlinearities, the DMMP problem is difficult to solve. The common methods relying on local gradient-based information may terminate in a local optimum that is not global, and we will typically not know how far a local optimum is from the global one. Therefore, we follow a different approach based on the inverse price demand-relationship and introduce demand as a control.
In particular, we are interested in approximations that make the problem more tractable. Furthermore, we discuss several techniques to generate promising starting solutions as well as upper bounds in order to evaluate the quality of a known local optimum.

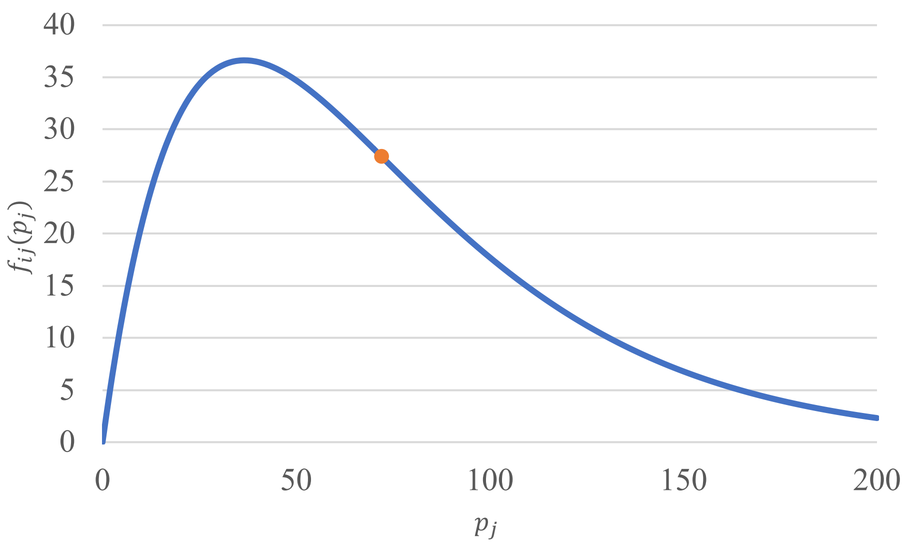

Figure 4: Example for $f_{ij}(p_j)$ with inflection point at $\bar{\bar{p}}_{ij}$ =73.26

3.2.2 An equivalent formulation with inverse price-demand relationship (P1)

We now reformulate the problem P0 by introducing new variables $\pi_{ij} \in [0,1]$ $(i \in I, j \in \tilde{J}_i)$ and $v_i > 0$ $(i \in I)$ with

$$\pi_{ij} = \frac{e^{\alpha_{ij}+\beta_{ij}p_j}}{A_{i0} + \sum_{n\in\tilde{J}_i} e^{\alpha_{in}+\beta_{in}p_n}} \quad (i \in I, j \in \tilde{J}_i) \tag{3.10}$$

being the probability that a consumer from segment $i \in I$ chooses product $j \in \tilde{J}_i$, and

$$v_i = \frac{1}{A_{i0} + \sum_{j \in \tilde{J}_i} e^{\alpha_{ij} + \beta_{ij} p_j}} \quad (i \in I) \tag{3.11}$$

being an auxiliary variable, which is proportional to the probability π_{i0} that segment $i \in I$ buys the outside option, i.e., $\pi_{i0} = A_{i0} v_i$.

Then, the two equations (3.10.) and (3.11) can be equivalently stated as

$$p_j = \frac{1}{\beta_{ij}} ln\left[\frac{\pi_{ij}}{v_i}\right] - \frac{\alpha_{ij}}{\beta_{ij}} \quad \left(j \in J, i \in \tilde{I}_j\right) \tag{3.12}$$

$$A_{i0} v_i + \sum_{j \in \tilde{J}_i} \pi_{ij} = 1 \quad (i \in I) \tag{3.13}$$

Constraints (3.12) represent the inverse price demand-relationship. For each product $j \in J$, they ensure price consistency across preference-based customer segments $i \in \tilde{I}_j$, since the left-hand side depends on index j only while the price inverse on the right-hand side depends on i and j. Therefore, we also call them price consistency (PC) constraints throughout the paper. Equations (3.13) state that the choice shares of all alternatives considered by segment $i \in I$ sum up to unity.

Accordingly, the problem P0 can be reformulated as the following problem with inverse price-demand relationship and price consistency, **P1**:

$$Max \sum_{i \in I} \omega_i \sum_{j \in \tilde{J}_i} \left(\frac{1}{\beta_{ij}} ln\left[\frac{\pi_{ij}}{v_i}\right] - \frac{\alpha_{ij}}{\beta_{ij}}\right) \pi_{ij} \tag{3.14}$$

$$p_j = \frac{1}{\beta_{ij}} ln\left[\frac{\pi_{ij}}{v_i}\right] - \frac{\alpha_{ij}}{\beta_{ij}} \quad \left(j \in J, i \in \tilde{I}_j\right) \tag{3.15}$$

$$\sum_{i \in I} \omega_i \sum_{j \in \ddot{J}_{ir}} \pi_{ij} \leq CAP_r \quad (r \in R) \tag{3.16}$$

$$A_{i0} v_i + \sum_{j \in \tilde{J}_i} \pi_{ij} = 1 \quad (i \in I) \tag{3.17}$$

$$e^{\alpha_{ij} + \beta_{ij} p_j^U} v_i \leq \pi_{ij} \leq e^{\alpha_{ij} + \beta_{ij} p_j^L} v_i \quad (i \in I, j \in \tilde{J}_i) \tag{3.18}$$

$$\pi_{ij} \in [0,1] \quad (i \in I, j \in \tilde{J}_i), \qquad v_i \geq 0 \quad (i \in I) \tag{3.19}$$

$$p_j \in [p_j^L, p_j^U] \quad (j \in J) \tag{3.20}$$

The objective function in (3.14) is the revenue as a function of $\boldsymbol{\pi}$ and $\boldsymbol{v}$. The variables in this formulation are $\boldsymbol{p}, \boldsymbol{\pi}$ and $\boldsymbol{v}$ as defined in (3.19) and (3.20). In (3.15), we have the price consistency constraints. The capacity constraints as a function of $\boldsymbol{\pi}$ are presented in (3.16) while (3.17) contains the sum-to-unity constraints. Constraints (3.18) correspond to lower and upper price bound constraints if price is expressed as a function of $\boldsymbol{\pi}$ and $\boldsymbol{v}$.

Obvious to see, a price vector $\boldsymbol{p}$ is an optimal solution to P0 if and only if $(\boldsymbol{p}, \boldsymbol{\pi}, \boldsymbol{v})$ is optimal to P1. Concerning the mathematical structure, it is straightforward to show that the objective function in (3.14) is jointly concave in $(\boldsymbol{\pi}, \boldsymbol{v})$, and all constraints but the PC constraints in (3.15) are linear. The PC constraints are non-linear non-convex and can thus be considered as complicating constraints. Note that the function $\frac{1}{\beta_{ij}} ln \left[\frac{\pi_{ij}}{v_i}\right] - \frac{\alpha_{ij}}{\beta_{ij}}$ on the right-hand side of each PC constraint is quasilinear.

Furthermore, note that if product $j \in J$ is contained in only one consideration set (i.e., $|\tilde{I}_j| = 1$, which is the case for example when segments have disjoint consideration sets), then the PC constraint for product j can be dropped from the problem formulation, along with the price p_j as decision variable.

3.2.3 Convex approximations yielding upper bounds

3.2.3.1 Segment-based pricing relaxation (P_REL1)

By relaxing (i.e., ignoring) the PC constraints (3.15) and thus dropping price as a decision variable in (3.20), we obtain a convex optimization problem whose optimal objective function value provides an upper bound to the original problem. We denote this problem as **P_REL1**:

Maximize (3.14) subject to (3.16)-(3.19).

From an optimal solution $(\pi_{ij}^{REL1}, v_i^{REL1})$ of P_REL1, we can obtain prices $p_{ij}^{REL1} := \frac{1}{\beta_{ij}} ln \left[\frac{\pi_{ij}^{REL1}}{v_i^{REL1}}\right] - \frac{\alpha_{ij}}{\beta_{ij}}$ $(j \in J, i \in \tilde{I}_j)$ that are potentially differentiated by segment $i \in \tilde{I}_j$.

Theorem 1: Problem P_REL1 is convex and its optimal objective function value is an upper bound on the optimal objective function value of problem P1. Furthermore, if $(\pi_{ij}^{REL1}, v_i^{REL1})$ is an optimal solution of P_REL1 and if inferred prices p_{ij}^{REL1} are homogeneous across segments $i \in \tilde{I}_j$ for all $j \in J$, then $(\boldsymbol{p^{REL1}}, \boldsymbol{\pi^{REL1}}, \boldsymbol{v^{REL1}})$ is optimal to P1.

Proof. For convexity proofs, see Schön (2010a) and Keller et al. (2014). The upper bound property is obvious since P_REL1 is a relaxation of P1. Further, if $p_{ij}^{REL1} = p_{i'j}^{REL1}\ \forall j \in J$, $i, i' \in \tilde{I}_j: i \neq i'$, the PC constraints (3.15) are satisfied and $(\boldsymbol{p^{REL1}}, \boldsymbol{\pi^{REL1}}, \boldsymbol{v^{REL1}})$ is obviously optimal to P1.

3.2.3.2 Segment-based pricing relaxation with envelopes for the price consistency constraint (P_REL2)

To decrease the variability of product prices p_{ij}^{REL1} for $i \in \tilde{I}_j$ and thereby the extent of infeasibilities with regard to the PC constraints (3.15), the following model variant P_REL2 does include the PC constraints but in a relaxed form. In particular, we derive two types of convex relaxations for the PC constraint. For this purpose, note that we can rewrite the PC constraints (3.15) equivalently as the following two inequalities:

$$\left[\frac{1}{\beta_{ij}} ln\left[\frac{\pi_{ij}}{v_i}\right] - \frac{\alpha_{ij}}{\beta_{ij}}\right] \leq p_j \quad (j \in J, i \in \tilde{I}_j) \tag{3.21}$$

$$\left[\frac{1}{\beta_{ij}} ln\left[\frac{\pi_{ij}}{v_i}\right] - \frac{\alpha_{ij}}{\beta_{ij}}\right] \geq p_j \quad (j \in J, i \in \tilde{I}_j) \tag{3.22}$$

3.2.3.2.1 Envelopes for the logarithm

Constraint sets (3.21) and (3.22) can be further rewritten as

$$\frac{1}{\beta_{ij}} ln[\pi_{ij}] \underbrace{- \frac{1}{\beta_{ij}} ln[v_i]}_{underestimate} - \frac{\alpha_{ij}}{\beta_{ij}} \leq p_j \quad (j \in J, i \in \tilde{I}_j) \tag{3.23}$$

and

$$\underbrace{\frac{1}{\beta_{ij}} ln[\pi_{ij}]}_{overestimate} - \frac{1}{\beta_{ij}} ln[v_i] - \frac{\alpha_{ij}}{\beta_{ij}} \geq p_j \quad (j \in J, i \in \tilde{I}_j) \tag{3.24}$$

respectively.

Note that for $\beta_{ij} < 0$, the function $\frac{1}{\beta_{ij}} ln[\pi_{ij}]$ is convex while $-\frac{1}{\beta_{ij}} ln[v_i]$ is concave. Therefore, we linearly underestimate $-\frac{1}{\beta_{ij}} ln[v_i]$ in (3.23) and linearly overestimate $\frac{1}{\beta_{ij}} ln[\pi_{ij}]$ in (3.24) in order to obtain a convex relaxation of the constraints.

Let v_i^L and v_i^U be lower and upper bounds on v_i $(i \in I)$, e.g.,

$$v_i^L := \frac{1}{A_{i0} + \sum_{j \in \tilde{J}_i} e^{\alpha_{ij} + \beta_{ij} p_j^L}}, v_i^U := \frac{1}{A_{i0} + \sum_{j \in \tilde{J}_i} e^{\alpha_{ij} + \beta_{ij} p_j^U}} \tag{3.25}$$

Then, we use the convex envelop of $ln[v_i]$ over $[v_i^L, v_i^U]$ to underestimate $ln[v_i]$:

$$ln[v_i] \geq \frac{ln[v_i^U] - ln[v_i^L]}{v_i^U - v_i^L}(v_i - v_i^L) + \ln[v_i^L] \tag{3.26}$$

Accordingly, we can relax the first price consistency constraint set (3.23) by

$$\frac{1}{\beta_{ij}} ln[\pi_{ij}] - \frac{1}{\beta_{ij}}\left(\frac{ln[v_i^U] - ln[v_i^L]}{v_i^U - v_i^L}(v_i - v_i^L) + \ln[v_i^L]\right) - \frac{\alpha_{ij}}{\beta_{ij}} \leq p_j \quad (j \in J, i \in \tilde{I}_j) \tag{3.27}$$

Similarly, we use $\ln[\pi_{ij}] \geq \frac{ln[\pi_{ij}^U] - ln[\pi_{ij}^L]}{\pi_{ij}^U - \pi_{ij}^L}(\pi_{ij} - \pi_{ij}^L) + \ln[\pi_{ij}^L]$ to relax constraint set (3.24) by

$$\frac{1}{\beta_{ij}}\left(\frac{ln[\pi_{ij}^U] - ln[\pi_{ij}^L]}{\pi_{ij}^U - \pi_{ij}^L}(\pi_{ij} - \pi_{ij}^L) + \ln[\pi_{ij}^L]\right) - \frac{1}{\beta_{ij}} ln[v_i] - \frac{\alpha_{ij}}{\beta_{ij}} \geq p_j \quad (j \in J, i \in \tilde{I}_j) \tag{3.28}$$

where π_{ij}^L and π_{ij}^U $(i \in I, j \in \tilde{J}_i)$ are lower and upper bounds on π_{ij}, e.g.,

$$\pi_{ij}^L := \frac{e^{\alpha_{ij} + \beta_{ij} p_j^U}}{A_{i0} + e^{\alpha_{ij} + \beta_{ij} p_j^U} + \sum_{n \in \tilde{J}_i \setminus \{j\}} e^{\alpha_{in} + \beta_{in} p_n^L}} \tag{3.29}$$

$$\pi_{ij}^U := \frac{e^{\alpha_{ij} + \beta_{ij} p_j^L}}{A_{i0} + e^{\alpha_{ij} + \beta_{ij} p_j^L} + \sum_{n \in \tilde{J}_i \setminus \{j\}} e^{\alpha_{in} + \beta_{in} p_n^U}} \tag{3.30}$$

3.2.3.2.2 McCormick envelopes for bilinear terms

The second relaxation of the PC constraints in (3.21) and (3.22) is based on McCormick envelopes for bilinear terms (McCormick 1976). To see how they can be applied to the PC constraints, we first consider (3.22) in Part 1, then (3.21) in Part 2.

PC constraint 1

For $\pi_{ij} > 0$, we can multiply the PC constraint (3.22) on both sides with π_{ij} and equivalently rewrite it as

$$\underbrace{\left[\frac{1}{\beta_{ij}} ln\left[\frac{\pi_{ij}}{v_i}\right] - \frac{\alpha_{ij}}{\beta_{ij}}\right]\pi_{ij}}_{concave} \geq \underbrace{p_j \pi_{ij}}_{r_{ij}} \quad (j \in J, i \in \tilde{I}_j) \tag{3.31}$$

While the left-hand side of the „≥"-inequality is concave and does not introduce any non-convex nonlinearities to the feasible set, the bilinear term on the right-hand side is complicating. Therefore, we introduce a new variable $r_{ij} \geq 0$ to replace the bilinear term and linearly approximate the requirement $r_{ij} = p_j \pi_{ij}$ by the following four McCormick envelopes:

$$r_{ij} \geq p_j^L \pi_{ij} + p_j \pi_{ij}^L - p_j^L \pi_{ij}^L \quad (j \in J, i \in \tilde{I}_j) \tag{3.32}$$

$$r_{ij} \geq p_j^U \pi_{ij} + p_j \pi_{ij}^U - p_j^U \pi_{ij}^U \quad (j \in J, i \in \tilde{I}_j) \tag{3.33}$$

$$r_{ij} \leq p_j^U \pi_{ij} + p_j \pi_{ij}^L - p_j^U \pi_{ij}^L \quad (j \in J, i \in \tilde{I}_j) \tag{3.34}$$

$$r_{ij} \leq p_j^L \pi_{ij} + p_j \pi_{ij}^U - p_j^L \pi_{ij}^U \quad (j \in J, i \in \tilde{I}_j) \tag{3.35}$$

As p_j^U and p_j^L converge to each other, the approximation becomes more accurate.

PC constraint 2

For $v_i \neq 0$, PC constraint 2 in (3.21) is equivalent to

$$\underbrace{\left[\frac{1}{\beta_{ij}} ln\left[\frac{\pi_{ij}}{v_i}\right] - \frac{\alpha_{ij}}{\beta_{ij}}\right] v_i}_{convex} \leq \underbrace{p_j v_i}_{u_{ij}} \quad (j \in J, i \in \tilde{I}_j) \tag{3.36}$$

The term on the left-hand side of the „≤"-inequality is convex for $\beta_{ij} < 0$ and thus easy to cope with. However, we have again a bilinear term on the right-hand side that is rendering the constraint non-convex. Therefore, we introduce a new variable $u_{ij} \geq 0$ as before to replace the bilinear term and linearly approximate the requirement $u_{ij} = p_j v_i$ by the following four McCormick envelopes:

$$u_{ij} \geq p_j^L v_i + p_j v_i^L - p_j^L v_i^L \quad (j \in J, i \in \tilde{I}_j) \tag{3.37}$$

$$u_{ij} \geq p_j^U v_i + p_j v_i^U - p_j^U v_i^U \quad (j \in J, i \in \tilde{I}_j) \tag{3.38}$$

$$u_{ij} \leq p_j^U v_i + p_j v_i^L - p_j^U v_i^L \quad (j \in J, i \in \tilde{I}_j) \tag{3.39}$$

$$u_{ij} \leq p_j^L v_i + p_j v_i^U - p_j^L v_i^U \quad (j \in J, i \in \tilde{I}_j) \tag{3.40}$$

Using both approximations for the PC constraints (i.e., the envelopes for the logarithm and the McCormick envelopes), we can define the problem **P_REL2:**

$$Max \sum_{i \in I} \omega_i \sum_{j \in \tilde{J}_i} r_{ij} \tag{3.41}$$

Subject to

Price consistency “≥” with McCormick:

$$\left[\frac{1}{\beta_{ij}} ln\left[\frac{\pi_{ij}}{v_i}\right] - \frac{\alpha_{ij}}{\beta_{ij}}\right]\pi_{ij} \geq r_{ij} \quad (j \in J, i \in \tilde{I}_j) \tag{3.42}$$

$$r_{ij} \geq p_j^L \pi_{ij} + p_j \pi_{ij}^L - p_j^L \pi_{ij}^L \quad (j \in J, i \in \tilde{I}_j) \tag{3.43}$$

$$r_{ij} \geq p_j^U \pi_{ij} + p_j \pi_{ij}^U - p_j^U \pi_{ij}^U \quad (j \in J, i \in \tilde{I}_j) \tag{3.44}$$

$$r_{ij} \leq p_j^U \pi_{ij} + p_j \pi_{ij}^L - p_j^U \pi_{ij}^L \quad (j \in J, i \in \tilde{I}_j) \tag{3.45}$$

$$r_{ij} \leq p_j^L \pi_{ij} + p_j \pi_{ij}^U - p_j^L \pi_{ij}^U \quad (j \in J, i \in \tilde{I}_j) \tag{3.46}$$

Price consistency “≤” with McCormick:

$$\left[\frac{1}{\beta_{ij}} ln\left[\frac{\pi_{ij}}{v_i}\right] - \frac{\alpha_{ij}}{\beta_{ij}}\right] v_i \leq u_{ij} \quad (j \in J, i \in \tilde{I}_j) \tag{3.47}$$

$$u_{ij} \geq p_j^L v_i + p_j v_i^L - p_j^L v_i^L \quad (j \in J, i \in \tilde{I}_j) \tag{3.48}$$

$$u_{ij} \geq p_j^U v_i + p_j v_i^U - p_j^U v_i^U \quad (j \in J, i \in \tilde{I}_j) \tag{3.49}$$

$$u_{ij} \leq p_j^U v_i + p_j v_i^L - p_j^U v_i^L \quad (j \in J, i \in \tilde{I}_j) \tag{3.50}$$

$$u_{ij} \leq p_j^L v_i + p_j v_i^U - p_j^L v_i^U \quad (j \in J, i \in \tilde{I}_j) \tag{3.51}$$

Price consistency "≤" with linearization of log:

$$\frac{1}{\beta_{ij}} ln[\pi_{ij}] - \frac{1}{\beta_{ij}}\left(\frac{ln[v_i^U] - ln[v_i^L]}{v_i^U - v_i^L}(v_i - v_i^L) + \ln[v_i^L]\right) - \frac{\alpha_{ij}}{\beta_{ij}} \leq p_j \quad (j \in J, i \in \tilde{I}_j) \tag{3.52}$$

Price consistency "≥" with linearization of log:

$$\frac{1}{\beta_{ij}}\left(\frac{ln[\pi_{ij}^U] - ln[\pi_{ij}^L]}{\pi_{ij}^U - \pi_{ij}^L}(\pi_{ij} - \pi_{ij}^L) + \ln[\pi_{ij}^L]\right) - \frac{1}{\beta_{ij}} ln[v_i] - \frac{\alpha_{ij}}{\beta_{ij}} \geq p_j \quad (j \in J, i \in \tilde{I}_j) \tag{3.53}$$

Standard constraints from the inverse model:

$$\sum_{i \in I} \omega_i \sum_{j \in \tilde{J}_{ir}} \pi_{ij} \leq CAP_r \quad (r \in R) \tag{3.54}$$

$$A_{i0} v_i + \sum_{j \in \tilde{J}_i} \pi_{ij} = 1 \quad (i \in I) \tag{3.55}$$

$$e^{\alpha_{ij} + \beta_{ij} p_j^U} v_i \leq \pi_{ij} \leq e^{\alpha_{ij} + \beta_{ij} p_j^L} v_i \quad (i \in I, j \in \tilde{J}_i) \tag{3.56}$$

Variables:

$$p_j^L \leq p_j \leq p_j^U \quad (j \in J) \tag{3.57}$$

$$\pi_{ij}^L \leq \pi_{ij} \leq \pi_{ij}^U \quad (i \in I, j \in \tilde{J}_i) \tag{3.58}$$

$$v_i^L \leq v_i \leq v_i^U \quad (i \in I) \tag{3.59}$$

$$u_{ij} \geq 0, \qquad r_{ij} \geq 0 \quad (i \in I, j \in \tilde{J}_i) \tag{3.60}$$

3.2.3.3 Piecewise approximations through interval partitioning for product prices (P_REL3)

3.2.3.3.1 Piecewise McCormick approximations

The McCormick approximations of $p_j \pi_{ij}$ in (3.32)-(3.35) (and of $p_j v_i$ in (3.37)-(3.40) respectively) are tighter, the tighter the lower and upper bounds on the two variables, p_j^L, p_j^U and π_{ij}^L, π_{ij}^U, are. To strengthen the approximation of the McCormick relaxation for bilinear terms, different *piecewise McCormick* approximations have been proposed. For example, to globally solve complex problems typically arising in the chemical process industry, Bergamini et al.

(2005) present a piecewise McCormick relaxation based on a univariate and uniform partitioning approach. In our case, we partition the domain $[p_j^L, p_j^U]$ of each price variable p_j into disjoint intervals $\mathfrak{I}_{jk} \coloneqq [\check{p}_{jk}, \hat{p}_{jk}]$ $(k \in K_j)$ with $\hat{p}_{jk} = \check{p}_{j,k+1}, \check{p}_{j1} = p_j^L, \hat{p}_{j,K_j} = p_j^U$. Then, the best price interval for product j is determined by introducing binary variables $x_{jk} \in \{0,1\}$ with $x_{jk} = 1$ if interval k is selected, 0 otherwise. To ensure that exactly one price interval is selected we impose

$$\sum_{k \in K_j} x_{jk} = 1 \quad (j \in J) \tag{3.61}$$

The interval partitioning leads to the following set of disjunctive constraints for the McCormick approximations of $p_j \pi_{ij}$ (replacing (3.43)-(3.46)):

$$\bigvee_{k \in K_j} \begin{bmatrix} x_{jk} \\ r_{ij} \geq \check{p}_{jk}\pi_{ij} + p_j\check{\pi}_{ijk} - \check{p}_{jk}\check{\pi}_{ijk} \\ r_{ij} \geq \hat{p}_{jk}\pi_{ij} + p_j\hat{\pi}_{ijk} - \hat{p}_{jk}\hat{\pi}_{ijk} \\ r_{ij} \leq \hat{p}_{jk}\pi_{ij} + p_j\check{\pi}_{ijk} - \hat{p}_{jk}\check{\pi}_{ijk} \\ r_{ij} \leq \check{p}_{jk}\pi_{ij} + p_j\hat{\pi}_{ijk} - \check{p}_{jk}\hat{\pi}_{ijk} \end{bmatrix} \tag{3.62}$$

where $\check{\pi}_{ijk}$ and $\hat{\pi}_{ijk}$ are lower and upper bounds on π_{ij} if the price of product j is bounded to interval $k \in K_j$, e.g.,

$$\check{\pi}_{ijk} := \frac{e^{\alpha_{ij}+\beta_{ij}\hat{p}_{jk}}}{A_{i0} + e^{\alpha_{ij}+\beta_{ij}\hat{p}_{jk}} + \sum_{n \in \tilde{J}_i \setminus \{j\}} e^{\alpha_{in}+\beta_{in}p_n^L}} \tag{3.63}$$

$$\hat{\pi}_{ijk} := \frac{e^{\alpha_{ij}+\beta_{ij}\check{p}_{jk}}}{A_{i0} + e^{\alpha_{ij}+\beta_{ij}\check{p}_{jk}} + \sum_{n \in \tilde{J}_i \setminus \{j\}} e^{\alpha_{in}+\beta_{in}p_n^U}} \tag{3.64}$$

Similarly, for $p_j v_i$ we have:

$$\bigvee_{k \in K_j} \begin{bmatrix} x_{jk} \\ u_{ij} \geq \check{p}_{jk}v_i + p_j\check{v}_{ijk} - \check{p}_{jk}\check{v}_{ijk} \\ u_{ij} \geq \hat{p}_{jk}v_i + p_j\hat{v}_{ijk} - \hat{p}_{jk}\hat{v}_{ijk} \\ u_{ij} \leq \hat{p}_{jk}v_i + p_j\check{v}_{ijk} - \hat{p}_{jk}\check{v}_{ijk} \\ u_{ij} \leq \check{p}_{jk}v_i + p_j\hat{v}_{ijk} - \check{p}_{jk}\hat{v}_{ijk} \end{bmatrix} \tag{3.65}$$

with lower and upper bounds $\check{v}_{ijk}$ and $\hat{v}_{ijk}$ on v_i $(i \in I)$ if the price of product j is bounded to interval $k \in K_j$, e.g.,

$$\check{v}_{ijk} := \left[A_{i0} + e^{\alpha_{ij} + \beta_{ij} \check{p}_{jk}} + \sum_{n \in \tilde{J}_i \setminus \{j\}} e^{\alpha_{in} + \beta_{in} p_n^L} \right]^{-1} \tag{3.66}$$

$$\hat{v}_{ijk} := \left[A_{i0} + e^{\alpha_{ij} + \beta_{ij} \hat{p}_{jk}} + \sum_{n \in \tilde{J}_i \setminus \{j\}} e^{\alpha_{in} + \beta_{in} p_n^U} \right]^{-1} \tag{3.67}$$

The disjunctive constraints can be algebraically expressed, e.g., through Big-M constraints or so-called Hull reformulations (Trespalacios and Grossmann 2014), resulting in (linear or non-linear) mixed-integer problem (MIP) formulations. While the Big-M formulations of the disjunctions generate a smaller MIPs, the Hull reformulation generates a tighter continuous relaxation of the MIP and is therefore applied in the following. In particular, for the Hull reformulation of (3.62) we define new continuous variables $\tilde{p}_{jk}, \tilde{\pi}_{ijk}, \tilde{r}_{ijk} \geq 0$ for each interval with

$$p_j = \sum_{k \in K_j} \tilde{p}_{jk} \ (j \in J), \qquad \check{p}_{jk} x_{jk} \leq \tilde{p}_{jk} \leq \hat{p}_{jk} x_{jk} \ \left(j \in J, k \in K_j\right) \tag{3.68}$$

$$\pi_{ij} = \sum_{k \in K_j} \tilde{\pi}_{ijk} \ (i \in I, j \in \tilde{J}_i), \check{\pi}_{ijk} x_{jk} \leq \tilde{\pi}_{ijk} \leq \hat{\pi}_{ijk} x_{jk} \ \left(i \in I, j \in \tilde{J}_i, k \in K_j\right) \tag{3.69}$$

$$r_{ij} = \sum_{k \in K_j} \tilde{r}_{ijk} \ (i \in I, j \in \tilde{J}_i), \check{r}_{ijk} x_{jk} \leq \tilde{r}_{ijk} \leq \hat{r}_{ijk} x_{jk} \ \left(i \in I, j \in \tilde{J}_i, k \in K_j\right) \tag{3.70}$$

with lower and upper bounds on $\tilde{r}_{ijk}$, $\check{r}_{ijk}$ and $\hat{r}_{ijk}$, respectively, if the price of product j is bounded to interval $k \in K_j$.

