UNIVERSITÉ DE MONTRÉAL

DEMAND FORECASTING IN REVENUE MANAGEMENT SYSTEMS

SHADI SHARIF AZADEH
DÉPARTEMENT DE MATHÉMATIQUES ET DE GÉNIE INDUSTREL
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

THÈSE PRÉSENTÉE EN VUE DE L'OBTENTION
DU DIPLÔME DE PHILOSOPHIÆ DOCTOR
(MATHÉMATIQUES DE L'INGÉNIEUR)
AOÛT 2013

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

Cette thèse intitulée :

DEMAND FORECASTING IN REVENUE MANAGEMENT SYSTEMS

présentée par : SHARIF AZADEH Shadi
en vue de l'obtention du diplôme de : Philosophiæ Doctor
a été dûment acceptée par le jury d'examen constitué de :

M. ADJENGUE Luc, Ph.D., président
M. SAVARD Gilles, Ph.D., membre et directeur de recherche
M. MARCOTTE Patrice, Ph.D., membre et codirecteur de recherche
M. LABIB Richard, Ph.D., membre et codirecteur de recherche
M. BASTIN Fabian, Doct., membre
Mme CIRILLO Cinzia, Ph.D., membre

To My Heroes, My Lovely Parents, My Dear Saeed , To The Love of My Life, Yousef
And To My Beloved Zari Who Always Believed in Me...

ACKNOWLEDGEMENT

I wish to express my sincere appreciation to many individuals and organizations who made this research possible.

Special acknowledgment is made to Professor Gilles Savard for directing the thesis with his well understanding, encouragement and guidance during the research.

Grateful acknowledgment is made to Professor Patrice Marcotte and Richard Labib for providing invaluable suggestions during my studies.

I would like to thank my industrial advisors, Dr. Anne Mercier, Dr. Louis-Philippe Bigras and Dr. Gerardo Berbeglia for their support and guidance.

I would like to express my deepest gratitude to my second family at GERAD who made this journey as enjoyable as possible. Special thanks to my dear colleagues and friends : Morad Hosseinalifam, Dr. Ahad Dehghani, Sanaz bararpour, Samin Sadrebazzaz.

Finally, I wish to extend my deepest appreciation and gratitude first and foremost to my parents who have shown me how to be passionate and motivated for getting better every day by example, to my best friend, the warmth and the love of my life, Yousef, to my adorable baby brother, Saeed and kind sister, Maryam, and my lovely baba Reza and maman Fariba not only for their encouragement, but also for their help and support in making this research possible.

This work has been supported by Natural Sciences and Engineering Research Council of Canada (NSERC)-Fonds Québécois de la Recherche sur la Nature et les Technologies (FQRNT) and ExPretio company in Montreal.

RÉSUMÉ

La gestion des revenues est l'art de développer des modèles mathématiques capables de déterminer quel produit offrir à quel segment de consommateurs à un moment précis dans le but de maximiser les profits. La prévision de la demande joue un rôle fondamental dans la gestion des revenues, car un manque de précision à cet égard engendrer une perte de profits.

Dans cette thèse, nous proposons une étude systématique et approfondie de différentes méthodes qui sont employées pour prévoir la demande. Tout d'abord, nous présenterons un nouveau schéma de classification détaillant les caractéristiques de ces différentes méthodes pour déterminer en quoi elles diffèrent les unes des autres. Dans ce but, nous ferons une analyse exhaustive de la littérature existant à ce sujet pour être à même de bien catégoriser ces méthodes dans notre schéma. Par la suite, nous investiguerons à propos des systèmes de gestion des revenue qui utilisent un réseau neuronal artificiel modifié combiné à un historique des données pour prévoir le nombre de passagers, selon les heures de départ, pour une importante entreprise européenne de transport ferroviaire. Après, afin de bien cerner les effets de saisonnalité et modéliser le comportement des consommateurs, nous proposerons un nouveau modèle non paramétrique.

La source de notre problématique part d'un modèle non-convexe et non linéaire composé de variables entières. Dans ce modèle, les variables représentent l'utilité de chaque produit ainsi que la demande potentielle de chaque jour et les variables binaires qui sont utilisées afin d'assigner chaque jour à chaque groupe des jours selon ses caractéristiques. Nous avons linéarisé et rendu convexe ce modèle avec succès en utilisant des techniques de linéarisation. Puis, nous avons présenté les caractéristiques de la disponibilité pour un temps donné afin d'extraire les corrélations entre les probabilités générées par ces choix. De plus, nous avons déterminé pour chaque journée un nombre prédéfini de blocs selon les caractéristiques spécifiques de la demande. Ainsi, nous avons pu déterminer une solution initiale basée sur laquelle on serre l'amplitude des variables. Ensuite, nous avons représenté un algorithme séparation et évaluation impliquant des techniques d'optimisation globale pour estimer les utilités et la demande potentielle à chaque jour. Le prétraitement des données a nécessité l'implémentation de plusieurs nœuds avant effectuer le branchement. Ce processus utilise des solveurs linéaires et non linéaires. Les résultats sont représentés par données synthétiques et données réelles.

Par ailleurs, ces résultats sont comparés à deux modèles non linéaires d'optimisation globale bien connus. Le modèle que nous proposons offre une performance nettement supérieure. Dans la dernière partie de cette dissertation, nous étudierons l'impact de ce modèle de demande sur la performance des revenues générées. Les résultats sont représentés à l'aide des données synthétiques générés par une programmation linéaire déterministe basée sur les modèles de choix discret.

Mots clés : Système de gestion des revenues, Modèle de choix discret, Réseau de neurones artificiel, Optimisation globale

ABSTRACT

A revenue management system is defined as the art of developing mathematical models that are capable of determining which product should be offered to which customer segment at a given time in order to maximize revenue. Demand forecasting plays a crucial role in revenue management. The lack of precision in demand models results in the loss of revenue. In this thesis, we provide an in-depth and systematic study of different methods that are applied to demand forecasting. We first introduce a new classification scheme for them and propose the characteristics that differentiate the methods from one another. All existing papers are reviewed and many of them have been categorized based on our classification scheme. After, we investigated a demand prediction model that uses a modified neural network method and historical data to forecast the number of passengers at the departure time for a major European railway company. Afterwards, in order to capture seasonal effects and taking customer behavior into account, we proposed a new, non-parametric mathematical model. The original problem is a nonconvex nonlinear model with integer variables. The variables in this model are the product utilities, the daily demand flow and binary assignment variables. We successfully linearized and convexified the model by using linearization techniques. Then, we used the characteristics of product availabilities for a given time to extract logical relations between choice probabilities. Moreover, we have classified each day to one of the predefined numbers of clusters based on their related daily demand flow. We represent a branch and bound algorithm, which uses global optimization techniques to find the estimated utilities and daily potential demand. Several node preprocessing techniques are implemented before branching. Both linear and nonlinear solvers are used in the branching process. The computational results are represented by using synthetic data. Also, they are compared to two well-known nonlinear and global optimizers and our proposed model outperforms both solvers. In the final part of this dissertation, we investigate the impact of the suggested demand model on revenue performance. The numerical results are presented using synthetic data produced by a modified Deterministic Choice-Based Linear Programming approach.

Keywords: Revenue Management, Choice-Based Demand Modeling, Uncensoring Methods, Neural Networks, Global Optimization Approach

TABLE OF CONTENTS

DEDICATION iii

ACKNOWLEDGEMENT iv

RÉSUMÉ v

ABSTRACT vii

TABLE OF CONTENTS viii

LIST OF TABLES xi

LIST OF FIGURES xii

INTRODUCTION 1

CHAPTER 1 ARTICLE 1 : A TAXONOMY OF DEMAND UNCENSORING METHOD IN REVENUE MANAGEMENT 4
1.1 Introduction 4
1.2 Features of demand models in RM systems 6
1.2.1 Supply-side features 7
1.2.2 Demand characteristics 8
1.3 Estimation of unconstrained demand 10
1.3.1 Basic methods 11
1.3.2 Statistical methods 12
1.3.3 Choice-based models 15
1.3.4 Optimization methods 16
1.4 A taxonomy 20
1.5 Conclusion 21

CHAPTER 2 ARTICLE 2 : RAILWAY DEMAND FORECASTING IN REVENUE MANAGEMENT USING NEURAL NETWORKS 24
2.1 Introduction 24
2.2 Problem definition 26
2.3 Pre-processing 28

2.4 Model . . . 30
2.4.1 Architecture of the neural network . . . 30
2.4.2 Learning algorithm and parameter adjustments . . . 31
2.4.3 Model improvements . . . 32
2.4.4 Validation . . . 33
2.5 Results . . . 34
2.6 Conclusion . . . 39

CHAPTER 3 ARTICLE 3 : A NON-PARAMETRIC APPROACH TO DEMAND FORECASTING IN REVENUE MANAGEMENT . . . 44
3.1 Introduction . . . 44
3.2 Problem formulation . . . 46
3.3 Algorithmic framework . . . 49
3.3.1 Linearization . . . 49
3.3.2 Relaxation and convexification . . . 51
3.4 Solution algorithm . . . 52
3.4.1 Preprocessing . . . 52
3.4.2 Branch-and-bound . . . 54
3.5 Computational results . . . 57
3.5.1 Data generation . . . 57
3.5.2 Numerical experiments . . . 58
3.6 Conclusion . . . 61

CHAPTER 4 ARTICLE 4 : THE IMPACT OF CUSTOMER BEHAVIOR MODELS ON REVENUE MANAGEMENT SYSTEMS . . . 64
4.1 Introduction . . . 64
4.2 Problem description . . . 66
4.3 Modified CDLP model . . . 67
4.4 Non-parametric approach . . . 69
4.5 Computational results . . . 70
4.5.1 Data instances . . . 70
4.5.2 Numerical results . . . 71
4.5.3 Discussion . . . 73
4.6 Conclusion . . . 73

CHAPTER 5 GENERAL DISCUSSION . . . 74

CONCLUSION AND RECOMMENDATION 76

REFERENCES 78

LIST OF TABLES

Table 1.1 Uncensoring demand : basic methods . . . 13
Table 1.2 α_3- non-choice based methods . . . 22
Table 1.3 α_3- choice based methods . . . 23
Table 2.1 Table of inputs and outputs . . . 40
Table 2.2 Table of frequencies of January 2005 for the number of bookings and cancellations . . . 41
Table 2.3 Proportion of outliers of output variables . . . 41
Table 2.4 Error at each step by adding each method for improving the results . . . 41
Table 2.5 7-Fold cross validation (January 2005) . . . 42
Table 2.6 3-Fold cross validation (January 2005) . . . 42
Table 2.7 5-Fold cross validation (January 2005) . . . 42
Table 2.8 5-Fold cross validation (Monthly data of 2007) . . . 43
Table 3.1 Demand censorship . . . 47
Table 3.2 Summary of notation . . . 48
Table 3.3 Clustering algorithm for determining initial solution . . . 55
Table 3.4 Non-Perturbed instances . . . 59
Table 3.5 Perturbed instances . . . 62
Table 3.6 Comparison Framework . . . 63
Table 4.1 Revenue Comparison . . . 72

LIST OF FIGURES

Figure 1.1 Demand censorship 6
Figure 1.2 Elements of Demand Forecasting Revenue Management (DFRM) 10
Figure 1.3 Methods Applied to Uncensor Demand (MAUD) 20
Figure 2.1 Sample pattern of the number of passengers at departure 27
Figure 2.2 Final structure of the network having two hidden layers with 5 neurons each 31
Figure 2.3 Fitting exponential distributions to given dataset 35
Figure 2.4 Process of determining the number of hidden neurons according to the method of network growing 36
Figure 2.5 Network training through iterative process 37
Figure 2.6 Prediction accuracy before imposing improvement methods 37
Figure 2.7 Improvements impact error reduction 38
Figure 2.8 Prediction accuracy after imposing improvement methods 38
Figure 2.9 Cross validation with March 2005 data 39
Figure 3.1 Algorithm effectiveness in reducing the gap 60
Figure 4.1 Registered bookings of a given product for successive departure days . . 66
Figure 4.2 Non parametric method of preference estimation and its impact on revenue 71

INTRODUCTION

Revenue Management (RM) is the application of disciplined tactics that predict consumer behavior and optimize product availability and price to maximize revenue (Cross (1997)). These systems include two main components. The first one is an optimization tool that finds the best price and allocation scenario and the second is a demand forecasting tool. The demand modeling aspect provides the essential input to the optimization model, which has been neglected in the literature compared to research that has been done on optimization models. Traditionally, demand for different products in RM systems is assumed to be independent. Many methods have been used in order to provide precise demand models. A detailed study of the application of statistical methods and probabilistic demand models in the airline industry was performed by Lee (1990). Usually, statistical methods are used in capturing seasonal effects of demand ; however, they often fail to properly respond to sudden changes in the set of available products at a given time. van Ryzin (2005) has stated that in revenue management systems, it is beneficial to use customer behavior models instead of product demand models. Afterwards, more optimization methods have been incorporated with choice probabilities in order to better monitor customer behavior (see van Ryzin et Vulcano (2008) ; Talluri (2010) ; Škurla Babić *et al.* (2011) ; Vulcano *et al.* (2010) ; Zeni et Lawrance (2004) ; Haensel et Koole (2010) ; Farias (2007)). This assumption about demand independence is convenient because it simplifies the computations while using historical data. However, in reality, the demands of different products are dependent. As soon as one of the products at a given time interval is no longer available, the data collection system stops gathering information. Therefore, historical data represents only the registered bookings (censored demand). If the censored data is ignored in forecasting models, it causes an underestimation of demand, which results in loss of revenue (Cooper *et al.* (2006) ; Weatherford et Belobaba (2002)). Hence, there is an essential need to uncensor demand in order to provide reliable forecasting models in revenue management systems. In the literature, there has been some research that has tackled the problem of censored demand by using different methods. That is, research has found the demand that one would have observed if unavailable products had still been available (Liu *et al.* (2002) ; Haensel et Koole (2010) ; Weatherford et Polt (2002) ; Queenan *et al.* (2009)).

The content of this dissertation can be categorized in four main parts. In the first part, we introduce a state-of-art taxonomy on uncensoring methods in demand forecasting in revenue management. In the second part, we apply a modified statistical method (Artificial Neural Networks) in order to predict the number of passengers at departure time for a major railway company. The third part introduces a new algorithm that is able to capture the seasonal effects of departure days and estimate the utilities of offered products. The final part tests the effect of our proposed model on revenue by using synthetic data.

Chapter 1 introduces a complete study of the existing uncensoring methods in demand forecasting in RM systems. More than two hundred articles in this field were reviewed and categorized based on our proposed criteria. The existing articles are classified based upon a tuple notation technique, which is represented in this chapter. Our main contributions are as follows :

- Representing the main features of demand from both supplier and consumer sides in Revenue Management Systems (RMS).
- Introducing uncensoring methods applied to revenue management problems from different mathematical aspects.
- Defining a new tuple notation method to classify research in this domain.

Chapter 2 shows a modified artificial neural network that is used in order to predict the demand of each product for a major railway company. In this research we incorporate statistical techniques (for preprocessing data) and neural networks in transportation demand forecasting. The model used is an improved Multi-Layer Perceptron (MLP) that describes the relationship between the amount of passengers and factors that affect this quantity based on historical data. The main contributions can be classified as follows :

- Proposing a tailored neural network to predict number of passengers for a railway company.
- Introducing a relevant pre-processing approach to make the learning process efficient.
- Testing the generalization ability of the network using real data.

Chapter 3 presents an original model of demand forecasting that uses a least-square method of optimization to predict the demand of each product at a given time. One of the original aspects of the model is that it avoids a parametric representation of the product utilities. Unlike most of the models in the literature, we consider utilities to be variables of our mathematical formulation. Therefore, utilities are defined based on products, which is the main difference between our work and classical choice-based models. We briefly present the main nonconvex nonlinear problem and its variables. We then introduce a new non-parametric algorithm that is able to forecast demand under the change of product availabilities. Afterwards, we linearize and convexify the original problem and propose a series of properties that enable us to increase the quality of our solutions. A branch and bound is introduced to solve the problem. Both linear and nonlinear solvers are used at the same time in our branch and bound. This model estimates the expected demand of each product at a given time in addition to product utilities. Below, the main contributions that shape this chapter's framework are summarized :

- Proposing a new mixed integer nonlinear formulation for modeling demand.
- Capturing customer behavior and demand seasonal effects simultaneously.
- Introducing a global optimization method with a tailored branch and bound strategy to solve the problem.

Finally, in Chapter 4, we examine the impact of our proposed prediction model on revenue by using a modified CDLP problem and providing a small simulation study. Our main contributions in this chapter are summarized as follows :

- Finding product utilities using directly historical data.
- Proposing a non-parametric method to obtain a customer preference vector.
- Comparing the impact of parametric and non-parametric method of preference estimation on revenue.

We analyze the outcomes of our research in Chapter 5 and we discuss possible future works.

CHAPTER 1

ARTICLE 1 : A TAXONOMY OF DEMAND UNCENSORING METHOD IN REVENUE MANAGEMENT

Chapter Information : An article based on this chapter is submitted for publication. Sh. Sharif Azadeh, P. Marcotte, and G. Savard.

In this article, more than two hundred papers were reviewed and categorized based on a state-of-art classification technique.

Abstract Revenue management systems rely on customer data, and are thus affected by the absence of registered demand that arises when a product is no longer available. In the present work, we review the uncensoring (or unconstraining) techniques that have been proposed to deal with this issue, and develop a taxonomy based on their respective features. This study will be helpful in identifying the relative merits of these techniques, as well as avenues for future research.

Keywords Revenue management, Demand forecasting, Uncensoring, Statistical methods, Optimization, Customer choice behaviour.

1.1 Introduction

The purpose of Revenue Management (RM) is to enhance the profitability of a firm through the optimal management of its inventory. In the service industry (airlines, railways, hotels), this can be achieved by controlling the availability of products, in order to redirect customers to "products" with high profit margins. Throughout this process, a trade-off must be stricken between the sale of low cost products when resources are plentiful, and the protection of high fare products towards the end of the booking horizon. Any such strategy is highly dependent on historical demand forecasts, and must cope with the lack of information resulting from censored demand, i.e., virtual demand for products that have been withdrawn, due to their "booking limits" being reached. This demand may either be lost ("spill") or recaptured by a more expensive ("buy-up") or cheaper ("buy-down") available product. In either case, the observed demand does not match the true behaviour of the customers, and may yield unreliable estimates. According to Weatherford et Belobaba (2002), underestimating demand by 12.5%

to 25% can result in a loss of revenue from 1% to 3%, which is significant. The main goal of this paper is to review and propose a taxonomy for the techniques that have been developed to address the issue of missing data. The remaining of this introductory section, following a simple illustration of the censored data issue, will put it into the proper context of the RM literature.

Let us consider a service company that sells a high fare "product" A and a low fare product B. An arriving customer may wish to purchase A, B, or renege. As long as both products are available, i.e., the booking limits have not been reached, sale figures (registered demand) reflect actual demand. If the booking limit set by the RM policy is reached for product B first, the upcoming demand for B is either be transferred to A (buy-up) or lost (spill). This is illustrated in Figure 1.1 for given streams of arrivals. Note that, if streams A and B are independent, then the native demand for A should not change once B is closed. Even in such simple case, one realizes the difficulties of retrieving the true demand from incomplete historical data, and of striking the right balance between accuracy and practicality in real-life instances. This leads to a variety of approaches, which have been investigated from different viewpoints :

- Wickham (1995) probed statistical forecasting methods for short-term demand in the airline industry. Time series, linear regression and booking pickup models were considered to estimate demand where some historical data is missing.
- Lee (1990) and McGill et van Ryzin (1999) introduced a variety of statistical methods to extract demand features using registered booking data.
- Zeni (2001) and Weatherford (2000) investigated statistical unconstraining techniques at a micro-level. They integrated techniques such as imputations or expectation maximization (EM) within the framework of exponential smoothing, time series, linear regression, or pickup models.
- In van Ryzin (2005), the focus shifted from traditional product demand models to the analysis of customer behaviour, based on the theory of discrete choice (random utility).
- For an airline application, Ratliff *et al.* (2008) integrated product dependencies, and proposed a hierarchical classification of previous unconstraining models. Three frameworks were considered : (i) single-class models, where product demand is assumed to be independent, (ii) multi-class, with up-sell and down-sell among different fare classes, (iii) Multi-flight methods, which include the most general unconstraining approaches. All models take into account the interactions between the various fare products.

Although the above mentioned studies cover important subsets of uncensoring methods, there yet exists a need to structure the field, so that adequate methods be easily matched to areas of application. Hence our proposal for a flexible and expandable taxonomy that should

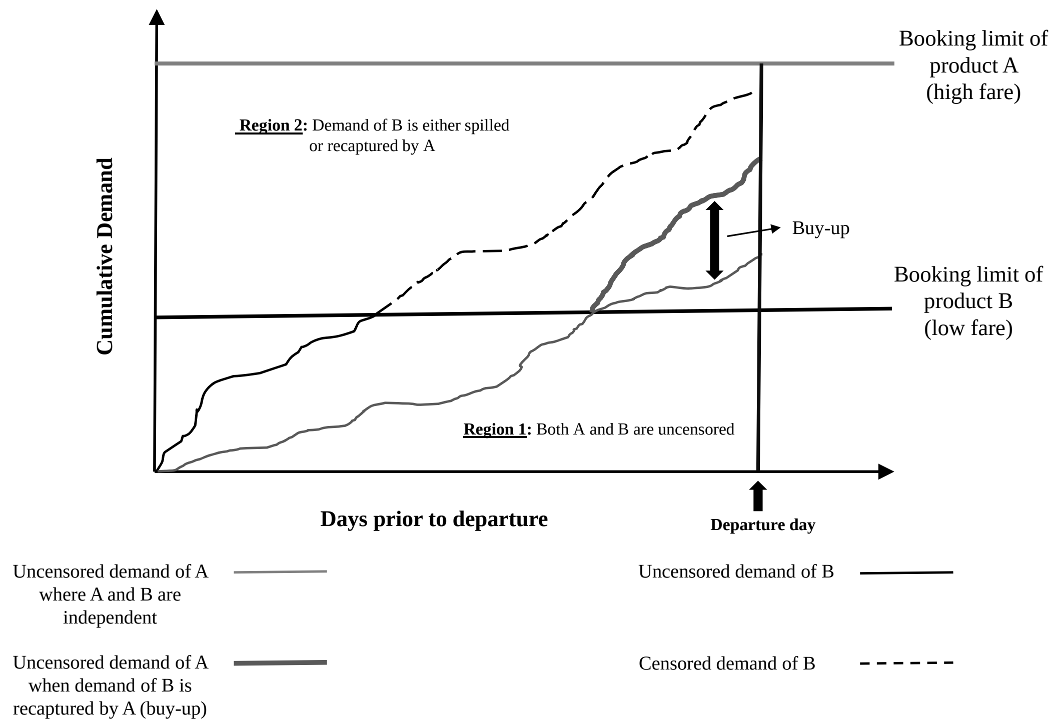


Figure 1.1 Demand censorship

prove useful in future settings.

Our classification makes use of a tuple notation, which allows for a concise review of existing models, identifying the key elements that distinguish them from one another. In this framework, a model is represented as a tuple $[\mu|\delta|\alpha]$, where μ is the set of attributes of the supplier, δ is the set of demand features, and α identifies uncensoring approaches. Whenever an element is not considered in a specific model, it does not appear in the tuple.

The remainder of this paper is structured as follows. Section 2 introduces the main elements of demand models in RM. Section 3 presents uncensoring methods, whose efficiencies are assessed in Section 4, together with guidelines for use in future applications. Avenues for further research are outlined in the concluding Section 5.

1.2 Features of demand models in RM systems

In this section, we introduce the features according to which the demand models will be classified.

1. Supply side
   - *Customer type (for data gathering)*
   - *Domain of application*

2. Demand
   - *Dependencies among products*
   - *Diversion* (spill or recapture)
   - *Seasonality*
   - *Segmentation (internal or external)*
   - *Competition*

They are detailed, together with their respective domains of application, in the following subsections.

