EVŞEN KORKMAZ

Bridging Models and Business

Understanding heterogeneity in hidden drivers of customer purchase behavior

Bridging Models and Business

Understanding heterogeneity in hidden drivers of customer purchase behavior

Bridging Models and Business

Understanding heterogeneity in hidden drivers of customer purchase behavior

Bruggen slaan tussen modellen en bedrijven
Het begrijpen van heterogeniteit in verborgen drijfveren in het koopgedrag van klanten

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board.

The public defense shall be held on

Friday 12 September 2014 at 11:30 hrs

by

Evşen Korkmaz
born in Niğde, Turkey.

ERASMUS UNIVERSITEIT ROTTERDAM

Doctoral Committee

Promoters:	Prof.dr. S.L. van de Velde
	Prof.dr. D. Fok
Other members:	Prof.dr. L.G. Kroon
	Prof.dr. G.H. van Bruggen
	Prof.dr. F. von Wangenheim
Copromoter:	Dr. R. Kuik

Erasmus Research Institute of Management - ERIM
The joint research institute of the Rotterdam School of Management (RSM)
and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam
Internet: http://www.erim.eur.nl

ERIM Electronic Series Portal: http://hdl.handle.net/1765/1

ERIM PhD Series in Research in Management, 316
ERIM reference number: EPS-2014-316-LIS
ISBN 978-90-5892-373-8

Design: B&T Ontwerp en advies www.b-en-t.nl and Les Nouveaux Voisins www.lesnouveauxvoisins.fr

This publication (cover and interior) is printed by haveka.nl on recycled paper, Revive®.
The ink used is produced from renewable resources and alcohol free fountain solution.
Certifications for the paper and the printing production process: Recycle, EU Flower, FSC, ISO14001.
More info: http://www.haveka.nl/greening

To my parents Nevin Zehra and Ahmet

Acknowledgments

Approaching the end of my PhD journey presents me the opportunity to thank those who have contributed to my dissertation and supported me throughout the process.

First and foremost, I would like to express my gratitude to my supervisors Roelof Kuik, Dennis Fok and Steef van de Velde. Even though neither of you were directly involved in my hiring process, I am truly grateful for your acceptance of supervising me at different stages of my PhD research. Roelof, thank you for your guidance and inspiration over the years. Your scientific curiosity is fascinating. Each time I decided to go into new and unfamiliar territory, you encouraged me enormously by having even more enthusiasm than I had. I believe one can conduct research in any field with your mentorship. I am thankful that you were there every time I needed your advice. I hope to have captured your scientific attitude over the years that I worked with you. Dennis, your involvement in my PhD research was without a doubt a milestone. Thank you for helping me shape up my work, providing clear guidelines and guiding me through Bayesian Econometrics. I look up to you as an academic role model with great research skills. Your professionalism has been exemplary, which I am determined to remember in my future path in research. I am grateful to you for accepting to be my co-author and then my promoter. Steef, your contribution came at a pivotal point in my PhD career. Thank you for your interest in my work, and your support to continue my PhD research which lead to the finalizing of this dissertation.

I am also grateful to the late Jo van Nunen and to Moritz Fleischmann who gave me the opportunity to pursue a PhD at Erasmus University Rotterdam. I wish we could have worked longer together. Moritz, I appreciate the warm hospitality when visiting your group in Mannheim and thank you for being a member of my doctoral committee.

Part of the research described in this dissertation has been carried out at Johns Hopkins University Carey Business School where I was a visiting scholar for a three month period. I am grateful to both Özge Şahin and Ricard Gil for their collaboration and hospitality. It has been a pleasure to work with you, and I really enjoyed my time in Baltimore.

I am indebted to the members of my inner committee, Leo Kroon, Florian von Wangenheim and Gerrit van Bruggen, for their valuable time to evaluate this thesis.

I am blessed to have so many wonderful friends who have supported me enormously while I was experiencing ups and downs of various sizes during my PhD studies. I would like to express my gratitude to my colleagues in the department, in particular my office mates Gabor, Lars, Paul, Irina and Shuguang. Gabor and Lars, I am so happy that we have built a long-lasting friendship. Paul, thank you for accepting to stand by my side as one of my paranymphs. I also feel very privileged to get to know Başak, Tim, Melek, Barış, Thomas, Nüfer, Anne, Nalan and Nilüfer. Thank you for being there to get a coffee or a beer any time a nice chat was needed.

Furthermore, I want to thank my friends from the other side of the campus. Even though I stepped away from Operations Research a longtime ago, I am very happy to keep in close contact with you Remy, Kristiaan, Zahra, Ilse, Judith, Twan, Willem and Mathijn. I have always enjoyed the fun times we had after work and also playing basketball against and with you (I must confess the latter was more enjoyable ;)).

I'd also like to express a special thank you to Ingrid and Cheryl for always being an amazing support in every possible way. I also appreciate all the help from Tineke, Kim, Miho and Natalia at the ERIM office.

I'd like to thank my friends Esra, Hasene, Fatma, Nico and Willem for their precious friendship. Hasene, it has been great to have you in my life for more than 20 years now. Esra, your creativity and enthusiasm always inspires me. Thank you for accepting to be my paranymph and coming to my defense from wherever you will be on the surface of the earth. Fatma, thank you for the photo of '*the bridge*' that you took for the cover of my dissertation. Nico, I appreciate the time you spent to design the cover. Willem and Fiona, we should definitely "Machu Pichu"!

My long lasting friends Gülçin and Ozan, such a great coincidence that we ran into each other after all those years in Istanbul on a cloudy Rotterdam day which turned out to be a milestone for my years here in the Netherlands. Both of you, Aslin, Meliha, Arnold, Özden, Peter, Melike, Stefaan, Peren, Zeyda and Pascal, you provide warmth of an endearing friendship. I enjoyed every single moment that I have spent with you; and you always make me feel at home.

I also want to thank my friends Sandrine, Togay, Gaia, Lukas, Juho, Laura, Federica and Pablo whom I met in Eindhoven during the Erasmus exchange program. If I had not had such a great time with you back then, I might have not considered coming back to the Netherlands. Therefore, you deserve a big thanks for indirectly contributing to this thesis. I also feel very privileged to enjoy our '*traditional trips*' to Budapest, Vienna, Lisbon and Stockholm. This thesis is the very reason of missing the trips to Malta, Tenerife, Barcelona, Rome and Marrakech which I will never be able to stop regretting.