Then, the McCormick approximations of $p_j \pi_{ij}$ can be stated as

$$\tilde{r}_{ijk} \geq \check{p}_{jk} \tilde{\pi}_{ijk} + \tilde{p}_{jk} \check{\pi}_{ijk} - \check{p}_{jk} \check{\pi}_{ijk} \quad \left(i \in I, j \in \tilde{J}_i, k \in K_j\right) \tag{3.71}$$

$$\tilde{r}_{ijk} \geq \hat{p}_{jk} \tilde{\pi}_{ijk} + \tilde{p}_{jk} \hat{\pi}_{ijk} - \hat{p}_{jk} \hat{\pi}_{ijk} \quad \left(i \in I, j \in \tilde{J}_i, k \in K_j\right) \tag{3.72}$$

$$\tilde{r}_{ijk} \leq \hat{p}_{jk} \tilde{\pi}_{ijk} + \tilde{p}_{jk} \check{\pi}_{ijk} - \hat{p}_{jk} \check{\pi}_{ijk} \quad \left(i \in I, j \in \tilde{J}_i, k \in K_j\right) \tag{3.73}$$

$$\tilde{r}_{ijk} \leq \check{p}_{jk} \tilde{\pi}_{ijk} + \tilde{p}_{jk} \hat{\pi}_{ijk} - \check{p}_{jk} \hat{\pi}_{ijk} \quad \left(i \in I, j \in \tilde{J}_i, k \in K_j\right) \tag{3.74}$$

Similarly, for the Hull reformulation of (3.65) we define new continuous variables $\tilde{v}_{ijk}, \tilde{u}_{ijk} \geq 0$ for each interval with

$$v_i = \sum_{k \in K_j} \tilde{v}_{ijk} \ \ (i \in I, j \in \tilde{J}_i), \check{v}_{ijk} x_{jk} \leq \tilde{v}_{ijk} \leq \hat{v}_{ijk} x_{jk} \ \ \left(i \in I, j \in \tilde{J}_i, k \in K_j\right) \tag{3.75}$$

$$u_{ij} = \sum_{k \in K_j} \tilde{u}_{ijk} \ \ (i \in I, j \in \tilde{J}_i), \check{u}_{ijk} x_{jk} \leq \tilde{u}_{ijk} \leq \hat{u}_{ijk} x_{jk} \ \ \left(i \in I, j \in \tilde{J}_i, k \in K_j\right) \tag{3.76}$$

with lower and upper bounds on $\tilde{u}_{ijk}$, $\check{u}_{ijk}$ and $\hat{u}_{ijk}$, respectively, if the price of product j is bounded to interval $k \in K_j$, and McCormick approximations of $p_j v_i$ as follows:

$$\tilde{u}_{ijk} \geq \check{p}_{jk} \tilde{v}_{ijk} + \tilde{p}_{jk} \check{v}_{ijk} - \check{p}_{jk} \check{v}_{ijk} \ \ \left(i \in I, j \in \tilde{J}_i, k \in K_j\right) \tag{3.77}$$

$$\tilde{u}_{ijk} \geq \hat{p}_{jk} \tilde{v}_{ijk} + \tilde{p}_{jk} \hat{v}_{ijk} - \hat{p}_{jk} \hat{v}_{ijk} \ \ \left(i \in I, j \in \tilde{J}_i, k \in K_j\right) \tag{3.78}$$

$$\tilde{u}_{ijk} \leq \hat{p}_{jk} \tilde{v}_{ijk} + \tilde{p}_{jk} \check{v}_{ijk} - \hat{p}_{jk} \check{v}_{ijk} \ \ \left(i \in I, j \in \tilde{J}_i, k \in K_j\right) \tag{3.79}$$

$$\tilde{u}_{ijk} \leq \check{p}_{jk} \tilde{v}_{ijk} + \tilde{p}_{jk} \hat{v}_{ijk} - \check{p}_{jk} \hat{v}_{ijk} \ \ \left(i \in I, j \in \tilde{J}_i, k \in K_j\right) \tag{3.80}$$

Finally, we also reformulate (3.56) as

$$e^{\alpha_{ij} + \beta_{ij} \hat{p}_{jk}} \tilde{v}_{ijk} \leq \tilde{\pi}_{ijk} \leq e^{\alpha_{ij} + \beta_{ij} \check{p}_{jk}} \tilde{v}_{ijk} \quad \left(i \in I, j \in \tilde{J}_i, k \in K_j\right) \tag{3.81}$$

In summary, denote the problem to determine variables p_j, $\tilde{p}_{jk}$, π_{ij}, $\tilde{\pi}_{ijk}$, v_i, $\tilde{v}_{ijk}$, r_{ij}, $\tilde{r}_{ijk}$, u_{ij}, $\tilde{u}_{ijk}$, and x_{jk} $\left(i \in I, j \in \tilde{J}_i, k \in K_j\right)$ such that (3.41) is maximized subject to constraints (3.42), (3.47), (3.54), (3.55), (3.61), (3.68)-(3.81) by **P_REL3**. Note that P_REL3 is a mixed-integer convex optimization (MICP) problem that, compared to the convex optimization problems P_REL1 and P_REL2, includes additional binary variables but is tighter. Obviously, the optimal objective function value of P_REL3 provides an upper bound to the optimal objective function value of P1.

3.2.3.3.2 Piecewise envelopes for the logarithm

Based on the partitioning of the price intervals introduced in the subsection above, we can also tighten the envelopes of the logarithm in (3.52) and (3.53), respectively, as follows and add them to P_REL3 ($i \in I, j \in \tilde{J}_i, k \in K_j$):

$$\frac{1}{\beta_{ij}} ln\big[\tilde{\pi}_{ijk}\big] - \frac{1}{\beta_{ij}}\left(\frac{ln\big[\hat{v}_{ijk}\big] - ln\big[\check{v}_{ijk}\big]}{\hat{v}_{ijk} - \check{v}_{ijk}}\big(\tilde{v}_{ijk} - \check{v}_{ijk}\big) + \ln\big[\check{v}_{ijk}\big]\right) - \frac{\alpha_{ij}}{\beta_{ij}} \leq \tilde{p}_{jk} \tag{3.82}$$

$$\frac{1}{\beta_{ij}}\left(\frac{ln\big[\hat{\pi}_{ijk}\big] - ln\big[\check{\pi}_{ijk}\big]}{\hat{\pi}_{ijk} - \check{\pi}_{ijk}}(\tilde{\pi}_{ijk} - \check{\pi}_{ijk}) + \ln\big[\check{\pi}_{ijk}\big]\right) - \frac{1}{\beta_{ij}} ln\big[\tilde{v}_{ijk}\big] - \frac{\alpha_{ij}}{\beta_{ij}} \geq \tilde{p}_{jk} \tag{3.83}$$

3.3 Heuristics yielding lower bounds

3.3.1 Outer approximation heuristic

In this heuristic, the optimal price vector $\mathbf{p}$ from the outer approximation problem P_REL2 or P_REL3 is used. If $\mathbf{p}$ is capacity feasible (i.e., satisfying constraints (3.7)), the price vector can be used directly as an approximate solution to P0. If $\mathbf{p}$ is not capacity feasible, we use it as a (slightly infeasible) warm start solution for a gradient-based search procedure (e.g. an interior point method implemented in Knitro) to determine a (potentially locally) optimal and thus feasible solution of P0.

3.3.2 Price interval heuristic (3-step approach)

For large-scale networks such as passenger railway networks, mixed-integer problem formulations are typically intractable due to the exponential number of integer variables. For this case, we propose in the following a heuristic that is based on the models developed in 3.2 but relies on continuous decision variables to make an application for large-scale instances possible.

As already mentioned in 3.2.3.3.2, the approximations of the PC constraints are tighter the smaller the price ranges p_j^L and p_j^U are. While price ranges can be quite large in a practical application, we can dynamically use the optimal segment-specific prices from the solution of P_REL1 and derive approximate, updated lower and upper price bounds (p_j^{L*} and p_j^{U*}, respectively) to increase the accuracy of the approximation for the PC constraints in P_REL2. To keep the number of processed constraints as low as possible, we only apply the envelopes for the logarithm as in (3.52) and (3.53) and drop the McCormick constraints (incl. the corresponding variables $\boldsymbol{u}$ and $\boldsymbol{r}$) from P_REL2 for this heuristic. Hence, the objective function is replaced by the original inverse formulation as in (3.14). The reason for choosing the envelopes for the logarithm instead of the McCormick approximation is a good balance between solution quality and time as we will show later in the numerical study. Finally, the solution of P_REL2 is used as a warm start to solve P0 and find a good locally optimal solution.

Since both P_REL1 and P_REL2 can be solved in polynomial time and globally optimal with appropriate solution techniques such as the interior-point method, they are reliable in terms of solution time and quality. The optimal values for p_j from P_REL2 serve as a warm start for P0 and decrease the solution time of P0 significantly as we will show later in the numerical study.

In the following, more technical details and the processed steps for the heuristic are presented:

Step 1 - Derive tighter lower and upper price bounds:

Step 1a: Solve the relaxed inverse model with segment-specific prices (P_REL1) to derive an optimal solution for v_i^{REL1} and π_{ij}^{REL1}.

Step 1b: Compute the segment-specific prices $p_{ij}^{REL1} := \frac{1}{\beta_{ij}} ln\left[\frac{\pi_{ij}^{REL1}}{v_i^{REL1}}\right] - \frac{\alpha_{ij}}{\beta_{ij}}$ $(j \in J, i \in \tilde{I}_j)$ and determine the new lower and upper price bounds $p_j^{L*} := \min_{i \in \tilde{I}_j} p_{ij}^{REL1}$ and $p_j^{U*} := \max_{i \in \tilde{I}_j} p_{ij}^{REL1}$ and define $A_{ij}^{max} := e^{\alpha_{ij}+\beta_{ij}p_j^{L*}}$ and $A_{ij}^{min} := e^{\alpha_{ij}+\beta_{ij}p_j^{U*}}$.

Step 1c: Determine the lower and upper bounds on market shares π_{ij} and v_i with $\pi_{ij}^L := \frac{A_{ij}^{min}}{A_{i0}+A_{ij}^{min}+\sum_{n \in \tilde{J}_i \setminus \{j\}} A_{in}^{max}}$, $\pi_{ij}^U := \frac{A_{ij}^{max}}{A_{i0}+A_{ij}^{max}+\sum_{n \in \tilde{J}_i \setminus \{j\}} A_{in}^{min}}$ and $v_i^L := \frac{1}{A_{i0}+\sum_{j \in \tilde{J}_i} A_{ij}^{max}}$, $v_i^U := \frac{1}{A_{i0}+\sum_{j \in \tilde{J}_i} A_{ij}^{min}}$ $(i \in I, j \in \tilde{J}_i)$.

Step 1d: Derive the demand weighted average price for each product $j \in J$ that serves as a warm start for the uniform price in step 2 by $\bar{p}_j := \frac{\sum_{i \in \tilde{I}_j} \pi_{ij}^{REL1} \omega_i p_{ij}^{REL1}}{\sum_{i \in \tilde{I}_j} \pi_{ij}^{REL1} \omega_i}$.

Step 2 - Solve model with approximate uniform price constraints (P_REL2):

Apply the bounds from step 1c on variables v_i and π_{ij} and the price bounds p_j^{L*} and p_j^{U*} from step 1b on variables p_j. Furthermore, use the lower and upper bounds from step 1c in the approximate PC constraints in P_REL2 (i.e., the envelopes for the logarithm). Finally, use the warm start $\bar{p}_j$, v_i^{REL1}, π_{ij}^{REL1} from step 1a and 1d for the variables p_j, v_i and π_{ij} and solve P_REL2.

Step 3 - Solve P0:

Solve P0 under the consideration of the original price bounds p_j^L and p_j^U $(j \in J)$ with the warm start for p_j using the optimal solution for the uniform product prices from P_REL2 derived in step 2.

End.

3.4 Numerical study

In this section, we report the application of different models in the context of a systematic numerical experiment to illustrate the overall performance of our approach with regard to solution quality and computation time. Furthermore, we present a case study of pricing in the German long-distance railway network to demonstrate the real-world applicability of our approach.

3.4.1 Systematic experiment

3.4.1.1 Experimental setup

Products: We assume that products are differentiated by market (origin destination pair), departure time window and quality level. With presumably 5 markets, 5 departure time windows and 2 quality levels, the number of products is determined as the product of the number of the three factors, i.e., 50 products. For each product $j \in J$, the lower and upper price bounds, p_j^L and p_j^U, respectively, are set to the minimum and maximum segment-specific optimal prices obtained from P_REL1 (i.e., $p_j^L := \min_{i \in \tilde{I}_j} p_{ij}^{REL1}, p_j^U := \max_{i \in \tilde{I}_j} p_{ij}^{REL1}$) within an initial price range of $p_j \in [10,200]$.

Customer segments: We assume that customers are first of all segmented by markets and desired departure time window. Furthermore, for each market and time window, there are heterogeneous preference segments that differ with regard to the parameters α_{ij} and β_{ij} where we assume $\alpha_{ij} \sim U(0,2)$ and $\beta_{ij} \sim U(-0.1, -0.01)$. Thus, the number of segments equals the product of the number of markets (5), time windows (5) and preference segments (3), resulting in 125 customer segments. For the market size, we assume that $\omega_i \sim DF \cdot U(10,30)$ where $DF \in \{1, 10, 20\}$ is a demand factor representing low, medium, and high demand scenarios, respectively. The attraction value of the outside option is randomly drawn from a uniform distribution with $A_{i0} \sim U(1,10)$.

For each segment, we generate overlapping considerations sets of 2-8 products as follows. Each segment has a random time tolerance $\sim \lfloor U(0,4) \rfloor$ that denotes the number of time windows that a segment is willing to depart later than its desired departure time window. The consideration set $\tilde{J}_i$ of segment i includes all connections in the same market within the segment's time tolerance.

Resources: The network of resources is generated as follows. First, we derive the total number $|R|$ of resources from the number of connections. Assuming one connection per market and time window, the number of connections corresponds to the product of the number of markets and the number of time windows (i.e., 25). To derive a reasonable number of resources for a given number of connections, we make use of the algorithm of Hohberger and Schön (2019) that generates test data for large-scale railway network revenue management models based on schedules of transportation companies. In particular, we use the test data presented by the authors to estimate a functional relationship between the number of connections and the number of resources in typical railway networks. A regression turns out that the number of resources can well be approximated by a square root function of the number of connections (see Figure 5).

Then, we assume that each product j uses a random number of resources, $\#res_j \sim \lfloor \mathrm{U}(1, \lceil 0.5 * |\mathrm{R}| \rceil) \rfloor$. The specific resources that a product uses are selected randomly where each resource $r \in R$ has an equal probability of being assigned. The available capacity CAP_r of resource $r \in R$ is drawn from a discrete Uniform distribution or set to the required capacity at the upper bound price vector, $\mathbf{p}^U$, whatever is larger: $CAP_r \sim \max\{\lceil \mathrm{U}(400{,}800) \rceil, CAP_r^{req}(\mathbf{p}^U)\}$.

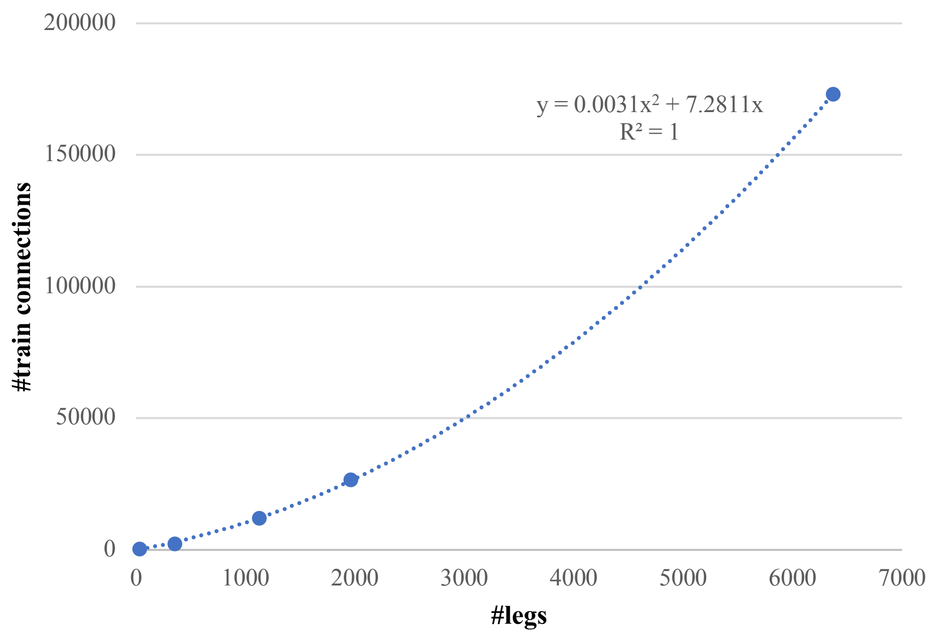

Figure 5: Relation between #train connections & #legs in a railway network (Hohberger and Schön 2019)

For each of the three demand scenarios, we randomly generate 5 instances. All tests were performed in AMPL with instances solved with CPLEX 12.8[3] on a Workstation with Windows 7, 2x Intel Xeon E5-2687W v4 @ 3GHz, 32 cores, 64GB RAM. The time limit to solve an instance was set to 10 minutes.

3.4.1.2 Results

Table 5 shows the results of the experimental study. The first three columns display the number of products, the number of segments and the demand scenario. The fourth column shows the (potentially locally optimal) objective function value in absolute terms that the warm started Knitro solver achieved for problem P0. The blue numbers below indicate the time to solve P0 (including warm start) which were neglectable for these instances. Columns 5 and 6 show the performance of the two convex approximations in terms of the relative gap between the optimal objective function value of the convex approximation and P0. As expected, P_REL2 provides slightly smaller upper bounds than P_REL1 since McCormick and log approximations of the PC constraints are incorporated in P_REL2. The improvement of the bound is achieved without significantly increasing run times. The red numbers indicate the average price range over segment-specific prices.

Columns 7-9 show the relative gap of the upper bound obtained from different versions of P_REL3 with three equidistant price intervals (PI): "P_REL3 BND" denotes a version of P_REL3 where only interval price bounds in (3.81) with the disaggregation of variables according to (3.68), (3.69) and (3.75) are considered, but PC constraints are neglected. "P_REL3 log" additionally includes log PC constraints (3.82) and (3.83) while "P_REL3 McC" incorporates the McCormick approximations (3.70)-(3.74) and (3.76)-(3.80) of the PC constraints. Both the log and the McC PC constraint approximations significantly improve the upper bound over BND at the expense of solution time. The McC bound is close to the objective function value of P0 for the base and the high demand scenario (with an average gap of 1.47% and 0.18%, respectively) indicating that the potentially local solution of P0 is close to globally optimal. However, for the low demand scenario, the McC bound exhibits a relatively large average gap of 8.45% where the time limit is reached. Here, the log bound dominates the McC bound not only with respect to average solution time but also regarding solution quality.

[3] In theory, the problems could be solved exactly with some MINLP solvers like Knitro or Bonmin. However, to our experience, the branch-and-cut method of CPLEX was much more efficient on P_REL3, although applied to a linearized version of the MICP, where all convex functions were approximated with piecewise linear underestimators.

Overall, the log PC constraint approximations show a good balance between solution time and quality.

#prod	#segm	demand	P0 Obj.	P_REL1 (1PI)	P_REL2 (1PI)	P_REL3 BND (3PI)	P_REL3 log (3PI)	P_REL3 McC (3PI)
				Upper Bounds: Gaps, avg. price range, solution time in sec. (limit 600s)				
50	125	base	245,818	6.30%	5.56%	3.98%	2.83%	1.47%
			-	53.6	53.4	20.4	15.2	17
			0	1	2	23	292	350
50	125	high	349,800	5.64%	4.14%	2.93%	2.01%	0.18%
			-	54.8	54.8	28.4	24.4	25.6
			0	1	1	6	45	169
50	125	low	32,321	15.80%	14.92%	9.65%	6.93%	8.45%
			-	66.2	53.2	22	16	14.4
			0	1	1	526	433	603

Table 5: Results of the experimental study

Looking at a selected instance of the low demand scenario, Table 6 shows the improvements of the P_REL3 bound and the run time behavior as more price intervals are added. To maintain focus, the “P_REL3 log” bound is considered with 3, 10, 15 and 50 price intervals. As before, a time limit of 600 seconds was set for solving the instance of P_REL3 with 3 price intervals, while 3600 seconds were allowed to solve instances with 10 or more price intervals per product. Interesting to note is that the solution times are not necessarily monotone increasing in the number of price intervals. The gap between the log bound and the objective function value of P0 decreases to 0.62% as the number of PIs is increased to 50. Again, the bound proves that the potentially locally optimal solution of P0 found by the warm started Knitro solver is close-to-optimal.

Our experiments demonstrate that the proposed problem formulation and relaxation approach underlying P_REL2 and P_REL3 can be very powerful for approximately solving pricing problems in capacitated transportation networks under the MMNL model. While the convex problem formulation P_REL2 can be used to quickly generate (possibly slightly infeasible) warm start solutions for solving P0 at least locally, the upper bound obtained from P_REL2 (and from P_REL3, respectively) simultaneously provides valuable information on the quality of the local solution. The comparably long run times for solving P0 reported in Keller et al. (2014) suggest that the warm start solution from P_REL2 obviously helps to find “good” or even close-to-optimal solutions of P0 quickly.

Instance 1				Upper Bounds: Gaps, solution time in sec. (limit: 3 PI = 600s ; 10+ PI = 3,600s)					
#prod	#segm	demand	P0 Obj.	P_REL1 (1PI)	P_REL2 (1PI)	P_REL3 log (3PI)	P_REL3 log (10PI)	P_REL3 log (15PI)	P_REL3 log (50PI)
50	125	low	30,646	16.43%	15.53%	9.48%	2.34%	1.93%	0.62%
			0	1	1	605	68	1,668	1,484

Table 6: Results for solving the first instance in the low demand scenario

3.4.2 Railway network revenue management application

The results so far have shown that a warm started model P0 can achieve solutions close to the global optimum. While the proposed MIP (i.e., P_REL3) works well for medium-sized instances, large-scale networks such as the railway network in Germany need solution techniques that are able to deal with thousands of products and customer segments. Therefore, we apply and test the proposed price interval heuristic (3.3.2) in an experiment using data from Deutsche Bahn.

3.4.2.1 Experimental setup

Dynamic network pricing under the MMNL has many applications in practice. Railway companies like Deutsche Bahn (DB) for example sell a unique, capacitated product (a specific train ride from A to B in a given comfort class) to different customer segments that differ in the evaluation of price ($\dot{\beta}_i$) and quality attributes (α_{ij}). The general idea of revenue management (i.e., how many seats to sell early to passengers with low WTP vs. how many seats to protect for a late booking segment with high WTP) can also be translated to the choice-based dynamic pricing case with multiple customer segments. The goal is to find an optimal uniform price for each product at each point in time during the booking horizon where the product is considered by different customer segments and the price should therefore be set on a level that maximizes total ticket revenue under capacity constraint considerations.