1.2.1 Supply-side features

Forecasting techniques can address demand either at the macro or micro level. Macro-level analyses consider total demand, whereas micro-level forecasting is typically conducted on a booking date and fare-class basis (see Zeni (2001) ; Lee (1990)). In the context of micro-level forecasting (which is the main focus of this review), supply-side assumptions influence the choice of uncensoring method to a great extent. We consider the following classification :

1. Customer type (μ_1)
   - *Myopic*
   - *Strategic*
2. Domain of application (μ_2)
   - *Airline*
   - *Rental-Retail*
   - *Railway*
   - *Hotel*

The parameter $\mu_1 \in \{myop, strat\}$ refers to inter-temporal substitutions that involve (or not) delaying one's purchase (see Shen et Su (2007)). In the standard models, *myopic* customers make their final decision at the time of arrival, whereas more recent models allow *strategic* customers to reconsider their choice in the future (see Liu et van Ryzin (2008) ; Bansal (2012) ; Cachon et Swinney (2009) ; Su (2007) ; Yang *et al.* (2010) ; Levin *et al.* (2010) ; Cachon et Swinney (2011) ; Yin *et al.* (2009) ; Swinney (2011)).

The second attribute $\mu_2 \in \{air, rent - ret, rail, hotel\}$ refers to application domains. Although the initial research focused on airlines, RM has subsequently made its way into the realms of rental and retail (Ja *et al.* (2001) ; Stefanescu *et al.* (2004) ; Stefanescu (2009) ; Vulcano *et al.* (2010) ; Talluri (2009) ; Ratliff *et al.* (2008) ; Haensel et Koole (2010) ; Haensel *et al.* (2011)), and unconstraining methods have been applied to these domains (see also domains to which unconstraining methods are applied Zhu (2006) ; Conlon et Mortimer (2008) ; Huh *et al.* (2011)). This is also the case of the rail industry, especially in Europe, where competition with low cost airlines that operate "point-to-point" is fierce (Armstrong et Meissner (2010) Crevier *et al.* (2012)). In the hotel industry, Queenan *et al.* (2009) have assessed unconstraining techniques using actual data, whereas Haensel et Koole (2010) have done so for the hotel industry. Other proposals can be found in the recent literature (see Meissner *et al.* (2012) ; Ferguson et Queenan (2009) ; Bodea (2008)).

1.2.2 Demand characteristics

The impact of demand representation over the choice of an unconstraining technique is important. We characterize the demand process through the following five attributes δ_i :

1. Product dependency (δ_1)
   - *Dependent*
   - *Independent*
2. Diversion (δ_2)
   - *Spill*
   - *Recapture*
3. Seasonality (δ_3)
   - *Seasonal effects*
4. Segmentation (δ_4)
   - *Internal* (latent characteristics, such as income)
   - *External* (time dependent arrivals)
5. Competition (δ_5)
   - *Competition*

We now discuss the parameters in some detail. First, we note that the independence assumption facilitates the estimation process and makes it possible to address larger problem instances (Haensel et Koole (2010) ; Zeni (2001) ; Queenan *et al.* (2009) ; Meissner et Strauss (2012a) ; Meissner et Strauss (2012b)). However, it is clearly an over-simplification, and recent studies have explicitly considered correlations, either linear or nonlinear, either inter-temporal or not (McGill (1995)).

In most situations, customers who are denied their preferred choice have recourses within service companies products. If products are nested, i.e., a discontinued low fare cannot be reactivated further in time, then customers can either purchase at a higher fare (*buy-up*) or renege (*spill*) (Swan (1979) ; Swan (1999)). Some researchers have considered mass balance equations that link spill and recapture (Andersson (1998) ; Ja *et al.* (2001) ; Ratliff *et al.* (2008)). Vulcano *et al.* (2010)-(2012) have considered a method for estimating substitute and turned away demand for the case of incomplete data. Similar research has been conducted by Talluri et Van Ryzin (2004) and Haensel et Koole (2010).

In service companies such as rentals, airlines, railways and hotels, reservations are made days or weeks in advance. These periods of time are divided into *booking intervals* during which customers register for a specific day. Reservations usually face a considerable degree of seasonality, which may be inadequately captured if only a small portion of data is used to estimate the parameters of the model. In order to capture *seasonal effects*, the time scope of the historical data needs to be specified, i.e., one must determine the number of booking intervals included in the historical data. Too much information makes the forecasting model inflexible, whereas too little does not allow to capture seasonality in a meaningful fashion. To address the issue, a time series approach (ARIMA) has been adopted by Lee (1990), Sa (1987) and Queenan *et al.* (2009).

Market segmentation can affect both the choice of uncensoring and optimization approaches in RM systems. Ideally, one would tailor the fare of a product to the willingness-to-pay of each individual (see Gurbuz *et al.* (2011) ; Meissner et Strauss (2009) ; Talluri (2010)). In this framework, *internal segmentation* refers to customer features (income, purpose, age, etc.), while *external segmentation* refers to time-based customer behaviour. For instance, customers who book late are more likely to be business travellers who opt for high fare products, while weekenders are more likely looking for economy fares.

The last parameter δ_5 refers to *competition*. Surprisingly, this feature of revenue management, which actually motivated the very field, has only recently been paid close attention (see Jiang (2007) ; Jiang et Pang (2011) ; Perakis et Sood (2006) ; Kwon *et al.* (2009) ; Martínez et Talluri (2011) ; Gallego et Hu (2008) ; Belobaba (1987)). Including competition within a RM system can significantly modify the demand model, which could embed competition between different products of the same company, or competition between companies that offer similar products. For instance, Netessine et Shumsky (2005) have considered quantity-based games of booking controls under horizontal and vertical competition, and Liu et Zhang (2011) have addressed the issue of dynamic pricing competition between two firms offering vertically differentiated products to strategic consumers.

The tree-like Figure 1.2 summarizes the elements of demand models and their related components.

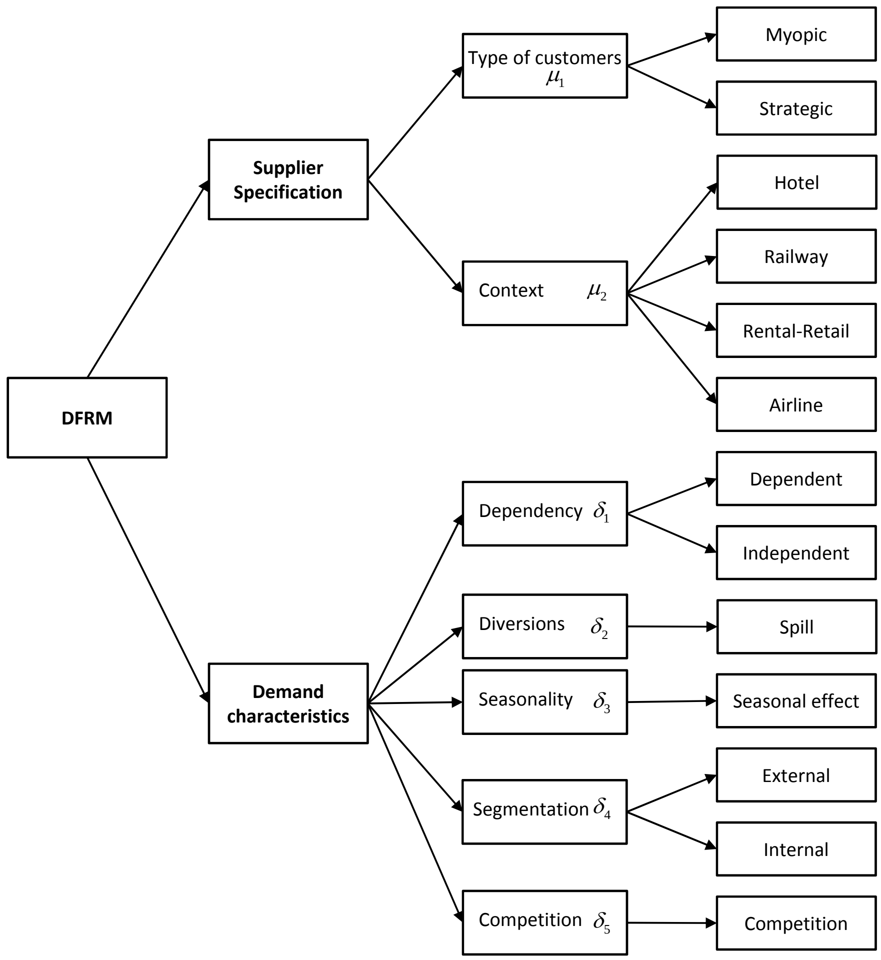


Figure 1.2 Elements of Demand Forecasting Revenue Management (DFRM)

1.3 Estimation of unconstrained demand

At the heart of the revenue management is a twofold process that consists in parameter estimation, and optimization. These can be conducted sequentially (estimate then optimize) or in parallel (estimate and optimize). Depending on the strategy adopted, and also on demand specification, different unconstraining methods can be applied, either parametric or non-parametric. Note that classical forecasting methods, such as time series and linear regression, are unable to properly capture customer behaviour and product availability at a given time.

Uncensoring methods can be classified into four main categories, each one associated with a symbol α_i : basic methods (α_1), statistical methods (α_2), choice-based models (α_3), and optimization methods (α_4). The first three categories fit the "estimate then optimize" framework, whereas optimization methods are of an "estimate and optimize" nature.

1.3.1 Basic methods

Basic methods are nonparametric and may either (i) limit themselves to observed booking data (ii) ignore censorship altogether (iii) discard censored data (iv) use "imputations" to make up for missing data. They are now discussed in more detail.

– Direct observation

One of the simplest methods to tackle the problem of censored demand is not to tell customers that their required product is unavailable. Rejected requests are then appended to registered bookings, resulting in an unbiased estimation of the true demand. In practice, the existence of several booking outlets (online or not) makes this "ideal" approach unsuitable, notwithstanding the additional burden of processing this data, and the impediment on the perceived quality of service. Moreover, this strategy could not cope with dynamic variations of customer behaviour (Queenan *et al.* (2009) ; Orkin (1998)).

– Ignoring censorship

Assuming that data is uncensored is tantamount to setting demand estimates to their booking limits, whenever these are reached, and will obviously lead to underestimation (Cooper *et al.* (2006) ; Saleh (1997) ; Little et Rubin (2002)).

– Discarding censored data

This strategy limits the size of the sample and may yield either over or underestimation, depending whether products with low or high demand levels are censored. This method usually performs adequately when the arrival process is totally random and the number of sell-outs is small. If these conditions are not fulfilled, a negative bias can occur (Zeni (2001) ; Saleh (1997)).

– Imputations

The term "imputation" refers to methods that fill in censored demand. A commonly used method is "mean imputation", whereby censored data is replaced by the mean of registered booking data, whenever the latter is less than the average (Zeni (2001) ; Little et Rubin (2002) ; Farias (2007)). In a similar fashion, one obtains an imputation based on the median of the historical unconstrained demand, in place of its mean.

Each approach is illustrated on an example involving 3 available products whose data is displayed in Table 1.1. In the first part of the table general information about these products are provided. Registered demand for "direct observation" method differs from the other three basic methods. Hence, we have represented registered and uncensored demand of this method, separately from "ignoring censorship", "discarding censorship" and "mean imputation".

We close this subsection with a list of the acronyms corresponding to each uncensoring approach.

$$\alpha_1 = \begin{cases} dir.obs & \text{Directly Observed booking data} \\ ign.cen & \text{Ignore Censorship} \\ dis.cen & \text{Discard Censored data} \\ imp & \text{Imputations} \end{cases}$$

1.3.2 Statistical methods

In revenue management, statistical methods are broadly expressed in three categories (Weatherford et Kimes (2003) and Lee (1990)) : historical, advanced, and combined booking models.

– Historical booking models

Historical booking models resort to traditional parametric forecasting such as time series, exponential smoothing, or linear regression (see Sa (1987) ; Littlewood (2005) ; Pölt (2000) ; Weatherford (2000) ; Kachitvichyanukul *et al.* (2012)).

Time series describe the random nature of the data, and are based on final booking numbers. Despite their relatively simple mathematical structure, they are rich enough to embody a wide range of data features. For one, the ARIMA model comprises autoregressive and moving average components (Box *et al.* (2011)). It can be mathematically expressed as follows :

$$Y_t = \mu + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q} \tag{1}$$

where Y_t represents the demand at time t, μ is the mean of a stationary process, θ_i's are coefficients, and ϵ_t's are uncorrelated random terms with zero mean and common variance σ_ε^2. The first terms in the above equation represent the autoregressive component, and the second set of linear combinations the moving average. Time series explicitly exploit the correlations between successive data points to improve forecasts.

Based on data observed up to time $t-1$, *Simple exponential smoothing* adjusts the next value $\hat{Y}_t$ through the formula

$$\hat{Y}_{t+1} = \hat{Y}_t + \alpha(Y_t - \hat{Y}_t) \tag{2}$$

where the parameter α lies between zero (no adjustment) and one ("strong" adjustment). This method, which relies on a weighted average of the most recent observations (Hyndman *et al.*

Table 1.1 Uncensoring demand : basic methods

Product	**A**	**B**	**C**	**D**
General information				
Availability	1	1	0	1
Booking limit	8	6	0	23
Actual demand	15	0	9	22
Uncensored demand by direct observation				
Registered demand (Direct observation)	12	0	0	22
Direct observation	12	0	0	22
Uncensored demand by :				
Registered demand (Other three methods)	8	0	0	22
Ignoring censorship	8	0	0	22
Discarding censorship	-	0	-	22
Mean imputation	10	0	0	22

(2008)), is not recommended for the analysis of time series characterized by a large number of null values and a high variability among the non-zero data.

Linear regression assumes a linear trend of registered bookings in successive time periods, the key issue being to properly select the number and nature of the descriptive variables entering the model. The parameters of the regression are usually estimated via least squares. For a case involving two descriptive variables over two successive booking intervals, we have that

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \epsilon_t \tag{3}$$

where Y_t is the current booking and Y_{t-1}, Y_{t-2} represent the total bookings for the two preceding time intervals. The drawback of this regression model is the underlying linearity assumption, which may not always hold.

– Advanced bookings

Advanced booking models (including *pickup*, *advanced pickup*, *booking profile*) are based on registered bookings over time, and can be of the *additive* or *multiplicative* type. Both types have been considered in the transportation literature (see L'Heureux (1986) ; Skwarek (1996a) ; Skwarek (1996b) ; Weatherford et Polt (2002) ; Zickus (1998) ; Gorin (2000) ; Mishra (2003) ; Lee (1990) ; Wickham (1995) ; Zakhary *et al.* (2008)). Classical or advanced pickups are additive models that differ in their treatment of historical data (Lee (1990) ; Wickham (1995)). In classical method, an overall average on the products that are no longer available is used to show uncensored demand. However, advanced method applies an incremental average over time, to better depict small changes in demand. In general, they assume no proportionality relationship between current registered bookings (at the time when the product is no longer available) and final bookings. Instead, they assume that the absolute growth (pickup) in bookings between the current time interval and the last interval of other open similar products is a good indicator of the booking history, had the closed products been still open. This yields

$$Y_0^i = Y_t^i + \frac{1}{J} \cdot \sum_{j=1}^{J} (Y_0^j - Y_t^j), \tag{4}$$

where Y_0^i is an estimate of the final uncensored booking of product i, J represents the number of remaining booking intervals, and Y_t^i is the current booking (at time t) of the closed product. In this formula, the average term corresponds to the mean number of pickups following the closure of available products.

L'Heureux (1986) has suggested that the inclusion of data drawn from all reservation intervals (i.e., taking incremental pickups into account) provides valuable information about demand behavior.

Multiplicative pickup models operate in a similar fashion, but base their forecasts on the "pickup ratio", defined as

$$pick - ratio(t, 0) = \frac{1}{J} \cdot \sum_{j=1}^{J} (Y_0^j - Y_t^j) \times \frac{1}{Y_t^i} \tag{5}$$

$$Y_0^i = Y_t^i \times pick - ratio(t, 0). \tag{6}$$

It is important to point out that these methods only rely on historical data, and neglect socioeconomic or behavioural features of the population. They are of course highly dependent on the quality of the data collection process.

– Combined models

Combined models use regression or weighted average of historical and advanced booking models to produce forecasts. In order to achieve high accuracy, they may resort to parametric regression, neural networks, or distribution based demand models. The use of *weighted moving average* allows to emphasize the most recent bookings. Given a set of weights summing up to one, we have

$$\hat{Y}_{t+1} = w_1 Y_t + w_2 Y_{t-1} + w_3 Y_{t-2} + \cdots + w_N Y_{t-N+1} \tag{7}$$

In this context, Wickham (1995) has implemented both simple and weighted averages and found that they were outperformed by pickup methods (see Van Ryzin et McGill (2000) ; Liu (2004) ; Ja *et al.* (2001)).

More advanced techniques, such as *supervised learning neural networks*, are akin to complex nonlinear regressions, and are able to process large and complex data sets. A neural network comprises an input layer, one or several hidden layers, and an output layer. Individual inputs are processed through the network, and their weighted combination is compared to the neuron's threshold value.

In the "training phase" one iteratively adjusts each weight until the difference between expected bookings and actual data falls below a predefined threshold value. Following this phase, the network is used to predict future demand from a data set that should not differ too widely from the training set. Although neural networks have been applied successfully to transportation demand forecasting (Weatherford *et al.* (2003a) ; Sharif Azadeh *et al.* (2012) ; Dantas *et al.* (2000)), supervised learning is yet unable to produce accurate forecasts when a large proportion of historical data is censored.

In *Distribution based demand models*, it is assumed that the statistical distribution underlying the demand process (usually *Normal* or *Gamma*) is known, and that its parameters (mean, variance, etc.) are estimated based on historical data.

Alongside the Normal or Gamma assumptions, Brummer *et al.* (1988) has considered log-normal distributions, while Logistic, Gamma, Weibull, Exponential and Poisson distributions have been advocated (see Guo (2008) ; ZF Li et Hoon Oum (2000) ; Swan (2002) ; Kaplan et Meier (1958) ; Huh *et al.* (2011) ; Popescu *et al.* (2012) ; Eren et Maglaras (2009)).

In the following, the statistical methods are partitioned according to the parameter α_2 :

$$\alpha_2 = \begin{cases} hbm & \text{“historical booking models” : } \textit{time series (tseries), exponential smoothing (exp.smooth)} \text{ and } \textit{linear regression (lin.reg)} \\ abm & \text{“advanced booking models” : } \textit{Additive, Multiplicative (pickup)}, \text{ and } \textit{Booking Profile (BP)} \\ cm & \text{“combined models” : } \textit{weighted average (weight.ave), parametric regression (par.reg), Neural Networks (NN),} \text{ and } \textit{distribution based demand (dist.dem)} \end{cases}$$

1.3.3 Choice-based models

The integration of a discrete choice framework (McFadden (2001)) within RM systems has provided the flexibility required to take into account strategic customers. In these models, these make their decision based on the set of available alternatives (*Choice sets*), under the following restrictions (Train (2009)) :

- only one choice can be made at any given time period ;
- all available choices are included in the choice set ;
- the number of alternatives is finite.

In the discrete choice framework, a customer selects the product that maximizes his expected utility, the latter being expressed as the sum of deterministic and a stochastic terms that are related to the features of each product. The choice of the random term results in different models : Probit (normal), Logit (Gumbel), Mixed Logit, etc., and their parameters are typically estimated via maximum likelihood. In the much touted Multinomial Logit model, which involves a Gumbel-distributed random term, the probability that a product i with utility u_i be selected is given by the closed form formula

$$P_i(S_t) = \frac{\exp(u_i)}{\sum\limits_{j \in S_t} \exp(u_j) + 1}, \tag{8}$$

where S_t denotes the subset of products available at time t. We will assign the acronym *cb* (choice-based) to the parameter α_3 when discrete choice models are considered.

The embedding of discrete choice models within an optimization process has been considered by Talluri et Van Ryzin (2004) ; Vulcano *et al.* (2010) ; Vulcano *et al.* (2012) ; Haensel et Koole (2010) ; Haensel *et al.* (2011) ; Conlon et Mortimer (2008) and Zhang et Cooper (2005). In particular, Belobaba et Hopperstad (1999) have studied the impact of customer behavior on traditional RM systems, while Talluri et Van Ryzin (2004) have characterized optimal control policies in a very general discrete choice setting.

1.3.4 Optimization methods

In recent years, techniques that focus on optimization have been introduced in choice-based RM. These can be broadly divided into four main categories : *Expectation-Maximization* (EM), *Projection-Detruncation* (PD), *Double Exponential Smoothing* (DES) and *Nonlinear Programming* (NLP). The first three methods are parametric, while nonlinear programming covers most nonparametric estimation methods.

– Expectation Maximization (EM)

After its introduction to revenue management in the late 1990's by Salch (1997), the two-stage EM process has quickly become one of the most popular unconstraining methods. In the first step, *E-step*, unobserved demand of an unavailable product is replaced by its average observed demand, prior to its reaching the capacity. In the subsequent *M-step*, the parameters of the demand distribution (mean and variance) are estimated via maximum likelihood. The first step is then repeated, and the fixed point process is halted when no significant progress is observed. In this setting, seasonality is usually ignored.

For a given product [1], let $Y_1, \ldots Y_{N_1}, Y_{N_1+1}, ..., Y_{N_1+N_2}$ denote a stream of registered bookings consisting of N_2 uncensored and N_1 censored realizations, the latter obtained after the product has reached its booking limit. Following common practice, the index of booking intervals decreases from n (in this case, $N_1 + N_2$) to 0, which corresponds to departure time in transportation RM.

Assuming that demand follows a normal distribution with mean μ and variance σ^2, when $\mu^{(0)}$ shows an initial value for the expected value, the procedure goes through the following steps.

Initialization : Estimate μ and σ, based on N_2 uncensored observed data :

$$\mu = \frac{1}{N_2} \sum_{i=N_1+1}^{N_1+N_2} Y_i \tag{9}$$

$$\sigma = \sqrt{\frac{1}{N_2} \sum_{i=N_1+1}^{N_1+N_2} (Y_i - \mu^{(0)})^2} \tag{10}$$

E-Step : For a given number C or constrained observations, the first and second moments of the censored data, required to form the log-likelihood function), are estimated according to the formula : iteratively (assuming that random variable of demand, Y) to replace the missing data to form the complete log-likelihood function where C represents registered constrained observation.

$$\hat{Y}_i^{(+)} = E[Y|Y > C, Y \sim N(\mu, \sigma)] \tag{11}$$

$$(\hat{Y}_i^2)^+ = E[Y^2|Y > C, Y \sim N(\mu, \sigma)] \tag{12}$$

for $i = 1, ..., N_1$, $Y_i, ..., Y_{N_1}$ and $Y_1^2, ..., Y_{N_1}^2$ are replaced by the above values to complete the data set.

M-Step : Maximize the log-likelihood function with respect to μ and σ to obtain μ^+ and σ^+.

Stopping criterion : Repeat steps E and M until the difference between successive iterates is less than some predetermined threshold value δ.

Several researchers have recognized the EM method as one the most efficient for uncensoring demand in RM (see Talluri et Van Ryzin (2005) ; Guo (2008) ; Pölt (2000) ; Weatherford (2000) ; Zeni (2001) ; Zeni et Lawrance (2004) ; Chen et Luo. (2005) ; He et Luo (2006) ; Karmarkar *et al.* (2010) ; McGill (1995) ; Haensel et Koole (2010) ; Haensel *et al.* (2011) ; Vulcano

1. Products are assumed to be independent.

et al. (2012) Little et Rubin (2002) ; Stefanescu *et al.* (2004) ; Stefanescu (2009) ; Hopperstad (1996) ; Hopperstad (1997) ; Hopperstad *et al.* (2007)). The drawback is that, if large and correlated data are involved, the maximum likelihood step is difficult to implement (Xu (1997) ; Naim et Gildea (2012)).

– Projection Detruncation (PD)

This method is similar to the EM method, but uses the median instead of the mean. Also, a weighting constant may be used to yield aggressive demand estimates. Recently, this method has been applied to RM systems (see Hopperstad (1995) ; Skwarek (1996a) ; Skwarek (1996b) ; Weatherford et Ratliff (2010) ; Chen et Luo. (2005) ; Zickus (1998) ; Gorin (2000) ; Zeni (2001) ; Guo (2008) and Queenan *et al.* (2009)).