A big thanks goes to my awesome volleyball team of Esther, Sander, Nina, Arnold, Dennis, Rianne and Sebastian. You cannot imagine how much you contributed to this thesis. Our playing together provided moments to all of a sudden solve quite some research related problems. You might have questioned my sanity if you caught any of those moments though :)

I am not only blessed with great colleagues and friends, but also with a great family. I would like to thank my family for being all one could ever wish for, and for their unconditional support, endless trust and love. Thank you Ceren, my one and only sister, for adding so much value and joy into my life. Thank you mom and dad for knowing when to hold me close and when to let me go. Sevgili anneciğim, babacığım ve biricik kardeşim, iyi ki varsınız, sizi çok seviyorum!

Even though I have encountered a couple of unfortunate moments in the beginning of my PhD trajectory, I count myself a lucky person with all the offerings of life.[1] But, if I need to name one, the luckiest moment, that would be the moment that I met you Tom. Thank you for being an amazing friend, a big inspiration, a great source of humor and laughter, and my love.

Evşen Korkmaz
Amsterdam, July 2014

[1]How one cannot be by being from the happiest city of Turkey (tuik.gov.tr, 2013) and living in the happiest country of the world (earth.columbia.edu, 2013).

Table of Contents

Chapter 1

Introduction

The success of many businesses lies within their ability to understand their customers. Insights in customer behavior as well as customer heterogeneity allow companies to improve their business strategy. The primary goal of this dissertation is to provide companies with tools to improve this understanding, and thus to support managerial decision making.

This chapter is organized as follows; in the next section, we expand on the reasons behind our research. In Section 1.2, we present the scope of our research along with the research goals. The particular research questions that we address in this dissertation are presented in Section 1.3. Section 1.4 presents the outline of this dissertation with a discussion of its contributions as well as a description of the research methodologies employed.

1.1 Motivation

Companies may not readily have relevant information they need on their customers' purchase behavior. Some of the challenging questions that they constantly deal with are whether customers will continue buying the company's service or its products; how much, how often and when they will buy; what their willingness to pay is for a specific part of the service offering. The more the company knows about customer purchase behavior, the better equipped it is to gain or retain customers, and to excel in its business. Therefore, understanding customer behavior in order to predict and consequently to steer future behavior is a never-ending challenge not only for companies, but also for academic researchers from various backgrounds.

The scientific process of transforming data into insight for making better decisions, further referred to as *analytics*,[1] provides one of the best tools to help companies to understand customer behavior. We see analytics as the assortment of analytic modeling techniques that enable firms to predict customer behavior; to understand customer heterogeneity; to develop business metrics that help to evaluate the success of marketing and operations effectiveness; and to transform these insights into business strategies.

Analytics is becoming more prominent for companies for three main reasons. First of all, data is growing exponentially, not only in size but also in the variety of its sources. New technologies enable collecting more data than ever before, yet many companies are still looking for ways to obtain value from their data. Therefore, companies adopt analytics to exploit their growing data potential to get smarter and more innovative.

Secondly, there have been advances in quantitative modeling techniques such as recent developments in econometric methods and also increases in computational resources, allowing the estimation of large-scale Bayesian and simulation-based algorithms. All these advances bring new opportunities, new ways to thoroughly analyze data; and thus, to better understand customer behavior.

Third and foremost, companies start to realize that analytics can improve their business performance not only by increasing their direct revenues but also by creating a longer term relationship with their customers. In a survey[2] of nearly 3,000 executives, managers and analysts working across more than 30 industries and 100 countries, half of the respondents said that improving information systems and adopting advanced quantitative models are top priorities for their organizations. Another striking result is that top-performing organizations use analytics five times more than lower performers. Overall, this survey underpins the widespread belief that analytics can offer value to companies. Therefore, managers increasingly adopt analytics to enhance their business performance.

Until recently analytics have mostly been associated with quantitative marketing techniques. However, companies have recognized that most business functions can be improved with this data-driven approach. We observe that companies have started to apply a so-called *enterprise analytics approach*. For example, UPS, which is counted among the world's most rigorous practitioners of Operations Research and Industrial Engineering with its sophisticated operations planning, extends its quantitative techniques to other business functions. Today, UPS applies quantitative techniques to anticipate and influence the actions of customers. The company currently predicts customer defection by

[1] This definition is provided by the Institute for Operations Research, the Management Sciences and Analytics [INFORMS].

[2] This survey was conducted by MIT Sloan Management Review partnered with the IBM Institute for Business Value (LaValle et al., 2014).

examining usage patterns and customer complaints so that they can manage their operations more effectively (Davenport, 2006). Similarly, Procter & Gamble recently created an analytics group consisting of more than 100 analysts from different functions including operations, supply chain, sales, consumer research and marketing (Davenport, 2006).

The holistic data-driven approach also influences various scientific disciplines such as Operations Management [OM] and Operations Research [OR].[3] For already a long time, many researchers have pointed out that OM has drifted far away from its original empirical source (Fisher (2007), Agatz (2009)). Therefore, re-introducing a more data-driven approach is very valuable in the field. Revenue Management [RM], being an OR methodology, represents a good example where a data-driven approach can increase its effectiveness. RM aims to maximize revenue by either pricing and/or inventory allocation decisions under constrained supply capacity conditions. It involves managing the firm's interface with the market; therefore, an understanding in customer behavior and customer heterogeneity should be seen as the core of an RM system. However, current RM systems use only a fraction of the relevant data which is made available by today's information technology systems (van Ryzin, 2005). Only recently we see that advanced demand methods have been introduced in RM to specify the probability of purchasing products and the expected timing of purchases. However, applications of these advanced demand models to different industries (such as online retail industry) are limited, and heterogeneity in choice behavior across different customers has mostly been ignored (Cirillo and Hetrakul, 2011).

This dissertation acknowledges the necessity of empirical input into inter-disciplinary business research. Therefore, in each chapter of the dissertation, customer transaction data forms the basis for building models and expanding on theories. The results aim to complement other decision-making tools, such as an RM system, by providing better input from customer behavior and customer heterogeneity.