A product in the railway context can be described as the combination of a specific train connection/itinerary b and the comfort class q (i.e., $(b, q) = j \in J$). Since segment demand and the corresponding willingness to pay is varying throughout the booking horizon (e.g., early booking customers with low WTP vs. late booking customers with high WTP), the price for each product should be able to vary over the booking horizon $t \in T$ and therefore be defined as p_{jt}. Furthermore, a customer segment is defined by four attributes: the travel purpose *(tp)*, considered O&D *(od)*, desired departure time *(dt)* and desired booking time *(t)*, leading to a definition of $(tp, od, dt, t) = i \in I$. This split is for instance reflected in the definition of the segment-specific market demand parameter $\omega_{(tp,od,dt,t)} = \omega_i$, e.g., the number of business travelers *(tp)* who want to travel from Frankfurt to Berlin *(od)* on a Monday morning between 7 a.m. and 9 a.m. *(dt)* and would like to book 5-7 days in advance *(t)*. The consideration of booking time makes it possible to consider information about future demand in the today's pricing decision. The market share parameter π_{ij} then defines the probability that a specific segment i chooses product j. An important part in the choice model is the definition of the consideration set $\tilde{J}_i$ since it defines the set of products that are considered by segment i. In the railway case, it could be defined as the set of products that serve the same O&D within a given departure time tolerance (e.g., all connections during the day or all connections with departure time +/- two hours) and also the other available comfort classes. This leads to a choice model that not only covers switching effects between train connections from a departure time perspective but also between comfort classes. The concept of limited time windows that define the maximum deviation between desired and scheduled departure time is also used by, e.g., Proussaloglou and Koppelman (1999) in their study about the choice of a flight. Switching effects between booking dates are ignored in our study due to minor information about the strategic behavior over time, i.e., customers that decide not to book in period t are considered as lost demand and not backlogged for future periods. The underlying idea of a choice model in railway business is that customers do not have a demand for a specific train connection, they rather have a demand for going from A to B within a given departure time frame and therefore choose the product that maximizes their utility or choose not to buy at all.

We were able to test our approach by using data from our project partner Deutsche Bahn with the focus of solving large-scale instances to prove the tractability of practical problems. To our knowledge, this is the first study that solves large-scale dynamic pricing models under the MMNL. To generate the set of train connections in the rail network (incl. transfer connections), we apply the network data generation algorithm NDGA (Hohberger and Schön 2019). The NDGA provides also information about the resource consumption for each itinerary (the set of resources $r \in R$ is defined as the combination of a leg and comfort class), the available capacities for each resource and product attribute values such as travel time, number of transfers etc. that are necessary to derive the non-price product utilities α_{ij}. Therefore, NDGA helps to set-up a large proportion of the realistic test data. Remaining data, i.e., the segment-specific demand ω_i, the current lower and upper price bounds for each product and the necessary sensitivities (β-values, including price beta $\dot{\beta}_i$) were provided by Deutsche Bahn.

The challenge that arises with large-scale models is not only the nonconvexity of the original model but also the solution time that should lie within a given range to make frequent recalculations in practice possible. A carefully considered tradeoff between accuracy and performance is therefore of high interest. One powerful lever to decrease the complexity of the model is the limitation of the consideration set. Fewer products that are considered to be an alternative for product j help to decrease the model size. We use a maximum deviation of +/- two hours, i.e., a segment has a preferred departure time window but also considers train connections that lie outside of the preferred window. Since departure time preference is part of α_{ij}, the value of α_{ij} will be lower for a train connection that lies outside of the preferred window compared to a train connection in the preferred window (ceteris paribus). In addition, the number of desired departure time windows and booking time periods significantly influence the total number of segments and therefore the size of the model. We use a split into 60 minutes departure time windows and consider only two points in time for bookings, $t = 1 = today$ and $t = 2 = "the\ sum\ of\ all\ future\ periods"$.

Different sized railway networks are used in the numerical study (see Table 7). The north-south axis network consists of all major O&Ds and the corresponding train connections on a north-south axis of DB's rail network. The other two networks are defined by choosing the most important O&Ds based on past sales data. The largest network covers all daily train connections in a network of the Top 400 O&Ds, creating approx. 10,000 train connections and over 10,800 resources. This network already accounts for >50% of DB's revenue and therefore covers a big portion of ticket sales. Detailed information about the structure of the three different test networks is given in Table 7. The numbers shown correspond to a specific, full departure day in future, i.e., we optimize one full departure day incl. all products within the given network.

Network	**#train connections**	**#train stations**	**#O&Ds**	**#resources**	**#prices**	**#segments**
(1) North-South Axis	667	8	21	1,080	2,668	3,784
(2) Top 50 O&Ds	1,523	12	50	6,384	6,092	9,296
(3) Top 400 O&Ds	10,003	53	400	10,844	40,012	72,240

Table 7: Statistics for tested railway networks

For each of the three networks, we construct four demand cases with low to very high demand realizations following the idea in Zhang and Adelman (2009), e.g., the low demand case corresponds to a market demand that leads to a load factor of 40% in the unconstrained case.

The other cases are medium with load factor 0.7, high with load factor 1.0 and very high with load factor 1.3. To achieve meaningful and stable results independent from numerical outliers, we solve the networks (1) and (2) 30x and the large-scale network (3) 10x for each demand-case, where the demand is randomly varied by +/− 5% to construct slightly different test instances. Table 8 summarizes the setup of the experiment.

Network	**Demand cases**	**Unconstrained avg. load factor**	**# test instances**
North-South Axis	low	40%	30x each
	medium	70%	
	high	100%	
	very high	130%	
Top 50 O&Ds	low	40%	30x each
	medium	70%	
	high	100%	
	very high	130%	
Top 400 O&Ds	low	40%	10x each
	medium	70%	
	high	100%	
	very high	130%	

Table 8: Experimental setup for railway study

All problem instances were coded in AMPL and solved using Knitro 10.3 on a compute node of the Baden-Württemberg High Performance Cluster with 2x 2.6 GHz Intel Xeon E5-2640v3 processors using 8 threads and 128 GB RAM on a Linux operating system.

3.4.2.2 Results

Results for P0 without warm start

The decision of model P0 in the railway case can be defined as: Find the revenue maximizing uniform price for each product j in each booking period t where the accepted demand of all products j that use resource r must be smaller than CAP_r where p_{jt} must lie within a given price range. Note that p_{jt} in the railway setting depends also on the booking period $t \in T$, i.e., different price levels within the booking horizon are possible.

Keller et al. (2014) have already shown that P0 for the classic MNL and also the multi-segment case shows poor results in terms of solution time. Our tests can confirm that finding. We implemented P0 in a slightly adapted formulation, with an additional variable v_i $(i \in I)$ with $v_i := \left(A_{i0} + \sum_{j \in \bar{J}_i} e^{\alpha_{ij} + \dot{\beta}_i p_{jt}}\right)^{-1}$ representing the division by all considered product alternatives of segment i in the objective function and capacity constraint of P0. Therefore, an additional constraint is introduced to make sure that total market shares sum up to one (similar to (3.17)). The benefits of that reformulation are improved solution times for P0. Still with the modified version, solving P0 directly for large instances is intractable. Table 9 shows the solution times for a medium demand case scenario for the three different railway networks without a warm start solution.

Network	**Solution time of P0 (in sec.)**
North-South Axis	669
Top 50 O&Ds	3,350
Top 400 O&Ds	72,810

Table 9: Results for P0 without warm start

It is obvious that solution times >10h (=36,000 sec.) are not manageable for a practical application, especially for the time periods close to departure. Hetrakul and Cirillo (2014) have already shown that most of the bookings in the railway context show up very late and practical data from DB can confirm that, i.e., resolving a deterministic model is especially crucial in the last days prior to departure to account for (unforeseen) demand realizations.

Keller et al. (2014) have already presented that the inverse formulation solves much faster than the original formulation. As mentioned by Schön (2010a), the traditional inverse formulation assumes the possibility of segment-specific pricing, which is unrealistic in the railway business. A railway company typically cannot charge different prices for a specific product from a leisure customer than from a business customer, or from a leisure customer willing to travel between 7 a.m. and 9 a.m. than from a leisure customer willing to travel between 9 a.m. and 11 a.m. In practice, there is just one uniform price for each product.

The solution times for medium-sized problems with our MIP approaches in 3.4.1.2 already demonstrate that large-scale instances with thousands of products and segments will not be manageable in reasonable time.

Therefore, we apply in the following the price interval heuristic (3-step approach) from 3.3.2 since it fully relies on continuous decision variables.

Results of the price interval heuristic (3-step approach)

The results in Table 10 indicate that even large instances can be solved in under 30 minutes, a solution time that is more than convincing from a practical perspective. Both convex models solve fast and therefore serve as a good way to derive a warm start solution for P0. The comparison with the solution time of P0 without a warm start (see Table 9) highlights the necessity for a good warm start when solving P0. Additionally, the small objective value gap between P_REL1 and P0 suggests that a solution close to the global optimum is achieved. Remember that P_REL1 serves as an upper bound since it assumes segment-specific pricing which is a relaxed version of P0. The average load factor (lf) after the constrained optimization (most right column) shows that demand had to be shifted between trains (or to the outside option) due to limited capacities. Next to the average solution times, the variance of solution times is important in practice. Figure 6 and Figure 7 show the solution times of all instances in the four demand cases as a boxplot diagram. Although the total solution time of the 3-step approach varies within each demand case scenario, the variance is in an acceptable range. Even in the slowest test instances the solution is derived in 54 sec. for network (1), in 164 sec. for network (2) and 2,049 sec. for network (3). More detailed analyses have also shown that the variance of solution times is smallest for P_REL1, slightly higher for P_REL2 and highest for P0, which supports the idea of using the two convex models to derive a warm start solution.

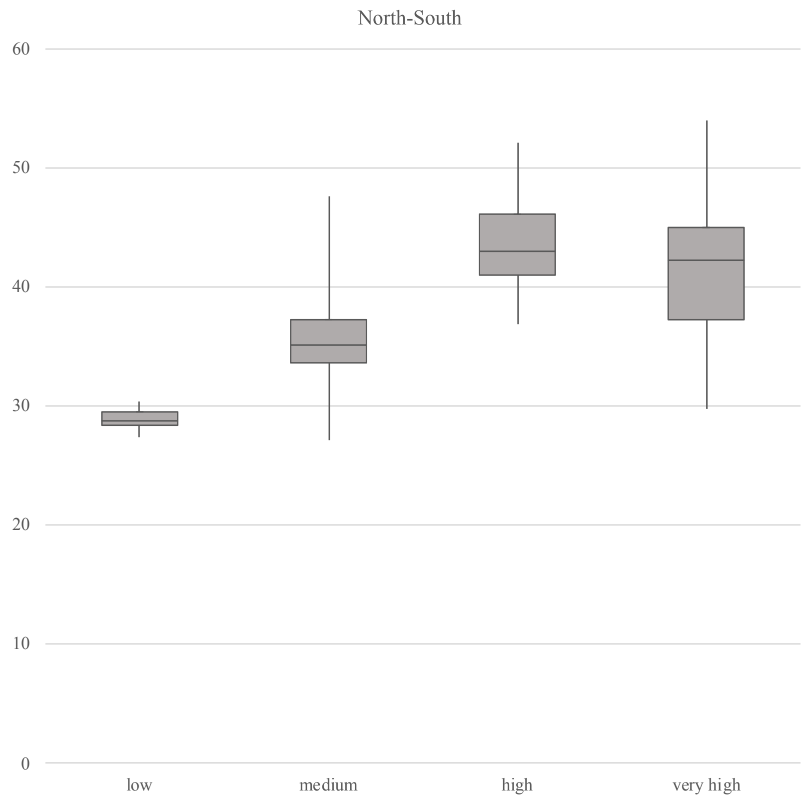

Figure 6: Solution times of all instances as boxplot for railway network (1)

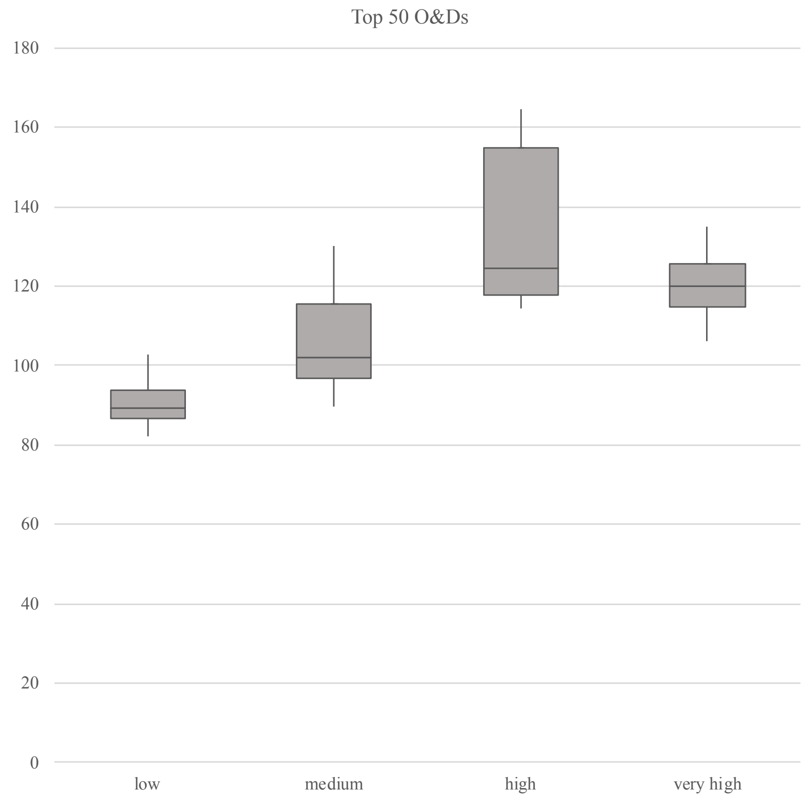

Figure 7: Solution times of all instances as boxplot for railway network (2)

Network	Demand case	Solution times for price interval heuristic (sec.)				Objective value gap	Avg. load factor
		P_REL1	P_REL2	P0	∑		
North-South Axis	low	8	10	11	29	2.5%	33%
	medium	8	12	15	36	2.4%	49%
	high	8	16	20	44	2.1%	59%
	very high	8	15	19	42	1.8%	64%
Top 50 O&Ds	low	27	23	40	90	2.6%	33%
	medium	30	34	42	105	2.5%	46%
	high	30	44	60	134	2.4%	56%
	very high	30	39	51	120	2.1%	64%
Top 400 O&Ds	low	283	255	555	1,093	2.5%	33%
	medium	308	316	704	1,328	2.4%	47%
	high	339	358	735	1,432	2.3%	56%
	very high	358	366	829	1,553	2.1%	63%

Table 10: Results for price interval heuristic (3-step approach)

A comparison with the data from P0 without a warm start shows significant improvements in terms of solution time. From the 30 instances in the medium demand case scenario, we select the instance with the smallest deviation from the average solution time and solve this directly with P0, i.e., without a warm start. The data in Table 11 show that the larger the network, the higher the advantage of the 3-step approach. The additional time that is necessary to solve P_REL1 and P_REL2 in the 3-step approach is a good investment to solve P0 with the warm start solution and therefore decreases the total solution time. The heuristic can decrease the solution time from over 20 hours to ~22 minutes for the Top 400 O&D problem, an improvement of factor 56. The advantage of the heuristic increases with the size of the network, meaning that the bigger the problem instance is, the higher are the solution time improvements.

Network	**Solution times**			**improvement factor**
	P0 without warm start	**P0 with warm start**	**Sum of 3-step heuristic**	
North-South Axis	669	16	35	19x
Top 50 O&Ds	3,350	46	106	32x
Top 400 O&Ds	72,810	652	1,292	56x

Table 11: Improvement by applying the price interval heuristic

Overall, the 3-step approach seems to be a promising way to derive prices of good quality in reasonable time, even for large-scale instances. Since it uses a single price interval for the linear approximation of the non-convex part of the original PC constraints, it works better the tighter the approximate lower and upper price bounds from P_REL1 are. For applications and data sets that achieve a large range of lower and upper price bounds after step 1, one should apply our proposed models with interval partitioning from 3.2.3.3.

3.5 Conclusion and future research

Summary

We considered the deterministic multi-product multi-resource pricing problem under discrete mixed multinomial logit demand. We discussed and analyzed the original model from a structural and mathematical perspective and presented various convex reformulations that serve as upper bounds and two heuristics yielding lower bounds for the original non-convex model. The numerical study showed that the model with the piecewise approximations can decrease the gap between the upper and lower bound significantly (< 1.5%), i.e., the objective value is close to the globally optimal solution. Further, for large-scale problem instances with thousands of products and customer segments (i.e., the railway case study), we have shown that the proposed price interval heuristic makes it able to derive a good solution in reasonable time and therefore serves as an approach for a practical application.

Future research

The current work of the authors is concerned with the proof that P_REL3 can serve as an upper bound arbitrarily close-to-optimum of the original problem by defining a sufficiently large number of price intervals $k \in K_j$. Further, our relaxation P_REL3 was based on piecewise McCormick approximations with Hull reformulations of disjunctive constraints. While the Hull reformulation generates a tighter continuous relaxation of the MIP, the Big-M reformulation would generate a smaller MIP. To obtain a stronger formulation with small growth in problem size, Trespalacios and Grossmann (2016) developed a cutting plane algorithm to solve generalized disjunctive convex optimization problems based on cuts for the Big-M formulation. It would be interesting to apply the method to the piecewise McCormick relaxation of our pricing problem and test its solution quality and speed.

Furthermore, it would be interesting to analyze the discrete pricing setting for P0 as an alternative for the continuous pricing case that we considered in this study. For the discrete pricing model, different solution techniques could be applied to achieve a convex optimization problem (e.g., linearization via Big-M constraints or the conic approach for the product assortment problem by Sen et al. (2018)).

From a practical perspective, it would be interesting to see how the DMMP under the MMNL performs compared to existing revenue management approaches in practice. Discussions with the project partner have also shown that there is another restriction on prices in railway practice, namely that longer itineraries should not be cheaper than shorter ones if the same train is used. This self-imposed rule is intended to prevent customers from buying a longer (potentially cheaper) ticket but leaving the train earlier. For this reason, an additional set of constraints should exclude this kind of price inconsistency.

Chapter 4 A simulation study for a choice-based dynamic pricing approach for large-scale railway networks

Single author:
Simon Hohberger
Chair of Service Operations Management, Business School, University of Mannheim

Abstract

During the past two decades revenue management has found its ways into the passenger railway business, in research as well as in practice. Still, advanced methods that represent the product choice of a rail customer by using a choice-based demand model, integrate network consideration including itineraries with transfers and railway specific constraints are missing. Therefore, this paper presents a choice-based dynamic pricing approach for large-scale railway networks that is developed in cooperation with the German railway company Deutsche Bahn (DB). An extensive simulation study is conducted to measure the revenue improvement over the existing approach at DB and an EMSR-b based method. The results show a significant revenue potential by the new model that accounts for many effects ignored in both benchmark models (e.g., customer choice effect, network effect, continuous pricing). Further, networks of realistic size with more than 3,000 products and over 6,000 resources can be solved in only a few minutes. Therefore, time requirements for an application in practice are fulfilled.

4.1 Introduction

4.1.1 Motivation and overview

The passenger railway market for long-distance travel has experienced high interest in public and politics during the last years in different regions all over the world. New high-speed routes in Asia (e.g., Beijing-Shanghai in China, the magnetic levitation train Chuo Shinkansen in Japan), USA (California High-Speed Rail) or Europe (Rotterdam-Genua axis) are already open or currently under construction. These are just a few examples of global efforts to expand rail transport. In addition to the increasing mobility of people in general, the environmental aspect of travelling becomes more and more important. Rail transport is considered to be a green means of transport (e.g., in Germany <1g CO_2 emissions per passenger kilometer for electrified trains, 32g for buses, 139g for cars or 201g for aircrafts (Umweltbundesamt 2018, Deutsche Bahn 2018)), contributing to reduce CO_2 emissions and thereby to mitigate climate change. The increasing interest is also reflected in the growing number of passengers, e.g., +20% passengers between 2009 and 2018 at DB Fernverkehr, the long-distance division of the German rail company Deutsche Bahn (Deutsche Bahn 2019). At the same time, the railway business is very capital intense since initial and ongoing investments in rail tracks and trains are high so that not only costs have to be tightly controlled but also revenue must be managed effectively for a profitable business model. Therefore, advanced revenue management (RM) techniques can help to increase revenues by making more informed pricing and capacity allocation decisions. This can be achieved by optimally matching supply (i.e., the limited seating capacity) with demand (i.e., different customer segments with heterogeneous willingness to pay). Like for airlines, the decision could be opening/closing booking classes for each itinerary or leg (quantity-based RM) or defining the prices for each train connection (price-based RM).

Hohberger (2019b) shows that RM is already widely applied in practice at European rail companies. Sibdari et al. (2008) and Bao et al. (2014) discuss RM applications to the rail sector in the USA and China, respectively. At the same time, the techniques applied are typically less advanced compared to, e.g., the airline sector. The survey results presented in Hohberger (2019b) show that RM managers see a high potential for better forecasting and optimization methods.

Therefore, the aim of this paper is to present a choice-based network dynamic pricing approach that covers the most important rail characteristics mentioned in Hohberger (2019b) and to compare the performance in terms of total revenue with two benchmark models: the well-known EMSR-b heuristic and DB's current RM approach from practice. The optimization model was developed in close cooperation with practitioners to make sure that all relevant constraints from practice are fulfilled. Since both benchmark models are leg-based models, the results provided can give interesting insights also for other rail companies using similar approaches. The performance of the three approaches is compared and analyzed in a systematic simulation study. In the experiments, the proposed model yields a revenue increase of 13%-15%, thus leading to a significant improvement over the current approach. Further analysis of the solution times shows that still large-scale networks with approx. 3,000 products and over 6,000 resources can be solved in only a few minutes so that time limitations from practice are fulfilled. Additionally, detailed information about the simulation framework is provided so that the aspects mentioned can help in the design of future simulation studies.

The paper is structured as follows: First, a short literature overview on railway revenue management and revenue management in general with a focus on price-based models is given. Second, the proposed choice-based network dynamic pricing model is introduced and the solution approach to solve large-scale instances is explained. Subsequently, the results of the extensive numerical simulation study are presented with the goal to evaluate the revenue potential in practice and to present statistics for solution times and quality of the new approach for different networks and demand cases.

4.1.2 Literature overview

Research that deals with revenue management for passenger railway companies is still in the minority compared to more prominent industries like airlines or hotels. Ciancimino et al. (1999) can be considered to be the first and still one of the most prominent examples for passenger rail RM research, i.e., more than 25 years after the famous contribution of Littlewood and his two fare class rule for BOAC (now British Airways) in 1972 (Littlewood 2005). In the past two decades, more papers dealing with railway RM have been published and most of them consider quantity-based models (e.g., You (2008), Dutta and Ghosh (2012), Wang et al. (2016)), i.e., an optimal decision about product availabilities is determined. You (2008) extends the one fare class approach of Ciancimino et al. (1999) to a two-fare class approach, i.e., a low and a high fare for each product. Both are dealing with a network approach, i.e., multiple products share the same resource. It is interesting to see that early RM models for railways already integrate the idea of network optimization but practical applications still rely on leg-based models (Hohberger 2019b). In recent years, a high interest in railway RM in China can be observed with models typically dealing with single-fare network models due to the restriction of the Chinese government of having fixed itinerary prices (Bao et al. 2014, Jiang et al. 2015, Wang et al. 2016). In addition to quantity-based methods, there exist a few examples that use price as a decision variable, e.g., Sibdari et al. (2008) or Sato and Sawaki (2012), which are both single-leg models, or Hetrakul and Cirillo (2014) who consider a rail network where price and availabilities are optimized jointly. They also integrate a discrete choice model that covers the booking time decision of customers. Since the applied choice model does not take alternative train connections into account and the considered network is of limited size, they propose to extend the model to cover the mentioned characteristics from practice.

In addition to models developed for the special purpose of railway RM, there exist many publications for the general RM problem. Since the proposed optimization problem in this paper is price-based and uses a choice-based demand function, the focus in the following is on research on choice-based pricing models. For a broad overview of general RM models (incl. assortment or availability control) under customer choice see Strauss et al. (2018) who give a current overview. Publications on choice-based pricing can be subdivided on the basis of the choice model applied. Examples of pricing models that use a classic multinomial logit model (MNL, i.e., one single, homogeneous customer segment) are, e.g., Dong et al. (2009), Akcay et al. (2010), Li and Huh (2011) or Keller et al. (2014). Due to the IIA axiom (independence of irrelevant alternatives) in the MNL, advanced models such as nested logit or mixed logit models are applied to the pricing problem. For nested logit models, Li and Huh (2011), Gallego and Wang (2014), Li et al. (2015), Rayfield et al. (2015) or Davis et al. (2017) can be mentioned. They typically differ in the formulation of the nested choice model. For example, Li and Huh (2011) consider a nested logit model with nest-specific price-betas; Gallego and Wang (2014) account for product-differentiated price-betas and arbitrary nest coefficients. Li et al. (2015) allow for multiple nest-levels (d-level nested logit). Finally, pricing models that integrate a choice model covering multiple customer segments (i.e., discrete mixed logit or latent class) can be found in Keller et al. (2014), van de Geer and den Boer (2018), Li et al. (2018) and Schön and Hohberger (2019). While Keller et al. (2014) and Schön and Hohberger (2019) consider the *capacitated* multi-product pricing problem, Li et al. (2018) and van de Geer and den Boer (2018) neglect capacity considerations so that their findings cannot be transferred universally to the classical RM setting. As a final remark on discrete mixed logit models, Sen et al. (2018) consider the assortment optimization under this type of demand model. Due to its integer formulation, the approach can only be transferred to a pricing model with discrete price levels.

As one can see, research on railway RM is rare but existing. At the same time, for the general pricing problem, there exist many high-quality publications for different kinds of choice-models. A model that covers the railway specific constraints and a solution approach that makes it possible to derive a good solution in reasonable time for railway networks of realistic size, is still missing. This study fills the gap by proposing a suitable model and additionally providing results of a simulation to evaluate the performance in terms of revenue against two benchmark models.

4.2 The choice-based railway network dynamic pricing model

Based on the example of DB Fernverkehr, Hohberger (2019b) has presented the major aspects of DB's business model that are relevant for a RM system. This analysis is used as the basis for the proposed model in this paper (see Hohberger (2019b) for detailed explanations).

Structural characteristics

Due to the structural characteristics of a railway network (a resource is used by multiple itineraries), a network-based formulation seems to be promising. This allows the decision maker to take the network interdependencies accurately into account, e.g., if short-haul and long-haul travelers compete for the same (critical) resource. Further, since train connections with transfers are common, a new model should also allow to consider itineraries with transfers.