– Double Exponential Smoothing (DES)

In DES, one predicts the total demand that would have been registered in the absence of booking limits. The first parameter is used for smoothing the *base component* of the demand pattern, and the second deals with the *trend component* (Queenan *et al.* (2009)). For each instance of censorship, a nonlinear optimization model estimates the two smoothing parameters while minimizing the forecasting error. This is achieved in the following manner. Let t be an instant when the booking limit of a given product has not been reached yet, i.e., registered demand matches observed demand up to t. Based on Queenan *et al.* (2009), let Y_t be the actual cumulative demand at time t, B_t the smoothed base component, T_t the smoothed trend component, and FT_t the cumulative forecast at time t, taking trend into account. The forecast for the upcoming time period $t+1$ then satisfies

$$FT_t = B_t + T_t \tag{13}$$

where

$$B_t = FT_{t+1} + \delta(Y_{t+1} - FT_{t+1}) \tag{14}$$

$$T_t = T_{t+1} + \beta(B_t - FT_{t+1}) \tag{15}$$

and the parameters δ and β are optimal solutions of the mathematical program

$$\min_{\delta,\beta} \sum_t (Y_t - FT_t)^2. \tag{16}$$

The procedure is initialized on historical data, and the nonconvex least-square problem may be solved via metaheuristics such as Tabu Search or Simulated Annealing. This framework has been applied to many demand uncensoring problems, and proved competitive with EM in most cases (Guo *et al.* (2008) ; Armstrong (2001)).

– Nonlinear Programming (NLP)

This class includes non-parametric methods such as *Least-Squares*, *discriminant analysis*, or *cluster analysis*. In the literature, Besbes et Zeevi (2006) have discussed a non-parametric algorithm that characterizes the underlying demand behavior. Farias *et al.* (2013) have considered non-parametric methods in the context of choice modeling with limited data. Lee *et al.* (2005) have used discriminant and cluster analysis to segment the customer population with respect to its preferences, with an application to the Taiwan Railway Administration.

Our classification of optimization is as follows.

$$\alpha_4 = \begin{cases} EM & \text{Expectation-Maximization} \\ PD & \text{Projection Detruncation} \\ DES & \text{Double Exponential Smoothing} \\ NLP & \text{Nonlinear Programming} \end{cases}$$

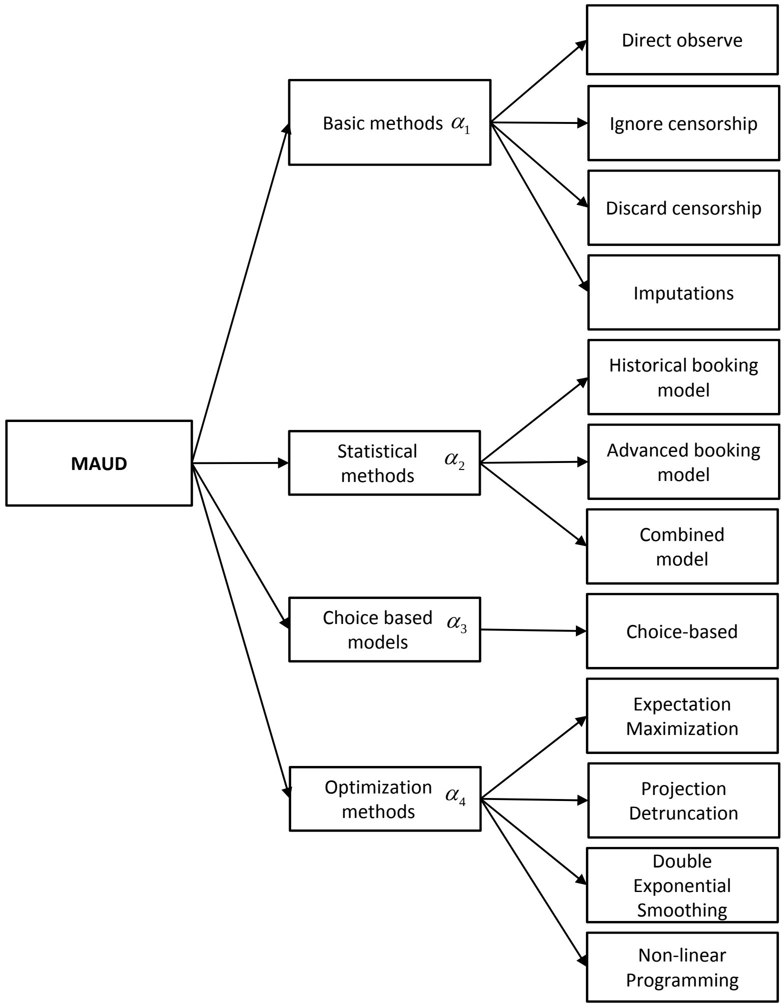


Figure 1.3 Methods Applied to Uncensor Demand (MAUD)

1.4 A taxonomy

In this section, the various techniques used in the revenue management literature to deal with missing data are classified with respect to their robustness, accuracy, applicability, and capability to deal with issues such as independence, stationarity, or seasonality. Throughout, each work is made to fit our tuple notation.

Table 1.2 focuses on non-choice based unconstraining methods. These ignore correlations between products, as well as the impact of availabilities on the demand for competing products, or competition from other firms. Within this class, EM and PD assume that demand distributions are known a priori, which is not the case in many practical situations.

Table 1.3 displays models where customers base their purchasing decisions upon product availabilities (choice set). Apart from a suitable representation of customer behaviour, the flexibility of these models may improve the accuracy of both the estimation and optimization processes. In this respect, two key issues that should be better addressed in the future are :

1. inter-product and inter-temporal correlations ($\delta_1 = dep$)
2. seasonal factors ($\delta_3 = season$)

When firms have the opportunity to set prices dynamically, it is natural to expect price variations to persist. In this environment, customers may react strategically to price fluctuations, and ignoring such responses may lead to sub-optimal pricing decisions. Note also that product diversity induces a significant correlation between demands for alternatives within the choice sets. If follows that the understanding of customer response to market mechanisms is an issue that should be addressed properly. In the same vein, the analysis of the optimal purchase timing (inter-temporal substitution) is being monitored more closely in both the industrial and academical worlds.

Other relevant issues include capacity rationing (creation of "artificial" scarcity to influence purchase timing), valuation uncertainty, and consumer learning effects. These relate the dynamics of consumer demand to the seller's dynamic pricing strategies, a dependency that is not captured by conventional models based on exogenous arrival processes.

With respect to seasonality, statistical methods show some promise, under simple assumptions. However, the nature of seasonality can be complex, as it may involve dimensions such as day of week, month of year, holidays, etc. Seasonality could actually be customer-dependent, and bedding this information within a combined estimation-optimization process is another challenge yet to be addressed. In this realm, note also that the very definition of reservation intervals is of importance.

1.5 Conclusion

Currently, demand forecasting may well be the most critical area in revenue management, and demand unconstraining clearly lies at the heart of the matter. The present paper aims at shedding some light on the latest developments in this area, through a novel taxonomy, in the hope of triggering research in this area.

Table 1.2 α_3- non-choice based methods

Author	Model Description
α_1 Basic methods	
Queenan *et al.* (2009)	$[\mu_1 = myop, \mu_2 = hotel \mid \delta_1 = ind, \delta_4 = in \mid \alpha_1^*]$
Zeni (2001)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind, \delta_4 = in \mid \alpha_1 = imp]$
Orkin (1998)	$[\mu_1 = myop, \mu_2 = hotel \mid \delta_1 = ind \mid \alpha_1 = dir.obs]$
Saleh (1997)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind \mid \alpha_1 = ign.cen, dis.cen, imp]$
α_2 : Statistical methods	
Sa (1987)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind, \delta_3 = season \mid \alpha_2 = hbm(\text{lin.reg}, \text{tseries})]$
Pölt (2000)- Weatherford (2000)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind \mid \alpha_2 = abm(BP)]$
Lee (1990)-Wickham (1995)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind \mid \alpha_2 = abm]$
Zeni (2001)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind, \delta_4 = in \mid \alpha_2 = abm(BP)]$
Van Ryzin et McGill (2000)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind \mid \alpha_2 = cm(LT)]$
Weatherford *et al.* (2003a)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind \mid \alpha_2 = cm(NN)]$
Dantas *et al.* (2000)	$[\mu_1 = myop, \mu_2 = rail \mid \alpha_2 = cm(NN)]$
ZF Li et Hoon Oum (2000)	$\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind, \delta_2 = spill, \delta_4 = in \mid \alpha_2 = cm(dist.dem)]$
Liu *et al.* (2002)	$[\mu_1 = myop, \mu_2 = hotel \mid \delta_1 = ind, \delta_3 = season \mid \alpha_2 = cm(param.reg)]$
Belobaba et Farkas (1999)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind, \delta_2 = spill, \delta_3 = season, \delta_4 = in \mid \alpha_2 = cm(dist.dem)]$
Mishra (2003)	$[\mu_1 = myop, \mu_2 = air, \mid \delta_1 = ind, \delta_3 = season, \delta_4 = in, \delta_5 = comp \mid \alpha_2 = abm(BP)]$
Huh *et al.* (2011)	$[\mu_1 = myop, \mu_2 = ret \mid \delta_1 = ind \mid \alpha_2 = cm(dist.dem)]$
α_4 : Optimization methods	
Queenan *et al.* (2009)	$[\mu_1 = myop, \mu_2 = hotel \mid \delta_1 = ind, \delta_4 = in \mid \alpha_4 = DES]$
Pölt (2000)-Weatherford (2000)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind \mid \alpha_4 = EM]$
McGill (1995)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = dep, \delta_3 = season, \delta_5 = comp \mid \alpha_4 = EM]$

Table 1.3 α_3- choice based methods

Author	Model Description
α_2 : Statistical methods	
Skwarek (1996a)-Hopperstad (1996)-Gorin (2000)-Zickus (1998)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind, \delta_2 = spill, \delta_4 = in, \delta_5 = comp \mid \alpha_2 = abm(BP)]$
Ja *et al.* (2001)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind, \delta_2 = spill, \delta_4 = in, \delta_5 = comp \mid \alpha_2 = cm(lin.reg)]$
Guo (2008)-Swan (2002)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind, \delta_2 = spill \mid \alpha_2 = cm(dist.dem)]$
Zhang et Cooper (2009)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind \mid \alpha_2 = cm(dist.dem)]$
Kunnumkal et Topaloglu (2010)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind, \delta_4 = in \mid \alpha_2 = hbm]$
Cachon et Swinney (2009)	$[\mu_1 = strat, \mu_2 = ret \mid \delta_1 = ind, \delta_3 = season, \delta_4 = in \mid \alpha_2 = cm(dist.dem.Gamma)]$
α_4 : Optimization methods	
Skwarek (1996b)-Hopperstad (1997)-Zickus (1998)-Gorin (2000)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind, \delta_2 = spill, \delta_4 = in, \delta_5 = comp \mid \alpha_4 = PD]$
Karmarkar *et al.* (2010)-Vulcano *et al.* (2012)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = dep, \delta_3 = season \mid \alpha_4 = EM]$
Stefanescu *et al.* (2004)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = dep \mid \alpha_4 = EM]$
Talluri et Van Ryzin (2004)	$[\mu_1 = myop, \mu_2 = air \mid \delta_1 = ind, \delta_4 = in \mid \alpha_4 = EM]$
Dempster *et al.* (1977)	$[\delta_1 = ind \mid \alpha_4 = EM]$
Conlon et Mortimer (2008)	$[\mu_1 = myop, \mu_2 = ret \mid \delta_1 = ind, \delta_3 = season, \delta_4 = in \mid \alpha_4 = EM]$
Bansal (2012)	$[\mu_1 = strat, \mu_2 = ret \mid \delta_1 = ind, \delta_3 = season, \delta_4 = in \mid \alpha_4 = NLP(DP)]$

CHAPTER 2

ARTICLE 2 : RAILWAY DEMAND FORECASTING IN REVENUE MANAGEMENT USING NEURAL NETWORKS

Chapter Information : An article based on this chapter was published in International Journal of Revenue Management. Sh. Sharif Azadeh, R. Labib, and G. Savard. Railway demand forecasting in revenue management using neural networks. Vol 7. No 1, (2013).

This paper investigates a statistical method for demand forecasting in revenue management systems.

Abstract This study analyzes the use of neural networks to produce accurate forecasts of total bookings and cancellations before departure, of a major European rail operator. Effective forecasting models, can improve revenue performance of transportation companies significantly. The prediction model used in this research is an improved Multi-Layer Perceptron (MLP) describing the relationship between number of passengers and factors affecting this quantity based on historical data. Relevant pre-processing approaches have been employed to make learning more efficient. The generalization of the network is tested to evaluate the accuracy prediction of the regression model for future trends of reservations and cancellations using actual railroad data. The results show that it is a promising approach in railway demand forecasting with a low prediction error.

keyword Demand forecasting, Pre-processing, Neural Network (NN), Revenue Management ; Transportation

2.1 Introduction

Revenue management (RM) refers to the collection of strategies and tactics that firms use to scientifically predict customer behavior and manage demand for their products and services Talluri *et al.* (2008).

In transportation industry, demand forecasting plays a critical role in pricing, overbooking and inventory control Xiaolong (2007). A poor estimate of demand causes inefficient inventory controls and sub-optimal revenue performance. Accurate forecasting considerably enhances the operation of capacity planning and inventory management Kandananond (2012). Based on demand models, decision makers know how many seats to make available at each of the listed fares or how much capacity to make available for each customer segment. Railway companies can set up booking limits for each pair of origins-destinations according to demand predictions Tsai *et al.* (2005). A 20 percent increase in demand forecast precision would result in revenue growth by one percent, which is highly significant in the transportation industry Talluri *et al.* (2008).

In the context of demand forecasting in Revenue Management (RM), forecasting methods could be divided into two main categories, statistical based and mathematical programming based techniques. Statistical methods of forecasting examine historical data to extract underlying process on which we can predict future trends. The selection of forecasting methods depends on several factors, such as the forecast format required, the availability of data, the desired accuracy and the ease of operation. Although these statistical methods are vastly applied in demand forecasting, they have some drawbacks that motivate us to turn our attention to Artificial Neural Network (ANN) Devoto *et al.* (2002) Chung et Lee (2002). For example, time series models are described as mathematical processes that can be extended into the future. Despite the capabilities of this approach in transportation, the models cannot respond rapidly to sudden changes in bookings and cancellations. Sudden changes in the demand of each product happen as soon as one product is no longer available as a result of capacity limitations Montgomery *et al.* (2008). They also permit the forecast to take recent demand over earlier demand into account. However, the proper selection of past periods to use is a key decision that can be subjective.

One of the most popular methods, exponential smoothing, has advantages over time series. This approach requires a smaller amount of stored data and calculations. However, a major drawback of exponential smoothing is that it is difficult to select an optimum value for the constant without making restrictive assumptions about demand behavior. This problem is compounded when the form of the underlying problem changes over time Widiarta *et al.* (2007) Snyder *et al.* (2002). Although the regression method is very popular as a prediction tool, in the context of railway demand forecasting, the use of regression analysis for a large dataset with numerous predictors and response variables can be complicated and computationally time consuming Wei et Hong (2004) Anderson *et al.* (2006) Varagouli *et al.* (2005). There are cases in which the Bayesian method has been applied to predict demand. Even though this method works accurately to define the parameters of a regression model, there is still the open problem of determining the distribution of historical bookings and cancellations data Miltenburg et Pong (2007).

To overcome some of the drawbacks of the mentioned conventional methods, in this study, we focus on the model construction of artificial neural networks. Neural networks denote an opportunity to solve numerous railway or airline specific problems more accurately Weatherford *et al.* (2003b) Zhang et Qi (2005). The output of traditional models is the linear sum of the weighted responses, whereas in a neural network, multiple linear combinations are processed in parallel ; that is, the activation in each neuron is a separate linear combination. The major advantage of the neural network approach is that it is flexible enough to model complex non-linear relationships in an automated fashion Mozolin *et al.* (2000). Moreover, the most valuable property of a multilayer feed-forward neural network is its ability to approximate, as accurately as desired, a function from training examples. In fact, a three-layer, fully connected feed-forward neural network with n input nodes, a sufficiently large number of hidden nodes and one output node, can be trained to approximate any n-1 mapping function Mozolin *et al.* (2000) Celikoglu et Cigizoglu (2007). Neural networks are powerful tools in cases in which we need to deal with large scale datasets. Based on the literature, this method outperforms classical methods, as previously mentioned, in demand forecasting Gutierrez *et al.* (2008) Kandananond (2011) Ekonomou (2010).

In this paper, we use artificial neural networks to forecast number of passengers for a major European railway company. The problem deals with a huge amount of data with missing information, for which the results of our model is promising. In addition, in this research, recommended pre-processing procedures are used, such as using exponential distribution in data normalization based on the problem characteristics, to help significantly improve performance. Demand flow may differ for each month ; thus, in order to capture the seasonal effects, the network is trained separately for each month to have more accurate forecasts.

The reminder of this paper is organized as follows. The next section briefly introduces the problem and its variables more precisely. Section 2.3 describes the pre-processing techniques implemented on the data. The model, including network architecture, learning algorithm and improvement techniques, are illustrated in Section 2.4. Computational experiments are reported in Section 2.5 and the conclusion follows in Section 2.6.

2.2 Problem definition

In this research, we investigate demand forecasting of a major railroad. The aim is to predict the number of reservations (bookings) and cancellations and, consequently, the number of passengers (bookings minus cancellations) at the time of departure. For example, in Figure 3.1, the number of passengers is represented for different departure times and departure days in different classes (i.e. business, economy for different types of clients : junior, senior, or VIP, etc.,). Each case is indicated by an observation on the horizontal axis. The number of passengers for a sample consisting of 100 observations is illustrated in this figure.

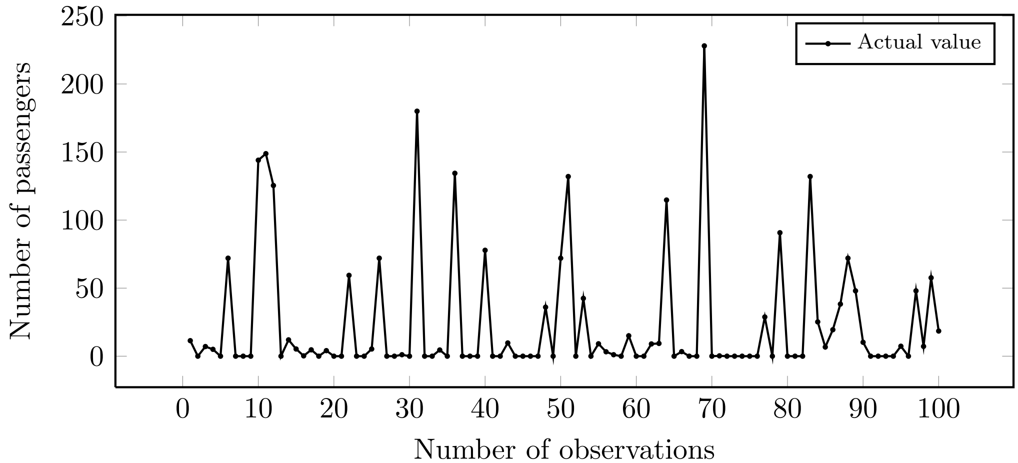


Figure 2.1 Sample pattern of the number of passengers at departure

The number of observations is equal to the number of passengers for a specific departure date and for a particular combination of class and product. Different classes are offered by the transportation company such as economy or business class, for which the prices and capacity differ. Moreover, the products are suggested by the company and are assigned to passengers, such as junior, senior, VIP and so on. We will expect our model to accurately estimate the number of passengers considered as response variables. The model proposed consists of two parts : a pre-processing phase and a regression phase. According to the transportation company's protocol, the reservation process starts 120 days before departure and there are 20 booking interval segments. During these 120 days, passengers register the information for their itinerary reservations. Some of the reservations may be cancelled throughout this period of time. There are several factors which affect the number of bookings and cancellations. We will investigate the impact of seven factors, which will serve as network inputs, on the number of passengers. This transport organization offers many departures every day at different hours. Depending on departure date and departure time the demand differs. The list of inputs and outputs is represented in Table 2.1.

During each week there are a lot of business travelers ; therefore, demand increases. During weekends there are more noticeable fluctuations in the quantity of passengers. Thus, we prefer to represent departure date in two codes : code 1 for weekdays and code 2 for weekends. Moreover, according to historical data, demand changes during the day. Therefore, the number of bookings and cancellations is affected by departure time. We express departure time using three codes. Code 1 represents the departures that are scheduled in the morning, code 2 for the afternoon and code 3 shows departures during the night. In this transportation context, tickets (i.e. products) are provided in 19 types that indicate the category of each passenger. Passengers can be students, employees, juniors, or seniors. Moreover, tickets are allocated in fourteen types of classes (i.e. business or economy). The output variables are reservations and cancellations. They vary between a minimum of zero passengers to a maximum of 330 passengers.

2.3 Pre-processing

The performance of the Multi-Layer Perceptrons (MLP) is directly influenced by the inputs that are fed to the network and the outputs which are used in the learning process. Therefore, an important part of the method is to deal with the data being used in the training procedure in which the parameters of the network are being fixed. Some data will have to be removed, some will be left unchanged and the rest will be transformed. This procedure is called pre-processing. As will be seen in the results, pre-processing contributes greatly in reducing the network generalization error. Inputs such as departure date, departure time, products and classes are determined by codes. That is, discrete numbers which are assigned to each category and do not follow a particular distribution. Thus, we may enter these data into the network without any transformations and they are left unchanged. Moreover, the daily prices for each class and itinerary differ. They are also used as inputs of the network and need to be pre-processed. But first, we have to detect the outliers of the outputs. Outliers impose a significant noise on the average and variance of the entire dataset. They can cause distortion in normalization and training. To discover them, we have divided the data into four distinct intervals. The reason is that the capacity of the train is limited ; therefore, the number of bookings and cancellations in a single request for each itinerary is rarely more than 300 passengers. On the other hand, there are a lot of departures with fewer than 100 passengers for each available combination of class and product. Thus, for a specific combination of class, product, and price, the number of reservations and cancellations could be zero, less than 100, between 100 and 300, or more than 300. Table 2.2 represents the distribution of output data in these four intervals. The range shows the number of passengers in each departure.

The first column specifies the number of bookings and the second the number of cancellations. The first row shows that for a specific departure date and time, and a particular combination of class and product, in 8907 cases during January of 2005 we had zero reservations. For example, there were some cases in which there were no juniors in Business class for a specific departure date and time. According to the tables, we find that there is significant variance in both datasets, which causes distortion in the normalization and training process ; it also creates noise in the performance of the network. To overcome this problem, we remove the outliers in an empirical way. As shown in the preceding table the majority of data is located in the first two intervals, motivating us to calculate the proportion of outliers for both bookings and cancellations. Table 2.3 represents the proportion of outliers for each set of data and for each output. These outliers are defined as the proportion of the departures with more than 100 passengers over the entire quantity of passengers. Although outliers contain a very small portion of the whole dataset, they still comprise information about demand behavior. The price we pay for keeping them during training process is greater than the price we pay for losing some degree in precision which will not affect the stability of the conclusion.

The first result denotes the proportion of outliers for the number of bookings for the January 2005 training data. In this case, the ratio of outliers is 2%. The second result is the proportion of outliers for the number of cancellations, which is 0.5%. The numbers are considered negligible and the outliers can be removed.