1.2 Scope and Research Goals

Throughout this dissertation, we examine the repeat-purchase behavior of customers in a retail context. Aside from simulated data for model testing, we use real data from companies operating in the retail industry. Even though most of our data comes from online retailers, brick-and-mortar retailers which apply customer loyalty programs can

[3]Following these trends in both business and research, INFORMS, the largest society in the world for professionals in the field of Operations Research and Management Science, expands into analytics which confirms the close relationship between OR and analytics.

also make use of our ideas and models. The companies that we obtain data from are grocery and CD retailers. Our contributions, however, can be extended to many other sectors, including digitalized products such as software, music and movies.

We particularly focus on the so-called *non-contractual* setting. Under such a setting, there is no tying contract for customers to continue buying from the company. In other words, customers are free to leave the company at any point in time without notifying the company. The unobserved defection time complicates the understanding and predicting the customer's repeat purchase behavior.

Our research is mainly situated in the fields of probabilistic customer base analysis modeling and two-part pricing schemes. Probabilistic customer base analysis literature deals with the understanding and predicting the customer behavior in a repeat purchase environment. In particular, we focus on Buy-Till-You-Defect [BTYD] modeling stream that concentrates on the non-contractual setting. These models specify a customer's transaction and defection processes. Even though these advanced models provide detailed predictions on both defection and purchase behavior on the individual level, it has been observed that not many of them have found their way into business practice (Wübben and Wangenheim, 2008).

Improved understanding in customer behavior should form the basis of decision-making that guides day-to-day operations and future strategies of companies. One of the most important decisions that a firm makes is pricing its services and products. Based on the insights we obtain from BTYD models, we next focus on pricing strategies for online retailers who derive their revenues from delivery fees and grocery sales. Therefore, they need to set prices on these two complementary services and products. In order to come up with optimal pricing strategies, we use the ideas from the two-part tariff pricing literature. Despite the detailed theoretical predictions and demand conditions discussed under this literature, there is a limited empirical work that checks whether these demand conditions are met in practice (Gil and Hartmann, 2009).

Given the time and costs associated with implementing advanced models and theories in managerial practice, the marketing executives need to be convinced by a clear demonstration of their contributions. The primary goal of our work is to bridge the gap between advanced models/theories and their business applications. To realize this goal, we first aim to broaden the scope of BTYD models and extend their uses. We believe that if marketing executives are clearly shown that they can obtain predictions on additional metrics and acquire more insights in customer behavior as well as in customer heterogeneity by applying BTYD models, the diffusion of such models in business will accelerate. In a similar vein, in order to bridge the gap between two-part tariff theory and its business practice, we target to extend its scope to a repeat-purchase environment. Moreover, we aim to develop empirical tests of two-part tariff that can be applied to transaction data in an online retail context.

To reach the primary goal of this dissertation, our research objectives are structured as follows:

- To improve the understanding of customers behavior by providing a deeper insight in the mechanics of their purchases;
- To extend the understanding in customer heterogeneity, especially in the hidden drivers of customer's purchase behavior;
- To better predict customer behavior;
- To propose pricing schemes that build on the customer insights.

1.3 Research Questions

This dissertation provides relevant practical and technical insights to support decision making by developing new quantitative models. In particular, these models aim at providing insights in customer's purchase behavior and customer heterogeneity, and guiding optimal pricing strategies. There is a vast body of literature on quantitative modeling in Marketing as well as in Economics and Operations Management. Even though the models developed in this thesis can be seen as quantitative marketing models, there are strong overlaps with Microeconomics as well as with RM literature.

To achieve the goals of this dissertation, we organize our research around four main research questions. The first step is to get a comprehensive overview of the BTYD models. Therefore, we analyze current practice and relevant literature on these models to answer the following research questions.

RQ1: Which of the state-of-the-art BTYD models perform better in predicting customer behavior under a non-contractual setting? How do we capture the differences on predictive results from these models?

These questions focus on providing an extensive validation and comparison study to guide managers on model choice. Once we applied the most established BTYD models on different data sets, we are able to identify relevant extension points for such models.

We identified two major extension points for BTYD models both of which aim at broadening the use of these models. The first extension concerns increasing the predictive output. These models are built upon sophisticated stochastic arrival processes on the individual customer level. Based on this, we can obtain another metric which was not associated with these models before. More specifically, we ask the following research question.

RQ2: Can BTYD models be used to predict the timing of next purchase for individuals?

The answer of this question is especially important because, to our knowledge, there is not a specific modeling stream that addresses the purchase timing prediction problem under a non-contractual setting. The timing predictions can directly be incorporated into a promotion planning or an RM model.

The second extension concerns modeling customer heterogeneity in a more flexible way. This leads us to the next research question.

RQ3: Does a heterogeneity distribution on the customer base that accommodates multimodality (customer segments) lead to a more extensive explanatory power as well as a better predictive performance for BTYD models?

To this end, we build a new BTYD model which extends the use of such models by further providing a customer segmentation. The results from this model help to transform customer insights into marketing actions. Applying the newly proposed BTYD model on a dataset from an online retailer, we identified customer segments with different purchase behavior. This finding leads us to the following research question.

RQ4: How do we link customer purchase behavior and heterogeneity insights to firms' marketing strategies such as pricing?

The aim of this step is to prescribe the way to deal with pricing of delivery fees for online retailers.

1.4 Overview of the Dissertation - Methodology and Contributions

The chapters of this dissertation are self-contained and can thus be read independently. According to the research question, different methodologies are employed in separate chapters. After this introductory chapter, Chapter 2 provides an in-depth study of BTYD models with an extensive comparison and validation study among the most established models in the field. In the same chapter, we show that with BTYD models one can also predict the timing of purchases on the individual customer level. In Chapter 3, we present a new BTYD model that provides additional insights on customer behavior and heterogeneity. Chapters 2 and 3 both rely on Bayesian hierarchical approaches. The former chapter also utilizes stochastic modeling techniques in order to provide analytical derivations on the newly proposed timing metric. The latter chapter expands on the BTYD models by employing a finite mixture probabilistic approach to analyze parameter heterogeneity within and across hidden segments in the customer base. Chapter 4 builds on the customer heterogeneity insights and presents a new model that aims to provide optimal pricing strategies for online retailers. This chapter differs from the previous chapters in terms of the literature stream that it fits in as well as the methodologies employed. Chapter 4 relies on microeconomic modeling techniques and non-Bayesian empirical methodologies. It, however, is linked to the previous chapter in terms of exploiting segments that we have identified in the customer base of an online retailer. Although each chapter concludes with its contribution to their respective areas of research, a concluding chapter summarizes the overall contributions that this dissertation makes to modeling customer behavior and customer heterogeneity literature. In the remainder of this chapter, we introduce the chapters in detail and summarize their contributions.