Representation of demand

From a demand perspective, multiple effects should be included. Due to the strong competition from cars, flights and long-distance buses (i.e., intermodal alternatives), adding these transport alternatives to the demand function allows a fast reaction to changes in the market (e.g., network expansion of an airline, sales campaigns of long-distance buses). Moreover, the demand function should not be based on the assumption of independent demand since customers do not make their product choice independently of other intramodal alternatives (i.e., other train connections earlier/later). Due to the missing fencing in DB's current system a demand model that accounts for buy-up and buy-down behavior can have additional positive impact. A discrete choice model that represents the product choice and uses choice probabilities to represent the demand is suitable for modelling both effects. This formulation would also help to determine a demand function that can automatically react to timetable changes by representing train connections on the basis of their attractiveness (e.g., travel time, no. of transfers etc.).

Further railway constraints

In addition, there are self-imposed business rules, such as the price consistency (i.e., a long itinerary should not be cheaper than a shorter one if they are overlapping) or having products other than the controlled tickets such as flexible tickets or annual tickets, which have an influence on the price decision. Furthermore, a second kind of price constraint can be applied with respect to different comfort classes, i.e., ensuring a minimum price difference between 1st and 2nd class.

Model formulation

Since a stochastic dynamic programming approach for the network case is impossible to solve for networks of realistic size due to the number of dimensions and the resulting calculation times (Talluri and van Ryzin 2005, p. 92), a multi-period, deterministic model could be a possible approximation. In addition to the prerequisites shown so far, there are also some aspects that may simplify a RM system. For example, the load factor is on average low and therefore many resources are not fully used (i.e., the capacity constraint is not binding) or that the capacity is only roughly fixed, since standing passengers are allowed at DB and therefore slight overloads do not lead to passenger rejections. Due to the missing fencing and also the lack of an external booking system (i.e., GDS such as Amadeus or Sabre for airlines) going from a traditional booking class approach to continuous pricing could help to further increase revenues by having the option to choose from more (in fact continuous) price points. Finally, from a practical perspective, a system should be able to react on (unforeseen) situations in reasonable time, i.e., frequent recalculations, especially during the days close to departure, should be possible.

Based on the practical requirements mentioned above, a network dynamic pricing approach that uses a discrete mixed multinomial logit demand model (MMNL) to represent the product choice of multiple discrete and heterogeneous customer groups is developed. It finds the optimal prices for each product (i, q) that is defined as the combination of an itinerary i and comfort class q (the mathematical representation of the model is introduced in 4.2.1). Since resources are limited, a capacity constraint that can also account for itineraries with transfers is integrated, which was a restriction in previous studies (e.g., Hetrakul and Cirillo (2014), You (2008)).
Furthermore, the optimization problem includes three different kinds of price constraints. First, the itinerary-price consistency constraints make sure that longer itineraries are more expensive than shorter ones (if travel paths are overlapping), second, comfort class price consistency constraints ensure a minimum price gap between a first- and second-class tickets. While the capacity considerations are also valid for, e.g., airline RM, the railway specific constraints, especially the price consistency between overlapping itineraries, are a special characteristic of a railway RM model. Furthermore, despite the fact that airline customers also have a choice between different flights, the number of offered alternative itineraries by a single company is typically smaller than in the railway example, especially for long-haul flights. Deutsche Bahn offers multiple itineraries between two cities within a day, with sometimes half-hourly services. That is the reason why the choice model plays a critical role in railway RM. The third kind of

price constraint ensures a lower and upper bound on the price variable, where the upper bound equals the flexible fare of the corresponding origin and destination (O&D) and the lower bound is, e.g., 19€ for a 2nd class ticket in the case of DB. Finally, prices can be chosen from a continuous set, thus offering more flexibility compared to traditional quantity-based approaches with discrete price points. The switch to a continuous pricing approach is already a development taking place in practice at other companies, e.g., Markus Brink, RM-Manager at Lufthansa Group, announced continuous pricing for 2019 (fvw 2018).

Currently, Deutsche Bahn is using a leg-based approach, which is typical for (European) railway companies as seen in the survey results in Hohberger (2019b). Demand is assumed to be independent, i.e., potential switching behavior between train connections as a result of price decisions are ignored. The proposed model has three main advantages over the current one: Decisions are derived on an itinerary-level to cover network effects, it accounts for customer choice behavior and can choose prices from a continuous set. The goal of the model and solution approach is to find a solution in reasonable time that is of good quality in order to fulfill the practical requirements. I.e., it can be applied to large and complex rail networks and is solvable within a given time limit so that frequent recalculations are possible. The general framework and optimization model is based on the 3-step approach proposed in Schön and Hohberger (2019) that is adapted to fit the railway setting and further improved to decrease solution times.

4.2.1 Model formulation

The starting point is the non-linear non-convex formulation of the choice-based network dynamic pricing model from Schön and Hohberger (2019). To adapt it to the railway setting, the railway specific constraints to cover the two types of price consistency (itinerary- and comfort class price consistency) are added as mentioned above. Furthermore, the following notations reflect the railway specific sets in the railway setting, so that the corresponding optimization model is denoted as RAIL_ORG. An itinerary $i \in I$ defines a train connection/itinerary between two train stations. The combination of i with a comfort class $q \in Q$ is denoted as product (i, q). Similarly, a leg $l \in L$ is defined as the edge between two train stations and the combination with comfort class q is denoted as resource (l, q). Set $z \in Z$ represents the combinations of different O&Ds and preferred departure time windows in the network, e.g., Frankfurt-Berlin between 9 a.m. and 11 a.m. (called *market segments*). $\dot{I}_z$ then represents the set of itineraries that passengers of market segment z consider as alternatives. Set $s \in S$ denotes the different *travel purposes* of passengers (e.g., business, leisure) and $t \in T$

the discrete booking periods. Following, a combination of s, z and t defines a *customer segment* (s, z, t). The role of the preferred departure time windows as part of the choice process of a customer segment is further explained at the end of this chapter. The set $\ddot{I}_{zl}$ includes all itineraries in market z that use leg l. The price for itinerary i in comfort class q if booked in time period t is denoted by p_{iqt} and represents the decision variable, ω_{szt} the market potential of customer segment (s, z, t), α_{izqs} the quality related utility in the logit model for product (i, q) in z and s, β_{sz} the price sensitivity parameter (with $\beta_{sz} < 0$) for s in market z and A^0_{szt} the outside option for s in market z and t. The parameter cap_{lq} represents the total seating capacity of a resource (l, q) that is reduced by the bookings on hand plus the forecast of travelers without a train-bounded ticket (e.g., flexible tickets, monthly passes) denoted by δ_{lq} (a number that increases during the booking horizon). Finally, p^{min}_{iq} defines the lower and p^{max}_{iq} the upper bound on the variable.

RAIL_ORG:

$$max \sum_{z\in Z}\sum_{i\in I_z}\sum_{s\in S}\sum_{q\in Q}\sum_{t\in T} p_{iqt}\omega_{szt}\frac{e^{\alpha_{izqs}+\beta_{sz}p_{iqt}}}{A^0_{szt}+\sum_{j\in I_z}\sum_{\bar{q}\in Q} e^{\alpha_{jz\bar{q}s}+\beta_{sz}p_{j\bar{q}t}}} \tag{4.1}$$

$$s.t. \sum_{z\in Z}\sum_{i\in \ddot{I}_{zl}}\sum_{s\in S}\sum_{t\in T} \omega_{szt}\frac{e^{\alpha_{izqs}+\beta_{sz}p_{iqt}}}{A^0_{szt}+\sum_{j\in I_z}\sum_{\bar{q}\in Q} e^{\alpha_{jz\bar{q}s}+\beta_{sz}p_{j\bar{q}t}}} \le cap_{lq} - \delta_{lq} \quad \forall\, l \in L, q \in Q \tag{4.2}$$

$$p_{iqt} \le p_{jqt} \quad \forall\, j \in I, i \in \ddot{I}_j, q \in Q, t \in T \tag{4.3}$$

$$p_{iqt} + \rho \le p_{i\bar{q}t} \quad \forall\, i \in I, q \in Q, \bar{q} \in \dot{Q}_q, t \in T \tag{4.4}$$

$$p^{min}_{iq} \le p_{iqt} \le p^{max}_{iq} \quad \forall\, i \in I, q \in Q, t \in T \tag{4.5}$$

$$p_{iqt} \in \mathbb{R}^+ \tag{4.6}$$

The objective function to be maximized (4.1) sums up the ticket revenue from all products (i, q) and all customer segments (s, z, t). The capacity constraints (4.2) make sure that trains are not overcrowded by summing up the demand on the left-hand side that uses resource (l, q). The railway-specific constraints (4.3) (called itinerary-price constraint) make sure that longer itineraries are not cheaper than shorter ones, if the travel paths are overlapping and therefore potential for arbitrage could exist. Set $\ddot{I}_j$ includes all itineraries that must fulfill itinerary-price consistency with respect to itinerary $j \in I$. Constraints (4.4) (called comfort class price constraint) ensures a minimum price gap (ρ) between different comfort classes $(\bar{q} \in \dot{Q}_q)$; a rule that is desired by the rail company from a price-strategic reason to distinguish higher and lower quality products also from a pricing perspective. Especially itinerary-price constraints cause strong network effects since many train connections are now interrelated not only by the capacity constraint but also by these additional price restrictions. Finally, constraints (4.5) define a lower and upper bound on the variable.

The demand function is represented by a discrete mixed multinomial logit choice model that can account for multiple customer segments. This is important in the railway case to account for different travel purposes $s \in S$ and also preferences for departure time windows ($z \in Z$, e.g., morning, noon, afternoon). The choice model describes the choice of a product from a set of alternative products, i.e., it reflects the switching behavior of customers depending on product attributes like travel time, number of transfers, comfort-level and most importantly price. Especially in a RM setting with capacity consideration, the integration of a discrete choice demand function into the price optimization model helps to anticipate the switching behavior as a result of price changes. This is a feature that a classic independent demand function cannot cover. The segmentation by preferred departure time helps to describe different preferences of customer groups from a departure time perspective. Customers preferring to travel in the morning probably do not consider a train connection in the evening as a possible alternative but consider a train connection at noon as possible (but less attractive) alternative. I.e., each customer (segment) has a unique departure time preference z but is also willing to deviate from its original plan to some (limited) extend so that also train connection outside z are considered. Thus, a portion of the utility parameter α_{izqs} covers this deviation from the preferred time window z. Figure 8 illustrates the idea of segmentation by departure time.

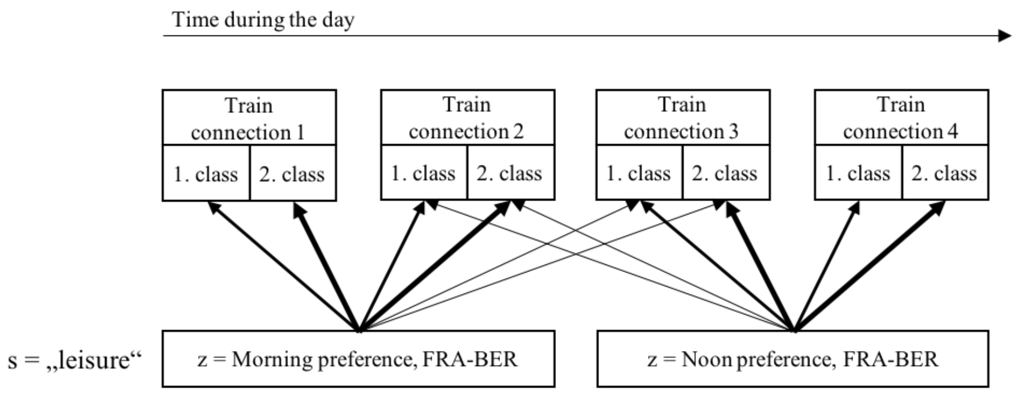

Figure 8: Segmentation by departure time preference. Bold arrows represent higher preference value.

Although RAIL_ORG is a multi-period model, the idea would be to reoptimize the model frequently to account for unforeseen demand realizations and to update the optimal pricing decisions.

The results of the numerical study in Schön and Hohberger (2019) for a railway example without the railway specific constraints have already shown that directly solving the non-linear non-convex model (here RAIL_ORG) is not practicable in terms of solution time and the evaluation of the quality of the solution due to the missing guarantee of a globally optimal solution is not possible. With the proposed 3-step approach, they are able to decrease solution times by a factor of up to 56 and can give information about the maximum gap to the (still unknown) globally optimal solution of the non-convex model. In the following only basic insights into the heuristic are given and the focus lies on the adjustments that were made to fit RAIL_ORG. Model formulations of RAIL_INV and RAIL_INV_LOG, that have to be solved as part of the heuristic, are given in the appendix. For further details of the 3-step approach and the general dynamic pricing model see Schön and Hohberger (2019).

4.2.2 Solution approach

The idea of the 3-step approach is to solve three slightly different models in a row, where RAIL_INV and RAIL_INV_LOG both are fast and convex and therefore able to derive a warm start solution for solving the original, non-convex model formulation (RAIL_ORG). Figure 9 gives an overview on the solution approach.

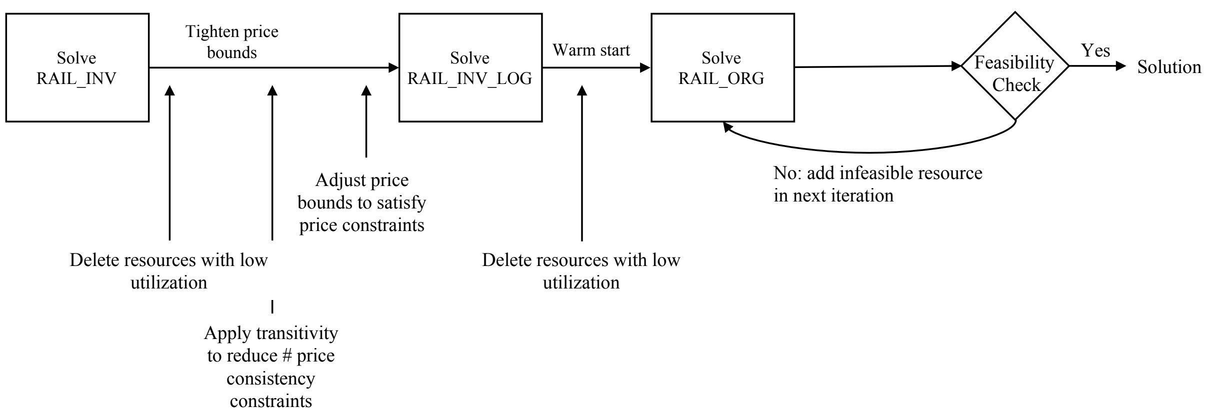

Figure 9: Overview of adapted 3-step approach for the railway model

RAIL_INV uses the inverse price-demand function so that the decision variables in that model are the choice probabilities for each product and customer segment (compared to p_{iqt} as in RAIL_ORG). This is a typical approach used by other researchers before to transform the non-convex model to a convex one (e.g., Schön (2010a), Keller et al. (2014)). The remaining challenge with a discrete mixed logit demand model in RAIL_INV is that the optimal product prices are potentially different for each customer segment (i.e., $\dot{p}_{iqtsz}$). Since segment-individual pricing is typically not of interest in practice, an additional constraint has to be added that ensures a uniform price for each product. This is done in RAIL_INV_LOG by a uniform-price constraint that ensures a unique price p_{iqt}. Since this new constraint is non-convex in its original form, the non-convex part of the constraint is approximated by a linear function. Thus, RAIL_INV_LOG includes only an approximate uniform price. The motivation of solving RAIL_INV beforehand is that the solution of RAIL_INV is used to improve the approximation in the uniform-price constraint in RAIL_INV_LOG. The solution of RAIL_INV_LOG can then be used as a warm start for solving RAIL_ORG as a last step of the solution approach to find a good, feasible solution since the outputs of both RAIL_INV and RAIL_INV_LOG must not deliver a feasible solution for RAIL_ORG. Additionally, since the optimal objective value of RAIL_INV is an upper bound to the (unknown) maximum revenue of RAIL_ORG, the former serves as a reference point to evaluate the quality of the (possibly local) solution of RAIL_ORG.

Since the 3-step approach is originally developed for the general dynamic pricing case, it is adjusted to fit the railway specific setting. Two main adjustments are considered: Since RAIL_INV_LOG introduces a specific price variable p_{iqt} , the two types of price constraints (itinerary-price and comfort class price constraints) can only be represented with a convex representation from this point on. To make sure that the tighter price bounds derived with RAIL_INV allow a feasible solution with respect to the newly introduced constraints, an additional step is applied before solving RAIL_INV_LOG. It ensures that two products which must fulfill price consistency have overlapping price ranges to make a feasible solution possible. The parameter that determines the size of the overlapping price range has to be defined by the user. The increase of the price range comes at the cost of a slightly worse linear approximation. The same kind of check and possible adjustment of overlapping price ranges is performed for the comfort class price consistency.

Furthermore, pretests have shown that the high number of additional price constraints (especially the itinerary-price consistency) makes it impossible to solve RAIL_INV_LOG and RAIL_ORG in acceptable time, even with a warm start. To further improve solution times, two ways to decrease the number of constraints in RAIL_INV_LOG and RAIL_ORG are proposed in the following section.

Itinerary price consistency constraints

Within a large-scale rail network, there exist multiple overlapping train connections and thus a high number of price consistency relations. Overlapping in this context means that an itinerary

B can also be used to travel itinerary A, since it includes all resources from A. Figure 10 shows that itinerary A is overlapping with B & C and B is overlapping with C so that the prices of these three products should fulfill the following three conditions: $p_A \leq p_B$ and $p_A \leq p_C$ and $p_B \leq p_C$.

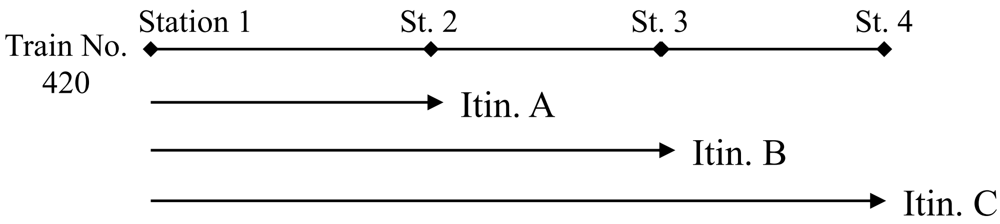

Figure 10: Example for itinerary price consistency

By applying the concept of transitivity, one can easily reduce the number from three to two constraints since $p_A \leq p_C$ can be expressed by the combination of the other two constraints ($p_A \leq p_B$ and $p_B \leq p_C$). It reduces the number of relations significantly. For a network that consists of DB's Top 400 O&Ds with approx. 10,000 itineraries, the number of price consistency constraints can be decreased from approx. 213,000 to 53,000 (-75%) without losing any accuracy.

Capacity constraints

The benefit of the 3-step approach is not only a good warm start solution for RAIL_ORG it can also be used to gain approximate information about the utilization of resources in the network from the intermediate results of RAIL_INV and RAIL_INV_LOG. Constraints that are not binding in the optimal solution can be removed without any change in the optimal solution. Typically, one has no knowledge about the capacity utilization before an optimal solution is derived and therefore has to solve a model with the full set of constraints. But in the 3-step approach, one can derive information about the utilization from RAIL_INV and therefore decrease the number of capacity constraints for RAIL_INV_LOG. In a similar way, the solution of RAIL_INV_LOG can be used to remove (probably non-binding) capacity constraints in RAIL_ORG.

With the solution of RAIL_INV, one can derive a utilization rate in % for each resource (l, q). Since the comfort class price constraint is only introduced in RAIL_INV_LOG and it can generate a significant number of passengers switching from first to second class, the rule that a resource (l, q) is deleted from the set of resources is applied if the utilization rates of (l, q) and $(l, \bar{q})$ are both $\leq \zeta$ (where ζ is a parameter to be defined). Similarly, the solution from RAIL_INV_LOG is used to derive updated resource utilization rates and a resource is deleted from the set if the utilization of (l, q) is $\leq \zeta'$ (where ζ' is a second parameter to define). The optimal solution from RAIL_ORG (with the reduced set of resources) is finally tested for feasibility for the full set of resources. In case of an infeasibility, RAIL_ORG is again solved with the additional constraints that cause the infeasibility. From a practical point of view, it is also possible to accept a slight overbooking in the feasibility check, e.g., overbookings of 1% of the capacity are considered to be fine and no recalculation is conducted. Especially in the railway case, where standing passengers are allowed, setting prices that yield an expected utilization slightly above 100% would be acceptable to avoid additional iterations of solving RAIL_ORG. To achieve a feasible solution directly in order to avoid frequent recalculations, one has to carefully choose the value of ζ' so that the number of constraints is significantly reduced but at the same time all critical resources are still considered. Historic demand data of DB show that there typically exists only a small amount of bottleneck resources and thus most of the resources can be ignored. In the simulation study values of $\zeta = 70\%$ and $\zeta' = 90\%$ are used.

4.3 The simulation framework

From a practical perspective, a new model is as good as it outperforms an existing approach. To evaluate the performance of the new choice-based network dynamic pricing model (denoted as *O&D*) an extensive market simulation is conducted that makes it possible to construct various, stochastic demand scenarios (explained below) and finally compare the total revenue of the new approach with alternative RM approaches. Therefore, prices for each product are derived under each considered RM model and a series of randomly generated arriving customers is constructed to derive a buying decision accordingly. In this study, the new *O&D* model is compared with two alternative RM-approaches: *Benchmark approach "DB"* reflects the current DB-approach and *benchmark approach "EMSR-b"* is based on the well-known EMSR-b heuristic. Since EMSR-b is a prominent heuristic in practice, the results of *"EMSR-b"* are probably more universally transferable to other rail companies and *"DB"* gives a practical insight about possible revenue improvements at the specific example of DB Fernverkehr. Since Hohberger (2019b) has shown that leg-based approaches are still dominant in practice (and both benchmarks rely on leg-controls) the results of the simulation provide interesting insights also for other rail companies.

4.3.1 Benchmark policies

Benchmark approach "EMSR-b"

EMSR (i.e., expected marginal seat revenue) is a well-known heuristic for the single-leg, quantity-based RM problem (Belobaba 1987). It is the extension of Littlewood's rule (assuming two fare classes) to the multiple fare class case and is widely applied in practice (Talluri and van Ryzin 2005, p. 45ff.). While there exist two versions of the heuristic (EMSR-a and EMSR-b), EMSR-b is chosen to be a benchmark approach since it is one of two models applied by DB Fernverkehr in practice (see explanations for the second benchmark *"DB"*). EMSR-b derives nested protection levels for each resource and booking class based on an independent leg forecast. A booking class (≠ comfort class) in DB's RM system is simply defined as a discrete price step without any fare rule. Following, the value of the booking class availability defines the maximum number of tickets to sell for the corresponding price level. Since EMSR-b can be implemented by using simple formulas, it is very fast in deriving the decision even for large networks and can also be parallelized very well. For detailed insights into the calculations for EMSR-b-based protection levels see Talluri and van Ryzin (2005).

Since EMSR-b is a single-leg heuristic, it does not derive a price for multi-leg products directly. To derive the final product prices, DB's way of transferring EMSR-b based booking class availabilities to product prices is applied. The following example illustrates the approach: First, for a product (e.g., a comfort class "2" ticket for a scheduled train connection departing from B at 10:00 a.m. and arriving at E at 11:35 a.m.) it is evaluated which of the three used resources (B-C, 2), (C-D, 2) and (D-E, 2) is the most restrictive one according to the EMSR-b booking class availabilities. Assuming (C-D, 2) is binding, where only booking class 3 is available, then $\hat{p}_{B-E,2}^{prelim} := \tilde{p}_3^{b.class}$, where $\tilde{p}_3^{b.class}$ is the price corresponding to booking class 3. Second, there exists for each product (i, q) a specific set A_{iq} of possible discrete prices, e.g., $A_{B-E,2} = \{19,29,39,59,79\}$. Both information combined is then used to derive the final O&D price $\grave{p}_{B-E,2}^{EMSRb} := \min_{a \ in \ A_{B-E,2}:\ a \geq \hat{p}_{B-E,2}^{prelim}} (a)$, meaning the smallest price larger than the booking class price is chosen. Or, in general form, for comfort class 2 itineraries: $\grave{p}_{i,2}^{EMSRb} := \min_{a \ in \ A_{i,2}:\ a \geq \hat{p}_{i,2}^{prelim}} (a)$. While the second-class price is derived as described, the first class price must additionally fulfill the requirement that a minimum price gap of ρ to comfort class 2 must hold, so that the following formula applies for first class prices: $\grave{p}_{i,1}^{EMSRb} := \min_{a \ in \ A_{i,1}:\ a \geq \hat{p}_{i,1}^{prelim} \wedge a \geq \grave{p}_{i,2}^{EMSRb} + \rho} (a)$.

Benchmark approach "DB"

The current system consists of two parallel optimization procedures, an EMSR-b based booking class availability control as described in the previous paragraph, and a rule-based approach that covers the willingness to pay of customers (also controlled on a leg level; in the following denoted as discount-control). The rules are defined manually by the RM personnel based on market tests and expert knowledge. The final price of a product is defined by the most restrictive approach (i.e., either the EMSR-b or the discount-control). The underlying idea is having the EMSR-b-control to increase prices in case of bottlenecks along the itinerary and second, the discount-control to optimize for willingness to pay of customers, depending on which approach dominates. Different from an EMSR-b-control, it does not control booking classes with fixed prices. Instead, it determines (also for each resource) the discount with respect to the flexible fare p_{iq}^{max} that is granted for any product that uses the resource, where the various discount levels are given by a discrete discount set, e.g., $D = \{70\%, 50\%, 20\%, 0\%\}$. To achieve lower prices, discount classes with higher discounts are opened and vice versa. Similar to the EMSR-b approach, the most constraining resource that a product uses determines the discount that is applied on the flexible fare of that product. Since the flexible fare p_{iq}^{max} is O&D/product-specific and therefore higher for longer itineraries than for shorter ones, the discount-control does take some O&D consideration into account, which is in fact the main reason why it is applied in practice. The availabilities of the discount classes in practice are controlled manually by the RM managers. Since this kind of manual control is hard to transfer to a computer driven simulation, an approach that automatically derives the resource-based discount levels was developed together with DB in order to represent the idea of the discount-control.