Nonlinear behavior is brought into the neural network by activation functions. The most commonly used activation function in multilayer perceptrons is the sigmoid function. In this case, outputs are images of this function producing values between 0 and 1. Outputs greater than 1 need to be transformed in order to have a more accurate mapping ; otherwise, the activation function will overweigh those features having larger values. It is preferable to fit a probability function that determines the characteristics of the data, which can also be reversed to calculate the error function. The Gaussian distribution is the most common one, used in normalization context but it does not have the property of being reversible because it is not bijective, that is, two different events may have the same probability of occurring. In our case, the output values are always positive, since the number of bookings minus cancellations is always positive. In addition, the shape of the data strictly decreases in terms of each interval ; for example, there are many data equal to zero and as the sample approaches 100 (after eliminating the outliers), the number of bookings and cancellations decrease. This suggests selecting the exponential distribution to map the output values into the interval [0, 1], which is also bijective. The general formulation of the exponential distribution for variable x is :

$$f(x) = \frac{1}{\lambda} e^{-\frac{x}{\lambda}} \; x \geq 0 \tag{1}$$

The estimated value for the parameter λ of the exponential distribution is the average of the data. Hence, for each set of outputs (i.e., bookings and cancellations) we have different exponential distributions with different parameters. The exponentialized outputs will be obtained via the following equation :

$$f(y_{RealValue}) = \frac{1}{\lambda_i} e^{-\frac{y_{RealValue}}{\lambda_i}} = y_{Exponentialized} \tag{2}$$

where $y_{RealValue}$ is the actual value corresponding to the specific input and $y_{Exponentialized}$ is the output transformed by the exponential distribution that was fed into the network during the learning process ; moreover, λ_i defines the parameter for each dataset. The weights are adjusted during the training process while the difference between the output of the network and the exponentialized real values are minimized. Because the learning process tries to approach the vector of the network outputs and the exponentialized real values as much as it can by using the least mean square algorithm, we suppose that the output of the network follows the same distribution as the real values. Once the predicted values have been generated, we reverse them in order to calculate the error. The absolute value of the transformed output and the corresponding error function are given by :

$$y_{Transformed} = \lambda_i \log(\lambda_i y_{NetworkOutput}) \tag{3}$$

$$Error = \vec{Y}_{RealValue} - \vec{Y}_{Transformed} \tag{4}$$

where $y_{NetworkOutput}$ is the prediction generated by the network and $y_{Transformed}$ represents the transformed predicted value. As before, λ_i is the parameter of the exponential distribution. The error function is interpreted as the difference between the vector of observed $y_{RealValue}$ and the vector of $y_{Transformed}$ produced by the network. The results will show that using an exponential distribution has a significant impact on improving the performance and generalization of the network.

2.4 Model

Throughout this section we consider the structure of the multilayer perceptron that will enable us to forecast the number of passengers. At first, we have to define the architecture of the network, and then choose an appropriate learning algorithm for training. Refining and fine-tuning the learning process will make it more efficient. After training, we will validate the accuracy of the network by the method of cross-validation. The results will be represented in the following section.

2.4.1 Architecture of the neural network

The typical network consists of an input layer, some hidden layers and an output layer. A neural network of minimum size is less likely to introduce noise into the training data and may result in better generalization. On the other hand, a large number of hidden neurons can mimic the phenomenon without understanding the underlying process. Therefore, finding a fairly convenient tradeoff between these two situations is critical. Choosing the number of hidden layers and hidden neurons is done empirically and there is no specific rule for it. A practical issue that arises in this context is that of minimizing the size of the network while maintaining good performance. In this study, we have chosen the network growing method, in which case we start with two neurons and then add progressively a new neuron or a new layer of hidden neurons. Preliminary results show that increasing the number of hidden neurons in the first layer does not reduce the error significantly. Thus, we add another hidden layer. At last, empirically we stop the growing network process at the point of two hidden layers each comprising five neurons. The final architecture of the network is illustrated in Figure 2.2.

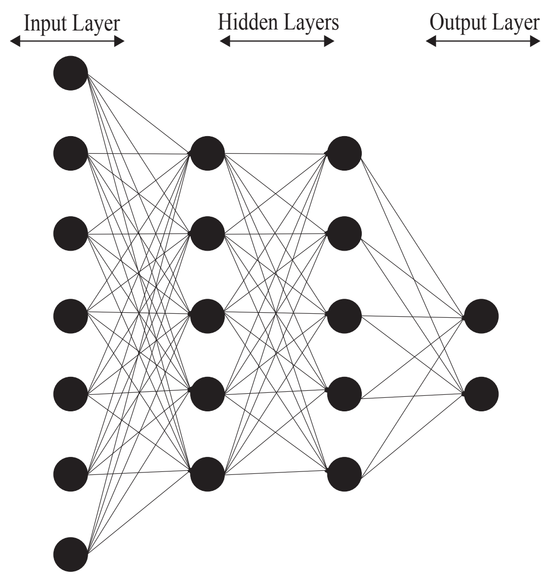


Figure 2.2 Final structure of the network having two hidden layers with 5 neurons each

2.4.2 Learning algorithm and parameter adjustments

In order to choose a learning algorithm for training and fixing the parameters we have to fully define the data set that will be fed to the network. In this study, we apply the data of January 2005 for training. To alleviate the overfitting, we use 80% of the data randomly to train the network and the remaining 20% to test the fixed parameters. This percentage is chosen empirically and it is the most common proportion for training and testing values. The neurons of each layer are connected via some coefficients, called weights, which have to be fixed during the training process. The most common learning algorithm for multilayer perceptrons, called back propagation, is applied in our case. This is an iterative process which will stop as soon as a local minimum is met with respect to a quadratic error function. After different trials we established 300 epochs to adjust the parameters during the network training. The reason is that after 300 iterations the performance of the network mostly remains constant and we cannot see any significant decrease in the error during the learning process. The learning process is maintained on an epoch-by-epoch basis until the synaptic weights of the network stabilize and the average squared error over the entire training set converges to a minimum target value. We have also chosen *batch-mode* learning, where weight updating is presented after entering all the training examples that constitute an epoch. The use of batch-mode training provides an accurate estimate of the gradient vector where convergence to a local minimum is thereby guaranteed under simple conditions Haykin (1998). Initial weights are chosen randomly. Hence, in training process, in order to calibrate the network, we repeat the learning process several times and use the average weights as initials. The primary focus of regression methods is to smoothen the predicted output variable, and in neural network, this task is accomplished with the use of sigmoid functions. The sigmoid function, whose graph is S-shaped, is by far the most common form of activation function used in the construction of artificial neural network mainly because it is differentiable. It is defined as a strictly increasing function that exhibits a graceful balance between linear and nonlinear behavior. The general format of the sigmoid function is as follows,

$$\varphi(x) = \frac{1}{1 + e^{-ax}} \tag{5}$$

where a is the slope parameter. When a is small, the network needs more data to be trained and when it is large, the generalization of the network is not good enough. In our study, after comparing the error of different trials we established a as being equal to 1.

2.4.3 Model improvements

Since back-propagation learning is basically a hill climbing technique, it runs the risk of being trapped in a local minimum where every small change in synaptic weights, w, increases the error function. The weight adjustments are done according to the following equation

$$w(n+1) = w(n) + \alpha[w(n-1)] + \eta\delta(n)y(n) \tag{6}$$

where δ represents the local gradients at each iteration n and y depicts the output of the corresponding neuron. η is the learning-rate parameter and α shows the *momentum* constant which increases the rate of learning yet avoids the danger of instability of training because the back-propagation algorithm provides an approximation to the trajectory in weight space computed by the method of steepest descent. Thus, the smaller we make the learning rate parameter η, the smaller the changes to the synaptic weights in the network will be from one iteration to the next and the smoother the trajectory will be in weight space. This improvement, however, is attained at the cost of a slower rate of learning. On the other hand, if we make η large in order to speed up the rate of learning, the resulting large changes in the synaptic weights assume such a form that the network may become unstable. Applying the momentum term helps us to avoid these problems. One technique that is often used to control the over-fitting phenomenon is that of regularization, which involves adding a penalty term to the error function in order to discourage the coefficients from reaching large values. The simplest such penalty term takes the form of a sum of squares of all of the coefficients, leading to a modified error function, E of the form

$$E(w) = \frac{1}{2}\sum_{n=1}^{N} \{y(x_n, w) - t_n\}^2 + \frac{\beta}{2}\|w\|^2 \tag{7}$$

where $\|w\|^2 \equiv w^T w = w_0^2 + w_1^2 + ... + w_M^2$, and t_i represents actual data. The coefficient β governs the relative importance of the regularization term compared with the sum-of-squares error term. In order to determine the parameter of the learning ratio and modify the training process, we employ the *adaptive learning* method. The performance of the steepest descent algorithm can be improved if we allow the parameter to change during the training process. An adaptive learning rate will attempt to keep the step size as large as possible while keeping the training process stable. This parameter is made responsive to the complexity of the local error surface and it requires some changes in the training procedure. First, the initial network output and error are evaluated. At each epoch, new weights and biases are calculated using the current parameter. New outputs and errors are then established. As with momentum, if the new error exceeds the old error by more than a predefined ratio, the new weights and biases are discarded and the learning rate is decreased; otherwise, the new weights are kept. If the new error is less than the old error, then the parameter is increased. This procedure increases the learning rate, but only to the extent of learning without large error increments. Thus, a near optimal value is obtained for the local terrain (Haykin (1998)).

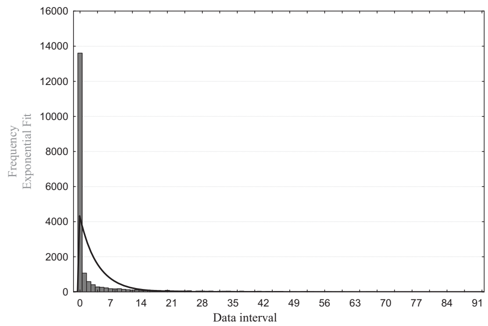
2.4.4 Validation

After training the network and fixing the parameters and also applying the improvement methods, we want to examine the generalization capability of the network. The motivation here is to validate the model on a different dataset than the one used for parameter estimation. Generalization is influenced by three factors : (1) the size of the training set and how representative it is of the environment of interest; (2) the architecture of the neural network; (3) the physical complexity of the problem at hand. To examine the network's generalizing ability we use cross-validation. Cross-validation, sometimes called rotation estimation, is the statistical practice of partitioning a sample of data into subsets such that the analysis is initially performed on a single subset, while the other subset(s) is (are) retained for subsequent use in confirming and validating the initial analysis. There is, however, the possibility that the model with the best-performing parameter values may end up overfitting the validation subset. In this study, we use multifold cross-validation by dividing the set into K subsets. The model is trained on all but one of the subsets and the validation error is measured by testing it on the remaining one. This procedure is repeated for a total of K trials, each time using a different subset for validation. The performance of the model is assessed by averaging the squared error under validation over all of the trials of the experiment. If K gets too small, the error estimate is pessimistically biased because of the difference in training-set size between the full-sample analysis and the cross-validation analysis. In contrast, if K is too large, it may require an excessive amount of computation since the model has to be trained K times with $1 \leq K \leq N$ where N is the number of examples. A value of 5 or 10 for K is popular for estimating the generalization error. The network is tested on an independent dataset that has not been used for training to give an unbiased estimate of the network performance. We trained the network on a randomly chosen subset of January 2005 for learning and validated the network with the data of March 2005.

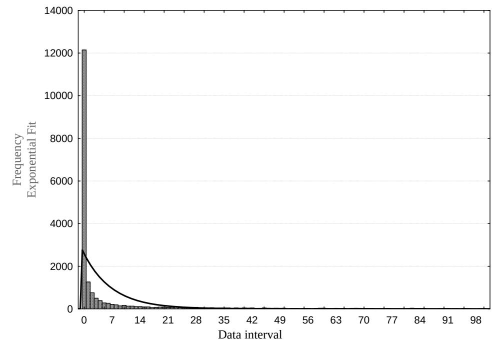
2.5 Results

Outlier elimination was one of the improvement methods that we have applied in order to reduce the forecasting error. As mentioned before, the data should be normalized before entering the network. This process is done according to the data structure. The exponential distribution is chosen as an appropriate distribution in order to normalize the data before feeding it to the network. In Figure 2.3(a) and 2.3(b) , the normalization process of the training set for both bookings and cancellations is presented. These figures show the exponential fit for the data that was used to train the network ; we consider bookings and cancellations in two separate graphs. The same was done for the other two datasets.

As can be seen, the fitted curve does not cover the whole dataset. This is due to the outlier elimination procedure, which we have already implemented in the pre-processing step. Therefore, the fitted distribution does not take the outliers into account. In order to determine the architecture of the network, we start from a network with one hidden layer in which there are two hidden neurons ; by increasing the number of hidden nodes, we consider the performance error of the network. As shown in Figure 2.4(a) and 2.4(b) , the error is not reduced significantly when the quantity of neurons increases. The minimum error obtained by using



(a) Exponential distribution fit for cancellations in January 2005(Train)



(b) Exponential distribution fit for bookings in January 2005(Train)

Figure 2.3 Fitting exponential distributions to given dataset

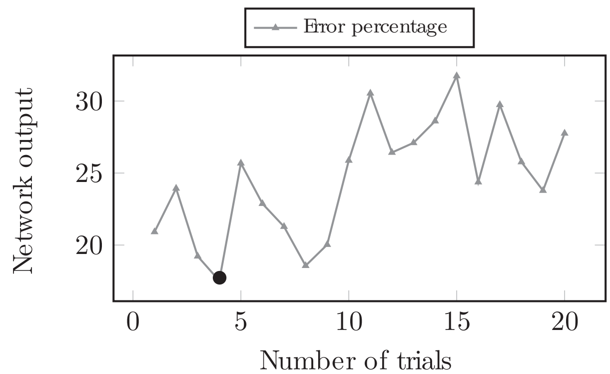
just one hidden layer remains over 15%, which motivates us to add another hidden layer to see if we can decrease the error empirically. Preliminary results show that two hidden layers predict better than single hidden layer networks.

Figure 2.5 depicts the training process of the network in an experiment with a specific predefined performance goal (i.e. the predefined error is 10^{-4}). The training process starts naturally with a large error and, during the adjustment phase, the error decreases gradually. The training process does not reach the predefined performance error or get stuck in a local or global minimum after 300 iterations.

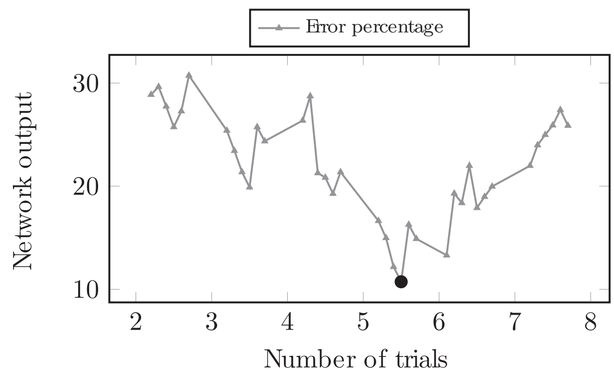
In this case, the predicted values are network outputs and the actual values are the numbers that were extracted from the transportation network. The preliminary results, without improvements, are shown in Figure 2.6. The results are clearly unsatisfactory because there are significant differences between the network outputs and the real values, making it necessary to employ some modification methods to improve the network's forecasting capability.

As we discussed earlier, some improvement methods were implemented to improve the results of the network. Before applying these methods, the error was in the 35%-45% interval, but after using exponential distribution, the results have improved dramatically giving an error rate lower than 28%. After applying the adaptive learning method to control the learning rate, the results are more stable and acceptable. However, our target error is about 8%-10%, and we are still far from this result. We used momentum and regularization to reduce the error and finally, reach our target by removing outliers. The summary of the error reduction process is represented in Table 2.4.

Moreover, Figure 2.7 illustrates the improvements, in terms of errors, obtained by these techniques. As we can see, the residuals have been reduced significantly and the network is capable of developing almost the same format as the actual values.



(a) Number of hidden neurons determination for a one-hidden-layer network



(b) Number of hidden neurons determination in a two-hidden-layers network

Figure 2.4 Process of determining the number of hidden neurons according to the method of network growing

Figure 2.8 illustrates the results obtained with the improvements. The figure shows that the network can reproduce, with accuracy, the actual data.

After developing the multi-layer perceptron and after applying the improvements, we expect that the network could generalize its ability of forecasting for unseen datasets as well. In order to validate the network, we performed a series of trial and error tests to determine how many folds give more appropriate results. To represent this analysis we have examined three different possibilities with K equal to 7, 3, and 5 folds, respectively.

As can be seen in Table 2.5, developing a 7-fold cross validation obtains an unrealistically low generalization error, which could cause unstable results when applied to large, new datasets. Here, we applied 86% of data to train the network and used the remaining 14% to test the generalization.

If we apply a completely new large dataset, the result will not remain the same, so we tried a 3-fold method, which is presented in Table 2.6.

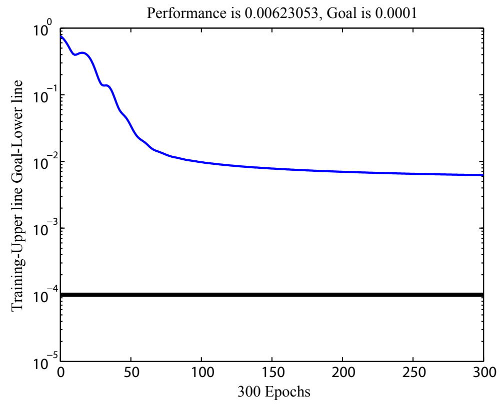


Figure 2.5 Network training through iterative process

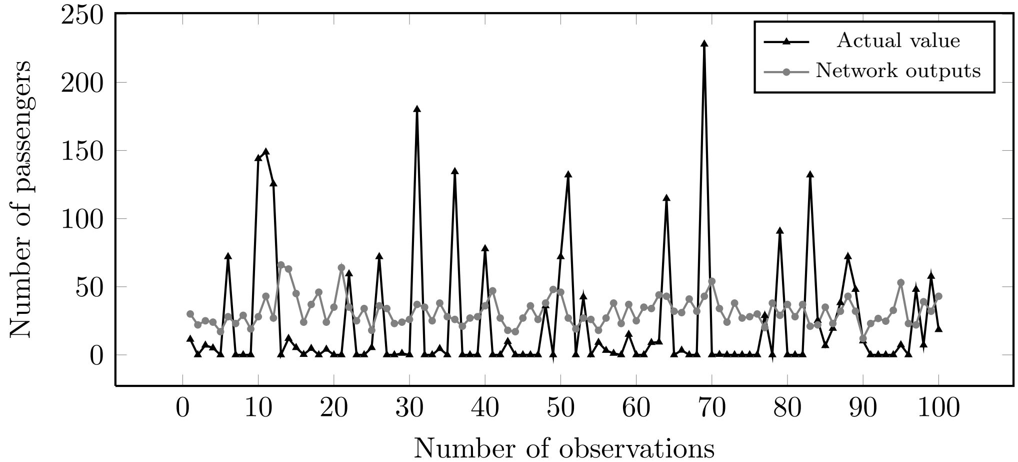


Figure 2.6 Prediction accuracy before imposing improvement methods

In 3-fold cross validation, the network could suffer from overfitting (i.e. while training the network, the error rate is low whereas the generalization error is high) because we use only 66% of the data to train the network. Also, the error is high because of the lack of training. Finally, we decided to choose a 5-fold method, in which we extract 80% of data randomly to train the network and use the remaining 20% to test it. We repeated this method five times and then we calculated the average value of the runs. The results are a good representation of the generalization error of the network. Table 2.7 shows the results of this experiment. The generalization error for booking is 9.19% and for cancellation is 8.84%.

As the second method of evaluating the generalization of the network, we used the dataset of March 2005 that had not been used in the training process. As illustrated in Figure 2.9, for 20 repetitions, the error is always steady in the 6%-12% interval. The fluctuations in the graph are due to the different subsets of the whole dataset that we have applied randomly. As shown

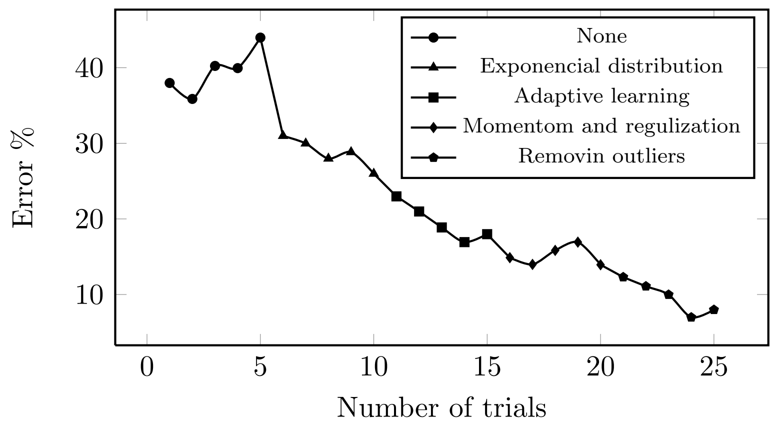


Figure 2.7 Improvements impact error reduction

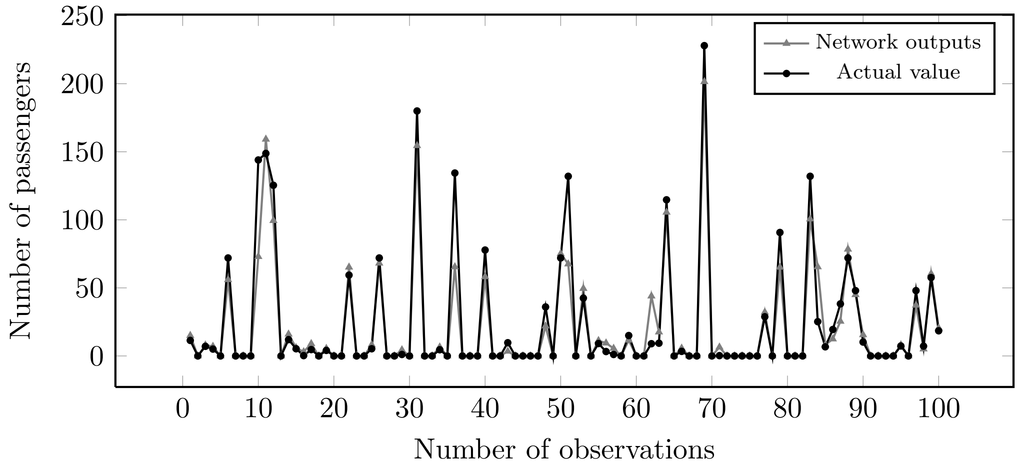


Figure 2.8 Prediction accuracy after imposing improvement methods

in the graph, the average generalization error is around 8%.

In order to capture the pattern of monthly demand, in addition to January and March of 2005, we have trained and tested the network via a 5-fold cross validation for each month separately (using data from 2007). This way, based on each month's characteristics, we estimate different parameters that enable us to predict demand in the future by taking the seasonal effects into account. Table 2.8 illustrates the corresponding results. The data from each month has been used individually as inputs of the network. The average prediction errors, which were obtained from the experiment, are satisfactory and demonstrate the ability of the network to produce acceptable demand predictions.

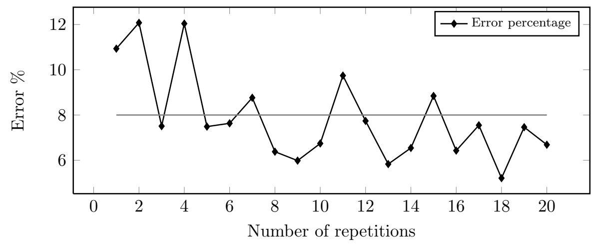


Figure 2.9 Cross validation with March 2005 data

2.6 Conclusion

Reliable passenger forecasting models play a crucial role in the transportation industry. For example, they help the transport organizations to determine seat availabilities, verify the quantity of crew members at each itinerary, and plan price settings. In this study, we have proposed a neural network to be used in transportation demand forecasting. Classical methods of statistics, such as regression or time series, struggle to cope with high dimensional data sets and sometimes refuse to respond accurately to sudden changes. In our proposed model we have chosen a Multi-Layer Perceptron (MLP) to circumvent the drawbacks of classical models. A neural network is more flexible when dealing with sudden changes in the format of data, missing information, and high dimensional data sets. We have opted to improve a typical MLP by using our knowledge of the transportation problem. This knowledge has helped us to accurately eliminate the outliers without losing too much information. Moreover, our understanding of the transportation problem has motivated us to apply exponential distribution in the process of data preparation, which reduced the forecasting error significantly. In addition, we have applied more technical approaches to improve the network performance. The efficiency of our model has been validated throughout this study. The results have shown a forecast error of around 8%, which is considered quite acceptable. Moreover, the network has been trained and tested for each month separately, using dataset from 2007. This will lead to predict future monthly demand more precisely. As a future work, the outcomes of our model can be integrated with time series. This new hybrid model is able to better express seasonal effects.