Chapter 2 addresses *RQ1* and *RQ2*. In this chapter, we provide a new way to validate and compare BTYD models. These models are typically used to identify active customers in a company's customer base and to predict the number of purchases. Surprisingly, the literature shows that models with quite different assumptions tend to have similar predictive performance.

We show that BTYD models can also be used to predict the timing of the next purchase for each customer (*RQ2*). Such timing predictions have a clear managerial purpose. To give an example, consider an online retailer implementing micro-marketing strategies. The most appropriate time to contact its customers depends on their expected timing of the next purchase. High quality timing predictions may contribute to achieving the full potential of micro-marketing (Zhang and Krishnamurthi, 2004). Likewise, online retailers may use purchase timing predictions to improve their operations planning. For example, they can use these predictions as input for RM models.[4] Given that even crude efforts aiming at understanding customer demand can have a significant impact on RM applications (Bell and Chen, 2006), detailed predictions on purchase timing as well as on purchase value have a big potential to increase the effectiveness of RM applications. Using these predictions, the operations managers can prioritize valued customers for highly demanded delivery time slots (Talluri and Van Ryzin, 2005). In summary, we believe that the ability to predict the timing of future transactions accelerates research in various fields such

[4]Online retailers are in a unique position to apply RM as they have (1) a heterogeneous customer base (2) flexibility to tailor the product proposition to their customers, (3) limited delivery capacity at a given time.

as promotion calendar, pricing and capacity allocation decisions under a noncontractual setting.

Moreover, the predictive performance on the purchase timing can be informative on the relative quality of BTYD models (*RQ1*). For each of the established models, we discuss the prediction of the purchase timing. Next, we compare these models across three datasets on the predictive performance on purchase timing as well as purchase frequency. We show that while the Pareto/NBD model (Schmittlein et al., 1987) and its Hierarchical Bayes [HB] extension (Abe, 2009a) perform the best in predicting transaction frequency, the PDO (Jerath et al., 2011) and HB models predict transaction timing more accurately. Furthermore, we find that differences in a model's predictive performance across datasets can be explained by the correlation between behavioral parameters and the proportion of customers without repeat purchases. Chapter 2 is joint work with Roelof Kuik and Dennis Fok (see Korkmaz et al. (2013)).

Even though BTYD models tend to perform well in predicting transaction frequency, amount and timing of individual customers as well as customer lifetime, they sometimes predict extremely long lifetimes for a substantial fraction of the customer base. This obvious lack of face validity limits the adoption of these models by practitioners. Moreover, it highlights a flaw in these models. In **Chapter 3**, based on a simulation study and an empirical analysis of different datasets, we argue that such long lifetime predictions can result from the existence of multiple segments in the customer base. In most cases there are at least two segments: one consisting of customers who purchase the service or product only a few times and the other of those who are frequent purchasers. Customer heterogeneity modeling in the current BTYD models is insufficient to account for such segments, thereby producing unrealistic lifetime predictions.

We extend the current BTYD models by incorporating segments within the customer base. This not only solves the extreme lifetime prediction problem, but also leads to a more insightful description of the customer base. More specifically, we consider a mixture of log-normals distribution to capture the heterogeneity across customers. The proposed model allows us to relate segment membership and within segment customer heterogeneity to observed customer characteristics. Our model, therefore, increases the descriptive power of BTYD models to a great extent (*RQ3*). We are able to evaluate the impact of customers' characteristics on the membership probabilities of different segments. This allows one to, for example, a-priori predict which customers are likely to become frequent purchasers.

The proposed model is compared against the benchmark Pareto/NBD model and its HB extension on simulated datasets as well as on a real dataset from a large grocery e-retailer in a Western European country. Our BTYD model indeed provides a useful customer segmentation that allows managers to draw conclusions on how customers' purchase and defection behavior are associated with their shopping characteristics such as basket size and the delivery fee paid. Chapter 3 is joint work with Dennis Fok and Roelof Kuik (see Korkmaz et al. (2014)).

In Chapters 2 and 3, next to probabilistic modeling techniques, we mostly rely on Bayesian hierarchical approaches as they enable us to produce not only aggregate level estimates but also individual or unit-level parameter estimates. This is very important as today's marketing practices should be designed to respond to consumer differences. Moreover, optimal decision-making requires not only point estimates of unit-level parameters but also a characterization of the uncertainty in these estimates (Rossi et al., 2005). Bayesian hierarchical approaches are ideal as it is possible to produce posterior distributions for a large number of unit-level parameters. Last but not least, estimation of the complex models in these chapters are relatively straightforward using simulation based Bayesian methods. Chapter 3 additionally exploits finite mixture probabilistic approach in order to provide more flexibility on heterogeneity modeling in BTYD models. Mixture modeling approach can be very useful in defining hidden segments in the data.

In **Chapter 4**, we address *RQ4*. Building on the customer insights that we have gained in Chapter 3, we focus on pricing of two complementary products in an online retail setting. Online grocery retailers derive their revenue and profits from delivery fees and grocery sales. The retailer may consider selling goods at a discount but make up for its revenue loss with high shipping fees or vice versa. We base our optimal pricing discussion upon the well-grounded two-part pricing literature. We adapt the theoretical framework of Schmalensee (1981) and take repeat purchase occasions into consideration that create a substitution effect between number of visits and consumption per visit, following Phillips and Battalio (1983). We derive testable implications regarding changes in the price of deliveries (access/primary good) and revenues from goods (secondary good). We take these predictions to the data using a dataset detailing transaction information from an online grocery retailer in a Western European country.

One of our main findings shows that there is a positive relationship in the data between the number of transactions and the average size of grocery baskets purchased. We also observe two very different customer groups in our data with different willingness to pay. This observation together with robust evidence that price-sensitive customers buy larger baskets is consistent with an optimal pricing strategy that offers discounts for the business customers and charges higher prices to the households for the primary good.