4.3.2 Framework of the simulation

Williamson (1992) discusses and proposes a general framework for a simulation to evaluate the performance of RM-techniques. This simulation study is based on her findings and is adjusted so that it fits the railway environment and choice-based passenger behavior. First, Williamson proposes to split the booking period into discrete time intervals rather than just considering a single (long) booking period. It has the advantage of updating the availabilities/prices in case of unforeseen demand realizations at the beginning of every period by reoptimizing the model and also better representing an application in practice, where data are frequently updated over the booking horizon. It can therefore be described as a dynamic simulation. Second, one has to decide how the randomly distributed demand in each period is derived. Demand in that sense can be divided into 1) the random number and segment membership of arriving customers and 2) the random product choice of each simulated customer. For simulating the arriving customers, different distributions are discussed by Williamson (1992). The well-known and widely adopted normal distribution has the disadvantage of being continuous, which is an unrealistic behavior when modeling discrete passengers and it can lead to negative demand realization, so it has to be adjusted to avoid negative realizations. The gamma distribution has the disadvantage of allocating zero probability for a zero-demand realization and is also continuous. As mentioned by Williamson, especially in the O&D case, it is not uncommon to realize zero demand for a specific O&D product within a booking period and thus an unrealistic assumption of the gamma distribution in this context. Finally, the Poisson distribution seems to be a promising way to represent the random demand over time. Due to its discrete formulation it avoids continuous demand realizations and it is truncated at zero, so that negative draws are avoided but values equal to zero still possible. Williamson (1992) and Lee (1990) conclude that using a Poisson distribution is a valid approach to randomly model arriving passengers in an O&D simulation. For further information and a comprehensive discussion of creating a simulation study in RM see Williamson (1992, Ch. 5). More details on how the demand is randomly created for this simulation study are given below.

The simulation is constructed as follows: The booking horizon of 180 days is split into 9 discrete time intervals. At the beginning of each time interval, the booking class availabilities in *benchmark model "DB"* and *"EMSR-b"* and the prices based on the *O&D* model are derived, taking the remaining capacities into account. Within each time interval, a series of customer requests is simulated, each of it characterized by a specific desire for an O&D, departure time and travel purpose, so that customers differ in the consideration set and the evaluation of products. This segment membership is randomly created for each arriving customer based on the empirical mix. Following, one can derive the MMNL choice probabilities for each customer for all considered product alternatives with prices based on *"O&D"*, *"DB"* and *"EMSR-b"* and finally simulate the product choice for all three models separately. The generated revenue from ticket sales in each period and for each of the three approaches can then be easily calculated. After all buying decisions for a specific period are simulated, the simulation proceeds with the next period (i.e., the capacities are updated based on realized ticket sales and all three models are solved again). In case of a fully booked resource, the simulation immediately reacts (i.e., within a time interval) and closes the sales of all products that use this resource and includes the information into the following simulated choice situations by making these products not available. Although this increases the complexity of the simulation, it is necessary to avoid the occurrence of overbooking.

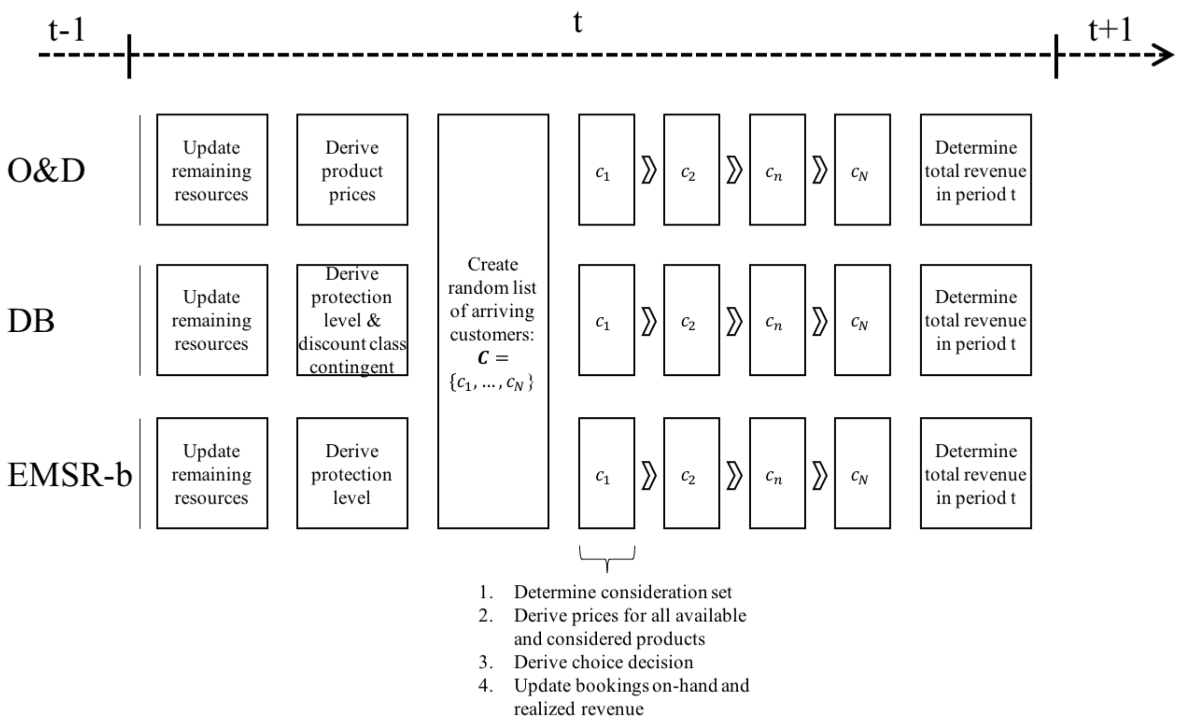

Figure 11: Overview on processed steps in each period

Figure 11 shows the flow chart of the simulation for one exemplary simulation period. In the following, further insights into the structure and assumptions of the simulation are explained.

Deriving leg based independent demand forecast

The goal of the simulation is to obtain insights into the performance of the three different optimization approaches, i.e., the results should not be driven by more accurate data for a specific model. Since the EMSR-b heuristic (that is part of *DB* and *EMSR-b*) is based on the independent demand assumption and uses a traditional forecast for booking classes for each resource, it differs from an O&D-choice-based forecasting approach where the O&D market demand and the choice model itself represents the price-sales relationship. Talluri and van Ryzin (2004) already mention in their study that there is no unique, correct way to derive forecasts from a choice-based demand model for an optimization approach that uses the independent demand assumption. To make sure that the forecasting quality is similar and the final comparison as fair as possible, the O&D-choice model is used to derive a leg-based forecast. I.e., first the demand for each product (i, q) in each booking class b and time period t is defined as

$\varphi_{iqbt} := \sum_{z \in \dot{Z}_i} \sum_{s \in S} \omega_{szt} \frac{e^{\alpha_{izqs} + \beta_{sz} \tilde{p}_b^{b.class}}}{A_{szt}^0 + e^{\alpha_{izqs} + \beta_{sz} \tilde{p}_b^{b.class}} + \sum_{\mathrm{j} \in \dot{I}_z, j \neq i} \sum_{\bar{q} \in Q} e^{\alpha_{\mathrm{j}z\bar{q}s} + \beta_{sz} \bar{p}_{\mathrm{j}\bar{q}t}}}$, where $\bar{p}_{\mathrm{j}\bar{q}t}$ is the price from the solution of RAIL_ORG for all alternative products. To derive the incremental demand for each booking class, define $\varphi_{iqbt}^{incr} = \varphi_{iqbt} - \varphi_{iq\,(b-1)\,t}$ and finally determine the leg-based forecast as $\mu_{lqb} = \sum_{i \in \tilde{I}_l} \sum_{t \in T} \varphi_{iqbt}^{incr}$. Further, a normally distributed demand is assumed which is in line with DB's approach. Since the procedure of deriving the leg-based demand as explained above does not deliver any knowledge about the variance, the variance is assumed to equal the expected demand for each leg. This approach of deriving a leg-based forecast has the advantage that the same source of data is used for *O&D*, *DB* and *EMSR-b*. Although the leg-based forecast is derived by a choice model, the values are finally considered as fixed/independent, i.e., do not dependent on other product or booking classes availabilities and therefore fulfill the independent demand assumption.

Simulating the number of customer arrivals

The number of arriving customers (i.e., total market potential) in each booking period t is assumed to be Poisson distributed with mean $\hat{\omega}_t$, where $\hat{\omega}_t := \sum_{s \in S} \sum_{z \in Z} \omega_{szt}$ for $t = current\ period$, i.e., $\hat{\omega}_t^{realized}$ is the randomly generated market demand value based on a Poisson distribution with mean $\hat{\omega}_t$. Since $\hat{\omega}_t^{realized}$ does not include any information about the desired departure time, O&D and travel purpose, one has to randomly assign its market z and travel purpose s to each of the customers in $\widehat{\omega}_t^{realized}$.

The allocation of an arriving customer request to market z and travel purpose s is modelled as a discrete random variable where the corresponding probabilities are estimated based on the empirical mix. It means that both the realized number of arriving customers and also the segment allocation are determined randomly. This approach also ensures that customers arrive in a mixing order, which is especially important close to departure with limited remaining capacities. Furthermore, demand in all discrete periods is assumed to be independent, i.e., a higher random demand realization in period t has no influence on the random demand realization in the following periods $t+1,\ldots,T$.

Simulating the customer choice behavior

In addition to the randomly generated arrival process of customers, the choice of each customer must be modeled to derive the choice decision. Two different choice-settings are constructed. In the first one (called *perfect forecast*), the choice model and data that are used to derive forecasts for *DB* and *EMSR-b* and are used in *O&D* are applied. This case corresponds to the practical situation where choice data (i.e., the α and β values) can be perfectly estimated.

In a second setting (called *imperfect forecast*), it is assumed that choice data cannot be perfectly estimated in practice. Therefore, choice parameters α_{iqsz} and β_{sz} are adjusted for each (s,z) and A_{szt}^0 for each (s,z,t) by a factor that is randomly drawn from the uniformly distributed interval $[0.7\,,1.3]$ and the buying decision is derived accordingly (similar to Vulcano et al. (2010)). This case, which is probably the more realistic scenario in practice, can give insights about the robustness of the revenue gap between the models.

In addition to the randomization of the choice parameters, it would also be conceivable to use a choice model with a different structure to derive the choice decision in the simulation (e.g., nested logit). Talluri et al. (2010) explain that it is a common approach in research to use the exact same model in optimization and simulation. However, they also mention the fact that on the basis of this approach only the improvement by the algorithm can be tested and no statement can be made about the validity of the choice model itself. Since the focus is on the evaluation of the optimization methods and the MMNL is able to represent heterogeneity in product evaluation of different customer groups, this approach seems reasonable.

Further assumptions

In the simulation, one full, specific departure day in the future is simulated. In fact, also for a practical implementation solving a series of single departure days is recommended instead of solving one single model that reflects all 180 departure days in future. The seat capacity in a comfort class on a train is defined as the limit for ticket sales (i.e., in case a resource is fully booked, all products that use the resource are marked as sold out). For simplicity, the case of cancellations is ignored and single arriving customers are assumed (i.e., no group bookings). Further, for solving RAIL_ORG the remaining booking periods are defined as two periods where $t = 1$ is the current planning period and $t = 2$ is the aggregated future. It reduces the number of variables to optimize but still makes it possible to consider future high-fare demand. Additionally, when testing the solution of RAIL_ORG for feasibility in the 3-step solution approach, an overbooking of up to 1% of the resource capacity is accepted. To account for uncontrolled customers (i.e., customers with flexible tickets or annual passes) the current practice is applied, i.e., the capacity is decreased at the very beginning of the first booking period by δ_{lq} based on historical data.

4.3.3 Test instances and data

Test networks

To evaluate the performance of the new O&D model the simulation is run for two different test networks, both based on the schedule of DB Fernverkehr. The first network covers the top O&Ds on a north to south axis in Germany (from Hamburg to Stuttgart). I.e., all important combinations of cities/train stations (=O&Ds) along that route are chosen and further, all itineraries between these O&Ds are considered. This network is called "North-South axis". The second test network is constructed by taking only the O&Ds of the entire long-distance network into account that generated high revenue based on internal data of DB Fernverkehr. This network is called "Top 50 O&Ds" since the 50 most valuable O&Ds are chosen. The rationale behind the two networks is having one network ("North-South axis") that has a dense, sequential structure with many overlapping itineraries and a mix of short- and long-haul O&Ds and a second highly interconnected network ("Top 50 O&Ds") that represents a realistic network for a practical application. Using various networks gives an indication of how much the results depend on the network structure. Table 12 shows some statistics about both networks.

Network	**# itineraries *i***	**# products *(i,q)***	**# resources *(l,q)***
North-South axis	667	1,334	1,080
Top 50 O&Ds	1,523	3,046	6,384

Table 12: Statistics for test networks

Data

The necessary data to solve *O&D, DB* and *EMSR-b* can be split into two categories. First, operational data like the set of products, the itinerary-leg consumption matrix, available resources etc. and second, demand related data like the α and β values for modelling the customer choice or market demand parameter ω. For the operational data, the network data generation algorithm (NDGA) from Hohberger and Schön (2019) is applied to the full schedule of DB to derive itineraries and corresponding data for DB's full network. In a following step, only the relevant data for the test network are extracted, so that all itineraries and corresponding data are included in the data set for each network. For the demand model, internal choice model data of DB Fernverkehr are applied (currently used for medium-term planning purposes), i.e., sensitivities (β) with respect to price and product attributes like travel time or number of transfers. Together with the attribute values (e.g., travel time, number of transfers) from the data generation algorithm, the non-price utility parameters α can be calculated. Further, based on internal data about historic market shares and prices the values for the outside option A^0 are derived.

Demand cases

To evaluate the performance under different demand scenarios, three demand cases (low, base and high) are defined. The base case demand scenario is calibrated in a way that the number of bottleneck resources (in %) in the simulation results of *DB* (i.e., the current DB-approach) equals the numbers from DB's booking history. A resource (l, q) is defined as a bottleneck resource if the utilization is larger than 95%. Numbers from DB's booking system show that on a standard day in 2018 4.1% of the resources are utilized by more than 95%, so the demand ω_{szt} is calibrated in the base case of the simulation accordingly. The low demand case is defined by a demand decrease of 30% and the high demand case by an increase of 30% over the base case.

Another possibility to define demand cases would be using the average utilization (i.e., load factor) in the network. Since an average network utilization rate does not give any information about the distribution over the network, especially about the number of highly utilized capacities which is crucial in network RM, using the bottleneck definition seems to be more suitable to construct realistic demand cases.

4.3.4 Simulation of special O&D model versions

In addition to the main simulation explained above, two adapted versions of the O&D based approach are tested. The first special case is denoted as "***Discretized Prices***". In this approach, the optimal continuous product prices of *O&D* are rounded to the closest discrete price given by a predefined set by the company (see A_{iq} from *EMSR-b*). The motivation behind is that legacy RM systems and sales platforms in practice are often based on the logic of booking classes and implementing continuous pricing policies would require major systems changes. Instead, mapping continuous prices to booking class prices could be a viable workaround. An interesting question is therefore how big the effect of continuous pricing is and how big the sales potential of the O&D model would be if one were still bound to discrete price levels. Since limiting the price variable in RAIL_ORG to a set of discrete prices would lead to a non-linear integer problem and is thus intractable to solve for large networks, the rounding heuristic described above is used. Also, RM managers from DB Fernverkehr argued that it is a reasonable way for a first practical application and therefore interesting for practice.

The second special case is denoted as **"Approximate Prices"**. The idea is to use the approximate prices from the solution of RAIL_INV_LOG since they are feasible in all self-imposed business rules (i.e., lower and upper price bounds, itinerary-price consistency and comfort class price consistency). The advantage of using the prices of RAIL_INV_LOG is that solution times for different instances have a low variance for both convex models (RAIL_INV and RAIL_INV_LOG), while RAIL_ORG has a high variance in high demand situations (see 4.4.2). Since the available solution time is limited in practice and predictability is important, the solution of RAIL_INV_LOG could be used as a valuable fallback. The disadvantage is that these prices may not be feasible with respect to the capacity constraint and probably do not yield the largest revenue in the simulation. Since the simulation is constructed in a way that sold out products are immediately removed for all following passengers from the set of available products, the infeasibility can be realistically reflected in the simulation.

Figure 12 gives an overview of all different settings that are tested as part of the simulation study. In total $3 * 2 * 2 * 3 = 36$ settings are created and for each combination in the main simulation 10 randomized instances and in both special cases 5 randomized instances are solved. This is done to report average results over various random simulation instances.

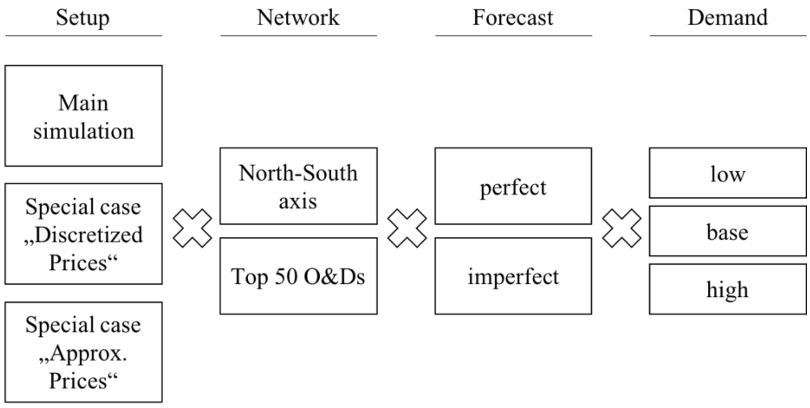

Figure 12: Overview on different settings in the simulation

4.4 Simulation results

In the following, the results of the simulation study are presented. First, the advantages of the new O&D based model over both benchmarks are qualitatively discussed. Second, quantitively results of the main simulation are presented. That includes the analysis of solution time and revenue improvement compared to the benchmark models. Finally, results of both special cases are presented and discussed.

4.4.1 Qualitatively discussion of O&D model

The overall improvement of the O&D model can be split into four main categories that are currently ignored in both benchmark models.

Controlling products/itineraries: The possibility to control itineraries instead of legs increases the number of control variables and thus makes more precise decisions possible. With a leg-based approach like *EMSR-b*, one single resource influences the price of multiple products simultaneously. One can imagine the *EMSR-b* control in the railway context as if a single RM manager was just trying to maximize revenue on that specific resource and ignoring sales and resources in the remaining part of the network. In fact, having the flexibility of controlling prices at the product level instead of the resource level can be seen as the enabler for the following three improvements.

(Traditional) network effect: Decisions on an itinerary level make it possible to account for (traditional) network effects, i.e., a resource is used by multiple products (short- and long-haul itineraries). Thus, in case of high demand, a decision which product request to accept or reject has to be made precisely. In the price-based formulation, the network consideration leads to accurate pricing decisions (i.e., increasing prices in various intensity) for the various products that use the same bottleneck resource.

Choice-based network effect: Integrating the discrete choice demand function into the optimization model allows for anticipating the switching behavior to alternative products (i.e., a switch to a different comfort class or to alternative train connections earlier/later). Especially in case of a bottleneck resource it can help to make the right decision, e.g., “Is it better to increase the prices of the products that use a bottleneck resource or is it better to decrease prices of alternative products, or both?”.

Continuous pricing: Finally, having the option of choosing prices from a continuous set increases the solution space and can thus increase revenue.

4.4.2 Results of the main simulation

Solution time

For a practical application, a model should be able to find a solution in reasonable time. Typically, reoptimization takes place over night. Additionally, close to the departure day also an intraday reoptimization should be possible to react immediately to unforeseen demand realizations. Assuming that for the future 180 departure days that are available for sale, daily reoptimization takes place only for the departure days close to departure (i.e., with ≤ 30 days remaining booking horizon) and every 3 days for all days with a longer remaining booking period, a company would have to solve 30 + 50 = 80 instances each night. With a nightly batch of 8 hours, a model should be solved in under $8 * 60/80 = 6$ minutes when using one machine. Since each model is separated (i.e., departure day 67 in future does not depend on departure day 64) one could easily parallelize the optimization. In addition to the (average) solution time, the variance is important from a practical perspective as well. A model that has a high variance for the solution time is somewhat unpredictable. The presented statistics are the averages of the ten random simulation instances for the perfect forecast scenario in the main simulation. In each simulation instance the O&D model is solved nine times (once for each booking period). Since the solution of the first period is the same for each of the ten instances (since randomization only has an impact afterwards) and the last period has approximately half the number of variables (since there is only t=1, and no future period t=2), the solution time results of booking period two to eight are averaged to give an unbiased overview. I.e., each number is an average over $7 * 10 = 70$ data points. For the simulation and optimization, AMPL and Knitro 11.0.1 under Windows 7, Intel Xeon E5 with 3 GHz 16 cores, 64 GB RAM were used. In total, the whole simulation for all instances in the different scenarios took approx. 105 hours (or four and a half days), while most of the time is caused by the simulation itself (and not the optimization of *O&D, DB* or *EMSR-b*).

For both networks in all demand cases, the proposed 3-step approach is able to find a feasible solution in under 6 minutes. Table 14 summarizes the solution times for the different settings. Solving the smaller, sequential network “North-South axis” takes on average 26 seconds in the base case, with slight decrease (increase) for the low (high) demand case and an acceptable standard deviation. Interesting to see is that RAIL_INV and RAIL_INV_LOG have larger solution times than RAIL_ORG. In addition to the warm start that is provided, also the reduction of capacity constraints helps to improve solution times (both information is a result of the previous solved model RAIL_INV_LOG). For the larger network (Top 50 O&Ds)

solution times are as expected larger but still fulfill the practical requirements. What is remarkable, however, is that the solution times for RAIL_ORG increase significantly with increasing demand, while the solution times of the two convex models (RAIL_INV and RAIL_INV_LOG) remain relatively stable (with also low variance). The same applies to the variance where RAIL_ORG has a larger standard deviation in all demand cases compared to the convex models, with strong outliers in the high demand scenario. In fact, this result was the motivation for constructing the special simulation case *"Approximate prices"* to test how the solution of RAIL_INV_LOG performs in case RAIL_ORG does not find a solution within a given time limit.

Finally, since RAIL_INV is a relaxed version of RAIL_ORG it serves as an upper bound and therefore enables the evaluation of the maximum possible deviation of a feasible solution from the globally optimal solution (under the consideration that a slight violation of the capacity constraint in RAIL_ORG is accepted). The gap is with approx. -2% quite small (i.e., compared to RAIL_INV the objective value in RAIL_ORG decreases by that percentage) and can also be explained by additional restrictions that apply in RAIL_ORG. Since the mentioned data are averages, a detailed look into single instances gives additional information. It can be observed that the largest gap achieved in the analyzed data is -3.3% gap for an instance in the North-South axis and -2.7% for the Top 50 O&D network.

Revenue impact

The analysis of the revenue impact provides an overview about the potential that can be generated in practice and serves as an important argument for an implementation decision. Benchmark models *DB* and *EMSR-b* are used as the basis for calculating the revenue improvement of the O&D model (in %). This means that positive percentages show an improvement through the application of *O&D.*

Table 15 summarizes the results and provides additional information about the utilization in the network. The load factor is the average over all resources with a utilization >20% and the no. of bottleneck shows the share of resources with utilization >95% in relation to the total number of resources. The data show a clear advantage of the O&D-based model. The revenue improvement compared to DB is significantly lower than compared to EMSR-b, which is in line with the assumption from practice. The poor performance of EMSR-b is the result of applying a methodology based on assumptions (e.g., fencing, low-fare demand arrives before high-fare demand, single resource) not found in DB Fernverkehr's pricing strategy. In fact, DB uses the manual discount-control to overcome these weaknesses. Overall, in most cases the

O&D model shows a higher capacity utilization compared to DB and EMSR-b, which suggests that the revenue increase is achieved through an improved network utilization. A look at the average prices from tickets sold (see Table 13) also shows that this is not achieved by a generally lower or higher average price level of sold tickets since $\ddot{p}_{EMSRb}^{avg} \leq \ddot{p}_{O\&D}^{avg} \leq \ddot{p}_{DB}^{avg}$. An improved steering between train connections and short- and long-haul travelers through the use of the choice-based network approach is one explanation. Interesting to see is also that compared to *DB* the improvement through *O&D* increases with the demand. A reason can be that bottleneck situations can be better solved and controlled by taking the customer choice and network effects in the optimization into account. The numbers also show that the revenue benefit (compared to *DB*) is relatively similar for both kind of networks. From a practical point of view, this is a satisfying result as it shows that the revenue potential is hardly dependent on the network structure to which the model is applied.

It is also striking that the revenue advantage in both comparative models is lower in the imperfect forecast case. The O&D approach thus appears to be less robust in the case of imperfect forecasts. Since an imperfect forecast is the more realistic case in practice, a revenue increase of approx. 13-15% (base case) by implementing *O&D* is considered to be the most appropriate result.

		Avg. price of tickets sold		
Network	**Demand**	**O&D**	**DB**	**EMSR-b**
North-South axis	low	27.52	28.09	24.62
	base	27.79	28.60	26.01
	high	28.36	29.24	27.43
Top 50 O&Ds	low	30.77	33.64	24.43
	base	31.67	35.53	27.94
	high	32.89	37.50	30.71

Table 13: Average prices from tickets sold

Network	Demand	RAIL_INV		RAIL_INV_LOG		RAIL_ORG		Total		Obj. value gap to upper bound
		mean	std.dev.	mean	std.dev.	mean	std.dev.	mean	std.dev.	
North-South axis	low	10	0.2	10	0.7	3	0.3	23	0.9	-2.01%
	base	10	0.2	12	2.9	4	0.4	26	2.9	-2.10%
	high	10	0.1	15	1.7	6	0.6	30	1.9	-2.59%
Top 50 O&Ds	low	41	1.6	29	3.2	12	7.6	82	9.4	-2.30%
	base	44	0.9	33	4.5	14	5.6	92	6.9	-2.01%
	high	45	1.1	35	3.9	54	45.0	133	43.5	-1.82%

Table 14: Statistics for solution times in seconds and objective value gap between RAIL_ORG and upper bound model RAIL_INV.