Table 2.1 Table of inputs and outputs

Inputs		
Name	Data type	Inputs
Departure date	Code	$\{1, 2\}$ 1 : Weekday 2 : Weekend
Departure time	Code	$\{1, 2, 3\}$ 1 : [6 :25 - 11 :55] 2 : [12 :25-17 :55] 3 : [18 :25-21 :55]
Product	Code	19 types
Class	Code	14 types
Class average price	Real value	Max=117.5, Min=24.5
Itinerary average price	Real value	Max=117.5, Min=70.86
Day average price	Real value	Max=93.61, Min=78.49
Outputs		
Reservations	Positive real value	[0,340]
Cancellations	Positive real value	[0,340]

Table 2.2 Table of frequencies of January 2005 for the number of bookings and cancellations

	Dataset			
	Training		Testing	
Range	Booking	Cancellation	Booking	Cancellation
x = 0	8907	10497	1906	2262
0 < x < 100	10502	9211	2248	1955
100 < x < 300	407	117	92	32
x > 300	10	1	3	0

Table 2.3 Proportion of outliers of output variables

	January 2005	
	(training, 80% of data)	(testing, 20% of data)
Bookings	2%	2%
Cancellations	0.60%	0.70%

Table 2.4 Error at each step by adding each method for improving the results

Method	Average error
Before improvement	40%
Exponential distribution	28%
adaptive learning	18%
momentum, regularization	15%
removing outliers	8%

Table 2.5 7-Fold cross validation (January 2005)

No. of folds	No. of incorrent estimations		Prediction error	
	Booking	Cancellation	Booking	Cancellation
1	120	143	3.55	4.24
2	132	123	3.89	3.63
3	151	134	4.46	3.96
4	92	170	2.72	5.02
5	180	127	5.32	3.75
6	141	172	4.17	5.08
7	201	172	5.94	5.08
Average			4.29	4.53

Table 2.6 3-Fold cross validation (January 2005)

No. of folds	No. of incorrent estimations		Prediction error	
	Booking	Cancellation	Booking	Cancellation
1	1027	982	13.02	12.45
2	628	826	7.96	10.47
3	923	889	11.7	11.27
Average			10.89	11.39

Table 2.7 5-Fold cross validation (January 2005)

No. of folds	No. of incorrent estimations		Prediction error	
	Booking	Cancellation	Booking	Cancellation
1	364	314	8.76	7.55
2	364	273	8.76	6.57
3	439	695	10.56	16.73
4	367	283	8.83	6.81
5	376	272	9.05	6.54
Average			9.19	8.84

Table 2.8 5-Fold cross validation (Monthly data of 2007)

Average prediction error (%)		
2007	**Bookings**	**Cancellations**
January	8.98	9.72
February	7.36	8.05
March	7.08	6.81
April	9.38	8.48
May	6.42	6.66
June	7.95	7.32
July	9.04	9.28
August	7.76	7.55
September	9.11	9.08
October	8.70	8.46
November	6.55	6.49
December	7.32	7.30

CHAPTER 3

ARTICLE 3 : A NON-PARAMETRIC APPROACH TO DEMAND FORECASTING IN REVENUE MANAGEMENT

Chapter Information : An article based on this chapter is submitted for publication Sh. Sharif Azadeh, P. Marcotte, and G. Savard.

In this paper, we propose a global optimization technique to estimate demand in revenue management systems.

Abstract In revenue management, the profitability of the inventory and pricing decisions rests on the accuracy of demand forecasts. However, whenever a product is no longer available, true demand may differ from registered bookings, thus inducing a negative bias in the estimation figures, as well as an artificial increase in demand for substitute products. In order to address these issues, we propose a behavioral model that solely rests on daily registered bookings and product availabilities. Its outputs are the product utilities and daily potential demands, together with the expected demand of each product in any given time interval.

Keyword Revenue management, Forecasting, Integer programming, Branch-and-bound, Heuristics.

3.1 Introduction

According to Cross (1997), Revenue Management (RM) is the research area that focuses on the study of disciplined tactics for making product availability and pricing decisions, with the aim of maximizing revenue growth. In the service industry, this goal can only be achieved through accurate demand forecasting, which must take into account the volatility of product availabilities over the booking horizon. Clearly, registered bookings alone are not sufficient to depict the true demand. Indeed, as soon as a product reaches its capacity (*booking limit*), true demand is constrained (*censored*) and cannot be observed. Upcoming customers can then either switch to a higher fare product (*buy-up*), switch to a lower fare product (*buy-down*), or renege (*spill*). According to Weatherford et Belobaba (2002), ignoring the data censorship phenomenon can lead to demand underestimation ranging from 12.5% to 25%, and negatively affect revenue by 1% to 3%, a significant amount for major rail or airline operators.

Although unconstraining techniques may have a big impact on the success of revenue management systems, this topic has not been paid much attention in the literature. Broadly, two frameworks have been considered to deal with the issue : statistics and optimization. Statistical techniques such as time series, exponential smoothing, or linear regression have been considered. All of these are able to include seasonal effects within their demand forecasts. Zeni (2001) and Queenan *et al.* (2009) have provided a comprehensive study of these methods, and have compared their respective impact on revenue. Their main drawback is that they cannot respond to sudden changes in customer behavior when a product becomes unavailable (see Sa (1987) ; Littlewood (2005) ; Pölt (2000) ; Weatherford (2000) ; Lee (1990)). Actually, authors such as van Ryzin (2005) have claimed that revenue management systems should focus on customer behavior and choice probabilities, rather than blindly estimating demand from historical booking data.

Choice-based models were introduced by Andersson (1998), and analyzed by Talluri et Van Ryzin (2004) and Vulcano *et al.* (2010) within the framework of discrete choice theory. In the latter two works, the parameters of the model have been estimated by maximum-likelihood techniques. In another research, Ratliff *et al.* (2008) have integrated historical demand data within a multi-flight heuristic procedure. Also, Vulcano *et al.* (2012) have applied customer choice models to the estimation of product primary demand (first-choice demand). In all the abovementioned optimization models, a parametric method of estimation (Expectation-Maximization, or EM in short) is used to estimate the parameters of the choice model, under demand independence assumptions. Although the approaches have been used for many years with some success, several issues still need to be addressed :

- Demand across fare products is not independent. Dealing with dependency yields a complex parameter estimation process that has been considered and tested by (Stefanescu (2009)) on small instances.
- As the proportion of censored demand in historical data grows, the accuracy of the standard estimation methods decreases (see Talluri et Van Ryzin (2004) ; Vulcano *et al.* (2012) ; Haensel et Koole (2010)).
- Several statistical methods fail to accurately capture seasonal effects.
- Choice probabilities should enter the optimization process as variables, not as parameters to be estimated. Indeed, these probabilities depend on the set of products available within each time period.

All these issues have motivated us to develop a non-parametric and distribution-free estimation procedure that, based upon historical bookings, takes explicitly into account the set of available products. The contribution of this work is twofold. First, we formulate a model for minimizing the difference between estimated and registered bookings. In order to obtain a realistic representation of customer behaviour, cross-temporal utilities enter the model as variables, and seasonal effects are captured by classifying daily demand flows into a predefined number of clusters. Next, we formulate the problem as a MINLP (mixed-integer nonlinear program), for which we develop a semi-global optimization algorithm.

We close this introductory section with an outline of the paper's structure. Following the description of the problem, together with its underlying assumptions and mathematical formulation (Section 2), we provide a detailed description of the solution algorithm, including the node selection strategy and the valid inequalities used for enhancing the branch-and-bound framework (Section 3). Computational results on synthetic data are analyzed in Section 4, while the concluding Section 5 opens avenues for future research.

3.2 Problem formulation

To illustrate demand censorship, let us consider the two-product example involving the data displayed in Table 3.1. As soon as demand for product A exceeds its booking limit 35, which it does since true demand is equal to 40, the data collection system stops counting the number of upcoming customers. As a result, the real demand for A is censored and may exceed 35. In the present case, one A-customer switched to B, while the other 4 reneged.

The main objective of our mathematical model is to minimize the difference between temporal registered bookings and their estimates. Let us introduce its main elements : a **product** i corresponds to a fare class offered at a given period [1], and is endowed with a utility u_i. The set of products available at a given period j is the **choice set** S_j. A **cluster** c denotes the set of periods that share common features based on the demand flow, such as weekdays, weekends, holidays, etc. Each **daily potential demand** d_j is associated with a unique cluster. For given utilities u_i and choice sets S_j the **choice probability** p_{ij} of selecting product i on day j is computed according to the multinomial logit (MNL) formula (Liu et van Ryzin (2008)) :

$$p_{ij}(S_j, u_i) = \begin{cases} \exp(u_i)/(\sum_{k \in S_j} \exp(u_k) + \exp(u_0)) & \text{if } i \in S_j \\ 0 & \text{otherwise,} \end{cases} \tag{1}$$

where u_0 represents the utility of the no-choice option.

For a given time horizon, $d_1, d_2, \ldots, d_{|J|}$ and a set of products I, we wish to minimize the discrepancy e_{ij} between the expected bookings w_{ij} of each available product i at a given day j and its associated observed registered booking O_{ij}, thus simultaneously capturing seasonal effects and customer behavior. We will therefore have achieved the three following goals :

- external segmentation (classification of days within clusters) ;

1. Throughout, the terms 'time period' and 'day' are used interchangeably.

Table 3.1 Demand censorship

Product	A	B
Availability status	yes	yes
Observed demand	35	5
Booking limit	35	6
Real demand	40	4

– estimation of daily potential demand ;
– estimation of product utilities.

A summary of the notation used in the model is displayed in Table 3.2.

Table 3.2 Summary of notation

sets	
Product	$i \in I = \{1, \dots, \lvert I \rvert\}$
Day	$j \in J = \{1, \dots, \lvert J \rvert\}$
Cluster	$c \in C = \{1, \dots, \lvert C \rvert\}$
Choice set	S_j, set of products available on day j
parameters	
O_{ij}	observed bookings for product $i \in I$ on day $j \in J$
A_{ij}	availability status of product $i \in I$ on day $j \in J$
R_c^U	upper bound on potential demand for cluster $c \in C$
R_c^L	lower bound on potential demand for cluster $c \in C$
D_j^U	upper bound on potential demand on day $j \in J$
D_j^L	lower bound on potential demand on day $j \in J$
P_{ij}^U	upper bound on choice probability for product $i \in I$ on day $j \in J$
P_{ij}^U	lower bound on choice probability of product $i \in I$ on day $j \in J$
variables	
e_{ij}	difference between estimated demand w_{ij} and observed bookings O_{ij}
w_{ij}	expected demand for product $i \in I$ on day $j \in J$
d_j	daily potential demand (integer)
d_{jc}^N	normalized daily potential demand $\in [0, 1]$
p_{ij}	probability of selecting product $i \in I$ on day $j \in J$
z_{jc}	cluster membership variable (binary)
r_c^N	normalized potential demand for each cluster $\in [0, 1]$
u_i	utility of product i
δ_c	potential demand of cluster c

The objective of the model is to minimize the difference between estimated and observed reservations, through the estimation of potential demand, product utilities, and cluster membership. This is achieved by solving the following mathematical model :

$$\text{MINLP :} \quad \min_{\delta_c,u,z} \quad \sum_{j\in J}\sum_{i\in S_j}(p_{ij}(S_j,u_i)d_jA_{ij}-O_{ij})^2 \tag{2}$$

$$\text{subject to} \quad p_{ij}(S_j,u_i)=\frac{\exp(u_i)}{\sum_{k\in S_j}\exp(u_k)+\exp(u_0)} \quad i\in S_j \tag{3}$$

$$d_j=\sum_c \delta_c z_{jc} \quad j\in J \tag{4}$$

$$\sum_{c\in C} z_{jc}=1 \quad j\in J \tag{5}$$

$$z_{jc}\in\{0,1\} \quad j\in J, c\in C. \tag{6}$$

In the objective function of the model, predicted booking for product i is set to the product of the relevant choice probability, potential demand of day j and products availability status.

3.3 Algorithmic framework

While the mathematical formulation of the problem is concise, its numerical resolution is challenging, due to both its combinatorial and nonlinear (fractional or multiplicative) nature. As we will observe later in the paper, it is not amenable to solution by global optimization software, its continuous relaxation being itself a difficult nonconvex program.

The line of attack that we have pursued is based on an approximate mixed integer *linear* reformulation, which was strengthened by valid inequalities. Provided with an appropriate initial solution, and through the application of efficient branching rules, quasi-optimal solutions could be obtained from an off-the-shelf MIP software such as CPLEX. Algorithm 1 represents a summary of the general resolution approach. The key elements of the algorithmic framework are presented in more details as follows :

3.3.1 Linearization

Let us assume that the potential demand for cluster c lies within predetermined bounds, i.e. $\delta_c\in[R_c^L,R_c^U]$, and let us introduce the normalized variable

$$r_c^N=\delta_c/R_c^U\in[R_c^L/R_c^U,1]. \tag{7}$$

Equation (4) then takes the form

$$d_j=\sum_c(r_c^N R_c^U)z_{jc} \tag{8}$$

Algorithm 1 General Solution Approach

Input : Registered bookings O_{ij}, set of available products, S_j

Output : Daily demand flows d_j, cluster memberships z_{jc}, utilities u_i

(1) Transformation into a MIP

i : Linearization

ii : Relaxation

ii : Convexification

(2) Preprocessing at root node

iii : Valid inequalities

iv : Initial solution

v : Domain reduction

(3) Branch-and-bound

vi : Branching strategy

vii : Adjustment of bounds at branching nodes

and, upon the change of variable

$$d_{jc}^N = r_c^N z_{jc}, \tag{9}$$

one derives the linear equation

$$d_j = \sum_c d_{jc}^N R_c^U. \tag{10}$$

Using the fact that z_{jc} is binary-valued, Equation (9) can be linearized. Indeed, if $z_{jc} = 1$, then $d_{jc}^N = r_c^N$. Since $z_{jc} \leq 1$, we have that

$$z_{jc} + d_{jc}^N \leq r_c^N + 1 \tag{11}$$

and

$$z_{jc} + r_c^N \leq d_{jc}^N + 1 \tag{12}$$

If $z_{jc} = 0$, then $d_{jc} = 0$, i.e.,

$$d_{jc}^N \leq z_{jc} \qquad i \in I, j \in J. \tag{13}$$

To tighten the feasible domain, the lower bound R_c^L on cluster demand d_j has been set to the minimum value of daily cumulative registered reservations.

3.3.2 Relaxation and convexification

Three sources of nonconvexity occur in the MINLP mathematical model :

- the choice probabilities, p_{ij} derived from MNL model involve a fractional term ;
- the estimated demand for a given product on a given day involve the bilinear term $(p_{ij}d_j)$;
- variables z_{jc} are binary-valued.

To deal with the first source of nonconvexity, we base our relaxation on choice probabilities p_{ij} (versus utilities), we approximate the bilinear terms w_{ij} by their convex envelopes, and resort to classical continuous relaxations for the binary variables z_{jc}.

Note that substituting the independent choice probability variables p_{ij} to the utilities u_i in the original problem may induce infeasibilities. To address the issue we check that, for each product i and day j, the inequality

$$P_{ij}^L \leq p_{ij} = \frac{\exp(u_i)}{\sum_{k \in S_j} \exp(u_k) + \exp(u_0)} \leq P_{ij}^U \qquad i \in I, j \in J \tag{14}$$

holds in the MINLP model solved by a nonlinear solver.

To deal with the bilinear term $w_{ij} = p_{ij}d_j$, we contrast the concave and convex envelopes of these functions against the relaxations introduced by McCormick (1976). Each bilinear term is relaxed independently. Making use of the bounds

$$\mathcal{E} = \left\{ w_{ij} = p_{ij}d_j \in [P_{ij}^L, P_{ij}^U] \times [D_j^L, D_j^U] \times R \right\}, \tag{15}$$

we locally convexify the bilinear term based on the following inequalities, which are only valid if product i is available on day j, i.e., $A_{ij} = 1$:

$$\begin{array}{lcll}
w_{ij} & \geq & D_j^U p_{ij} + P_{ij}^L d_j - P_{ij}^L D_j^U & i \in I, j \in J \\
w_{ij} & \geq & D_j^L p_{ij} + P_{ij}^L d_j - D_{ij}^L P_{ij}^L & i \in I, j \in J \\
w_{ij} & \leq & D_j^U p_{ij} + P_{ij}^L d_j - P_{ij}^L D_j^U & i \in I, j \in J \\
w_{ij} & \leq & D_j^L p_{ij} + P_{ij}^U d_j - P_{ij}^U D_j^L & i \in I, j \in J.
\end{array}$$

We then iteratively update, for each subproblem and at each node of the enumeration tree, the upper and lower bounds of the choice probabilities of available products. This yields the convex quadratic program

$$
\begin{aligned}
\text{RELAX :} \quad & \min_{p,d,z} \quad \sum_{i\in I}\sum_{j\in J} e_{ij}^2 && && (16)\\
& w_{ij}A_{ij} - O_{ij} = e_{ij} && i \in I, j \in J \\
& \\
& P_{ij}^U d_j + D_j^U p_{ij} - P_{ij}^U D_j^U \leq w_{ij} && i \in I, j \in J \\
& P_{ij}^L d_j + D_j^L p_{ij} - P_{ij}^L D_j^L \leq w_{ij} && i \in I, j \in J \\
& P_{ij}^U d_j + D_j^L p_{ij} - P_{ij}^U D_j^L \geq w_{ij} && i \in I, j \in J \\
& P_{ij}^L d_j + D_j^U p_{ij} - P_{ij}^L D_j^U \geq w_{ij} && i \in I, j \in J \\
& \\
& d_{jc}^N \leq z_{jc} && j \in J, c \in C \\
& z_{jc} + r_c^N \leq d_{jc}^N + 1 && j \in J, c \in C \\
& z_{jc} + d_{jc}^N \leq r_c^N + 1 && j \in J, c \in C \\
& d_j = \sum_{c\in C} d_{jc}^N R_c^U && j \in J \\
& R_c^L / R_c^U \leq r_c^N && c \in C \\
& \sum_{c\in C} z_{jc} = 1 && j \in J \\
& 0 \leq z_{jc} \leq 1 && j \in J, c \in C
\end{aligned}
$$

where the second to fifth constraints express the McCormick inequalities of the bilinear terms, while the next six constraints assign each day to a specific cluster.

3.4 Solution algorithm

The algorithm for globally solving the original problem is a branch-and-bound based on RELAX, and where branching is performed with respect to the binary variables δ_c, the integer variables d_j, as well as the continuous variables p_{ij}. While a linear solver is put to contribution for the first two sets of variables, a nonlinear solver is required for computing the choice probabilities. We now provide a detailed description of the main elements of the algorithm.

3.4.1 Preprocessing

The performance of the enumeration scheme can be greatly enhanced through three procedures : introduction of valid inequalities at the root node, warm-starting the algorithm with a feasible solution provided by a heuristic algorithm, and tightening the feasible domain at each node of the branch-and-bound tree.

Valid inequalities

Based on the set of available products, one can derive logical relations that must be satisfied by any optimal solution. In general, considering the choice sets of two separate days, S_j and $S_{j'}$, three cases may happen : (i) the choice probabilities of two products are equal, $p_{ij} = p_{ij'}$ (ii) one of them is less than the other one, $p_{ij} > p_{ij'}$ (iii) we cannot establish a logical relation between two probabilities.

Valid inequality 1. It is a property of the multinomial logit that, if the choice set of day j is a subset of the choice set of day j', that is $S_j \subseteq S_{j'}$, then we have

$$p_{ij} = \frac{\exp(u_i)}{\sum\limits_{k \in S_j} \exp(u_k) a_{kj} + \exp(u_0)} \geq \frac{\exp(u_i)}{\sum\limits_{k \in S_{j'}} \exp(u_k) a_{kj'} + \exp(u_0)} = p_{ij'} \tag{17}$$

Valid inequality 2. In order to discard symmetric and equivalent solutions we order, without loss of generality, the demands associated with the cluster indices, i.e.,

$$\delta_1 < \delta_2 \leq \ldots \leq \delta_k \leq \ldots \leq \delta_{|C|}. \tag{18}$$

Equivalently :

$$r_1^N R_1^U \leq r_2^N R_2^U \leq \ldots \leq r_k^N R_k^U \leq \ldots \leq r_{|C|}^N R_{|C|}^U. \tag{19}$$

Initial solution

At the root node, initially, we find estimated daily potential demand, d_j, by solving RELAX problem. Then, an integer initial solution is obtained via a K-nearest neighbor algorithm to fix class membership variables, z_{jc}.

First, one matches each day to its own cluster. Then, one iteratively merges the two clusters having the closest averages, until the required number of clusters is attained. Since ties are broken arbitrarily, different choices could yield different partitions of the set of days into clusters. Table 3.3 shows the progression of the algorithm corresponding to the vector of daily potential demands $\{36, 6, 30, 14, 42\}$, and a number of final clusters set to two. In this example, the same solution would have been achieved if 36 and 42 had been merged at the first iteration. Of course, this result does not hold in general, as can be readily verified on the demand vector $\{1, 3, 5\}$ with two clusters yielding either the partition $\{1, 3\}$ $\{5\}$ or $\{1\}$ $\{3, 5\}$.

By using fixed z_{jc}s, we again solve the RELAX model to obtain estimated potential demand of each cluster δ_c. Finally, MINLP model is solved to find initial solution for product utilities u_i.

Domain reduction

Prior to branching, the respective ranges of the variables w_{ij}, p_{ij} and d_j can be tightened. For example, the sum of registered bookings on a given day d_j provides the lower bound

$$d_j \geq D_j^L = \sum_i O_{ij} \qquad j \in J. \tag{20}$$

When $\mathcal{F}^*$ shows the best integer solution, an upper bound $D_j^U(0)$ on d_j can be set to the optimum of the convex optimization problem

$$\begin{aligned} \max_d \quad & d_j \\ \text{subject to} \quad & \sum_i \sum_j e_{ij}^2 \leq \mathcal{F}^* \qquad i \in I, j \in J \\ & \text{constraints of RELAX.} \end{aligned}$$

In a similar fashion, upper bounds $P_{ij}^U(0)$ on the choice probabilities p_{ij} are obtained by solving the convex program

$$\begin{aligned} \max_p \quad & p_{ij} \\ \text{subject to} \quad & \sum_i \sum_j e_{ij}^2 \leq \mathcal{F}^* \qquad i \in I, j \in J \\ & \text{constraints of RELAX.} \end{aligned}$$

A total of $2|I| + 2|I||J|$ optimization problems are solved to derive the above upper bounds. Finally, it follows from the inequality

$$\sum_{i \in I} \sum_{j \in J} e_{ij}^2 \leq \mathcal{F}^*$$

that w_{ij} can be upper bounded by $\sqrt{\mathcal{F}^*} + O_{ij}$.

3.4.2 Branch-and-bound

The optimum of the relaxed program provides a lower bound on the true optimal value, while the corresponding solution can be used to construct a feasible solution that yields an

Table 3.3 Clustering algorithm for determining initial solution

iteration						
0	clusters	$\{36\}$	$\{6\}$	$\{30\}$	$\{14\}$	$\{42\}$
	averages	36	6	30	14	42
1	clusters	$\{36, 30\}$	$\{6\}$	$\{14\}$	$\{42\}$	
	averages	33	6	14	42	
2	clusters	$\{36, 30, 42\}$	$\{6\}$	$\{14\}$		
	averages	36	6	14		
3	clusters	$\{36, 30, 42\}$	$\{6, 14\}$			
	averages	36	10			

upper bound on the optimum. Note that the performance of the partial enumeration process rests in large part on the quality of the upper bounds on the variables, hence the importance of tightening these.

At each node of the enumeration tree, we implement a series of range reductions with respect to daily potential demand, choice probabilities, potential demand of each cluster $R_c^L < \delta_c < R_c^U$, and the bilinear term w_{ij}. Several techniques, such as interval arithmetic, have been implemented. For instance one can fix the value of z_{jc} without branching. Indeed, if for a given node n, the set $[D_j^L(n), D_j^U(n)] \cap [R_c^L(n), R_c^U(n)]$ is empty, then z_{jc} must be zero.