We conclude that firms may increase profits by implementing alternative and simpler price discrimination strategies by combining second and third-degree price discrimination schemes. Chapter 4 is joint work with Özge Şahin and Ricard Gil.

Chapter 5 is the last chapter of this dissertation in which we summarize our findings and give our concluding remarks.

Chapter 2

"Counting Your Customers": When will they buy next?

2.1 Introduction

Many firms routinely store data on customer transactions. However, processing this data in order to provide managerially relevant information can still be a challenge. The customer base analysis literature provides a number of methods to use such data to gain a good understanding of the customer's transaction behavior. In the literature, a distinction is made between a contractual and a noncontractual setting. The latter is especially challenging as one does not observe the moment at which a customer leaves the company. In this setting, it is interesting to predict the number of future purchases, and to infer from observed behavior whether a customer has already left the company. A wide variety of models is available for these purposes.

The online retail industry is an important example of an industry operating in a noncontractual setting. Retailers never know which customers are *active*, or in other words, which customers will continue buying from the firm. Thus, the customer database of an online retailer is likely to contain many inactive customers. For example, in October 2005, eBAY reported 168 million registered customers but only 68 million of them were counted as active by the company (Gupta et al., 2006). It is, therefore, very useful to develop a method to identify active customers under a noncontractual setting.

It has been widely recognized in the literature that models that ignore defection, like the early NBD model by Ehrenberg (1988), do not provide good predictions for this type of industry. They generally overestimate future transaction frequencies (Schmittlein and Peterson, 1994). Schmittlein et al. (1987) proposed one of the first models that does account for defection. Since then, there has been a strong focus on the so-called buy-till-you-defect [BTYD] model. Several extensions of the model by Schmittlein et al. (1987) have been introduced (Fader et al. (2005a), Abe (2009a) and Jerath et al. (2011)). Some of these models have also been used to generate managerially relevant insights (Reinartz and Kumar (2000), Reinartz and Kumar (2003), and Wübben and Wangenheim (2008)). However, little attention has been paid to providing a rigorous empirical comparison of the growing number of BTYD models. The models have mainly been compared on their performance in predicting a customer's number of purchases in a time interval.

In this paper, we suggest to include another measure in the comparison, namely the timing of the purchases. The existing models mainly differ in the distribution that governs the defection process. However, differences in the shape of this distribution may not directly lead to substantial differences in the expected number of purchases. Other measures, such as the customer being active at the end of the observation interval, directly involve the (unobserved) time of defection. If we want to use such measures for validation, we require additional assumptions or heuristics. The timing of the purchase is, however, observed and critically depends on the interplay between the transaction and defection processes. Yet, predicting the timing of the next purchase is not straightforward. We develop methods for all state-of-the-art BTYD models. Based on these predictions, we provide an extensive empirical validation and comparison of these models where we go beyond the typical comparison that mainly considers only purchase frequency.

We present the in-sample and out-of-sample performance on predicting the transaction frequency as well as the transaction timing of each customer for three datasets. The first dataset is from an online grocer in a Western European country. The second is the well-known CDNOW dataset which has been commonly used as a benchmark set. The third dataset is also used by Batislam et al. (2007), and Jerath et al. (2011) and is from a Turkish grocery retailer.

Our results show that different models can lead to different predictions on timing and frequency. It is important to understand how the underlying behavioral assumptions of the models lead to differences in performance. It turns out that certain data characteristics such as the correlation between behavioral parameters favor use of certain models.

The remainder of this chapter is structured as follows. Section 2.2 gives an overview of the existing literature on BTYD models. We discuss the main features of and differences across the models, and present our contribution in more detail. In Section 2.3, we provide technical details of the considered models and present new results that deal with the timing of transactions. Section 2.4 gives a detailed description of the datasets. After presenting results of the empirical study in Section 2.5, general conclusions are discussed in Section 2.6.

2.2 BTYD Models

In this section, we briefly review the main ideas underlying the BTYD models. We also discuss the similarities and differences across the most established BTYD models. Next, we review earlier empirical validation studies. Table 2.1 gives a summary of the related empirical work. We omit from this table studies that employ the Pareto/NBD model without testing its predictive performance in a holdout period (Reinartz and Kumar (2000), Reinartz and Kumar (2003) and Wu and Chen (2000)). Finally, we discuss lifetime estimation using these models.

2.2.1 Models in Comparison

The Pareto/Negative Binomial Distribution [Pareto/NBD] model (Schmittlein et al., 1987) is one of the first models that considers the customer defection. This model assumes that, while alive, customers make purchases according to a Poisson process with heterogeneous rates. The lifetime of a customer is modeled using an exponential distribution, also with a heterogeneous rate. The individual-specific rates of both processes are next treated as random effects and modeled using independent gamma distributions. This model allows for individual-level calculations on the probability of being active and the number of future purchases. The structure of the model leads to closed-form expressions for such predictions given the (hyper)parameters of the heterogeneity distributions. This feature has made this model useful for today's personalized marketing concepts such as direct marketing, one-to-one marketing and customer relationship management.

Three important extensions of the Pareto/NBD model have been introduced in the literature. Fader et al. (2005a) suggested replacing the continuous time defection process by a discrete time process. After each purchase, the customer defects with an individual-specific probability. The resulting model is called a Beta-Geometric/Negative Binomial Distribution [BG/NBD] model. The disadvantage of this model is that frequent purchasers have more "opportunities" to defect. In some cases this may not correspond to reality. To solve this problem, Jerath et al. (2011) introduced the Periodic-Death-Opportunity [PDO] model. This model is very similar to the BG/NBD, but defection opportunities are defined in calendar time. In other words, defection can only occur at certain time intervals, independent of the transaction timing.

Another extension of the Pareto/NBD model deals with the relation between the purchase rate and the defection rate. In the Pareto/NBD model, and in the above-mentioned extensions, the behavioral rates are assumed to be independent. In practice, this assumption may be violated as, for example, frequent shoppers tend to have a longer lifetime. This would imply a negative correlation between both rates. Abe (2009a) recently suggested a Hierarchical Bayes extension of the Pareto/NBD model that incorporates such correlation. In this model, the two gamma distributions are replaced by a bivariate lognormal distribution. Next to the possibility to capture correlations, another advantage of this model is that individual-specific covariates can be used. A disadvantage of this extension is that for some quantities, closed-form expressions are no longer available. As a result, the proposed model by Abe (2009a) needs Bayesian (simulation) techniques. We will refer to this model as the HB model.