			Revenue improvement w.r.t.		Load factor (>20%)			Bottleneck resources (>95% load factor)		
Network	Demand	forecast quality	DB	EMSR-b	O&D	DB	EMSR-b	O&D	DB	EMSR-b
North-South axis	low	perfect	13.1%	20.0%	51.1%	48.5%	51.2%	4.6%	0.8%	1.5%
	base	perfect	16.5%	20.9%	58.6%	55.3%	57.7%	10.6%	3.9%	4.7%
	high	perfect	21.2%	24.4%	64.4%	59.0%	60.4%	17.3%	7.0%	7.7%
	low	imperfect	9.8%	14.8%	50.5%	48.8%	51.4%	3.9%	0.9%	1.6%
	base	imperfect	13.1%	16.2%	57.8%	55.5%	57.9%	9.1%	4.2%	5.0%
	high	imperfect	16.2%	18.6%	63.8%	59.7%	61.2%	15.2%	7.6%	8.2%
Top 50 O&Ds	low	perfect	15.0%	30.6%	53.8%	46.6%	53.2%	7.7%	2.4%	4.2%
	base	perfect	19.5%	29.0%	61.0%	51.8%	58.3%	14.6%	4.0%	7.0%
	high	perfect	22.5%	29.0%	66.4%	56.4%	62.0%	21.2%	5.4%	9.1%
	low	imperfect	11.0%	23.7%	52.9%	46.6%	52.9%	6.7%	2.5%	4.1%
	base	imperfect	14.7%	22.0%	60.0%	52.1%	58.1%	13.2%	4.2%	7.1%
	high	imperfect	17.1%	22.1%	65.5%	56.9%	62.0%	19.3%	6.4%	9.9%

Table 15: Revenue improvement of the O&D model in the different demand cases and corresponding statistics about network utilization.

4.4.3 Results of special cases

Discretized Prices

A heuristic to derive discretized prices based on *O&D* is to round the continuous prices to the closest discrete price after optimization. Under these conditions one could claim that RAIL_ORG is the relaxed version of a discrete pricing version of *O&D*. This special case has two underlying motivations: First, from a practical perspective it serves as a benchmark for the expected revenue improvement in a first, live market test (where *O&D* can only be tested with the existing IT-infrastructure) and second, the share of continuous pricing in the overall effect can be approximately estimated. The numbers in Table 16 compare the revenue improvement (again with both alternative approaches as benchmark) under the special case with rounded prices. As expected, rounding the prices leads to decreased revenue compared to the continuous O&D model and thus smaller revenue improvement (compare the numbers with the results from Table 15). The gap varies depending on the applied network and demand case and the %-deltas (in absolute terms) lie in the range of 2.0% and 3.3% for the imperfect case when using *DB* as benchmark model (deltas are quite similar for a comparison with EMSR-b). Thus, the loss in revenue in this special case is significant but the overall revenue effect is still positive and large enough to justify an advantage from the adapted version of *O&D*.

Approximate Prices

The results in Table 16 show that using the approximate prices leads to a smaller revenue improvement in all scenarios (compared to the full *O&D* instances). In other words: Solving RAIL_ORG as a third step is beneficial because it finds a solution that generates higher revenue while ensuring capacity limitations. Further, a detailed look into the data shows that it is true for each simulation instance, i.e., the positive effect of solving RAIL_ORG is not only seen on average numbers but also for single instances. Differences between both networks are quite low, i.e., the structure of the network seems to have a low influence on the effect. Interestingly, the absolute difference of the percentage improvement factors is smaller at high demand, suggesting a better approximation of RAIL_INV_LOG at high demand cases. Still, the loss in revenue is large enough to justify solving RAIL_ORG in practice. A possible way in practice would be to use the prices from RAIL_INV_LOG if RAIL_ORG does not provide a solution in a predefined time limit.

			Special case "Discretized Prices"		Special case "Approx. Prices"	
			Revenue improvement w.r.t.		Revenue improvement w.r.t.	
Network	**Demand**	**forecast quality**	**DB**	**EMSR-b**	**DB**	**EMSR-b**
North-South axis	low	perfect	10.0%	17.0%	9.2%	16.0%
	base	perfect	13.5%	17.7%	13.3%	17.6%
	high	perfect	17.2%	20.7%	18.0%	21.6%
	low	imperfect	7.6%	13.0%	6.9%	12.3%
	base	imperfect	10.7%	13.9%	10.5%	13.7%
	high	imperfect	14.2%	16.5%	14.3%	16.7%
Top 50 O&Ds	low	perfect	10.6%	25.8%	11.0%	26.2%
	base	perfect	15.1%	24.4%	16.3%	25.8%
	high	perfect	17.9%	24.4%	19.4%	26.1%
	low	imperfect	7.7%	20.5%	8.3%	21.1%
	base	imperfect	11.6%	19.0%	12.3%	19.8%
	high	imperfect	14.4%	19.2%	15.3%	20.1%

Table 16: Results for both special cases

4.5 Insights from practice

Defining a core network

The O&D approach should be applied to a so-called *core network* in order to keep solution times reasonably short in practice and since accurate forecasting requires a minimum amount of historic booking data. During the joint project, different criteria are discussed and evaluated to define such a network. As potential criteria, there exists using 1) only the top itineraries, 2) top train lines, 3) top train stations or 4) top O&Ds.

Since one crucial benefit of the model *O&D* is the integrated discrete choice model, using top itineraries and top train lines is not a suitable criterion. Both would solve *O&D* for only a part of the itineraries, thus the consideration set would be incomplete and the switching effects not accurately represented. Further, considering a network of all itineraries between top train stations would work with a full set of itineraries (i.e., customer choice is represented correctly). At the same time, choosing top stations does not mean that all the combinations of these top stations (i.e., O&Ds) have a high revenue contribution. Therefore, many unpopular O&Ds and itineraries could become part of the optimization process. Finally, defining a core network based on the top O&Ds appears to be the most appropriate approach. It has the advantage to cover a large part of the revenue and at the same time to have a complete consideration set. Although there is the disadvantage that unpopular connections (e.g., at off-peak times) are also optimized, it is necessary for consideration set reasons. This is one reason why the Top 50 O&D network is used in the simulation since this could be a potential network for a practical implementation.

Understanding the pricing decision

One part of the analysis is to understand the price decision for a single product (i, q). This is particularly important for prices that deviate from the general price structure at first glance (i.e., outliers). Due to the large number of effects, it is usually hardly possible to determine the exact reason. Figure 13 shows an example of a connection in 2nd class (bold printed, underlined product). Compared to a basic network pricing model where the demand for a product (i, q) is assumed to be independent of alternatives, a single product in RAIL_ORG is linked to many other product decisions (i.e., $d_{iq}(\boldsymbol{p})$). The choice model links the 2nd class product both with the 1st class product of the same connection (see circle 1) as well as with alternative products of both classes, which depart earlier or later (see 2).

In addition, the capacity constraint connects the price decision with other connections that use the same 2nd class resource (see 3). The price consistency constraint now creates a further dependency on an additional (here: longer) connection (see 4). Finally, the comfort class price consistency constraint creates an additional dependency to the 1st class of the same connection (see 5).

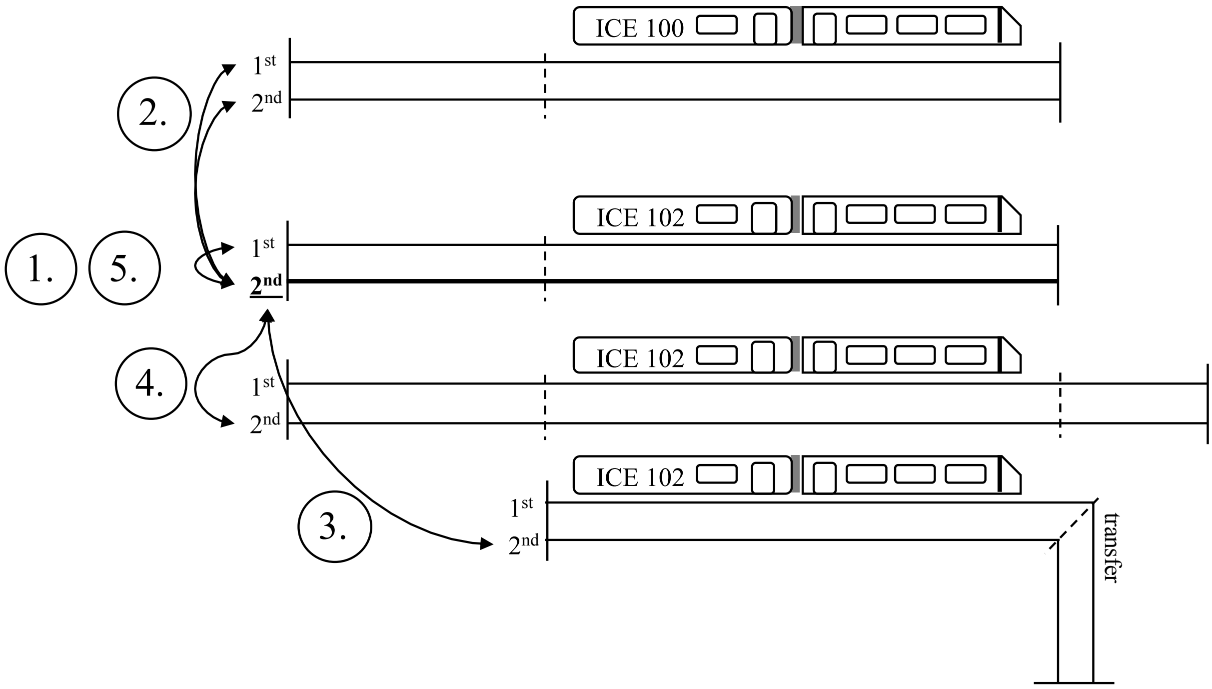

Figure 13: Price decision and its dependency

First, this example shows that an optimal decision for such a network of dependencies cannot be optimally controlled by RM managers manually. This reason is a strong argument for applying mathematical optimization techniques as proposed with RAIL_ORG. Second, it comes at the cost of introducing a black box so that extensive live tests are necessary in practice to convince the management of the idea and see whether the revenue improvement from the simulation can be realized in practice as well. In fact, applying the proposed model in practice as part of live market tests is the next step of the project at DB Fernverkehr.

4.6 Conclusion

Revenue management for passenger rail companies is a small but active research field. While network approaches are popular in theory, leg-based models are still found in practice. In this paper, a choice-based network dynamic pricing model was presented that accounts for the main rail specific constraints from practice and uses a discrete mixed multinomial logit model to represent the customer choice. The extensive simulation study that covers different rail networks and demand cases has shown that the revenue improvement can be estimated to be approx. +14% by using the *O&D* model compared to the current approach of DB Fernverkehr. Additionally, building on the approach of Schön and Hohberger (2019) a good solution can be derived in a few minutes even for networks of realistic size. Two special cases were analyzed to evaluate the effect of continuous pricing and using approximate prices from a relaxed, convex model. The results showed that a significant revenue improvement is still given under these scenarios although the gap decreases. Since simulations cannot cover all special conditions, further live tests are necessary to validate the results in practice. Therefore, this paper lies the foundation for a bigger project at DB Fernverkehr.

Chapter 5 An algorithm to create test data for large-scale railway network revenue management models with customer choice

First author:

Simon Hohberger

Chair of Service Operations Management, Business School, University of Mannheim

Second author:

Cornelia Schön

Chair of Service Operations Management, Business School, University of Mannheim

Abstract

Large-scale railway network revenue management models with customer choice behavior are not only a challenge from an optimization perspective, it is also complex and time-consuming to collect and set up test data for large networks. To promote research in this field, we present an algorithm that generates test data based on the schedules of railway companies, e.g., the set of itineraries and corresponding data, such as the resource consumption or product attribute values like travel time, number of transfers, etc. The generated data are also useful for other fields of research, such as crew scheduling or delay management. We show that the algorithm generates realistic test data for large-scale networks in only a few seconds. To promote research in the field of large-scale railway revenue management, we make the programming code (incl. a small schedule dataset) publicly available.

An earlier version of this paper is available on SSRN:

Hohberger, S. and Schön, C. (2019) An Algorithm to Create Test Data for Large-Scale Railway Network Revenue Management Models with Customer Choice. Available at SSRN: 3439270.

5.1 Introduction

Revenue management (RM) is a well-studied discipline in Operations Research that deals with the question of how to sell your products so that revenue is maximized. Revenue management is usually divided into quantity-based or price-based RM. In the former case, the *availabilities* of products (i.e., booking classes of itineraries) are controlled to steer demand, while in the latter case, *prices* for products are controlled (Talluri and van Ryzin 2005, Erdelyi and Topaloglu 2011). Most revenue management problems in practice are so-called network problems, where multiple products and multiple resources are optimized jointly in order to account for product interdependencies with regard to demand and/or shared resources. While earlier models for network revenue management assumed that demand for one product is independent of the controls for other products (see McGill and van Ryzin (1999) for an overview of network models), there has been a recent focus in research on models that overcome this strong assumption by including discrete choice models such as the multinomial logit model (MNL) (e.g., Talluri and van Ryzin (2004), Liu and van Ryzin (2008), Ratliff and Vinod (2005)).

At the same time, revenue management has become more and more prominent not only in research but also in practice. There are many industries that apply RM-techniques in their daily business, such as airlines, railway companies or hotels (see Chiang et al. (2007) for an industry overview of revenue management applications). Usually, real-life product networks are bigger and more complex in terms of model and data size than the examples used in research. One reason is that it is typically very time consuming to construct realistic test data for big networks, e.g., a railway network with thousands of itineraries, legs, etc. The consideration of customer choice models intensifies the challenge, as for each itinerary, the corresponding attribute values (e.g., travel time, comfort level) are necessary to derive product utilities. This quickly leads to the challenge that test data for a network revenue management model with customer choice behavior can be manually constructed only for a network of very limited size. Additionally, past research has shown that new model formulations are typically tested by using self-defined (academic) networks and data sets, which makes a comprehensive evaluation and comparison of results in different publications much harder.

To overcome that challenge and promote research for large-scale revenue management, we present a "Network Data Generation Algorithm" (NDGA) that derives test data for railway companies based on schedules. We also make the corresponding programming code publicly available so that everyone can use it for their own small- or large-sized examples and adjust it according to personal needs. Additionally, our motivation with the NDGA is to do a first step for establishing a consistent database that can be used by various academics and practitioners that would improve the comparability of novel approaches in the context of revenue management. Although the NDGA was created for the purpose of creating railway test data, it also works for other transportation companies, such as airlines or intercity bus services. It automatically creates all promising itineraries between two cities/train stations (incl. up to one transfer) based on some predefined efficiency rules to avoid certain itineraries, e.g., with big detours. In other words, one goal of the NDGA is to create the set of candidate itineraries that a railway company would offer their customers when searching for connections from city A to city B. In a second step, it derives relevant data for RM models, such as the resource consumption of the itineraries and attribute values like travel time, number of transfers, etc., that determine customer choice. The NDGA is based on AMPL programming language (Fourer et al. 1990). Researchers that use different software for their optimization models can use it as a blueprint to copy the idea into their appropriate programming language or to transfer the data via the AMPL API to their specific software environment. The algorithm is designed specially to construct data for transportation networks, primarily for railway companies and especially for applications in the revenue management context. Some part of the generated data can also be interesting for other fields of research, such as delay management or crew scheduling. A similar idea of generating test instances in a different research area (project scheduling projects) is provided by Kolisch et al. (1995). We follow that way with the NDGA to support research in railway RM for large-scale networks.

The paper is structured as follows: First, we present a basic network RM model with customer choice as a starting point to explain the necessary data that are generated in the NDGA presented in Chapter 5.3. Chapter 5.4 presents examples and statistics for the runtime of the algorithm. Finally, Chapter 5.5 summarizes the findings.

5.2 Theoretical background

The basic idea of network revenue management (also multi-resource RM) compared to a leg-based approach (single-resource RM) is to consider the fact that some products use multiple resources rather than a single resource, and therefore, multiple products share one resource. Under limited capacity, network RM helps to determine the optimal mix of product availabilities (in capacity-based RM) or optimal prices (in price-based RM) for the products that are offered (Talluri and van Ryzin 2005).

Traditional RM models assumed the demand to be independent of other products. In reality, demand for products does of course not only depend on the own price but also on quality and prices of alternative products. *Example*: A customer is willing to travel from city A to city B, and a railway company offers two different connections, one in the morning (1) and one at noon (2). Then, the demand for train connection 1 will not only depend on its own price p_1 but also on the price of the alternative train connection 2 (p_2) and other quality-related attributes. To model this more realistic behavior, the implementation of customer choice models into RM optimization has increased in popularity during the last few years. A basic deterministic network-based pricing model that includes choice-based demand (CB-NPM) can be described as:

CB-NPM:

$$\max \sum_{k \in K} p_k d_k(\boldsymbol{p}) \tag{5.1}$$

$$s.t. \sum_{k \in \dot{K}_r} d_k(\boldsymbol{p}) \leq c_r \quad \forall\, r \in R \tag{5.2}$$

$$p_k \geq 0 \quad \forall\, k \in K \tag{5.3}$$

where $k \in K$ describes the set of products, $r \in R$ the resources/legs, and $k \in \dot{K}_r$ the products that use resource r. Let $d_k(\boldsymbol{p})$ denote the price-dependent demand function with the price vector $\boldsymbol{p}$ that includes p_k, the price for product k, and the prices of all products that are defined as alternatives for k. Finally, c_r defines the capacity of leg r. The objective function in (5.1) represents total revenue from ticket sales, expression (5.2) describes the set of capacity constraints for legs $r \in R$, and inequality (5.3) is the non-negativity requirement for all price variables.

There exist different approaches to represent choice-based demand functions. A well-known approach for a discrete choice model is the MNL model (following McFadden (2001)):

$$\pi_k = \frac{e^{V_k}}{e^{V_0} + \sum_{k' \in \widetilde{K}_k} e^{V_{k'}}} \quad \forall\, k \in K \tag{5.4}$$

where π_k is the choice probability, V_k is the utility of product k, and V_0 is the utility of the outside option (no purchase). $\widetilde{K}_k$ denotes the set of products that are considered as alternatives for k.

In transportation RM, we denote a product k as (j, q), i.e., the combination of an itinerary $j \in J$ and comfort class $q \in Q$, e.g., a 2nd class ticket for a specific itinerary j. Throughout this paper, we use the notions "connection" and "itinerary" interchangeably. As a result, one can describe the utility of a product as:

$$V_{jq} = \tau_{jq} + \alpha_j + \beta_{jq}^P p_{jq} = \tau_{jq} + \sum_{a \in A} \beta_a f_{ja} + \beta_{jq}^P p_{jq} \tag{5.5}$$

The total utility of product (j, q) consists of the utility for all comfort class-related attributes of a product τ_{jq} (e.g., seat comfort, on-board service), the utility of all schedule-related attributes α_j (e.g., travel time of itinerary j) and the price-dependent utility $\beta_{jq}^P p_{jq}$. The term f_{ja} represents the level of the schedule-related product attribute $a \in A$, e.g., travel time in minutes, number of interchanges and β_a the sensitivity parameter of attribute $a \in A$. The utility function in (5.5) is based on a vector model but can, of course, easily be replaced by a more flexible approach such as a part-worth utility function (Green and Srinivasan 1978). Following, e.g., Schön (2010a) or Zhang and Lu (2013), the MNL model multiplied by the market size can be used to replace the demand function $d_{jq}(\boldsymbol{p})$ so that CB-NPM becomes the MNL-NPM (with $m \in M$ representing markets, i.e., different origin-destination (O&D) pairs and w_m the market demand):

MNL-NPM:

$$\max \sum_{m \in M} \sum_{j \in \ddot{J}_m} \sum_{q \in Q} p_{jq} w_m \frac{e^{\tau_{jq} + \alpha_j + \beta_{jq}^P p_{jq}}}{e^{V_{0m}} + \sum_{j' \in \tilde{J}_j} \sum_{q' \in Q} e^{\tau_{j'q'} + \alpha_{j'} + \beta_{j'q'}^P p_{j'q'}}} \tag{5.6}$$

$$s.t. \sum_{m \in M} \sum_{j \in \ddot{J}_{m,r}} w_m \frac{e^{\tau_{jq} + \alpha_j + \beta_{jq}^P p_{jq}}}{e^{V_{0m}} + \sum_{j' \in \tilde{J}_j} \sum_{q' \in Q} e^{\tau_{j'q'} + \alpha_{j'} + \beta_{j'q'}^P p_{j'q'}}} \leq c_{rq} \quad \forall\, r \in R, q \in Q \tag{5.7}$$

$$p_{jq} \leq p_{j'q} \quad \forall\, j' \in J, j \in J_{j'}^{overlap}, q \in Q \tag{5.8}$$

$$p_{jq} \geq 0 \quad \forall\, j \in J, q \in Q \tag{5.9}$$

MNL-NPM also contains the price consistency constraints (5.8), which can be relevant for some railway or intercity bus companies to make sure that longer itineraries are never cheaper than shorter itineraries to avoid strategic customers that buy a ticket for a longer trip and disembark earlier (Hetrakul and Cirillo 2014). The relationship between a short and long itinerary that has to fulfill price consistency is represented by the set $J_{j'}^{overlap}$.

The challenge with MNL-NPM is not only that it is a non-convex optimization problem (Hanson and Martin 1996) and therefore not directly solvable globally optimal but also that a large amount of data is necessary to set up an optimization model for medium- or large-sized instances. The algorithm that we present automatically generates a significant proportion of data to set-up MNL-NPM and CB-NPM.

Given a schedule, not all connections that can be theoretically generated in the network are viable alternatives to be offered to customers (e.g., in case an itinerary includes unreasonable detours or an unreasonable number of changing activities for passengers). Therefore, our algorithm allows us to generate only those connections from the schedule that are promising with regard to some efficiency criteria. In revenue management, generating the set of connections from a given schedule deals with the question of which itineraries to ultimately offer to customers. And since an itinerary j determines a product (j, q) to a large extent, this decision is directly linked to the set of offered products. We assume that a company first determines the set of promising itineraries from a given schedule and in a second step optimizes its revenue by applying RM models like MNL-NPM for the resulting set of fixed products. This idea is strongly related to the two-step approach discussed in product line design (PLD), where first a set of *candidate products* is determined, and in a second step, based on the reduced set, a product line selection (assortment optimization) and pricing problem is solved to find the optimal product line (Steiner and Hruschka 2002, Green and Krieger 1989). As mentioned by Steiner and Hruschka (2002), most research on two-step approaches in PLD focuses on the second step, while the first step (i.e., generating the set of candidate products) seems to be underdeveloped. Still, in the current revenue management literature, models assume a given set of products, typically without detailed information about how these products are derived, and optimize revenues by setting optimal prices or booking class availabilities for the given products. NDGA can therefore be seen as an approach for *step one* for RM models in transportation companies. A challenge of a two-step approach in price-based RM is that the price of the product is determined in step two and thus unknown when determining the *candidate products.* Our algorithm ensures that itineraries of lower quality (with regard to non-price attributes, e.g., longer travel duration) can be kept in the candidate set if it is feasible to offset this lower quality by a lower price, such that the overall product offer may still be attractive from a customer's perspective as well as from a seller's revenue perspective. A balanced ratio between keeping itineraries and decreasing model size is therefore of high interest to generate a "good" candidate product set. NDGA is one possible way to derive a set of promising itineraries and thus products. There are also other approaches that could be adapted to the railway environment, e.g., the best-in heuristic proposed by Green and Krieger (1987). In their heuristic, a promising product is found by choosing the one product that generates the highest utility for one specific customer. The final product set is then the union of the highest-valued products over all customers. Our proposed efficiency rules are inspired by algorithms used in practice (e.g., on websites) by railway companies.

A deeper look into contributions on railway revenue management shows that models are typically tested only with a limited sized rail network. Ciancimino et al. (1999) can be seen as the starting point for the increasing interest in railway RM problems. They present a network seat allocation model and test it with different sized networks, from five up to 15 legs and 15 up to 120 itineraries, which represents, as they mention, a typical long trip in Italy. You (2008) applies the proposed seat allocation approach to networks of different sizes where the largest instance has eight legs and represents a serial network (i.e., train stations are sequentially connected). Hetrakul and Cirillo (2014) propose a latent-class based seat allocation and pricing model. They use a train line with nine stations, i.e., 36 itineraries to test the approach. They consider four trains covering the same pathway during the day and only differ in the departure time. They also conclude that for future research it would be interesting to look at more complex networks with transfers and a hub-and-spoke structure. Wang et al. (2016) present a choice-based seat allocation model for a Chinese rail provider. In their numerical study they focus on two trains going from Shanghai to Nanchang (via three intermediate stops), i.e., four legs for each train or in total $5 * 4/2 = 10$ itineraries. Since two trains are serving these cities, the network consists of 20 itineraries and 60 products (defined by three comfort classes). They also mention, that in reality a high number of trains are serving popular O&Ds such as Beijing to Shanghai with nearly 40 trains each day.

These examples show that railway RM models are typically tested with a limited number of train connections and with a serial rail network (e.g., one (long) train that serves multiple stations). Rail networks in practice are typically larger and more complex in terms of the structure, i.e., transfers are possible so that multiple single trains form a large network of itineraries and customers can choose from a large set of itineraries that serve the desired O&D.

We believe that next to runtime issues for larger networks the collection of data is a big issue. Our algorithm is therefore a first step in filling the gap and thus making it possible to solve large-scale railway RM models for a broader range of researchers.

5.3 The network data generation algorithm (NDGA)

The starting point for the NDGA is a resource schedule, as shown in Table 17, that represents the legs in the provider's transportation network. Together with the AMPL-files, we provide one small test schedule as a .dat-file so that everyone can test and understand how the algorithm works. For bigger examples, users have to create their own schedules, which can be easily done in Excel randomly or by transforming publicly available schedule data into the required format. The schedule has to include information in the following order: the vehicle/train number, vehicle type (e.g., high-speed train, intercity train), the train station, the corresponding arrival and departure time in minutes, a consecutive station number, the capacities for class 1 and class 2, and the cumulative travel distance (e.g., in km/miles) from the initial to the current station of the train (if the information about distances is not available, one can just fill out the column with zero-values). In the railway case in Table 17, a specific train has many intermediate stations and therefore fills many rows (e.g., ICE279 going from Berlin to Mannheim), whereas an airline example would typically contain only two lines for each flight number, corresponding to the departure and arrival of each flight.