For node n, the bounds on cluster demand δ_c can be set to

$$\begin{aligned} R_c^L(n) &= \max\left\{R_c^L(n), \min_j D_j^L(n)\right\} \\ R_c^U(n) &= \min\left\{R_c^U(n), \max_j D_j^U(n)\right\}. \end{aligned}$$

The lower bound can be updated according to the formula

$$D_j^L(n) = \max\left\{D_j^L(n), \min\left\{\frac{w_{ij}^U(n)}{P_{ij}^L(n)}, \frac{w_{ij}^U(n)}{P_{ij}^U(n)}, \frac{w_{ij}^L(n)}{P_{ij}^L(n)}, \frac{w_{ij}^L(n)}{P_{ij}^U(n)}\right\}\right\} \quad i \in I, j \in J \tag{21}$$

The upper bounds $D_j^U(n)$ on daily potential demand, as well as the bounds on choice probabilities are updated in a similar fashion. Next, we adjust the upper and lower bounds of daily potential demand d_j, and potential demand of each cluster δ_c. Meanwhile, we fix the value of assignment variables z_{jc}, whenever the ranges of d_j and δ_c intersect.

For a given node, if the range of a variable d_j obtained from the relaxation model overlaps with the ranges of two or more clusters, then its lower bound is updated to

$$D_j^L(n) = \max\left\{R_c^L(n), D_j^L(n)\right\}. \tag{22}$$

Similarly, the upper bound is set to

$$D_j^U(n) = \min\left\{R_c^U(n), D_j^U(n)\right\}. \tag{23}$$

Finally, feasibility conditions are verified by using (14) and (19). In addition, the solution of RELAX and MINLP problems are used to prune the partial enumeration tree.

We close this section with a description of the branching strategy. At node n, the binary variables z_{jc} are relaxed, and d_j can therefore 'partially' belong to more than one class. Let $\hat{d}_j(n)$ be the estimated potential demand of day j obtained from optimal solution of the RELAX problem. Let

$$I_c(\hat{d}_j(n)) = \begin{cases} 1 & \text{if} \quad \hat{d}_j(n) \in [R_c^L(n), R_c^U(n)] \qquad c \in C \\ 0 & \text{otherwise} \end{cases} \tag{24}$$

$$I(n) = \sum_{j \in J} I_c(\hat{d}_j(n)) \qquad c \in C. \tag{25}$$

In the branching scheme, **node selection** follows these rules :

– Branch on the node from which the relaxed optimum is minimal.
– In case of a tie, branch on the deepest node.
– In case of yet another tie (this rarely occurs in practice), branch on any node having the maximum number of overlapping intervals with respect to variables d_j and δ_c, i.e., $[D_j^L(n), D_j^U(n)] \cap [R_c^L(n), R_c^U(n)] \neq \emptyset$.

As far as **variable selection** is concerned, we prioritize the cluster demand variables δ_c for branching, but switch to daily potential demand d_j when all clusters are disjoint. A variable δ_c is selected if it achieves maximum interval length, ties being broken in favor of clusters with large $I(n)$-values in RELAX. The assignment of each day to a single disjunctive cluster is achieved by branching on d_j. To reduce each cluster to a singleton, we branch again on δ_c. As mentioned above, we branch on d_j to fix z_{jc}, and select variables for which the difference between lower and upper bounds is the largest.

Once all days have been assigned to clusters and δ_c is integer-valued, we solve MINLP to find product utilities u_i. Finally, if the gap between the RELAX and MINLP solutions is larger than a predefined threshold, we branch on variables p_{ij}, with the aim of either fathoming the current node or obtain a better feasible solution.

3.5 Computational results

The algorithm has been tested on a number of synthetic instances, and its performance assessed with respect to three criteria :

- **Calibration** : this criterion is used to verify whether the algorithm is able to recover exactly the data used to generate the synthetic instances, i.e., achieves a zero objective for MINLP.
- **Classification :** this criterion is used to determine the error level achieved by the algorithm on perturbed instances, and also to compare the performance of the algorithm against two well-known global and nonlinear solvers.
- **Generalization :** this criterion is used to assess the robustness of the estimation process, i.e., verifying how well the parameters calibrated on a set of controlled instances can generalize to distinct perturbed datasets, thus constituting a reliable tool for decision making.

3.5.1 Data generation

Each instance is characterized by a triple (C, J, I) where C denotes the number of clusters (2, 3, or 4), J the number of days (7, 14, 21, or 28) and I the number of products (4, 6, or 8). Observed bookings (O_{ij}) have been generated according to the formula

$$O_{ij} = A_{ij} p_{ij} dj, \tag{26}$$

which requires knowledge of the set of available products, as well as the utilities u_i from which the probabilities p_{ij} are derived. In this process, the product utilities and the potential demand δ_c of each cluster are exogenous. The availability parameters A_{ij} associated with product i on a given day j are generated according to a Bernoulli random variable. Finally, each day j has been randomly assigned to one of the clusters.

A first set of 33 unperturbed instances allowed to check whether the algorithm could actually replicate the original values z_{jc}, d_j and u_i. Next, a second set of 33 perturbed samples were created to test the generalization ability of the model. Keeping the other parameters (choice set, potential cluster demand, class membership, product utilities) fixed, the daily demand was modified according to the formula

$$d_j = d_j(1 + \gamma(2\epsilon_j - 1)) \qquad j \in J, \tag{27}$$

where the perturbation parameter γ was fixed to 0.1, and ϵ_j was uniformly distributed between 0 and 1.

The outcomes of our proposed model have been compared with those of two of the most acknowledged softwares : Knitro 8, a nonlinear solver, and Baron 11.0, a global optimization solver. Our algorithm has been halted whenever no improvement occurred within 60 minutes of CPU time. The computational experiments have been carried out on a Quad-core computer with 2.4 GHz CPU and 8 GB of RAM. The branching algorithm was implemented in C++, and we resorted to the Quadratic Solver of CPLEX 12.3 (sequential quadratic programming) and IPOPT 3.11 (interior point method) as nonlinear solvers. The software Baron was accessed through the NEOS server (see IBM (2013), BARON (2013), Ziena (2013), NEOS (2013), IPOPT (2013)).

3.5.2 Numerical experiments

The non-perturbed instances used for calibration purposes are displayed in Table 3.4. For all instances, the global optimum with zero value was reached. Moreover, the algorithm was able to reproduce the exact original product utilities u_i and cluster potential demand δ_c from which the data was initially generated.

The numerical results corresponding to the 33 perturbed instances are summarized in Table 3.5. The first three columns describe instances and their characteristics : number of clusters, number of days and number of products. The four 'Time' columns contain execution time (CPU time in seconds) of different parts of the algorithm : 'Total' (some of the next two columns plus the time spent to implement branching strategy), 'Relax' (time spent solving the RELAX model), 'NLP' (time spent solving MINLP using IPOPT), 'Pre-Proc.' (time spent implementing the pre-processing at root node). Cases where the run time is significantly less than 3 600 seconds attested to the efficiency of the branching strategy.

The next four columns under 'Node' provide statistics related to the branch-and-bound tree. The first column 'Gen.' represents the total number of nodes generated during the branch-and-bound procedure. Although reasonable for small instances, it increases quickly with the number of products and clusters. Column 'Br.' represents the number of branched nodes, which is significantly lowered by implementing the feasibility conditions and valid inequalities. The caption 'Dis.' refers to the number of nodes that have been discarded during the branching process, through the violation of the feasibility conditions (14) and (19). The heading 'Domin.' refers to the number of nodes dominated by the current best solution. Data in

Table 3.4 Non-Perturbed instances

			Time				Node						
Class	**Days**	**Product**	Total	Relax	NLP	Pre-Proc.	Gen.	Br.	Dis.	Dom.	#IPOPT	VI#1	Ini Sol.
		4	1.38	0.94	0.02	0.73	221	81	14	4	2	12	172.87
	7	6	0.28	0.28	0.00	1.98	76	30	46	0	2	32	419.99
		8	19.30	10.84	0.00	1.69	2367	1034	73	53	54	5	341.65
		4	3.00	1.72	0.00	3.27	339	135	12	32	4	67	276.54
	14	6	3.84	2.52	0.00	6.39	301	117	9	4	4	75	369.85
		8	9.80	5.09	0.06	9.47	657	457	6	178	179	54	1684.91
2		4	0.70	0.25	0.02	7.88	55	21	8	5	1	138	71.74
	21	6	32.27	17.17	0.00	12.75	1461	655	12	12	13	134	449.93
		8	5.92	2.98	0.00	21.72	242	109	19	3	4	116	1364.69
		4	4.19	2.31	0.00	13.69	179	79	6	27	3	218	314.34
	28	6	11.50	5.88	0.00	27.05	329	138	6	37	3	227	178.34
		8	45.27	20.44	0.00	136.34	1022	420	21	0	1	283	2720.01
		4	3.94	2.42	0.03	0.66	625	227	12	13	13	14	136.63
	7	6	82.34	1.72	0.13	2.56	1100	800	213	32	126	23	534.00
		8	17.42	9.88	0.05	2.48	1620	708	34	58	59	19	794.63
		4	2.45	1.44	0.00	3.55	189	77	6	0	1	80	484.42
	14	6	37.36	17.33	0.45	2.20	4026	2284	24	633	634	16	830.42
3		8	10.14	5.33	0.06	2.48	1313	615	50	106	107	7	896.69
		4	5.50	3.05	0.00	8.97	338	134	17	4	1	152	350.19
	21	6	14.00	7.27	0.00	14.55	594	236	19	0	1	139	912.96
		8	155.17	84.11	0.00	34.39	4385	1667	240	5	6	163	2956.55
		4	28.05	12.86	0.02	14.52	1102	453	24	1	2	233	2666.03
	28	6	149.88	72.53	0.02	31.11	3656	1463	84	26	27	277	929.66
		8	3396.59	1505.86	0.00	203.69	51052	21811	1981	0	0	302	2675.14
		4	1234.98	408.80	1.11	3.92	72210	31117	1339	2050	1951	76	358.54
	14	6	18.94	10.52	0.00	9.48	1172	433	55	0	1	106	1827.04
		8	179.41	81.27	0.05	17.39	8835	3640	39	56	57	141	2171.08
		4	136.69	67.05	0.02	11.72	6651	2044	113	0	1	143	1093.43
4	21	6	7.08	3.58	0.00	23.34	236	74	8	0	1	242	1555.92
		8	1669.09	733.95	0.02	151.38	40997	13503	1213	50	27	252	2479.78
		4	169.84	70.72	0.00	19.63	5367	1728	66	1	2	260	1307.35
	28	6	38.17	17.91	0.00	41.47	785	269	10	0	1	340	1937.09
		8	212.00	106.64	0.00	217.91	2813	1032	6	2	3	342	3822.17

these two columns attest to the efficiency of the algorithm, more precisely to the large number of subtrees that could be pruned.

Column '# NLP' refers to the number of times the algorithm resorted to IPOPT for solving MINLP, i.e., the number of times the relaxation problem reached an integer feasible solution. Numbers under the heading 'VI#1' correspond to the number of valid inequalities appended to the model, and thus are a measure of the contribution of constraint (17).

The next three columns show the initial solution 'Ini Sol' obtained from a variant of K-nearest neighbor algorithm, the best feasible solution 'Best' and the best bound 'Bound'. They illustrate the sharp improvement of the initial solution, which admittedly not very good. The iterative process halts if one of two conditions holds : either the gap between the best solution (MINLP) and the best bound (RELAX) is less than 1%, or the algorithm makes no improvement for a period of 3 600 seconds. The 'Gap', set to the value (Best bound-Best solution)/100, is displayed in the last column.

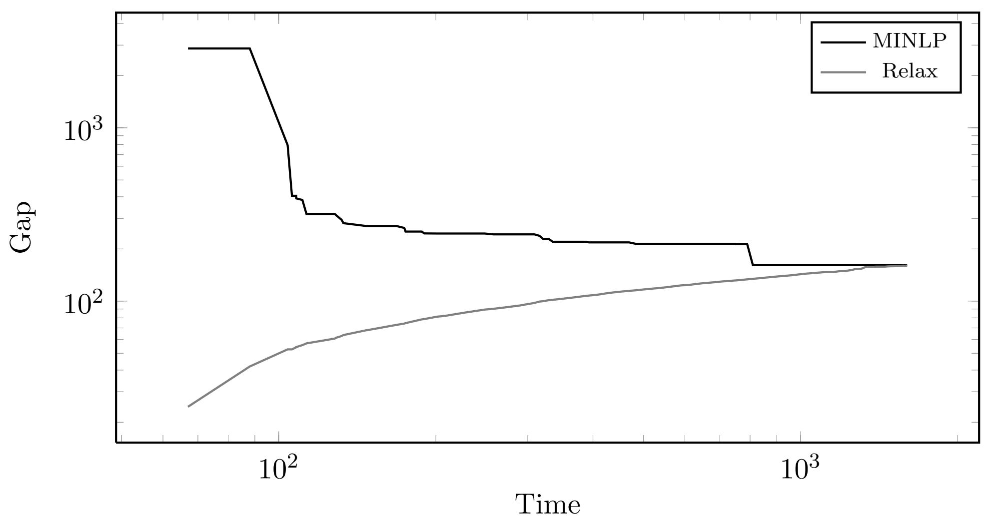


Figure 3.1 Algorithm effectiveness in reducing the gap

In Figure 1, for an instance involving three classes, 28 days and four products, we illustrate the effectiveness of the algorithm by plotting the best bound and the best solution against CPU time. The Figure is logarithmically scaled for the ease of presentation. The final gap between MINLP and RELAX models for this example is equal to 0.98%.

In Table 3.6, we contrast the performance of our algorithm against those of Knitro and Baron, on the 33 perturbed instances of Table 3.5. The first three columns specify the dimensions of the instances. The next three under 'BB Sol' show the outcomes of our demand model, initialized with 'Ini Sol', the prediction error (objective function) 'MSE', and the classification error 'Class.%'. The latter indicates whether each day has been properly assigned to one of the clusters, based on its demand flow. This information helps us to accurately predict demand seasonal features. We observe that the algorithm has been successful in correctly assigning each day to one of the clusters. From a theoretical point of view, small gaps between the solutions of MINLP and RELAX, as well as null classification errors, testify to the effectiveness of the algorithm.

Under the heading 'Baron', the next two columns show the prediction ('MSE') and classification ('Class') errors associated with the global optimizer Baron. While Baron is efficient on small instances, it is highly sensitive to the size of the problem. It actually fails to solve problems involving more than 14 days and four clusters.

Likewise, the two columns under 'Knitro' correspond to prediction and classification errors for that solver. Once again, this solver successfully classifies days into clusters for small instances, with prediction errors similar to those of Baron and our algorithm. Besides size (based on the number of clusters, days and products), Knitro is also sensitive to product availabilities. When the number of available products for a given day decreases, the values of MSE and classification error sharply increase.

The figures in the last column ('Generalization') of Table 3.6 illustrate the good performance of our algorithm on perturbed data. This stability actually depends on the availability of the products. In some cases, the error is slightly lower than the calibration error (MSE), due to the number of available products being small compared to the number used for calibrating the model. In all cases, the proposed algorithm outperformed by a large margin both Knitro and Baron.

3.6 Conclusion

In revenue management, dealing with censored demand is a complex issue, especially when a large proportion of products are not available to customers, a situation that occurs in practice. To address this problem, we have proposed a choice-based non-parametric method whose output is the demand for any given product within any given time period, and proposed for its solution an efficient algorithm that has been validated on synthetic data. Simultaneously, we obtained a variety of solutions for class potential demands, product utilities and daily assignments, depending on the features of historical data. Through the introduction of exogenous customer segmentation into the mathematical model, or through the implementation of a revenue maximization model, these sets could be reduced, and a small number of scenarios retained. In a future work, we intend to investigate situations involving more sophisticated segmentation models, with the aim of accurately estimating the probabilities of buy-ups or buy-downs.

Table 3.5 Perturbed instances

			Time				Node									
Class	**Days**	**Product**	Total	Relax	NLP	Pre-Proc.	Gen.	Br.	Dis.	Dom.	#IPOPT	VI#1	Ini Sol.	Best	Bound	Gap(%)
		4	0.70	0.48	0.00	1.14	125	54	26	24	6	32	154.55	5.70	5.68	0.02
	7	6	4.06	2.31	0.11	2.64	539	438	30	34	228	38	761.24	7.53	6.96	0.57
		8	4.23	2.59	0.17	4.19	468	362	36	51	173	44	561.02	6.66	6.30	0.36
		4	4.33	2.30	0.13	4.59	436	339	40	55	168	94	536.44	30.75	30.01	0.74
	14	6	15.69	7.53	0.36	8.77	1089	902	40	137	467	115	1589.76	28.70	27.90	0.80
2		8	9.89	4.80	0.14	70.59	590	476	61	46	234	109	1212.62	16.56	15.72	0.84
		4	19.02	9.05	0.36	8.41	1162	904	47	116	426	190	2095.22	38.80	38.20	0.60
	21	6	107.64	51.44	1.17	14.72	4687	4019	135	518	2052	136	2853.39	32.46	31.80	0.66
		8	133.75	55.98	1.50	65.63	4492	3988	133	313	2156	181	3046.96	48.83	47.94	0.89
		4	13.63	6.22	0.11	15.31	557	376	14	48	144	320	2521.28	48.69	48.46	0.23
	28	6	55.95	24.56	0.17	33.97	1593	1142	54	378	446	326	476.92	47.02	46.60	0.42
		8	124.27	62.31	0.38	179.95	2177	1599	61	413	631	413	2941.36	27.30	26.31	0.99
		4	5.16	3.23	0.08	0.89	728	359	83	54	83	16	363.41	11.27	10.28	0.99
	7	6	370.47	174.22	9.06	2.23	36309	29545	469	1530	13914	35	756.64	4.98	3.99	0.99
		8	31.39	16.63	0.92	4.59	3045	2446	188	129	1150	50	959.91	4.06	3.24	0.83
		4	127.33	62.52	2.03	4.59	10965	8029	282	1634	3485	91	1890.64	17.31	16.50	0.81
	14	6	3619.84	903.67	28.56	7.69	128032	112342	4118	1711	45030	81	216.42	15.07	10.89	4.18
		8	3709.91	1098.17	31.66	13.34	117360	102422	2920	2262	44936	87	1566.21	20.46	15.94	4.52
3		4	2362.35	962.11	4.19	20.81	48931	27561	1493	19653	6349	293	1768.88	71.78	70.99	0.79
	21	6	3119.17	1026.05	15.92	22.28	95000	66618	3366	17583	24113	252	103.92	25.10	24.37	0.72
		8	3954.83	1144.75	14.64	28.23	84114	64588	4043	2885	20754	137	1214.47	28.51	23.69	4.82
		4	1473.64	601.23	3.41	19.50	45650	24643	1361	18775	5151	300	2974.46	75.63	74.70	0.93
	28	6	3691.58	1225.94	13.97	40.33	70195	53171	3501	5069	18088	351	1837.06	36.76	32.73	4.03
		8	3670.41	1515.92	3.00	250.94	51928	34237	1822	5114	4210	260	2430.05	29.89	25.59	4.30
		4	4238.61	1243.47	0.55	3.94	121336	71764	5178	40387	916	90	1361.45	11.87	11.03	0.84
	14	6	4247.70	1500.89	0.84	7.80	110504	62968	3582	31329	1104	78	1688.36	12.60	9.43	3.17
		8	7302.53	1864.11	11.11	12.41	155860	93126	4153	41149	13749	50	859.01	12.13	7.45	4.68
		4	8083.58	2306.41	11.50	11.52	180764	102470	5470	60386	15246	169	248.52	39.63	37.23	2.40
4	21	6	6189.22	1774.72	9.89	17.52	118253	72650	4220	23518	13046	165	1328.48	34.23	29.72	4.51
		8	5645.56	2245.20	10.13	29.92	140178	94594	4552	27345	14177	102	1696.79	38.47	33.85	4.62
		4	4064.64	1112.59	15.48	23.00	94901	60890	2288	13886	19503	386	416.61	53.68	49.64	4.04
	28	6	9252.73	3292.19	7.67	43.23	132019	87375	5832	29557	13077	343	953.01	56.95	52.10	4.85
		8	11689.00	3746.00	9.83	296.54	147664	90428	6373	32346	15746	368	1234.02	45.63	40.93	4.70

Table 3.6 Comparison Framework

			BB Sol			Baron		Knitro		Generalization
Class	**Days**	**Product**	Ini Sol	MSE	Class.(%)	MSE	Class.(%)	MSE	Class.(%)	MSE
2	7	4	154.55	5.70	0.00	5.70	0.00	5.63	0.00	32.95
		6	761.24	7.53	0.00	7.53	0.00	7.53	0.00	22.17
		8	561.02	6.66	0.00	8.08	14.29	8.12	14.29	14.01
	14	4	536.44	30.75	0.00	3682.29	28.57	157.58	21.43	54.58
		6	1589.76	28.70	0.00	3441.98	50.00	147.56	42.86	37.94
		8	1212.62	16.56	0.00	456.94	14.29	232.32	21.43	53.14
	21	4	1768.88	71.78	0.00	17983.55	76.19	239.75	38.10	66.28
		6	2853.39	32.46	0.00	880.03	23.81	165.07	38.10	40.39
		8	3046.96	48.83	0.00	5830.66	66.67	180.49	47.62	83.47
	28	4	2521.28	48.69	0.00	499.01	14.29	548.69	25.00	43.28
		6	476.92	47.02	0.00	463.12	21.43	347.69	14.29	38.84
		8	2941.36	27.30	0.00	827.22	17.86	247.26	46.43	36.51
3	7	4	363.41	11.27	0.00	161.32	57.14	132.84	28.57	23.33
		6	756.64	4.98	0.00	149.63	42.86	171.26	42.86	11.72
		8	959.91	4.06	0.00	18.64	14.29	19.15	28.57	17.32
	14	4	1890.64	17.31	0.00	604.04	14.29	207.60	42.86	49.09
		6	216.42	15.07	0.00	1954.43	28.57	135.67	28.57	34.02
		8	1566.20	20.46	0.00	117.07	42.86	157.09	28.57	18.25
	21	4	2095.22	38.80	0.00	2314.24	28.57	91.38	14.29	36.77
		6	103.92	25.10	0.00	9283.33	71.43	124.95	52.38	33.79
		8	1214.47	28.51	0.00	9308.44	76.19	279.36	47.62	41.72
	28	4	2974.46	75.63	0.00	21006.68	67.86	198.67	32.14	93.28
		6	1837.06	36.76	0.00	15188.60	67.86	856.47	57.14	63.89
		8	2430.05	29.89	0.00	2054.74	25.00	1146.00	32.14	45.76
4	14	4	1361.45	11.87	0.00	796.87	64.29	456.36	42.86	23.67
		6	1688.36	12.60	0.00	5620.43	57.14	2345.00	57.14	69.58
		8	859.01	12.13	0.00	4235.06	71.43	4235.06	57.14	14.44
	21	4	248.52	39.63	0.00	n\a	n\a	n\a	n\a	41.12
		6	1328.48	34.23	0.00	n\a	n\a	n\a	n\a	44.48
		8	1696.79	38.47	0.00	n\a	n\a	n\a	n\a	33.67
	28	4	416.60	53.68	0.00	n\a	n\a	n\a	n\a	58.51
		6	953.01	56.95	0.00	n\a	n\a	n\a	n\a	69.58
		8	1234.02	45.63	0.00	n\a	n\a	n\a	n\a	58.63

CHAPTER 4

ARTICLE 4 : THE IMPACT OF CUSTOMER BEHAVIOR MODELS ON REVENUE MANAGEMENT SYSTEMS

Chapter Information : The article based on this chapter is submitted for publication.
Sh. Sharif Azadeh, M. Hosseinalifam, and G. Savard.

In this chapter, we present a comparative study which compares the impact of parametric and non-parametric demand models on revenue.

Abstract Revenue Management (RM) can be considered as an application of operation research in the transportation industry. For these service companies, it is a difficult task to adjust supply and demand. In order to maximize revenue, RM systems display demand behavior by using historical data. Usually, parametric methods are applied to estimate the probability of choosing a product at a given time. However, parameter estimation becomes challenging when we have a large dataset with a great proportion of unavailable products. In this research, we compare the impact of choosing a non-parametric method for probability estimation on revenue. The outcomes of this method have been compared to the total expected revenue using synthetic data.