2.2.2 Model Performance

The first empirical validation study in the field, which reports the predictive performance of a BTYD model in a holdout period, is presented by Schmittlein and Peterson (1994). This study not only provides an extensive empirical validation of the Pareto/NBD model, but also extends the model by adding the customer's spending decision. A major contribution of this paper is that it provides insights into the sampling properties of parameter estimates. For instance, the authors show how the accuracy of parameter estimation depends on the average observation time and on the number of customers in the sample (the space/time trade-off). Schmittlein and Peterson (1994) also examine whether customer characteristics can help in predicting transaction and defection behavior. In an application in the business-to-business market, they show that some groups of customers tend to have higher transaction rates while others have higher average dropout rates or a greater variation in dropout rates.

Fader et al. (2005a) also include a validation study. This study compares the performance of the BG/NBD and the Pareto/NBD models on a dataset from the online CD retailer CDNOW. They show that replacing the exponential dropout process (of Pareto/NBD) with a geometric one (BG/NBD) improves the model fit in the calibration period. The Pareto/NBD model, however, performs slightly better than the BG/NBD based on the quality of predictions of individual-level transactions in the forecast period. Fader et al. (2005a) argue that the BG/NBD model is a good alternative for the Pareto/NBD model as it has similar performance, but requires fewer resources for parameter estimation.

In a third study, Batislam et al. (2007) compare the Pareto/NBD and BG/NBD models in terms of predicting the future number of transactions and the accuracy of the probability of being active. The comparison is based on loyalty card data from a specific store of a large grocery chain in Turkey. The authors also present a slight variation on the BG/NBD model. In this modified BG/NBD [MBG/NBD] model, customers may also drop out at time zero that is directly after making their first purchase. The MBG/NBD model yields almost identical estimates for the expected number of repeat purchases to the BG/NBD model. The general conclusion is that both the Pareto/NBD and the MBG/NBD models show similar performance on customer's purchase and defection processes.

Wübben and Wangenheim (2008) compare the Pareto/NBD and the BG/NBD models against managerial heuristics. In general, these heuristics are easy to implement, but are less detailed in terms of their predictions. Wübben and Wangenheim (2008) focus on predicting the number of future transactions and classifying active versus inactive customers. In terms of this classification, the managerial heuristics perform at least as well as the models. However, the models perform better than the heuristics when predicting future transactions numbers. In this paper, the authors identify a potentially important problem of the BTYD models. On some datasets, the models produce extremely high probabilities of being active. Such high probabilities correspond to extremely long (residual) lifetime estimates.

Abe (2009a) compares his HB model to the Pareto/NBD model. He finds a similar fit and predictive performance. The disaggregate fit measures are the Mean Squared Error [MSE] of the predicted transaction numbers of individual customers, and the correlation between these predictions and the corresponding realizations. With regard to predicting future transaction numbers, the HB model performs slightly better than the Pareto/NBD model on two of the three datasets. The covariance matrix of the heterogeneity distribution is used to test the independence assumption of the Pareto/NBD. No significant dependency is found for any of the three datasets.

Finally, Jerath et al. (2011) compare their PDO model to the Pareto/NBD and BG/NBD models using two datasets. They pay more attention to the defection process, and check model's performance on the median of lifetime estimates for each model. Note that the median lifetime is considered here, not the mean lifetime. Previous research has shown that the former is a better descriptor of the lifetime distribution (Reinartz and Kumar, 2000) as using the median results in less extreme lifetime predictions. At a first glance, the Pareto/NBD and the PDO models produce similar results on the median lifetime. However, the PDO model predicts longer lifetimes for a randomly chosen customer than the Pareto/NBD model. The BG/NBD model's estimates are very different in that it predicts extremely long lifetimes. Based on these results, the authors suggest that the modeling of the defection process needs to be improved. Jerath et al. (2011) also compare the models with respect to their predictions of the number of transactions. The Pareto/NBD and the PDO models show similar predictive performance and generally outperform the BG/NBD model.

2.2.3 Lifetime Estimation

The BTYD models are usually compared on two dimensions: transaction frequency and lifetime related measures. Mostly, the first dimension is emphasized. An important challenge with the second dimension is that the exact lifetime is never observed. Even the state of a customer (active or inactive) can never be perfectly measured. There have been many attempts to validate predictions on customer lifetime or the active/inactive state. However, the majority of these studies acknowledge that the used indicators are not perfect.

Schmittlein and Peterson (1994) use telephone interviews to validate customer defection predictions. Customers are called and asked about their intentions to purchase from the company at an unspecified time in the future. However, even such a direct contact with a customer may not lead to the 'actual' defection information. It is known that customer's intentions are imperfect predictors of future behavior (Morwitz and Schmittlein, 1992).

Batislam et al. (2007), Reinartz and Kumar (2000) and Wübben and Wangenheim (2008) base the 'true' active status of a customer on observed purchase activity in a holdout period. The model's predictive performance in terms of the defection process is next evaluated on this active status. However, as acknowledged by Wübben and Wangenheim (2008), customers who have not purchased in the holdout period may still be active and make a purchase after that period. In this sense, such a comparison is not fair and leads to favoring models that underestimate the lifetime. This is especially true, if the holdout period is short and/or the purchase rate is low.

Apart from the complexity of validating lifetime predictions, the managerial relevance of the lifetime concept has also been questioned. Reinartz and Kumar (2000) challenge the implicitly assumed strong association between lifetime and profitability in the noncontractual setting. Contrary to the general claim that a long customer lifetime is always desirable, they find that revenues mainly drive the lifetime value of a customer, not the duration of customer tenure. This argument is particularly valid in industries where customer switching costs are small (Reinartz and Kumar, 2000). Furthermore, Jerath et al.