Train No.	Train Type	Station	Arrival time	Departure time	Station No.	Capacity 1. class	Capacity 2. class	Cum. distance
ICE279	ICE	Berlin	510	510	1	200	500	0
ICE279	ICE	Wolfsburg	578	580	2	200	500	50
ICE279	ICE	Braunschweig	596	598	3	200	500	140
ICE279	ICE	Hildesheim	623	625	4	200	500	240
ICE279	ICE	Goettingen	653	655	5	200	500	320
ICE279	ICE	Kassel_Wilh	674	676	6	200	500	450
ICE279	ICE	Fulda	706	708	7	200	500	470
ICE279	ICE	Hanau	746	748	8	200	500	510
ICE279	ICE	Frankfurt	764	770	9	300	750	530
ICE279	ICE	Mannheim	807	816	10	300	750	630
ICE108	ICE	Basel_Bad	683	683	1	150	650	0
ICE108	ICE	Freiburg	715	716	2	150	650	200
ICE108	ICE	Offenburg	748	749	3	150	650	280
ICE108	ICE	Karlsruhe	778	780	4	150	650	420
ICE108	ICE	Mannheim	803	816	5	150	650	480
ICE108	ICE	F_Flugh	846	849	6	150	650	560
ICE108	ICE	Siegburg_Bonn	886	887	7	150	650	660
ICE108	ICE	Koeln	905	905	8	150	650	720

Table 17: Exemplary schedule for two trains

The capacities have to be formulated from an outbound perspective, meaning that capacities in a row represent the levels from the station in the same row to the station in the next row (e.g., capacity in ICE 279 in first class from Hanau to Frankfurt is 200 and from Frankfurt to Mannheim is 300).

Based on the schedule that has to be delivered as a .dat-file, NDGA generates the following output:

- Sets J and $\dot{J}_m$, i.e., creating the set J of promising itineraries in the network, which here are itineraries with no transfer (=direct connection) or one transfer (=indirect connection) and the allocation to its O&D-market ($\dot{J}_m$).
- The set R, which includes all resources in the network and the set $\ddot{J}_r$, which describes which itineraries $j \in J$ use a resource $r \in R$.
- The set $\ddot{J}_{m,r}$, which adds the information of the market m to $\ddot{J}_r$ and would be necessary for a formulation as in (5.7) of MNL-NPM.
- The parameter c_{rq}, i.e., the capacity for each resource r in comfort class q.
- The set $j \in J_{j'}^{overlap}$, which describes if the (short) itinerary j is part of a (longer) itinerary j' and should therefore fulfill price consistency with respect to j'.
- For each itinerary $j \in J$, the specific schedule-related attribute levels (f_{ja}) that can be used to calculate α_j in a discrete choice model:
  - travel time in minutes
  - number of transfers
  - waiting time in minutes at the transfer station
  - used transport mode (e.g., train type)
  - length of each connection (e.g., km/miles, depends on how the data are provided in the schedule).

A brief overview of the algorithm is shown in Figure 14. NDGA is further explained in the following.

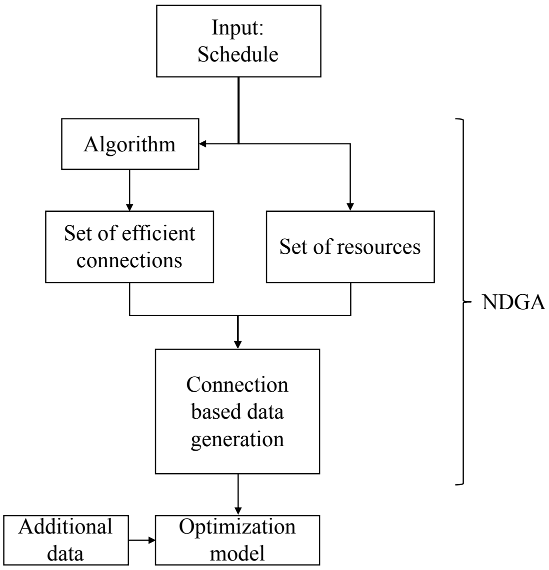

Figure 14: Process flow of NDGA

5.3.1 Constructing the set of efficient connections

In the following, we will explain the notion of an efficient connection and how the NDGA constructs the set of efficient connections. Please note again that we use the expressions "connection" and "itinerary" interchangeably. In the airline example, there are multiple ways to travel from city A to B (via intermediate transfer stations). Some ways are fast and convenient, some routes are long and inconvenient. It is theoretically possible to make big detours, e.g., flying from Paris to London via Hong-Kong. This would be a possible connection but not a reasonable one, and we consider it to be inefficient. Our goal is to define rules for identifying itineraries that are – though theoretically possible – inefficient in terms of travel time and transfers and will therefore be excluded from the offered set J.

The algorithm works as follows: it first constructs all possible direct connections in a network, i.e., for each individual vehicle number, it finds all possible combinations of stations in the direction of travel. In a second step, it combines two direct connections to an indirect connection (i.e., changing from one train to another) based on predefined transfer time limits. The parameters *dur_change_min* and *dur_change_max* define the minimum and maximum time that is allowed to create an itinerary with a transfer and can be easily adjusted by the user in the algorithm. Currently, NDGA only creates connections with no transfer or exactly one transfer. After finding all possible connections (i.e., also connections with big detours or connections going from A via B back to A), the algorithm identifies and deletes all inefficient itineraries based on predefined "efficiency rules" in a filtering process. It works by comparing a connection with all connections of the same O&D. The filtering is a sequential process and consists of multiple steps. If a criterion is fulfilled, the set member will be deleted from the set. After applying all filtering steps, a set of efficient itineraries remains.

The filtering steps are (in that specific order):

1. Delete connections that are inefficient with regard to travel time. A connection $j' \in J$ is considered to be inefficient with regard to travel time
   a. if there exists another connection $j \in J$ serving the same O&D that departs at the origin at a time t_j^{dep} no earlier than j' (i.e., $t_j^{dep} \geq t_{j'}^{dep}$), arrives at the destination more than *dep_arr_buffer* minutes earlier (i.e., at time $t_j^{arr} < t_{j'}^{arr} - dep\_arr\_buffer$) and has fewer or equal transfers.
   b. if there exists another connection $j \in J$ serving the same O&D that departs at the origin more than dep_arr_buffer later than j' (i.e., at time $t_j^{dep} > t_{j'}^{dep} + dep\_arr\_buffer$), arrives at the destination no later than j' (i.e., $t_j^{arr} \leq t_{j'}^{arr}$) and has fewer or equal transfers.
2. Delete itineraries with the same departure and arrival station.
3. Delete itineraries that are inefficient with regard to the transfer station. A connection $j' \in J$ with one transfer is considered to be inefficient with regard to the transfer station if there exists another connection $j \in J$ that uses the same two vehicle numbers but has a superior transfer station, i.e.,
   a. the transfer time is longer, or
   b. if both connections have the same transfer time, the transfer is taking place earlier.

In this filtering process, we assume that a rational passenger would never choose an itinerary if another connection exists that is much more comfortable with regard to travel time and transfers. Note that it can happen that a slow <u>direct</u> connection and a <u>fast indirect</u> connection exist. In that case, the algorithm keeps both connections as it only deletes connections if the number of transfers is higher or equal to the reference itinerary. In other words, an indirect connection can never define a direct connection as inefficient. The meaning of “much more” comfortable can be controlled by the user through the buffer parameter *dep_arr_buffer* (in minutes). It defines how strict the algorithm handles slower connections, e.g., if *dep_arr_buffer* is defined to be ten minutes, then a (slower) connection that departs up to ten minutes earlier and arrives at the same time than another (faster) connection, will still be kept in the set of train connections. The same is true for a delta in the arrival time up to the level of *dep_arr_buffer*.

The parameter defines how rigorous the filtering process in terms of departure and arrival times is carried out. The higher it is, the more connections will be kept in the final set. In the default setting, the parameter is set to zero.
The second step is quite obvious but necessary to delete all itineraries with the same departure and arrival station. These kinds of connections are theoretically possible but not realistic in practice and therefore not efficient.
The final step considers the case where two or more connections only differ in the transfer stations. This could happen, e.g., when two trains have overlapping travel paths and therefore more than one possible transfer station where a passenger can change from train 1 to train 2. To determine the itinerary with the best transfer station, we first apply the rule that the station with the largest time for the transfer is chosen to make the transfer convenient for the passenger (step 3a). As this rule has the possibility to still have two connections with the same time for the transfer, we finally apply a distinct rule that only the itinerary with the earliest transfer is kept (step 3b). After filtering the connections, only the efficient itineraries are left over. Figure 15 shows the evolution of the size of the set of connections for a large set of trains in the network that was provided by a European railway company. We first produce a set of approx. 850,000 theoretical connections and reduce it by applying the efficiency rules to finally approx. 170,000 efficient connections.

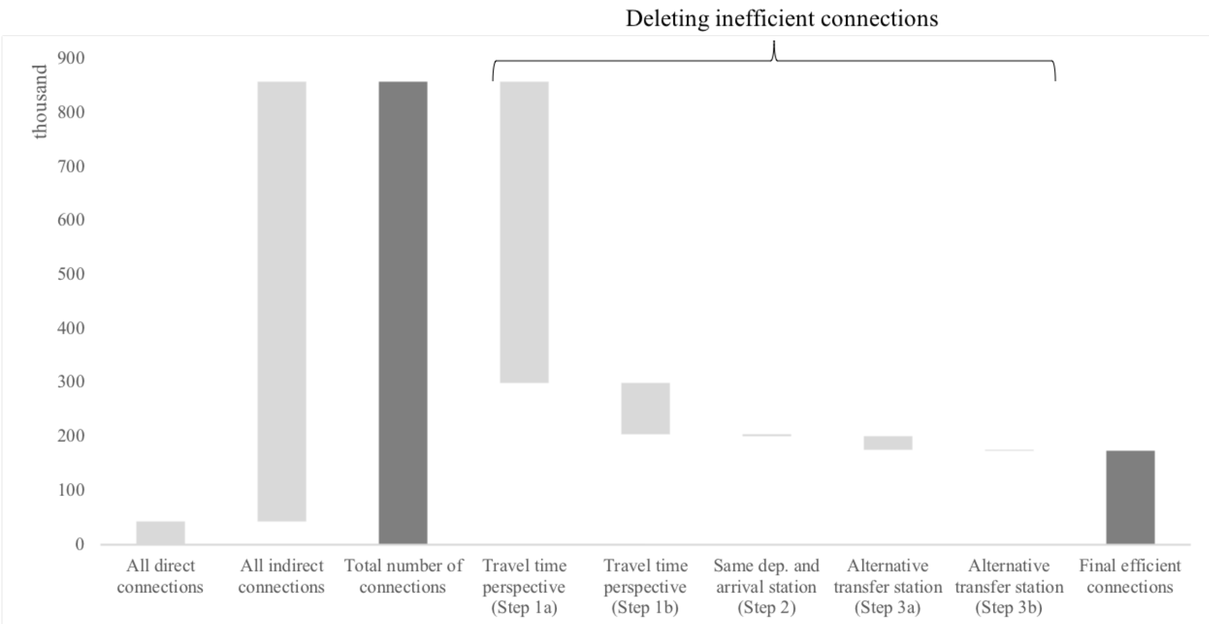

Figure 15: Evolution of efficient connections.

One could argue that the process of first setting up a large set of itineraries that will then be decreased by some efficiency rules is inefficient from a computational perspective. But we will see below that this approach works well for large networks in only a few seconds. Technically speaking, NDGA uses the powerful feature of set commands in AMPL. Set commands make it possible to compare different entries of a set easily. The possibility of creating multidimensional sets makes it possible to save up to 20 pieces of information into one set member. For example, you can define a set "Connection-Information" that has four dimensions $\{$(Connection, travel time (min.), train type, No. of transfers)$\}$, which could result in the specific set member $\{$(Frankfurt_Berlin_ICE580, 260, ICE, 0)$\}$. This not only makes it possible to link relevant information such as travel time to a connection, but it also helps to compare connections, as we can compare numbers (like travel time) or strings (like vehicle type) between connections. We will also see later that set operations work very well for large amounts of data and therefore make it possible to handle large networks.

5.3.2 Deriving data for discrete choice models

After the final definition of all train connections in the network, it is possible to extract additional information that can be used in a choice model. The algorithm creates a parameter f_{ja} that represents the values for eight different attributes that supposedly determine passenger convenience and customer choice behavior for each connection j with $a \in \{$Travel_Time_Section1, Travel_Time_Section2, Total_Travel_Time, Vehicle_Type_Section1, Vehicle_Type_Section2, No_of_Transfers, Waiting_Time_Transfer, Trip_Length$\}$.

Travel_Time_S1	127
Travel_Time_S2	176
Total_Travel_Time	313
Vehicle_Type_S1	2
Vehicle_Type_S2	1
No_of_Transfers	1
Waiting_Time_Transfer	10
Trip_Length	590

Table 18: Exemplary data for one connection

The split into Section 1 and Section 2 is relevant for itineraries with a transfer. Section 1 represents the first part of the itinerary (from the starting train station to the transfer station).

Section 2 represents the second part of the itinerary (from the transfer to the final station). The attributes travel time and vehicle type are split into these sections (S1 and S2). This is relevant, e.g., if the first part of the itinerary is served with a comfortable type of train/vehicle and the second part with an uncomfortable one. Then, a choice model should be able to account for that mix of comfort levels during the whole trip. An example for the output of f_{ja} for a specific train connection j can be seen in Table 18.

5.3.3 Deriving operational data

In addition to the set of offered products and the data on itinerary-specific attribute levels for the choice model, a network optimization problem needs further information such as the resource consumption of each itinerary or the relationship between itineraries to cover the price consistency requirements as in (5.8). Based on the schedule, the NDGA first creates the resource set R that defines all legs in the network. In a second step, it defines a set $\ddot{J}_r$ with two-dimensional elements that consist of all itinerary-leg combinations that are valid. For example, if connection j uses leg 15 and leg 16, then two set members $\{(j,15); (j,16)\}$ exist. For the capacity constraint as in (5.7), NDGA also derives the capacities of leg r in comfort class q as a parameter value c_{rq}. When using a choice model, such as MNL-NPM, the information about the O&D-market m that j serves is also necessary. Therefore, the set $\ddot{J}_{m,r}$ is derived. Depending on the model formulation, one can choose which of the sets $\ddot{J}_r$ or $\ddot{J}_{m,r}$ is applicable.

The relationship between two connections to represent the price consistency in the network is slightly more difficult to define. Price consistency constraints are typically reasonable for companies without a check-in process, e.g., railway or intercity bus companies. The algorithm works by comparing two connections section-wise and checking for the following criteria:

1. Is direct connection j part of another direct connection j' ?
2. Is direct connection j part of another indirect connection j' ?
   a. j is part of j' before the transfer in j'
   b. j is part of j' after the transfer in j'
3. Is indirect connection j part of indirect connection j' ?

The case where an indirect connection j is part of a direct connection j' can never be possible since the transfer in j means that an additional vehicle is used and therefore cannot be part of a direct connection j'. Finally, as an extract of $\dot{J}_m$, we define the minimum distance for each O&D-market m as the parameter $od\_km_{orig,dest}$. This information can be used to classify train connections into long-haul and short-haul routes or to simply derive statistics of the network.

The AMPL files for NDGA that are publicly available include more detailed information to understand the programming code better.

5.4 Practical example

To show a practical example of the NDGA, we apply it to a small but already complex network of three trains running in a network (see Figure 16). The trains are crossing in a station and therefore make connections with transfers possible. Since each train has many intermediate stations, the network consists of 26 train stations in total.

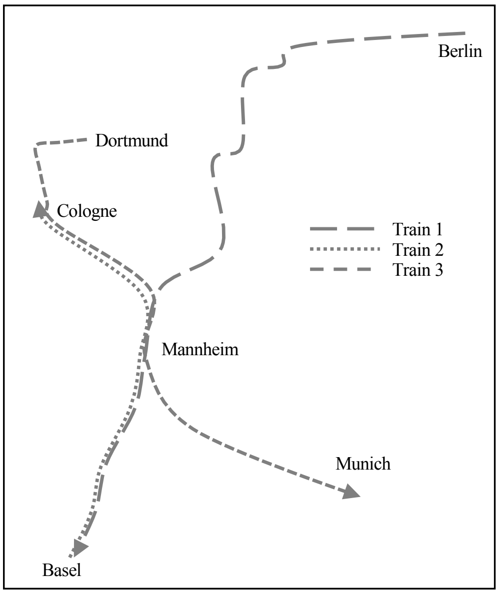

Figure 16: Network with three trains

Running the NDGA for that example takes 0.13 seconds. Table 19 shows some statistics about the amount of data that is generated. It first creates 348 connections, of which 40 are inefficient and therefore deleted so that 308 connections remain in the set J. The set R consists of 32 legs, each leg utilized by some of the 308 connections.

Example: Network of three trains	
# O&Ds	650
# unfiltered connections	348
# efficient connections	308
# legs	32
# capacity values	64
# itinerary-resource relations	1,666
# itinerary-itinerary relations	6,492
# choice parameter values	2,464

Table 19: Statistics regarding the generated data

...
Ulm_HbfAugsburg_HbfICE517_11_12
Ulm_HbfMuenchen_HbfICE517_11_13
Augsburg_HbfMuenchen_HbfICE517_12_13
...
Berlin_HbfMannheim_HbfICE279_Mannheim_Hbf_F_Flughafen_Fern__ICE108_1_10_5_6
Berlin_HbfMannheim_HbfICE279_Mannheim_Hbf_Siegburg_Bonn_ICE108_1_10_5_7
Berlin_HbfMannheim_HbfICE279_Mannheim_Hbf_Koeln_Hbf_ICE108_1_10_5_8
...

Table 20: Set of efficient connections

Table 20 shows an extract of the list of efficient connections including some direct and indirect connections. A connection is described by the departure and arrival station, the train number and the station numbers. For example, Table 20 contains a connection from Berlin via Mannheim (with train number ICE279) to Koeln/Cologne (with ICE108). From a capacity utilization perspective, $\left|\bigcup_{r \in R} \ddot{J}_r\right| = 1{,}666$, i.e., 1,666 itinerary-resource relations exist in this network. Further, each connection is described by the eight attributes (travel time, #transfers, etc.), and therefore, $308 * 8 = 2{,}464$ parameter values are created. Table 21 contains the choice data output for the connection from Berlin via Mannheim to Cologne.

Travel_Time_S1	297
Travel_Time_S2	89
Total_Travel_Time	395
Vehicle_Type_S1	1
Vehicle_Type_S2	1
No_of_Transfers	1
Waiting_Time_Transfer	9
Trip_Length	1026

Table 21: Choice data for connection Berlin - Cologne

Such a small example of three trains can already illustrate that it is hardly possible to derive the possible connections and corresponding data manually. In particular, creating efficient itineraries and deriving resource consumption matrix information would be both very time-consuming and error-prone.

The NDGA was created to construct data for much larger networks with more trains running in the system. To show how it performs for larger networks, we applied different medium- and large-sized practical examples. We had access to schedules from a European railway company that made it possible to test NDGA with realistic data. Table 22 shows the results of four test instances (on a Mac OS X, 2.9 GHz Intel Core i5, 8 GB RAM).

The second column shows the number of trains in the test instance. The third and fourth columns show the number of generated connections and the corresponding run time (sec.). Columns 5-9 show the amount of generated data for the efficient connections, and the last column shows the total run time (sec.) of NDGA.

Basic data		Connections		Data generation					Total run time
Network	**# trains**	**# eff. connec-tions**	**Runtime**	**# O&Ds**	**# legs**	**# leg consump.**	**# price consis.**	**# choice param.**	
1	35	2,240	<1	240	358	9,913	32,905	17,920	<1
2	159	11,886	<1	1,206	1,128	64,211	242,971	95,088	5
3	230	26,666	1.3	4,780	1,961	165,936	721,067	213,328	16.4
4	752	172,982	11	37,895	6,371	1,396,736	7,347,427	1,383,856	322

Table 22: Overview of the amount of generated data and run time (in sec.) for different network sizes

Networks such as No. 1 are constructed in less than a second. Larger networks are, especially in the data generation phase, more time consuming but still practical. Even for very large networks like No. 4, it takes only 11 seconds to create the efficient connections and approximately 5 additional minutes to generate the corresponding data.

5.4.1 Comments on run time

When you plan to solve CB-NPM or MNL-NPM with large networks, such as No. 3 or No. 4, one should run the NDGA just once and store the generated data in a .dat file by using display commands. For smaller networks (and therefore smaller run times) such as No. 1, it would also be possible to combine an optimization model with NDGA directly so that the data are first generated by NDGA and then the optimization model is solved using the data. Table 22 shows that the process of data generation takes most of the time. The calculation of the leg consumption and price consistency sets are especially time-consuming.

Typically, there are multiple ways to write a code for a specific outcome. Different formulations were tested, and an unorthodox solution worked best. In some parts of the code, the *printf* command in AMPL is used to build up a set of data. It adds rows/set members into a file and loads this file at the end. Therefore, every time the NDGA is executed, a file called NDGA_data_temp.dat is created/updated. This file is of no further use for the end consumer; it is just used to speed up the data generation process.

Another formulation that differs from usual model formulations in RM is the definition of resource consumption. Some models in literature have a capacity constraint as in (5.10) where $\rho_{jr} \in \{0,1\}$:

$$\sum_{j \in J} d_{jq}(\boldsymbol{p})\rho_{jr} \leq c_{rq} \quad \forall\, r \in R, q \in Q \tag{5.10}$$

The problem with this binary formulation is that in a railway RM network, most of the entries in the incidence matrix are zero. The network No. 4 has a resource consumption incidence matrix with size $A = J * R = 1{,}102{,}068{,}322$ entries, where only 1,396,736 entries equal 1 (see # Leg Consumption in Table 22) and all others are zero. That is, only 0.13% of the entries of A equal 1 and are relevant for capacity constraints. A problem when solving an optimization model occurs, since AMPL first has to detect all zero-values and cancel out the irrelevant demand values. This overwhelms the model and makes it much slower to solve. This also results in a solution failure for larger networks due to memory constraints.

It is better to use a subset formulation ($\ddot{J}_r$) that only sums up the demand on the left-hand side that corresponds to the resource r. This leads to a slightly different capacity constraint equal to:

$$\sum_{j \in \ddot{J}_r} d_{jq}(\boldsymbol{p}) \leq c_{rq} \quad \forall\, r \in R, q \in Q \tag{5.11}$$

The same idea is used for the set of price consistency relationships where the subset $J_{j'}^{overlap}$ only contains the connections j that should fulfill price consistency with respect to j'.

5.5 Conclusion

In this paper, we presented an algorithm called NDGA that automatically generates a set of offered itineraries and corresponding data based on the schedule of a railway company. It could also be applied to other areas like airlines or intercity bus companies to derive consistent and realistic test data. The objective of NDGA is to generate the majority of set and parameter values that are necessary to set up network revenue management optimization models under customer choice behavior. The generated information includes, for example, the set of connections with up to one transfer, the resource consumption and product attribute values such as the travel time or number of transfers that are necessary to compute utilities in a discrete choice model. NDGA is formulated for use in AMPL and creates data even for large networks in a reasonable amount of time.

Our goal is to promote research for large-scale revenue management models, especially in railway business, and to make a first step in establishing a consistent database for researchers to improve the comparability of novel RM approaches. Therefore, we make the NDGA publicly available so that everyone can test and use it for their own research purposes.

Chapter 6 Conclusion and outlook

This thesis has shown that railway revenue management is still an underdeveloped field of research but with an increasing interest in the past years. Due to the growing demand, revenue management methods play not only an important role for revenue maximization but also for an improved passenger steering to avoid overcrowded trains. Choice-based demand functions, such as the MMNL, can help to better represent the demand shift as a result of the price control and are therefore one possibility to overcome that challenge. In addition, rail networks are ideally suited to use network-based models that make the right pricing decisions for different itineraries sharing the same critical resource. While network-based approaches are already common in railway RM research, there are still many companies that use a leg-based control. With the choice-based network dynamic pricing model, this thesis presented an approach that covers the most important conditions from practice using DB Fernverkehr as an example. The new model has four main advantages over DB's existing approach. The most important improvement is the possibility to control product prices (or itineraries) instead of legs. This is the enabler to cover the traditional network effect and to integrate the MMNL that represents the itinerary choice of customers. The fourth innovation is the possibility of choosing prices from a continuous set since this increases the solution space and thus offers additional revenue potential compared to models that rely on booking classes.

Since the resulting model turns out to be non-linear and non-convex, different convex reformulations were presented that approximate the original model and derive a good upper bound. Numerical experiments using synthetic data have shown that the upper bound can be significantly decreased by applying the model using a piecewise approximation technique. Furthermore, the presented heuristic (3-step approach) makes it able to derive a good solution in reasonable time even for large-scale railway instances. This heuristic was applied and extended in the simulation study (e.g., to account for price consistency between long and short itineraries). Additional improvements were implemented to increase the solution time by, e.g., considering only critical resources in the optimization and thus reducing the number of processed constraints. The results showed that the new price-based approach generates higher revenues than the two leg-based benchmark models from practice, with a total revenue increase of approximately 14% under a realistic demand scenario compared to DB's existing policy. The numbers shown serve as a motivation to take the next step and test the model in real market tests.

Outlook

Future research directions can be split into two categories, a theoretical and a practical perspective.

From a theoretical perspective, dynamic pricing under the MMNL remains a promising research area since a procedure that derives a proven globally optimal solution does still not exist. While some of the current research considers a setting without capacity restrictions, it would be necessary to integrate it to apply the model in a traditional railway/airline revenue management context.

Furthermore, combining the pricing problem with assortment optimization is a promising research area from both a theoretical and practical perspective. I.e., prices as well as the set of offered products are decision variables and optimized jointly. Pricing models typically assume a fixed set of itineraries that is determined in a preprocessing step (e.g., by an algorithm that finds all comfortable/promising train connections). Integrating the decision into the pricing model can lead to additional revenues and further help in practice to ensure a feasible solution even in high-demand scenarios.

From a practical perspective, it would be interesting to see whether the simulation results can be confirmed in a real application since a simulation can never capture all conditions from practice. An open challenge regarding the choice model is still how the sensitivities (the β values), the outside option and arrival rates for a short-term, high-quality demand forecast can be regularly updated and derived from realized booking data without observing the no-buy customers for a real application in practice. DB has currently only access to realized bookings (i.e., customers that chose to buy a ticket) but cannot observe customers who decided to take the car/flight etc. or did not travel at all. It exists some recent research about that issue (e.g., Newman et al. (2014) for the MNL), but the application and accuracy for railway choice models is still an open question and needs to be further analyzed, especially for the more advanced MMNL that was applied in this study.