Keyword Revenue management, Parametric and non-parametric demand models, Customer choice behaviour.

4.1 Introduction

Revenue management systems rely on the expected demand of each fare class. Therefore, the accuracy of demand models can affect revenue significantly. Most research in the literature has focused on the optimization methods based on which booking limits and fare products are determined. In 2005, van Ryzin shifted the focus from traditional product demand models to the analysis of customer behaviour in revenue management systems, based on the theory of discrete choice models (random utility) van Ryzin (2005). This change of paradigm has made it possible to blend the concept of revenue maximization with customer behavior analysis in recent research. Cooper has shown that ignoring customer behavior in RM systems results in loss of revenue Cooper *et al.* (2006).

An RM system has to decide whether to accept or reject a request from an arriving customer for a given product. Usually, products that are purchased in advance belong to the price-sensitive customer segments. However, the higher fare products are more likely to be bought right before departure. Therefore, product availabilities at a given time before departure have a direct impact on customer behavior. It is assumed that an arriving customer from each segment has a *consideration set* (a set of products among which the customer selects his choice) and he is willing to purchase the most attractive product based on a *preference vector* Liu et van Ryzin (2008). This vector is expressed in terms of product-based utilities according to which the probability of choosing a product by an arriving customer is determined.

In revenue maximization literature, the demand of a given product is often assumed to be independent from the others. That is, every client chooses a product independently from other ones (Talluri et Van Ryzin (2004), Weatherford (2000), Cooper *et al.* (2006)). One of these revenue maximization techniques is Deterministic Linear Programming, DLP, that was first introduced by Simpson et al. Simpson (1989). In their research, the expected demand has been determined by using the mean forecasted value. Afterwards, a linear program was suggested to define the optimal demand based on the capacity constraints for a given time period.

A more advanced model was proposed by Gallego and Liu et al (Gallego *et al.* (2004), Liu et van Ryzin (2008)). They have suggested a Choice-based Deterministic Linear Program (CDLP) to maximize revenue by defining at a given time, which product should be offered to the arriving customers from different segments to maximize revenue. However, in reality, the demand has a stochastic nature and the only information available in transportation companies is the registered booking of products during different time periods. Therefore, we need to extract customer behavior based on historical registered data.

The major contribution of this paper is to study a network revenue management problem with discrete customer choice behavior via a non-parametric algorithm that helps us to estimate customer preferences by directly using historical data. The revenue impact of this model has been compared to an upper bound resulting from a modified CDLP model and the outcome of the expected revenue of a simulation model. Our numerical experiments show that the proposed method of preference vector approximation performs as well as a parametric method with less computational cost.

The remainder of the paper is organized as follows. The main problem and the concept of customer preferences are described in Section 2. In Section 3, a modified CDLP problem is reviewed, which is used to generate synthetic data and an upper bound to the revenue in the comparison scheme. In Section 4, a non-parametric mathematical model is represented . In Section 5, numerical results are presented that suggest the non-parametric method of preference vector approximation can produce results close to those obtained with the original model. Avenues for further research are outlined in the concluding Section 6.

4.2 Problem description

In revenue management systems, customer behavior is expressed by their choices. In this research, each choice is made based on three basic rules : 1) each customer can choose only one of the available products, 2) the selected product has the maximum utility (or maximum weight) compared to the other available alternatives, and 3) only myopic customers are considered (customers who make their final decision at the time of arrival).

An arriving customer has a personal preference when purchasing a product. A subset of available offered products, considered by the client, is called a *consideration set*. Each customer selects an available alternative from his consideration set based on his preferences. *Preference vector* illustrates the vector of weights of all available products. As soon as one of these products is no longer available the probability of choosing a substitute changes, that suggests a conditional probability for choosing another product.

From a theorical point of view, the number of possible scenarios for preference orders are numerous for arriving customers from different segments. This makes the problem of revenue maximization computationally difficult to be solved. On the other hand, the only available information about choices made by clients is the historical data (registered bookings). In this research, we aim to use the information that historical data provides us in order to estimate choice probabilities. Figure 4.1 shows the registered booking of a given product for different departure days.

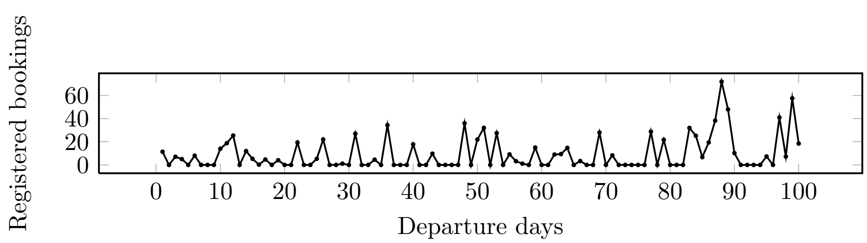


Figure 4.1 Registered bookings of a given product for successive departure days

In discrete choice models framework, the probability of choosing a given product by an arriving customer from a given segment is calculated by using product utilities. These utilities are defined as the sum of deterministic and stochastic terms that are related to the features of each product (McFadden (2001), Train (2009)). The choice of the random term results in different models. The one that is frequentlty used in RM systems is the Multi-Nomial Logit (MNL) model. The parameters of this model are typically estimated via maximum likelihood. The abovementioned *preference vector* in our model are also calculated by MNL.

Parameter estimation becomes more challenging when we have to deal with large volume of registered bookings with censored data which is a result of the unavailability of some products at different time intervals. In this research, we tackle this problem via a non-parametric method of estimation. The probability of choosing a product at a given time by an arriving customer is expressed as a variable of the mathematical model. We represent this model in Section 4. Then, we propose a learning process that helps us to estimate choice probabilities and consequently customer preferences in a more realistic way.

4.3 Modified CDLP model

In this section, we present a modified version of a choice-based linear program introduced by Liu and van Ryzin (Bront *et al.* (2009),Liu et van Ryzin (2008)). Consider a network with m resources (legs) providing n products. The set of products is expressed by $N = \{1, 2, .., n\}$ and vector $r = (r_1, r_2, ..., r_n)$ denotes associated revenue to the products. Vector $c = (c_1, c_2, ..., c_m)$ shows the initial capacities of resources. More than one resource unit can be used by a given product. The usage of each unit of capacity related to each product is described with an incidence matrix $A = [a_{ij}] \in B^{m\times n}$. The matrix entries are defined by :

$$a_{ij} = \begin{cases} 1, & \text{if resource } i \text{ is used by product } j \\ 0, & \text{otherwise} \end{cases}$$

Time is expressed in discrete periods. The total number of periods is defined by τ, $t = 1, 2, ..., \tau$. A customer arrival rate, λ, is considered for a given time interval. We suppose that at most one customer arrives during each period of time and he can buy only a single product or decide not to purchase at all.

Customers are divided into $l = \{1, ..., L\}$ different segments with corresponding consideration set c_l. If we have one arrival, p_l represents the probability that an arriving customer belongs to segment l with $\sum_{l=1}^{L} p_l = 1$. We consider a Poisson process of arriving streams of customers from segment l with rate $\lambda_l = \lambda p_l$ and total arriving rate of $\lambda = \sum_{l=1}^{L} \lambda_l$.

In each period of time t, the firm should decide about the offer set (*i.e.* a subset of products, $S \subseteq N$, that the firm makes available to arriving customers). If set S is offered, the deterministic quantity $P_j(S)$ indicates the probability of choosing product $j \in S$ and $P_j(S) = 0$ if $j \notin S$. We have $\sum_{j\in S} P_j(S)+P_0(S) = 1$, where $P_0(S)$ indicates the no-purchase probability.

Customers' choice probabilities are derived from a Multi-Nomial Logit (MNL) model which is one of the most commonly used models to study how customers make their choices. In the

MNL choice model, $v_l \geq 0$ represents a customer's preference vector for available products in consideration set C_l and v_{l0} represents the no-purchase preference. We let $P_{lj}(S)$ denote the probability of selling product $j \in C_l \bigcap S$ to a customer from segment l when set S is offered. According to the MNL choice model, this choice probability can be expressed as follows :

$$P_{lj}(S) = \frac{v_{lj}}{\sum_{h \in C_l \bigcap S} v_{lh} + v_{l0}}. \tag{1}$$

It can be obtained from formulation (1) that $P_{lj}(S) = 0$ if $v_{lj} = 0$ which can be a result of $j \notin C_l$ or $j \notin C_l \bigcap S$. We assume that $v_{l0} > 0$ for all segment $l = 1, 2, ..., L$.

In the more general case, as a firm cannot recognize the corresponding segment of an arrival in advance, the probability that the firm sells product j to an arriving customer is described as follows,

$$P_j(S) = \sum_{l=1}^{L} p_l P_{lj}(S). \tag{2}$$

Therefore, if a set S is offered the corresponding expected revenue is given by :

$$R(S) = \sum_{j \in S} r_j P_j(S). \tag{3}$$

Given $P(S) = (P_1(S), ..., P_n(S))^\top$ be the vector of purchase probabilities, the vector of capacity consumption probabilities $Q(S)$ is denoted by :

$$Q(S) = AP(S), \tag{4}$$

where $Q(S) = (Q_1(S), ..., Q_m(S))^\top$ and $Q_i(S)$ indicates the probability of using a unit of capacity on leg i, for $i = 1, 2, ..., m$.

Let binary variable $X_t(S)$ indicate whether set S at time t is offered. After obtaining the values of $R(S)$ and $Q_i(S)$, we can embed these functions in the following mathematical programming model to obtain the optimal resource allocation by taking into account the time and capacity constraints while maximizing revenue.

$$CDLP' = \max_X \sum_{t=1}^{\tau} \sum_S \lambda R(S) X_t(S) \tag{5}$$

subject to

$$\sum_t \sum_S \lambda Q_i(S) X_t(S) \leq c_i \qquad \forall i$$
$$\sum_S X_t(S) = 1 \qquad \forall t$$
$$X_t(S) \in \{0, 1\} \qquad \forall t, S$$

Model (5) is a modified formulation of the customer choice-based deterministic linear programming model (CDLP′) where in (5) the decision variable $X_t(S)$ indicates offering set S at booking period t instead of the total time periods during which S is offered (Bront *et al.* (2009) ,Kunnumkal et Topaloglu (2008)).

4.4 Non-parametric approach

Using the notation defined in the preceding section, we now exhibit the optimization model presented in this paper. The goal is to determine choice probabilities by using historical data. The objective function is to minimize the *prediction error* (the difference between estimated demand of each product at a given time and the related registered booking).

$$\mathcal{R} = \min_p \quad \sum_{j \in J} \sum_{t \in T} e_{jt}^2 \tag{6}$$
$$P_{jt}(S) B_{jt} D - O_{jt} = e_{jt} \qquad \forall j, t$$

where $P_{jt}(S)$ is the model variable that gives the probability of choosing product j at time t, B_{jt} depicts parameters to show availability status of this product, O_{jt} represents the observed booking for product j at time t, D is the total demand and finally, e_{jt} presents the difference between estimated demand and the registered booking.

During the customer decision making process, whenever a product is not available at a time interval, its demand can be either transferred to the other available products offered by the company (*recapture*) or it can be lost (*spill*). This changes the choice probabilities. We extract relations between these probabilities based on daily *choice sets* (the set of available products for each day).

Proposition. For a given customer segment, $l \in L$, if the choice set of time interval t is a subset of time period t', that is $S_t \subseteq S_{t'}$, then we have $P_{jt}(S) \geq P_{jt'}(S)$.

Input : Choice sets and registered bookings
Output : Valid inequalities

For all time periods $(t \in T)$
    For all days $(t' \in T)$
    Comapare the choice sets
    If $S_t \subset S_{t'}$ **then**
        If $S_t = S_{t'}$ **then**
            For all products $(j \in S_t, j \in S_{t'})$
                Write $P_{jt}(S) = P_{jt'}(S)$
        Else
            For all products $(j \in S_t, j \in S'_t)$
                Write $P_{jt}(S) > P_{jt'}(S)$

Algorithm 2 Valid inequalities on choice probabilities

We add the proposed set of valid inequalities to (6) in order to estimate the choice probabilities. By solving this model, we obtain upper and lower bounds to $P_{jt}(S)$. However, to provide a customer preference vector, we need to estimate product utilities, u_j. Thus, we solve the following system of inqualities based on the MNL model to obtain product utilities.

$$P^L_{jt}(S) \leq \frac{\exp(u_j)}{\sum\limits_{k \in C_t} \exp(u_k) + \exp(u_0)} \leq P^U_{jt}(S) \quad \forall j \in J \tag{7}$$

After estimating the utility of each product, we obtain the preference vectors based on a multi-nomial logit model. In the next section, we present the computational results in order to compare the outcome of our proposed preference estimation method on revenue with the revenue obtained from the CDLP$'$ model.

4.5 Computational results

4.5.1 Data instances

In this research, synthetic data is used to show the impact of our proposed demand model on revenue performance. Twenty-four generated instances are distinguishable based on three main elements : number of products, J (6,8), number of booking intervals, T (7,14,21,28), and the number of customer segments, L (3,4).

For a given number of products, number of segments and number of booking intervals, 3 different random instances have been produced. In order to generate registered historical data, O_{ij}, we use CDLP′ via the model presented in (5) to first derive the optimal offer sets, $X_t(S)$, which provide us the choice sets of each time interval, S_j, in the $\mathcal{R}$ model .

In the CDLP′ model, utilities and preference vectors are given. They are usually obtained by implementing an offline study of customer characterictics using parametric methods. Afterwards, for randomly arriving customers from different segments, we generate historical data.

4.5.2 Numerical results

The computational results have been carried out on a computer with 2.4 GHz CPU and 8 GB of RAM and 4 cores. We have used the Quadratic Solver of CPLEX 12.3 (sequential quadratic programming) to solve problem $\mathcal{R}$. To solve problem CDLP′ (column generation approach), we have utilized FICO Xpress-Mosel 7.2.

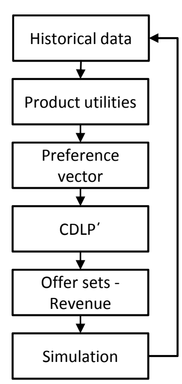


Figure 4.2 Non parametric method of preference estimation and its impact on revenue

Figure 4.2 presents our comparison scheme. Solving CDLP′ results in finding an upper bound to the revenue. We perturb the preference vector of each customer segment (by adding a Gumbel distributed error term to the utility function), then, we find the expected revenue of the simulated model. In this method, utilities and preferences are known.

Now, by directly using generated synthetic data, we estimate product utilities with our proposed nonparametric method. According to these estimated values, we can reproduce preference vector of arriving customers. The revenue resulting from both methods are compared.

Table 4.1 presents the comparative study on 24 instances. These examples are generated based on random customer arrival rates from different segments.

Table 4.1 Revenue Comparison

	Product	Segment	Book.Int	Instance #	UB	ER(CDLP′)	ER($\mathcal{R}$)	Gap(%)
1				1	3954	3439	3417	0.64%
2			7	2	3954	3439	3431	0.23%
3				3	3954	3439	3448	-0.26%
4				1	7500	6581	6472	1.66%
5			14	2	7500	6581	6743	-2.46%
6	6	3		3	7500	6581	6503	1.19%
7				1	12080	10570	10305	2.51%
8			21	2	12080	10570	10619	-0.46%
9				3	12080	10570	10437	1.26%
10				1	17371	15822	15062	4.80%
11			28	2	17371	15822	15801	0.13%
12				3	17371	15822	15796	0.16%
13				1	4817	4389	4310	1.80%
14			7	2	4817	4389	4347	0.96%
15				3	4817	4389	4371	0.41%
16				1	9634	9234	9246	-0.13%
17			14	2	9634	9234	9194	0.43%
18	8	4		3	9634	9234	9282	-0.52%
19				1	14064	13262	13264	-0.02%
20			21	2	14064	13262	13238	0.18%
21				3	14064	13262	13125	1.03%
22				1	16681	15783	15597	1.18%
23			28	2	16681	15783	15742	0.26%
24				3	16681	15783	15663	0.76%

Column "Book.Int" shows the number of booking periods. "UB" represents the revenue resulting from CDLP$'$ that provides us an upper bound on the expected revenue. Columns "ER(CDLP$'$)" and "ER($\mathcal{R}$)" respectively show the expected revenue resulting from simulation (with perturbed preference vector) and our non-parametric demand model represented in the $\mathcal{R}$ model.

The gap, "Gap(%)", between "ER(CDLP$'$)" and "ER($\mathcal{R}$)" is small. Even though there are cases where $\mathcal{R}$ has slightly outperformed the simulated model, this is not surprising. The reason for this is that while using a non-parametric method, the degree of freedom of the model increases which can result in slight outperformance. The numerical results suggest that the non-parametric method of preference vector approximation can produce revenues close to those obtained with the original CDLP$'$.

4.5.3 Discussion

From a practical point of view, this method can be helpful to revenue management systems in transportation companies. By directly using historical data, we can take product availabilities into account, which brings more dynamism to the decision making process in order to find the optimal product offer sets.

Moreover, by using this method, we avoid using other offline studies to capture customer behavior. For example, to find preference vectors, we usually need to gather information about customers' characteristics, such as, income, purpose of travel and comfort preferences, which can be time consuming and costly.

4.6 Conclusion

In this research, we have introduced a non-parametric method to capture customer behavior in revenue management systems. We have used historical data in order to extract logical relations between choice probabilities and solve an optimization problem in order to estimate these probabilities along with product utilities, which results in finding customer preferences.

A modified choice-based deterministic linear programming model has been chosen to generate synthetic data. We have obtained the expected revenue associated with two methods of utility estimation : 1) simulated CDLP$'$ model (with perturbed preference vector) and 2) approximated preference vector based on estimated product utilities by $\mathcal{R}$ model. The results testify to remarkable revenue performance. The gap between these two methods is slight, moreover, there are cases where our model outperforms the simulated revenue of CDLP$'$.

CHAPTER 5

GENERAL DISCUSSION

In this thesis, we have studied demand forecasting in revenue management systems in transportation. First, in Chapter 2, we have introduced new characteristics. We can classify and categorize the existing literature on the subject of demand uncensoring based on these characteristics. We have considered both supplier and demand factors. Supplier specifications comprise the customer type (myopic and strategic) and the domain of application that includes airline, railway, rental and hotel industries. On the other hand, demand characteristics are defined based on several factors : product dependency, diversion, seasonality and competition. Afterwards, we have introduced four main categories for the methods used in order to uncensor demand in RM systems : basic methods, statistical methods, choice based models and optimization tools. In Chapter 3, we have used a modified neural network in order to predict the number of passengers at departure time. The historical data we have applied in order to train the network belonged to a major European railway company. This method can be expressed according to our proposed classification scheme as follows : $[\mu_1 = myop, \mu_2 = rail|\delta_1 = ind|\alpha_2 = cm(NN)]$. The results are promising, suggesting a low error of prediction.

However, this proposed model can be used for datasets that do not have a high proportion of missing data. In addition, it does not take customer behavior into account. These shortcomings have motivated us to introduce a non parametric choice-based optimization model, which is able to take both customer behavior and seasonal effects into consideration. In Chapter 4 and 5, we have proposed an optimization tool that has minimized the differences between the registered bookings and the estimated demand of each product at a given time. The original problem suggests a nonconvex nonlinear model with integer and binary variables. This model included two main phases : Estimation and Clustering. The estimation part predicted the daily demand flow based on its related choice set and it has estimated product utilities. Afterwards, the clustering part has extracted the seasonal effect of each interval of time (in this case departure days) into one of the predefined number of clusters (external segmentation). Based on our proposed classification scheme in Chapter 2, this method can be expressed as follows : $[\mu_1 = myop, \mu_2 = rail|\delta_1 = dep, \delta_2 = spill, , \delta_3 = season, , \delta_4 = ex|\alpha_4 = NLP(LS)]$.

First, we linearize and convexify the problem by using modified McCormick inequalities. Afterwards, we implemented a series of preprocessings by setting a set of variable inequalities according to the characteristics of the choice sets and logical relations between choice probabilities. Moreover, by using a suggested mathematical model, we have reduced the range of variables before branching. Then, we proposed a branch and bound algorithm that used both linear (CPLEX) and nonlinear (IPOPT) solvers to find estimations and to classify departure days into one of the clusters. The computational results have been provided by applying synthetic data. The results have been compared to two nonlinear (KNITRO) and global (BARON) optimizers.

In Chapter 6, we examined the effect of our proposed model on revenue. We generated synthetic data based on a modified CDLP model. Then, we used our prediction model in order to estimate the product utilities. Subsequently, we reproduced the vector of customer preferences by calculating choice probabilities. The upper bound has been set by the original CDLP′ problem. Afterwards, we perturbed the original preference vectors by using a Gumbel distributed error term. The outcomes of our model have been compared to both the upper bound and perturbed model. The gap between these two methods testifies to the efficiency of our demand model and represents its positive impact on maximizing revenue.

CONCLUSION AND RECOMMENDATION

In this dissertation, we have analyzed the problem of demand unconstraining and forecasting in revenue management systems. In this chapter, we summarize our main findings and we point out future research directions.

In Chapter 2, we have proposed a state of art review framework on demand modeling in revenue management. We have analyzed the problem of censored demand in this field. Two hundred papers were reviewed and classified based on our proposed tuple notation technique. We have classified existing uncensoring methods based on the key elements introduced. Finally, we have suggested new directions in this domain of research in order to tackle the common problems that often occur while using each type of these methods.

In Chapter 3, we have applied a statistical method - more specifically, a modified neural network - which is used in order to predict the number of passengers at the departure time. The prediction model used in this research is an improved Multi-Layer Perceptron (MLP) describing the relationship between the amount of passengers and factors affecting this quantity based on historical data. The results show this approach is promising in railway demand forecasting.

In Chapter 4, an optimization model is shown to estimate the predicted demand of each product at a given time by minimizing the difference between estimated values of demand and registered bookings. Our variables have included the utilities of all products, daily potential demand and the binary values that assign each day to one of the predefined number of clusters. For a given booking interval and a given origin destination, we have classified departure days into one of the clusters based on its daily demand flow. The presented mathematical model has suggested a nonconvex nonlinear program with integer variables. Several definitions and propositions have been presented based on which, in Chapter 5, we have introduced a new algorithm to model customer behavior and demand.

In Chapter 5, we introduce a new algorithm that is capable of solving the problem of demand forecasting by using historical data in order to calibrate the model. We have linearized and convexified the original problem in two parts : estimation and classification. In the first part, we have defined a modified version of Mccormick inequalities, which linearizes the bilinear term of estimated demand. In the second part, we have reformulated the problem in order to avoid using the big M.

We have implemented a preprocessing approach before starting to branch on variables. Then, a series of valid inequalities were introduced to tighten the feasible region. A mathematical model has been represented, based on which we have reduced the range of variables.

By using a specific node and variable selection approach, we have used branch and bound. In the process, each node has been prepared before selecting the next node in the sense that by using interval arithmetic, we have readjusted the range of choice probabilities and daily potential demand. As soon as the relaxation problem found a binary solution for assignment variables on CPLEX, we implemented a nonlinear problem on IPOPT in order to estimate the product utilities. The iterative process stopped whenever the gap between the relaxation and nonlinear problem was less than 1% or there had been no improvement in the solution for more than 60 minutes.

Computational results were satisfactory for different batches of data of different sizes. The sizes of these datasets are determined based on the predefined number of clusters, number of days and number of products. The results have been compared to both nonlinear (KNITRO) and global (BARON) optimizers.

In Chapter 6, we have decided to examine the impact of our proposed demand model on revenue. We used a modified Choice-based Deterministic Linear Programming (CDLP) to find an upper bound to the revenue. Then, we produced the synthetic data based on this model for different sizes (based on the number of products, booking intervals and customer segments). Customers have arrived randomly from different segments with predefined arrival rates that follow the poisson process. Synthetic data was based on and produced from the choice model, a multinomial logit. Then, we applied our proposed prediction model and extracted the utilities related to each product offered. Based on these utilities, we have calculated the preference vector according to which arriving customers have made their decisions. Moreover, in order to compare the results, we have perturbed the preference vector of customers by using a Gumbel distributed error. The revenue of this model has been compared to the upper bound and perturbed model. The results testify to the efficiency of the proposed model on revenue.