Table 2.1: Literature on the empirical validation and comparison of BTYD models

Paper	Model(s)	Dataset(s)	Measures/Metrics	Results + Insights + Notes
Schmittlein, Peterson, (1994)	Pareto/NBD	1 (a B2B office products supplier)	•Individual and aggregate level # future transaction •Customer's active/inactive status •Dollar volume of transactions	•Dollar volume of transactions is added to model. •Customer's actual active status is designated by telephone interviews and significant evidence on model's ability to distinguish active customers is found. •Sampling properties are added; # customers and observed time (T) tradeoff. •Dropout process is validated by comparing to NBD model. •Pareto/NBD performs better in predicting future transaction # than a simple heuristic.
Fader, Hardie, Lee, (2005)	Pareto/NBD - BG/NBD	1 (CDNOW)	•Chi-square goodness-of-fit test •Individual and aggregate level # future transaction	•The transition from exponential to geometric distribution improves model fit performance (without a significant loss in prediction power). •BG/NBD is a good alternative to Pareto/NBD requiring fewer resources for parameter estimation.
Batislam, Denizel, Filiztekin, (2007)	Pareto/NBD - BG/NBD - MBG/NBD	1 (a store of a large grocery retail chain)	•Chi-square goodness-of-fit test •Individual and aggregate level # future transaction •Active status of customers by computation of being active probability	•Pareto/NBD and MBG/NBD have similar estimates of # future transactions. •Pareto/NBD model assigns slightly smaller active probabilities.
Wübben, Wangenheim, (2008)	Pareto/NBD - BG/NBD[a]	3 (an apparel retailer, a global airline, CDNOW)	•Individual and aggregate level # future transaction •Active status of customers	•Only Pareto/NBD model is compared against hiatus heuristic to distinguish active customers. Hiatus heuristic performs better. A sensitivity analysis on the threshold gives a similar result. •Both BTYD models outperform heuristics in predicting # transactions
Abe, (2009)	Pareto/NBD - HB	3 (CDNOW, a department store, music CD chain)	•Model fit assessment with correlation, MSE and MAPE •Individual and aggregate level # future transaction	•Pareto/NBD and HB model show similar fit. •The marginal log-likelihood suggests that HB model with covariates is better than without. •HB model performs slightly better than Pareto/NBD model on 2 (out of 3) datasets. •Independence assumption of Pareto/NBD has been examined and no significant dependency has been found.
Jerath, Fader, Hardie, (2011)	Pareto/NBD - BG/NBD - PDO	2 (CDNOW, a store of a large grocery retail chain[b])	•Chi-square goodness-of-fit test •Individual and aggregate level # future transaction •Median lifetime	•PDO shows significant improvement on calibration-period model fit compared to Pareto/NBD. Similar fit performance with BG/NBD model. •Pareto/NBD and PDO models show similar predictive performance on # transactions and outperform BG/NBD model. •Pareto/NBD and PDO show similar results on median lifetime whereas BG/NBD model predicts extreme lifetimes. •PDO model suggests modeling of defection process can be improved.

[a]Simple managerial heuristics are included in the comparison study.
[b]The same dataset from Batislam et al. (2007) is used.

(2011) show that lifetime estimations from various BTYD models can vary to a large extent.

As aforementioned, in some cases, the BTYD models give extremely high active probabilities, which correspond to the extreme lifetime estimations (Wübben and Wangenheim, 2008). Such clearly incorrect predictions could lead to a reluctance to use these models in practice. Perhaps with this in mind, Reinartz and Kumar (2000) strongly suggest firms not to neglect the transaction orientation of their business and to manage the short term accordingly.

2.2.4 Our Contribution

Based on the discussion above, the only theoretically valid measure that is available to compare the BTYD models seems to be the accuracy of the predicted (future) transaction frequency. However, although the existing models are quite different in terms of their specification, they produce similar predictions on this measure. In other words, this measure is not sensitive to differences among the models. In this paper, we introduce a new performance metric for BTYD models to overcome this problem and provide more insight on the relative predictive performance of these models.

Our measure is based on the timing of transactions and represents an observable value. Given the memoryless property on interarrival times of transactions in the considered BTYD models, we can predict the timing of the first and the last transaction in a certain period. As an in-sample metric, we propose the timing of the last in-sample transaction; as a holdout metric, we propose the minimum of the timing of the first out-of-sample transaction and the end of the holdout period.

In this paper, we compare the existing models' predictions on the timing of purchases as well as on the number of purchases. To make this possible, we derive formulas on the timing of transactions for each of the BTYD models. The methodology to calculate these timing predictions is also an important contribution of this paper. Besides providing a more rigorous comparison among BTYD models, these predictions also have managerial relevance. Predictions on the timing of the next purchase for each customer could be important information for both marketing and operations managers.

To our knowledge, our paper is the first to bring all the following models together: the Pareto/NBD, BG/NBD, the Hierarchical Bayes extension of the Pareto/NBD, and the recently proposed PDO model. Next, we are the first to compare these models based on also the timing of purchases. A challenge in this comparison study is that the models exhibit differences in their estimation procedures. The Pareto/NBD, BG/NBD and PDO models have closed-form expressions on some statistics for a 'randomly' chosen customer, such as the probability of being active and the expected number of future purchases. These models also yield closed form expressions for some statistics conditional on the observed transaction pattern of a customer. On the other hand, the HB model does not provide an analytical expression for important quantities due to the log-normal heterogeneity distribution. For this model, there is no closed-form expression for any relevant statistic not even for a randomly chosen customer. However, the complete distribution on any statistic can be obtained for each customer using MCMC methods. In order to overcome the difficulty of comparing the models, we bring the Pareto/NBD, BG/NBD and PDO models to the level of the HB model. More exactly, we obtain the complete individual-level distribution on the behavioral parameters for each model conditional on observed behavior. This provides great flexibility when computing various individual-level performance metrics.

2.3 Timing of Transactions with BTYD Models

In this section, we present the BTYD models in technical terms. All models provide a representation of individual behavior by considering two arrival processes: one on purchase and one on defection. Individuals are assumed to make transactions according to a purchase process until they defect. The defection and transaction processes for individual i depend on individual-specific parameters which we denote by θ_i. On the population-level, all models specify a heterogeneity distribution for (the elements of) θ_i. This distribution is parameterized by hyperparameters which are denoted by ξ. Below, we give the details for each model, and present expressions for the last transaction timing in the calibration period and the first transaction timing in the holdout period. The timing expressions vary depending on the assumptions of the models. To our knowledge, these expressions have not been presented before.