Additionally, it was shown that the heuristic can solve large-scale instances covering a large proportion of revenue and O&Ds. However, even if it is possible to derive optimal prices for the majority of products, there will still remain a set of (unpopular) products that need a price as well. Therefore, it is necessary to have some kind of a second heuristic that makes it able to derive prices for the remaining, unoptimized set of products. Furthermore, as already mentioned in Chapter 4, the pricing decision becomes more incomprehensible with a choice-based pricing approach since many effects have an influence on the price. A high degree of confidence in the approach is therefore required that can only be achieved through extensive and ongoing testing.

In addition, the result of the model are prices for the products in each booking period under the assumption of deterministic demand. Especially in cases of unexpectedly high demand, when a high amount of bookings occurs in only a few minutes, this can lead to revenue loss (e.g., when a soccer team from Frankfurt wins the half-final of a national cup and all fans want to travel to Berlin for the final match). If the algorithm cannot derive new prices automatically in a small period of time, this could lead to fully booked trains with low-priced tickets. This problem does typically not occur with booking class-based systems where prices automatically increase when a booking class is sold out (i.e., its availability decreases to zero) and the next-higher booking class determines the new price. Therefore, a price-based system should include a fallback solution that automatically increases prices until the model is reoptimized and updated prices are available.

Finally, going from a leg- to a network-based approach means a big change for both IT and the tasks of the RM staff. For this reason, the support of the management and the employees is essential in order to successfully carry out such a large project.

Appendix

The 3-step approach for RAIL_ORG

The 3-step approach for solving RAIL_ORG follows the heuristic as in Schön and Hohberger (2019) and the adjustments as described in Chapter 4.2.2. In step 1, the relaxed, convex model with the inverse demand function (RAIL_INV) and segment-specific prices is solved (the notations are related to Chapter 4):

$$\max_{\pi,v} \sum_{z\in Z}\sum_{i\in \dot{I}_z}\sum_{s\in S}\sum_{q\in Q}\sum_{t\in T} \pi_{izqst}\omega_{szt}\frac{ln\left(\frac{\pi_{izqst}}{v_{zst}}\right)-\alpha_{izqs}}{\beta_{sz}} \tag{A.1}$$

$$s.t. \sum_{z\in Z}\sum_{i\in \ddot{I}_{zl}}\sum_{s\in S}\sum_{t\in T} \pi_{izqst}\omega_{szt} \leq cap_{lq}-\delta_{lq} \quad \forall\, l\in L, q\in Q \tag{A.2}$$

$$v_{zst}A^0_{szt}+\sum_{i\in \dot{I}_z}\sum_{q\in Q}\pi_{izqst}=1 \quad \forall\, z\in Z, s\in S, t\in T \tag{A.3}$$

$$\pi_{izqst}-v_{zst}e^{\alpha_{izqs}+\beta_{sz}p^{min}_{iq}} \leq 0 \quad \forall\, z\in Z, i\in \dot{I}_z, q\in Q, s\in S, t\in T \tag{A.4}$$

$$\pi_{izqst}-v_{zst}e^{\alpha_{izqs}+\beta_{sz}p^{max}_{iq}} \geq 0 \quad \forall\, z\in Z, i\in \dot{I}_z, q\in Q, s\in S, t\in T \tag{A.5}$$

$$v_{zst}\in \mathbb{R}^+ \quad \forall\, z\in Z, s\in S, t\in T \tag{A.6}$$

$$\pi_{izqst}\in \mathbb{R}^+ \quad \forall\, z\in Z, i\in \dot{I}_z, q\in Q, s\in S, t\in T \tag{A.7}$$

In step 2, the convex formulation that includes the approximate uniform price constraints (RAIL_INV_LOG) is solved (incl. the two kind of railway specific price consistency constraints):

$$\max_{\pi,v,p} \sum_{z\in Z}\sum_{i\in \dot{I}_z}\sum_{s\in S}\sum_{q\in Q}\sum_{t\in T} \pi_{izqst}\omega_{szt}\frac{ln\left(\frac{\pi_{izqst}}{v_{zst}}\right)-\alpha_{izqs}}{\beta_{sz}} \tag{A.8}$$

$$s.t. \sum_{z\in Z}\sum_{i\in \ddot{I}_{zl}}\sum_{s\in S}\sum_{t\in T} \pi_{izqst}\omega_{szt} \leq cap_{lq}-\delta_{lq} \quad \forall\, l\in L, q\in Q \tag{A.9}$$

$$v_{zst}A^0_{szt}+\sum_{i\in \dot{I}_z}\sum_{q\in Q}\pi_{izqst}=1 \quad \forall\, z\in Z, s\in S, t\in T \tag{A.10}$$

$$\frac{1}{\beta_{sz}}ln(\pi_{izqst})-\frac{1}{\beta_{sz}}\left(\frac{ln(v^{UB}_{zst})-ln(v^{LB}_{zst})}{v^{UB}_{zst}-v^{LB}_{zst}}(v_{zst}-v^{LB}_{zst})+ln(v^{LB}_{zst})\right)-\frac{\alpha_{izqs}}{\beta_{sz}} \leq p_{iqt} \quad \forall\, z\in Z, i\in \dot{I}_z, q\in Q, s\in S, t\in T \tag{A.11}$$

$$\frac{1}{\beta_{sz}}\left(\frac{ln(\pi_{izqst}^{UB})-ln(\pi_{izqst}^{LB})}{\pi_{izqst}^{UB}-\pi_{izqst}^{LB}}(\pi_{izqst}-\pi_{izqst}^{LB})+ln(\pi_{izqst}^{LB})\right)-\frac{1}{\beta_{sz}}ln(v_{zst}) \tag{A.12}$$

$$-\frac{\alpha_{izqs}}{\beta_{sz}} \geq p_{iqt} \quad \forall\, z \in Z, i \in \dot{I}_z, q \in Q, s \in S, t \in T$$

$$\pi_{izqst} - v_{zst}e^{\alpha_{izqs}+\beta_{sz}p_{iq}^{min}} \leq 0 \quad \forall\, z \in Z, i \in \dot{I}_z, q \in Q, s \in S, t \in T \tag{A.13}$$

$$\pi_{izqst} - v_{zst}e^{\alpha_{izqs}+\beta_{sz}p_{iq}^{max}} \geq 0 \quad \forall\, z \in Z, i \in \dot{I}_z, q \in Q, s \in S, t \in T \tag{A.14}$$

$$p_{iqt} \leq p_{\mathrm{j}qt} \quad \forall\, \mathrm{j} \in I, i \in \ddot{I}_{\mathrm{j}}, q \in Q, t \in T \tag{A.15}$$

$$p_{iqt} + \rho \leq p_{i\bar{q}t} \quad \forall\, i \in I, q \in Q, \bar{q} \in \dot{Q}_q, t \in T \tag{A.16}$$

$$v_{zst}^{LB} \leq v_{zst} \leq v_{zst}^{UB} \text{ with } v_{zst} \in \mathbb{R}^+ \;\forall\, z \in Z, s \in S, t \in T \tag{A.17}$$

$$\pi_{izqst}^{LB} \leq \pi_{izqst} \leq \pi_{izqst}^{UB} \text{ with } \pi_{izqst} \in \mathbb{R}^+ \quad \forall\, z \in Z, i \in \dot{I}_z, q \in Q, s \in S, t \in T \tag{A.18}$$

$$p_{iqt}^{LB} \leq p_{iqt} \leq p_{iqt}^{UB} \text{ with } p_{iqt} \in \mathbb{R}^+ \quad \forall\, i \in I, q \in Q, t \in T \tag{A.19}$$

Compared to RAIL_INV, this model includes two constraints (A.11 and A.12) that approximate the uniform price p_{iqt}. Additionally, it is now possible to integrate the price consistency constraints (A.15) and (A.16) into the model since a specific variable for the price is introduced (which was not the case in RAIL_INV). The parameter values for the lower bounds (LB) and upper bounds (UB) on variables $\boldsymbol{v}, \boldsymbol{\pi}$ and $\boldsymbol{p}$ in (A.17)-(A.19) and the PC constraints in (A.11) and (A.12) are derived following the 3-step approach in Schön and Hohberger (2019) and the adjustments explained in 4.2.2.

Finally, RAIL_ORG is solved using the warm start solution from RAIL_INV_LOG. Note that both models RAIL_INV_LOG and RAIL_ORG are solved using the reduced set of resources following the constraint-reduction rules as described in Chapter 4.2.2.

References

Abe, I. (2007) Revenue Management in the railway industry in Japan and Portugal: a stakeholder approach. Doctoral dissertation, Massachusetts Institute of Technology.

Akcay, Y., Natarajan, H. P. and Xu, S. H. (2010) Joint dynamic pricing of multiple perishable products under consumer choice. *Management Science* 56(8): 1345-1361.

Anderson, S.P., de Palma, A. and Thisse, J.F. (1992) Discrete Choice Theory of Product Differentiation. MIT Press, Cambridge.

Armstrong, A. and Meissner, J. (2010) Railway revenue management: overview and models. *Working Paper*, Lancaster University Management School.

Avriel, M., Diewert, W. E., Schaible, S. and Zang, I. (1988) Generalized concavity (Vol. 36). Kluwer Academic Publishers Group, Dordrecht.

Aydin, G. and Porteus, E. L. (2008) Joint inventory and pricing decisions for an assortment. *Operations Research* 56(5): 1247-1255.

Bao, Y., Liu, J., Ma, M. S. and Meng, L. Y. (2014) Seat inventory control methods for Chinese passenger railways. *Journal of Central South University* 21(4): 1672-1682.

Belobaba, P. (1987) Air travel demand and airline seat inventory management. Doctoral dissertation, Massachusetts Institute of Technology.

Benson, H. P. (2002) Using concave envelopes to globally solve the nonlinear sum of ratios problem. *Journal of Global Optimization* 22(1-4): 343-364.

Bergamini, M. L., Aguirre, P. and Grossmann, I. (2005) Logic-based outer approximation for globally optimal synthesis of process networks. *Computers & chemical engineering* 29(9): 1914-1933.

Bharill, R. and Rangaraj, N. (2008) Revenue management in railway operations: A study of the Rajdhani Express, Indian Railways. *Transportation Research Part A: Policy and Practice* 42(9): 1195-1207.

Bild (2018) So werden 5 Mio. BahnCard-Kunden abgezockt. Available at https://www.bild.de/bild-plus/geld/mein-geld/sparfochs/versteckte-preiserhoehung-bei-der-bahn-56533458 (accessed October 3, 2019).

Bitran, G. and Caldentey, R. (2003) An overview of pricing models for revenue management. *Manufacturing & Service Operations Management* 5(3): 203–229.

Boyd, S. and Vandenberghe, L. (2004) Convex optimization. Cambridge University Press, Cambridge.

Brenck, A., Berndt, A. and Naumann, K. P. (2003) Neues Tarifsystem der Deutschen Bahn: Sinnvolle Preisgestaltung im öffentlichen Verkehr? *ifo Schnelldienst* 56(16): 3-13.

Carrier, E. (2008) Modeling the choice of an airline itinerary and fare product using booking and seat availability data. Doctoral dissertation, Massachusetts Institute of Technology.

Chen, Z. and Haynes, K. E. (2015) Impact of high speed rail on housing values: An observation from the Beijing–Shanghai line. *Journal of Transport Geography* 43: 91-100.

Chiang, W. C., Chen, J. C. and Xu, X. (2007) An overview of research on revenue management: current issues and future research. *International Journal of Revenue Management* 1(1): 97-128.

Ciancimino, A., Inzerillo, G., Lucidi, S. and Palagi, L. (1999) A mathematical programming approach for the solution of the railway yield management problem. *Transportation Science* 33(2): 168-181.

Davis, J. M., Topaloglu, H. and Williamson, D. P. (2017) Pricing problems under the nested logit model with a quality consistency constraint. *INFORMS Journal on Computing* 29(1): 54-76.

DB Netz (2017) Berlin–München unter vier Stunden: Kanzlerin, Bundesminister und Länderchefs eröffnen modernste Strecke Deutschlands. Available at http://www.vde8.de/---_site.news..ls_dir._type.press_cat.1_id.858_likecms.html (accessed October 3, 2019).

Désir, A., Goyal, V. and Zhang, J. (2014) Near-optimal algorithms for capacity constrained assortment optimization. Available at SSRN: 2543309.

Deutsche Bahn (2018) Treibhausgas-Emissionen. Available at https://www.deutschebahn.com/resource/blob/260060/9bb6cfd97d3e77700ea9d0088957f95b/klimaschutz-infografik_emissionen-data.pdf (accessed October 3, 2019).

Deutsche Bahn (2019) Daten & Fakten 2018. Available at https://www.deutschebahn.com/resource/blob/3992278/1d136e983334750ef4f24072d49c7cdf/20190325_bpk_2019_daten_fakten-data.pdf (accessed October 3, 2019).

Dong, L., Kouvelis, P. and Tian, Z. (2009) Dynamic pricing and inventory control of substitute products. *Manufacturing & Service Operations Management* 11(2): 317-339.

Dutta, G. and Ghosh, P. (2012) A passenger revenue management system (RMS) for a National Railway in an Emerging Asian Economy. *Journal of Revenue and Pricing Management* 11(5): 487-499.

Erdelyi, A. and Topaloglu, H. (2011) Using decomposition methods to solve pricing problems in network revenue management. *Journal of Revenue and Pricing Management* 10(4): 325-343.

Feldman, J. and Topaloglu. H. (2015) Bounding optimal expected revenues for assortment optimization under mixtures of multinomial logits. *Production and Operations Management* 24(10): 1598-1620.

Feng, X. (2018) Revenue Management Techniques in Practice: A Survey of Railway Companies. Master thesis, University of Mannheim.

Fourer, R., Gay, D. M. and Kernighan, B. W. (1990) A modeling language for mathematical programming. *Management Science* 36(5): 519-554.

Freund, R. W. and Jarre, F. (2001) Solving the sum-of-ratios problem by an interior-point method. *Journal of Global Optimization* 19(1): 83-102.

Fvw (2018) Dieser Mann baut das Lufthansa-Preissystem um. Fvw 18: 36-37, 31.08.2018.

Gallego, G. and van Ryzin, G. (1997) A multiproduct dynamic pricing problem and its applications to network yield management. *Operations Research* 45(1): 24-41.

Gallego, G. and Wang. R. (2014) Multiproduct price optimization and competition under the nested logit model with product-differentiated price sensitivities. *Operations Research* 62(2): 450-461.

Gao, L. and Shi, J. (2010) A numerical study on B&B algorithms for solving sum-of-ratios problem. Advances in Computer Science and Information Technology: 356-362. Springer, Berlin, Heidelberg.

Gensch, D. and Ghose, S. (1997) Differences in independence of irrelevant alternatives at individual vs aggregate levels, and at single pair vs full choice set. *Omega* 25(2): 201-214.

Gopalakrishnan, R. and Rangaraj, N. (2010) Capacity management on long-distance passenger trains of Indian Railways. *Interfaces* 40(4): 291-302.

Green, P. E. and Krieger, A. M. (1987) A consumer-based approach to designing product line extensions. *Journal of Product Innovation Management* 4(1): 21-32.

Green, P. E. and Krieger, A. M. (1989) Recent contributions to optimal product positioning and buyer segmentation. *European Journal of Operational Research* 41(2): 127-141.

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

Hanson, W. and Martin, K. (1996) Optimizing multinomial logit profit functions. *Management Science* 42(7): 992-1003.

Hausman, W. and Chen, K. (2000) Technical note: mathematical properties of the optimal product line selection problem using choice-based conjoint analysis. *Management Science* 46(2): 327-332.

Hensher, D. A. and Greene, W. H. (2003) The mixed logit model: the state of practice. *Transportation* 30(2): 133-176.

Hetrakul, P. and Cirillo, C. (2013) Accommodating taste heterogeneity in railway passenger choice models based on internet booking data. *Journal of Choice Modelling* 6: 1-16.

Hetrakul, P. and Cirillo, C. (2014) A latent class choice based model system for railway optimal pricing and seat allocation. *Transportation Research Part E: Logistics and Transportation Review* 61: 68-83.

Hohberger, S. (2019a) A simulation study for a choice-based dynamic pricing approach for large-scale railway networks. *Working Paper*.

Hohberger, S. (2019b) An overview on passenger railway revenue management. *Working Paper*.

Hohberger, S. and Schön, C. (2019) An Algorithm to Create Test Data for Large-Scale Railway Network Revenue Management Models with Customer Choice. Available at SSRN: 3439270.

Jiang, X., Chen, X., Zhang, L. and Zhang, R. (2015) Dynamic demand forecasting and ticket assignment for high-speed rail revenue management in China. *Transportation Research Record* 2475(1): 37-45.

Keller, P. W., Levi, R. and Perakis, G. (2014) Efficient formulations for pricing under attraction demand models. *Mathematical Programming* 145(1-2): 223-261.

Kimes, S. E. (1989) The basics of yield management. *Cornell Hotel and Restaurant Administration Quarterly* 30(3): 14-19.

Kolisch, R., Sprecher, A. and Drexl, A. (1995) Characterization and generation of a general class of resource-constrained project scheduling problems. *Management Science* 41(10): 1693-1703.

Kunnumkal, S. (2015) On upper bounds for assortment optimization under the mixture of multinomial logit models. *Operations Research Letters* 43(2): 189-194.

Lee, A. O. (1990) Airline reservations forecasting: Probabilistic and statistical models of the booking process. Doctoral dissertation, Massachusetts Institute of Technology.

Li, G., Rusmevichientong, P. and Topaloglu, H. (2015) The d-level nested logit model: Assortment and price optimization problems. *Operations Research* 63(2): 325-342.

Li, H. and Huh, W. T. (2011) Pricing multiple products with the multinomial logit and nested logit models: Concavity and implications. *Manufacturing & Service Operations Management* 13(4): 549-563.

Li, H., Webster, S., Mason, N. and Kempf, K. (2018) Winner – 2017 M&SOM Practice-Based Research Competition – Product-Line Pricing Under Discrete Mixed Multinomial Logit Demand. *Manufacturing & Service Operations Management*: Published online in Articles in Advance 16 May 2018.

Link, H. (2004) Pep–a yield-management scheme for rail passenger fares in Germany. *Japan Railway & Transport Review* 38(3): 50-55.

Littlewood, K. (2005) Special Issue Papers: Forecasting and control of passenger bookings. *Journal of Revenue and Pricing Management* 4(2): 11-123.

Liu, Q. and van Ryzin, G. (2008) On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management* 10(2): 288-310.

Maglaras, C. and Meissner, J. (2006) Dynamic pricing strategies for multiproduct revenue management problems. *Manufacturing & Service Operations Management* 8(2): 136-148.

McCormick, G. P. (1976) Computability of global solutions to factorable nonconvex programs: Part I – Convex underestimating problems. *Mathematical Programming* 10(1): 147-175.

McFadden, D. (1974) Conditional logit analysis of qualitative choice behavior. In Zarmbka, P., ed., Frontiers in Econometrics, Chapter 4: 105-142. Academic Press, New York.

McFadden, D. (2001) Economic choices. *American economic review* 91(3): 351-378.

McFadden, D. and Train, K. (2000) Mixed MNL models for discrete response. *Journal of applied Econometrics* 15(5): 447-470.

McGill, J. I. and van Ryzin, G. (1999) Revenue management: Research overview and prospects. *Transportation Science* 33(2): 233-256.

Méndez-Díaz, I., Miranda-Bront, J. J., Vulcano, G. and Zabala, P. (2014) A branch-and-cut algorithm for the latent-class logit assortment problem. *Discrete Applied Mathematics* 164: 246-263.

Newman, J. P., Ferguson, M. E., Garrow, L. A. and Jacobs, T. L. (2014) Estimation of choice-based models using sales data from a single firm. *Manufacturing & Service Operations Management* 16(2): 184-197.

Pölt, S. (2019) Back to Leg? Conference presentation at GOR-AG Pricing & Revenue Management Annual Meeting 2019, Munich.

Proussaloglou, K. and Koppelman, F. S. (1999) The choice of air carrier, flight, and fare class. *Journal of Air Transport Management* 5(4): 193-201.

Ratliff, R. and Vinod, B. (2005) Future of revenue management: Airline pricing and revenue management: A future outlook. *Journal of Revenue and Pricing Management* 4(3): 302-307.

Rayfield, W. Z., Rusmevichientong, P. and Topaloglu, H. (2015) Approximation methods for pricing problems under the nested logit model with price bounds. *INFORMS Journal on Computing* 27(2): 335-357.

Rusmevichientong, P., Shmoys, D., Tong, C. and Topaloglu, H. (2014) Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management* 23(11): 2023-2039.

Sato, K. and Sawaki, K. (2012) Dynamic pricing of high-speed rail with transport competition. *Journal of Revenue and Pricing Management* 11(5): 548-559.

Schaible, S. (1976) Minimization of ratios. *Journal of Optimization Theory and Applications* 19(2): 347-352.

Schaible, S. and Shi, J. (2003) Fractional programming: the sum-of-ratios case. *Optimization Methods and Software* 18(2): 219-229.

Schön, C. (2010a) On the product line selection problem under attraction choice models of consumer behavior. *European Journal of Operational Research* 206(1): 260-264.

Schön, C. (2010b) On the optimal product line selection problem with price discrimination. *Management Science* 56(5): 896-902.

Schön, C. and Hohberger, S. (2019) Continuous Pricing in a Capacitated Network under Mixed Multinomial Logit Demand. *Working Paper*.

Sen, A., Atamtürk, A. and Kaminsky, P. (2018) A conic integer optimization approach to the constrained assortment problem under the mixed multinomial logit model. *Operations Research* 66(4): 994-1003.

Sibdari, S., Lin, K. Y. and Chellappan, S. (2008) Multiproduct revenue management: An empirical study of Auto train at Amtrak. *Journal of Revenue and Pricing Management* 7(2): 172-184.

Song, J. S. and Xue, Z. (2007) Demand management and inventory control for substitutable products. *Working Paper*.

Steiner, W. J. and Hruschka, H. (2002) A Probabilistic One-Step Approach to the Optimal Product Line Design Problem Using Conjoint and Cost Data. *Review of Marketing Science Working Papers* 1(4): 4.

Strauss, A. K., Klein, R. and Steinhardt, C. (2018) A review of choice-based revenue management: Theory and methods. *European Journal of Operational Research* 271(2): 375-387.

Talluri, K., Castejon, F., Codina, B. and Magaz, J. (2010) Proving the performance of a new revenue management system. *Journal of Revenue and Pricing Management* 9(4): 300-312.

Talluri, K. and van Ryzin, G. (2004) Revenue management under a general discrete choice model of consumer behavior. *Management Science* 50(1): 15-33.

Talluri, K. and van Ryzin, G. (2005) The theory and practice of revenue management. Springer, New York.

Terabe, S. and Ongprasert, S. (2006) Seat inventory control for intercity passenger rail: how it works with customer satisfaction. *Proceedings of European Transport Conference*.

Train, K. E. (2008) EM algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling* 1(1): 40–69.

Trespalacios, F. and Grossmann, I. E. (2014) Algorithmic approach for improved mixed-integer reformulations of convex generalized disjunctive programs. *INFORMS Journal on Computing* 27(1): 59-74.

Trespalacios, F. and Grossmann, I. E. (2016) Cutting plane algorithm for convex generalized disjunctive programs. *INFORMS Journal on Computing* 28(2): 209-222.

Umweltbundesamt (2018) Vergleich der durchschnittlichen Emissionen einzelner Verkehrsmittel im Personenverkehr Bezugsjahr 2017 (TREMOD 5.82). Available at https://www.umweltbundesamt.de/sites/default/files/medien/366/bilder/dateien/vergleich_der_durchschnittlichen_emissionen_einzelner_verkehrsmittel_im_personenverkehr_bezugsjahr_2017.pdf (accessed October 3, 2019).

Van de Geer, R. and den Boer, A. (2018) Price Optimization Under the Finite-Mixture Logit Model. Available at SSRN: 3235432.

Vulcano, G., van Ryzin, G. and Chaar, W. (2010) OM Practice – Choice-Based Revenue Management: An Empirical Study of Estimation and Optimization. *Manufacturing & Service Operations Management* 12(3): 371-392.

Wang, X., Wang, H. and Zhang, X. (2016) Stochastic seat allocation models for passenger rail transportation under customer choice. *Transportation Research Part E: Logistics and Transportation Review* 96: 95-112.

Warburg, V., Bhat, C. and Adler, T. (2006) Modeling demographic and unobserved heterogeneity in air passengers' sensitivity to service attributes in itinerary choice. *Transportation Research Record: Journal of the Transportation Research Board* 1951(1): 7-16.

Wedel, M. and Kamakura, W. A. (2012) Market segmentation: Conceptual and methodological foundations, Volume 8. Springer Science & Business Media, New York.

Williamson, E. L. (1992) Airline network seat inventory control: Methodologies and revenue impacts. Doctoral dissertation, Massachusetts Institute of Technology.

Xiaoqiang, Z., Lang, M. and Jin, Z. (2017) Dynamic pricing for passenger groups of high-speed rail transportation. *Journal of Rail Transport Planning & Management* 6(4): 346-356.

You, P. S. (2008) An efficient computational approach for railway booking problems. *European Journal of Operational Research* 185(2): 811-824.

Zhai, X., Zhao, J. and Chen, Q. (2016) Optimization of the assignment of tickets for railway networks with large passenger flows. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit* 232(2): 632-642.

Zhang, D. and Adelman, D. (2009) An approximate dynamic programming approach to network revenue management with customer choice. *Transportation Science* 43(3): 381-394.

Zhang, D. and Lu, Z. (2013) Assessing the value of dynamic pricing in network revenue management. *INFORMS Journal on Computing* 25(1): 102-115.

Curriculum vitae

Academic employment

01/2017 – 12/2019	**Research Assistant** University of Mannheim, Business School, Mannheim, Germany Chair of Service Operations Management

Education

01/2017 – 01/2020	**Doctoral Candidate** University of Mannheim, Business School, Mannheim, Germany Chair of Service Operations Management
09/2013 – 04/2015	**Mannheim Master in Management (M. Sc.)** University of Mannheim, Business School, Mannheim, Germany
09/2012 – 01/2013	**Semester abroad** Católica Lisbon School of Business and Economics, Lisbon, Portugal
10/2010 – 08/2013	**Economics and Business Administration (B. Sc.)** Goethe University, Frankfurt am Main, Germany