An interesting future research direction concerning the demand uncensoring algorithms is to add an index for utilities in order to take customer segmentations as well as clustering departure days into account. This helps us accurately calculate the spilled and recaptured demand for unavailable products. In addition, we will be able to obtain an estimation for the probability of substitutions. This is essential for revenue management systems to consider the buy ups whenever one of the products is no longer available for a given interval of time.

REFERENCES

ANDERSON, M. D., SHARFI, K. et GHOLSTON, S. E. (2006). Direct demand forecasting model for small urban communities using multiple linear regression. *Transportation Research Record*, 1981, 114–117.

ANDERSSON, S. (1998). Passenger choice analysis for seat capacity control : A pilot project in Scandinavian Airlines. *International Transactions in Operational Research*, 5, 471–486.

ARMSTRONG, A. et MEISSNER, J. (2010). Railway revenue management : Overview and models. Rapport technique.

ARMSTRONG, J. (2001). *Principles of forecasting : a handbook for researchers and practitioners*. Springer, April 2001 édition.

BANSAL, M. (2012). *Modeling Customer Behavior for Revenue Management*. Thèse de doctorat, Columbia University.

BARON (2013). Baron. `http://www.gams.com/dd/docs/solvers/baron.pdf`.

BELOBABA, P. (1987). Air travel demand and airline seat inventory management. Rapport technique, Cambridge, MA : Flight Transportation Laboratory, Massachusetts Institute of Technology.

BELOBABA, P. et FARKAS, A. (1999). Yield management impacts on airline spill estimation. *Transportation Science*, 33, 217–232.

BELOBABA, P. et HOPPERSTAD, C. (1999). Boeing/MIT simulation study : PODS results update. *AGIFORS Reservations and Yield Management Study Group Symposium Proceedings*.

BESBES, O. et ZEEVI, A. (2006). Blind nonparametric revenue management : Asymptotic optimality of a joint learning and pricing method.

BODEA, T. (2008). *Choice-based revenue management : a hotel perspective*. Thèse de doctorat, Georgia Institute of Technology.

BOX, G., JENKINS, G. et REINSEL, G. (2011). *Time series analysis : forecasting and control*, vol. 734. Wiley, Fourth édition.

BRONT, J., MÉNDEZ-DíAZ, I. et VULCANO, G. (2009). A column generation algorithm for choice-based network revenue management. *Operations Research*, 57, 769–784.

BRUMMER, M. *ET AL.* (1988). Determination of potential load capacity based on observed load factor data, a study for northwest airlines. Rapport technique.

CACHON, G. et SWINNEY, R. (2009). Purchasing, pricing, and quick response in the presence of strategic consumers. *Management Science*, 55, 497–511.

CACHON, G. et SWINNEY, R. (2011). The value of fast fashion : Quick response, enhanced design, and strategic consumer behavior. *Management Science*, 57, 778–795.

CELIKOGLU, H. B. et CIGIZOGLU, H. K. (2007). Public transportation trip flow modeling with generalized regression neural networks. *Advances in Engineering Software*, 38, 71–9.

CHEN, Y. et LUO., L. (2005). Approach of uncensored demand in airline revenue. *Journal of Chengdu University of Information Technology*, 20, 747–750.

CHUNG, J. et LEE, D. (2002). Structural model of automobile demand in korea. *Transportation Research Record*, 1807, 87–91.

CONLON, C. et MORTIMER, J. (2008). Demand estimation under incomplete product availability. Rapport technique, National Bureau of Economic Research.

COOPER, W., HOMEM-DE MELLO, T. et KLEYWEGT, A. (2006). Models of the spiral-down effect in revenue management. *Operations Research*, 54, 968–987.

CREVIER, B., CORDEAU, J. et SAVARD, G. (2012). Integrated operations planning and revenue management for rail freight transportation. *Transportation Research Part B : Methodological*, 46, 100–119.

CROSS, R. (1997). *Revenue management*. Universities Press, December 29, 1997 édition.

DANTAS, A., YAMAMOTO, K., LAMAR, M. et YAMASHITA, Y. (2000). Neural network for travel demand forecast using GIS and Remote Sensing. Como, Italy. *INNS-ENNS International Joint Conference on Neural Networks*. IEEE.

DEMPSTER, A., LAIRD, N. et RUBIN, D. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society. Series B : Methodological*, 39, 1–38.

DEVOTO, R., FARCI, C. et LILLIU, F. (2002). Analysis and forecast of air transport demand in sardinia's airports as a function of tourism variables. WITPress, Seville, Spain, Advances in Transport, 699–709.

EKONOMOU, L. (2010). Greek long-term energy consumption prediction using artificial neural networks. *Energy*, 35, 512–517.

EREN, S. et MAGLARAS, C. (2009). A maximum entropy joint demand estimation and capacity control policy. Rapport technique, Columbia Business School, Center for pricing and revenue management.

FARIAS, V. (2007). *Revenue Management Beyond "Estimate, then Optimize"*. Thèse de doctorat, Stanford University.

FARIAS, V., JAGABATHULA, S. et SHAH, D. (2013). A nonparametric approach to modeling choice with limited data. *Management Science*, In press, accepted manuscript.

FERGUSON, M. et QUEENAN, C. (2009). Starting with good inputs : Unconstraining demand data in revenue management. *INFORMS Transactions on Education*, 9, 180–187.

GALLEGO, G. et HU, M. (2008). Dynamic pricing of perishable assets under competition. *AGIFORS Proceedings*.

GALLEGO, G., IYENGAR, G., PHILLIPS, R. et DUBEY, A. (2004). Managing flexible products on a network.

GORIN, T. (2000). *Airline revenue management : Sell-up and forecasting algorithms*. Thèse de doctorat, Massachusetts Institute of Technology.

GUO, P. (2008). *The research on unconstraining methods based on multi-distribution of demand in revenue management systems*. Dissertation. com.

GUO, P., XIAO, B. et LI, J. (2008). Research on the method of demand forecasting without constraint for airline passenger revenue management. *Journal of Transportation Engineering and Information*, 2, 71–78.

GURBUZ, E. *ET AL.* (2011). *Revenue Management Operations in Hotel Chains in Finland*. Mémoire de maîtrise.

GUTIERREZ, R., SOLIS, A. et MUKHOPADHYAY, S. (2008). Lumpy demand forecasting using neural networks. *International Journal of Production Economics*, 111, 409–420.

HAENSEL, A. et KOOLE, G. (2010). Estimating unconstrained demand rate functions using customer choice sets. *Journal of Revenue & Pricing Management*, 10, 438–454.

HAENSEL, A., KOOLE, G. et ERDMAN, J. (2011). Estimating unconstrained customer choice set demand : A case study on airline reservation data. *Journal of Choice Modelling*, 4, 75–87.

HAYKIN, S. (1998). *Neural networks : a comprehensive foundation*. Prentice Hall - Canada Toronto.

HE, D. et LUO, L. (2006). An improved winters model for airline demand forecast. *Journal of Transportation Systems Engineering and Information Technology*, 6, 103–107.

HOPPERSTAD, C. (1995). An alternative detruncation method. Rapport technique, Boeing Commercial Aircraft Company.

HOPPERSTAD, C. (1996). Passenger O/D simulator–PODS. *AGIFORS Proceedings*.

HOPPERSTAD, C. (1997). PODS update : simulation of O/D revenue management schemes. *AGIFORS Proceedings Yield Management Study Group*.

HOPPERSTAD, C., ZERBIB, G. et BELOBABA, P. (2007). Methods for estimating sell-up. *AGIFORS Yield Management Study Group Meeting, Jeju Island, Korea.*

HUH, W., LEVI, R., RUSMEVICHIENTONG, P. et ORLIN, J. (2011). Adaptive data-driven inventory control with censored demand based on kaplan-meier estimator. *Operations Research*, 59, 929–941.

HYNDMAN, R., KOEHLER, A., ORD, J. et SNYDER, R. (2008). *Forecasting with exponential smoothing : the state space approach.* Springer.

IBM (2013). Cplex. `http://www-01.ibm.com/software/commerce/optimization/cplex-optimizer/`.

IPOPT (2013). `https://projects.coin-or.org/Ipopt`.

JA, S., RAO, B. et CHANDLER, S. (2001). Passenger recapture estimation in airline revenue management. *AGIFORS 41st Annual Symposium.*

JIANG, H. (2007). Network capacity management competition. Rapport technique, Judge Business School at University of Cambridge, UK.

JIANG, H. et PANG, Z. (2011). Network capacity management under competition. *Computational Optimization and Applications*, 50, 287–326.

KACHITVICHYANUKUL, V., LUONG, H. et PITAKASO, R. (2012). Hotel room demand forecasting via observed reservation information.

KANDANANOND, K. (2011). Forecasting electricity demand in thailand with an artificial neural network approach. *Energies*, 4, 1246–1257.

KANDANANOND, K. (2012). Consumer product demand forecasting based on artificial neural network and support vector machine. *World Academy of Science, Engineering and Technology - The international scientific journal*, 63, 372–376.

KAPLAN, E. et MEIER, P. (1958). Nonparametric estimation from incomplete observations. *Journal of the American statistical association*, 53, 457–481.

KARMARKAR, S., GOUTAM, D. et TATHAGATA, B. (2010). Revenue impacts of demand unconstraining and accounting for dependency. *Journal of Revenue & Pricing Management*, 10, 367–381.

KUNNUMKAL, S. et TOPALOGLU, H. (2008). A refined deterministic linear program for the network revenue management problem with customer choice behavior. *Naval Research Logistics (NRL)*, 55, 563–580.

KUNNUMKAL, S. et TOPALOGLU, H. (2010). A randomized linear program for the network revenue management problem with customer choice behavior. *Journal of Revenue & Pricing Management*, 10, 455–470.

KWON, C., FRIESZ, T., MOOKHERJEE, R., YAO, T. et FENG, B. (2009). Non-cooperative competition among revenue maximizing service providers with demand learning. *European Journal of Operational Research*, 197, 981–996.

LEE, A. (1990). Airline reservations forecasting : Probabilistic and statistical models of the booking process. Rapport technique, Flight Transportation Laboratory, Dept. of Aeronautics and Astronautics, Massachusetts Institute of Technology.

LEE, C., LIN, T. et LIN, C. (2005). Pattern analysis on the booking curve of an inter-city railway. *Journal of the Eastern Asia Society for Transportation Studies*, 6, 303–317.

LEVIN, Y., MCGILL, J. et NEDIAK, M. (2010). Optimal dynamic pricing of perishable items by a monopolist facing strategic consumers. *Production and Operations Management*, 19, 40–60.

L'HEUREUX, E. (1986). A new twist in forecasting short-term passenger pickup. *AGIFORS Proceedings*.

LITTLE, R. et RUBIN, D. (2002). *Statistical analysis with missing data*. Wiley New York, Second édition.

LITTLEWOOD, K. (2005). Special issue papers : Forecasting and control of passenger bookings. *Journal of Revenue & Pricing Management*, 4, 111–123.

LIU, P. (2004). Hotel demand/cancellation analysis and estimation of unconstrained demand using statistical methods. *Revenue Management and Pricing : Case Studies and Applications*, Cengage Learning Business Press.

LIU, P., SMITH, S., ORKIN, E. et CAREY, G. (2002). Estimating unconstrained hotel demand based on censored booking data. *Journal of Revenue & Pricing Management*, 1, 121–138.

LIU, Q. et VAN RYZIN, G. (2008). On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management*, 10, 288–310.

LIU, Q. et ZHANG, D. (2011). Dynamic pricing competition with strategic customers under vertical product differentiation. Rapport technique, Hong Kong University of Science and Technology, Kowloon.

MARTíNEZ, V. et TALLURI, K. (2011). Dynamic price competition with fixed capacities. *Management Science*, 57, 1078–1093.

MCCORMICK, G. (1976). Computability of global solutions to factorable nonconvex programs Convex underestimating problems. *Mathematical Programming*, 10, 147–175.

MCFADDEN, D. (2001). Disaggregate behavioral travel demand's RUM side. *Travel Behaviour Research*. Pergamon Press : Oxford.

MCGILL, J. (1995). Censored regression analysis of multiclass passenger demand data subject to joint capacity constraints. *Annals of Operations Research*, 60, 209–240.

MCGILL, J. et VAN RYZIN, G. (1999). Revenue management : Research overview and prospects. *Transportation science*, 33, 233–256.

MEISSNER, J. et STRAUSS, A. (2009). Choice-based network revenue management under weak market segmentation. Rapport technique, Lancaster University Management School.

MEISSNER, J. et STRAUSS, A. (2012a). Improved bid prices for choice-based network revenue management. *European Journal of Operational Research*, 217, 417–427.

MEISSNER, J. et STRAUSS, A. (2012b). Network revenue management with inventory-sensitive bid prices and customer choice. *European Journal of Operational Research*, 216, 459–468.

MEISSNER, J., STRAUSS, A. et TALLURI, K. (2012). An enhanced concave program relaxation for choice network revenue management. Rapport technique, Lancaster University Management School.

MILTENBURG, J. et PONG, H. C. (2007). Order quantities for style goods with two order opportunities and bayesian updating of demand. part i : No capacity constraints. *International Journal of Production Research*, 45, 1643–1663.

MISHRA, S. VISWANATHAN, V. (2003). Revenue management with restriction-free pricing. *Proceedings of the AGIFORS Revenue Management and Distribution Study Group Meeting*.

MONTGOMERY, D., JENNINGS, C. et KULAHCI, M. (2008). *Introduction to time series analysis and forecasting*, vol. 526. Wiley - Hoboken New Jersey.

MOZOLIN, M., THILL, J. C. et USERY, E. L. (2000). Trip distribution forecasting with multilayer perceptron neural networks : a critical evaluation. *Transportation Research, Part B (Methodological)*, 34B, 53–73.

NAIM, I. et GILDEA, D. (2012). Convergence of the em algorithm for gaussian mixtures with unbalanced mixing coefficients. *arXiv preprint arXiv :1206.6427*.

NEOS (2013). `http://www.neos-server.org/neos/`.

NETESSINE, S. et SHUMSKY, R. (2005). Revenue management games : Horizontal and vertical competition. *Management Science*, 51, 813–831.

ORKIN, E. (1998). Wishful thinking and rocket science the essential matter of calculating unconstrained demand for revenue management. *Cornell Hotel and Restaurant Administration Quarterly*, 39, 15–19.

PERAKIS, G. et SOOD, A. (2006). Competitive multi-period pricing for perishable products : A robust optimization approach. *Mathematical Programming*, 107, 295–335.

PÖLT, S. (2000). From bookings to demand : the process of unconstraining. *AGIFORS Reservations and Yield Management Study Group, New York*.

POPESCU, A., BARNES, E., JOHNSON, E. et KESKINOCAK, P. (2012). Bid prices when demand is a mix of individual and batch bookings. *Transportation Science*, In press, accepted manuscript.

QUEENAN, C., FERGUSON, M., HIGBIE, J. et KAPOOR, R. (2009). A comparison of unconstraining methods to improve revenue management systems. *Production and Operations Management*, 16, 729–746.

RATLIFF, R., RAO, B., NARAYAN, C. et YELLEPEDDI, K. (2008). A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. *Journal of Revenue & Pricing Management*, 7, 153–171.

SA, J. (1987). Reservations forecasting in airline yield management. Rapport technique, Cambridge, MA : Flight Transportation Laboratory, Massachusetts Institute of Technology.

SALCH, J. (1997). Unconstraining passenger demand using the em algorithm. *Proceedings of the INFORMS Conference*.

SALEH, R. (1997). Estimating lost demand with imperfect availability indicators. *Proceedings of the AGIFORS Reservations and Yield Management Study Group*.

SHARIF AZADEH, S., LABIB, R. et SAVARD, G. (2012). Railway demand forecasting in revenue management using neural networks. *International Journal of Revenue Management*, In press, accepted manuscript.

SHEN, Z. et SU, X. (2007). Customer behavior modeling in revenue management and auctions : A review and new research opportunities. *Production and operations management*, 16, 713–728.

SIMPSON, R. W. (1989). Using network flow techniques to find shadow prices for market and seat inventory control. Rapport technique, Memorandum, M89-1, Flight Transportation Laboratory, MIT, Cambridge, MA.

ŠKURLA BABIĆ, R., GRGUREVIĆ, I. et MAJIĆ, Z. (2011). Comparison of air travel demand forecasting methods. *ICTS 2011 : 14th International Conference on Transport Science MARITIME, TRANSPORT AND LOGISTICS SCIENCE*.

SKWAREK, D. (1996a). Competitive impacts of yield management system components : Forecasting and sell-up models. Rapport technique, Cambridge, Mass. : Massachusetts Institute of Technology, Flight Transportation Laboratory.

SKWAREK, D. (1996b). Revenue and traffic impacts of alternative detruncation methods. *Proceedings of the AGIFORS Reservations and Yield Management Study Group.*

SNYDER, R. D., KOEHLER, A. B. et ORD, J. K. (2002). Forecasting for inventory control with exponential smoothing. *International Journal of Forecasting*, 18, 5–18.

STEFANESCU, C. (2009). Multivariate customer demand : modeling and estimation from censored sales. Rapport technique, London Business School.

STEFANESCU, C., DEMIGUEL, V., FRIDGEIRSDOTTIR, K. et ZENIOS, S. (2004). Revenue management with correlated demand forecasting. *American Statistical Association, Business and Economics Statistics Section.*

SU, X. (2007). Intertemporal pricing with strategic customer behavior. *Management Science*, 53, 726–741.

SWAN, M. (1999). Spill modeling for airlines. *8th World Conference on Tranport Research Proceedings.*

SWAN, W. (1979). A systems analysis of scheduled air transportation networks. Rapport technique, Massachusetts Institute of Technology, Flight Transportation Laboratory.

SWAN, W. (2002). Airline demand distributions : passenger revenue management and spill. *Transportation Research Part E : Logistics and Transportation Review*, 38, 253–263.

SWINNEY, R. (2011). Selling to strategic consumers when product value is uncertain : The value of matching supply and demand. *Management Science*, 57, 1737–1751.

TALLURI, K. (2009). A finite-population revenue management model and a risk-ratio procedure for the joint estimation of population size and parameters. Rapport technique, Universitat Pompeu Fabra.

TALLURI, K. (2010). A randomized concave programming method for choice network revenue management. Rapport technique, Universitat Pompeu Fabra.

TALLURI, K. et VAN RYZIN, G. (2004). Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50, 15–33.

TALLURI, K. et VAN RYZIN, G. (2005). *The theory and practice of revenue management.* Springer.

TALLURI, K., VAN RYZIN, G., KARAESMEN, I. et VULCANO, G. (2008). Revenue management : models and methods. *Proceedings of the 40th Conference on Winter Simulation.* Winter Simulation Conference, 145–156.

TRAIN, K. (2009). *Discrete choice methods with simulation.* Cambridge university press, Second édition.

TSAI, T., LEE, C. et WEI, C. (2005). Design of dynamic neural networks to forecast short-term railway passenger demand. *Journal of the Eastern Asia Society for Transportation Studies*, 1651–1666.

VAN RYZIN, G. (2005). Future of revenue management : models of demand. *Journal of Revenue & Pricing Management*, 4, 204–210.

VAN RYZIN, G. et MCGILL, J. (2000). Revenue management without forecasting or optimization : An adaptive algorithm for determining airline seat protection levels. *Management Science*, 46, 760–775.

VAN RYZIN, G. et VULCANO, G. (2008). Computing virtual nesting controls for network revenue management under customer choice behavior. *Manufacturing & Service Operations Management*, 10, 448–467.

VARAGOULI, E. G., SIMOS, T. E. et XEIDAKIS, G. S. (2005). Fitting a multiple regression line to travel demand forecasting : The case of the prefecture of xanthi, northern greece. *Mathematical and Computer Modelling*, 42, 817–36.

VULCANO, G., VAN RYZIN, G. et CHAAR, W. (2010). Om practice-choice-based revenue management : An empirical study of estimation and optimization. *Manufacturing & Service Operations Management*, 12, 371–392.

VULCANO, G., VAN RYZIN, G. et RATLIFF, R. (2012). Estimating primary demand for substitutable products from sales transaction data. *Operations Research*, 60, 313–334.

WEATHERFORD, L. (2000). Unconstraining methods. *AGIFORS Reservations and Yield Management Study Group.*

WEATHERFORD, L. et BELOBABA, P. (2002). Revenue impacts of fare input and demand forecast accuracy in airline yield management. *Journal of the Operational Research Society*, 52, 811–821.

WEATHERFORD, L., GENTRY, T. et WILAMOWSKI, B. (2003a). Neural network forecasting for airlines : A comparative analysis. *Journal of Revenue and Pricing Management*, 1, 319–331.

WEATHERFORD, L., GENTRY, T. et WILAMOWSKI, B. (2003b). Neural network forecasting for airlines : A comparative analysis. *Journal of Revenue and Pricing Management*, 1, 319–331.

WEATHERFORD, L. et KIMES, S. (2003). A comparison of forecasting methods for hotel revenue management. *International Journal of Forecasting*, 19, 401–415.

WEATHERFORD, L. et POLT, S. (2002). Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues. *Journal of Revenue and Pricing Management*, 1, 234–254.

WEATHERFORD, L. et RATLIFF, R. (2010). Review of revenue management methods with dependent demands. *Journal of Revenue & Pricing Management*, 9, 326–340.

WEI, F. et HONG, C. (2004). Cluster model for flight demand forecasting. IEEE, Piscataway, NJ, USA, vol. Vol.4 de *Fifth World Congress on Intelligent Control and Automation (IEEE Cat. No.04EX788)*, 3170–3.

WICKHAM, R. (1995). Evaluation of forecasting techniques for short-term demand of air transportation. Rapport technique, Massachusetts Institute of Technology, Flight Transportation Laboratory.

WIDIARTA, H., VISWANATHAN, S. et PIPLANI, R. (2007). On the effectiveness of top-down strategy for forecasting autoregressive demands. *Naval Research Logistics*, 54, 176–88.

XIAOLONG, Z. (2007). Inventory control under temporal demand heteroscedasticity. *European Journal of Operational Research*, 182, 127–44.

XU, L. (1997). Comparative analysis on convergence rates of the em algorithm and its two modifications for gaussian mixtures. *Neural Processing Letters*, 6, 69–76.

YANG, H., SONG, H. et ZHANG, S. (2010). Dynamic pricing with strategic and myopic consumers. *Business Intelligence and Financial Engineering (BIFE), 2010 Third International Conference on*. IEEE.

YIN, R., AVIV, Y., PAZGAL, A. et TANG, C. (2009). Optimal markdown pricing : Implications of inventory display formats in the presence of strategic customers. *Management Science*, 55, 1391–1408.

ZAKHARY, A., EL GAYAR, N. et ATIYA, A. F. (2008). A comparative study of the pickup method and its variations using a simulated hotel reservation data. *ICGST International Journal on Artificial Intelligence and Machine Learning*, 8, 15–21.

ZENI, R. (2001). *Improved forecast accuracy in airline revenue management by unconstraining demand estimates from censored data*. Thèse de doctorat, MIT.

ZENI, R. et LAWRANCE, K. (2004). Unconstraining demand data at us airways. *Revenue Management and Pricing : Case Studies and Applications*, Cengage Learning Business Press, chapitre 11. 124–136.

ZF LI, M. et HOON OUM, T. (2000). Airline spill analysis–beyond the normal demand. *European Journal of Operational Research*, 125, 205–215.

ZHANG, D. et COOPER, W. (2005). Revenue management for parallel flights with customer-choice behavior. *Operations Research*, 53, 415–431.

ZHANG, D. et COOPER, W. (2009). Pricing substitutable flights in airline revenue management. *European Journal of Operational Research*, 197, 848–861.

ZHANG, G. et QI, M. (2005). Neural network forecasting for seasonal and trend time series. *European journal of operational research*, 160, 501–514.

ZHU, J. (2006). Using turndowns to estimate the latent demand in a car rental unconstrained demand forecast. *Journal of Revenue & Pricing Management*, 4, 344–353.

ZICKUS, J. (1998). Forecasting for airline network revenue management : Revenue and competitive impacts. Rapport technique, Massachusetts Institute of Technology, Transportation Laboratory.

ZIENA (2013). Knitro. `http://www.ziena.com/docs/Knitro70_UserManual.pdf`.