Table 2.2 gives a summary of the assumptions and the dominant estimation method for each model. We distinguish modeling assumptions on individual behavior and on customer heterogeneity. All models have the same assumption on the purchase process of an individual, while active. The models do differ either in the defection process or in the heterogeneity distribution.

Before we present the models, we briefly discuss the general ideas used for calculating the predictions.

Table 2.2: Model comparison with respect to the assumptions and estimation process

	Pareto/NBD	BG/NBD	PDO	HB
Purchase process	Poisson	Poisson	Poisson	Poisson
Defection process	Exponential	Shifted geometric	Shifted geometric	Exponential
Defection timing	Continuous	On purchase moments	Fixed periods	Continuous
Purchase rate distribution	Gamma	Gamma	Gamma	Bi-variate log-normal
Defection rate distribution	Gamma	Beta	Beta	
Estimated parameters	Hyperparameters	Hyperparameters	Hyperparameters	Hyper & individual par.
Estimation procedure	MLE	MLE	MLE	MCMC

2.3.1 Conditional and Unconditional Inference

One can use the BTYD models to obtain predictions on different metrics. However, closed-form expressions for individual-level metrics conditional on the observed data are not always available. Below we indicate how to calculate such metrics. Suppose we want to predict a particular metric for customer i, we denote this as metric$_i$. There are two options: to include or not to include the purchase history of this customer. The latter case is mainly relevant for in-sample predictions (model calibration) and, the prediction can be seen as a prediction for a *randomly* chosen customer. We label this as unconditional inference. The former is relevant for out-of-sample predictions. These predictions are made conditional on data of the specific customer.

For conditional inference, we need to calculate $\mathbb{E}[\text{metric}_i|\text{all data}]$. We rewrite this expectation as

$$\begin{aligned}\mathbb{E}[\text{metric}_i|\text{all data}] &= \int_{\theta_i} \mathbb{E}[\text{metric}_i|\text{data}_i, \theta_i]\pi(\theta_i|\text{all data})\ \mathrm{d}\theta_i \\ &= \int_{\xi}\int_{\theta_i} \mathbb{E}[\text{metric}_i|\text{data}_i, \theta_i]\pi(\theta_i|\text{data}_i, \xi)\pi(\xi|\text{all data})\ \mathrm{d}\theta_i\,\mathrm{d}\xi,\end{aligned} \tag{2.1}$$

where θ_i denotes the individual-level parameters for individual i and ξ denotes the hyperparameters associated with the whole customer base in the focal BTYD model. In Sections 2.3.2 to 2.3.5, we provide closed-form expressions for $\mathbb{E}[\text{metric}_i|\text{data}_i, \theta_i]$ for each model. Calculating the integrals in (2.1) can still be very complex. However, samples from $\pi(\theta_i|\text{all data})$ can be obtained for all models. If the model relies on Maximum Likelihood Estimation [MLE], $\pi(\xi|\text{all data})$ is seen as a point mass at the Maximum Likelihood estimate $\hat{\xi}$, and draws are obtained by sampling from $\pi(\theta_i|\text{data}_i, \hat{\xi})$. For BG/NBD and PDO models, closed-form expressions are available for these conditional densities and we can apply direct sampling. For the other models, draws from the posterior are obtained using a Metropolis-Hastings MCMC sampler (Hastings, 1970). In general, we approximate the integral for all models using

$$\mathbb{E}[\text{metric}_i|\text{all data}] \approx \frac{1}{L}\sum_{l=1}^{L} \mathbb{E}[\text{metric}_i|\text{data}_i, \theta_i^{(l)}],$$

where $\theta_i^{(l)}$, $l = 1, \ldots, L$, are draws from the posterior $\pi(\theta_i|\text{all data})$.

In the case of unconditional inference we need to calculate

$$\begin{aligned}
\mathbb{E}[\text{metric}_i|\text{all data}_{-i}] &= \int_{\theta_i} \mathbb{E}[\text{metric}_i|\theta_i]\pi(\theta_i|\text{all data}_{-i})\ \mathrm{d}\theta_i \\
&= \int_{\xi}\int_{\theta_i} \mathbb{E}[\text{metric}_i|\theta_i]\pi(\theta_i|\xi)\pi(\xi|\text{all data}_{-i})\ \mathrm{d}\theta_i\,\mathrm{d}\xi \\
&\approx \int_{\xi}\int_{\theta_i} \mathbb{E}[\text{metric}_i|\theta_i]\pi(\theta_i|\xi)\pi(\xi|\text{all data})\ \mathrm{d}\theta_i\,\mathrm{d}\xi
\end{aligned} \tag{2.2}$$

where all data_{-i} denotes the available data ignoring the data for individual i. In the last line, we assume that enough data is available such that the contribution of a single individual to the conditional distribution of the hyperparameters can be ignored. In this case we approximate the expectation by

$$\mathbb{E}[\text{metric}_i|\text{all data}_{-i}] \approx \frac{1}{L}\sum_{l=1}^{L} \mathbb{E}[\text{metric}_i|\theta_i^{(l)}].$$

If hyperparameters are estimated using MLE, $\theta_i^{(l)}$ denotes a draw from $\pi(\theta_i|\hat{\xi})$, with $\hat{\xi}$ the Maximum Likelihood estimate. If Bayesian estimation is used, the draws are obtained by first sampling $\xi^{(l)}$ from $\pi(\xi|\text{all data})$ and next sampling $\theta_i^{(l)}$ from $\pi(\theta_i|\xi^{(l)})$.

In the sections below, we present the expressions for the conditional expectation of the timing of the last in-sample transaction and the next out-of-sample transaction together with the sampling schemes for the behavioral parameters.

2.3.2 Pareto/NBD Model

In the Pareto/NBD model, customer i remains active for a stochastic lifetime ($t_{\Delta,i}$) which has an exponential distribution with rate μ_i. While active, this customer makes purchases according to a Poisson process with rate λ_i. The purchase rate and the defection rate are assumed to be distributed according to two independent gamma distributions across the population. The distribution for λ_i has shape parameters r, and scale parameter α. The shape and scale parameters for μ_i are s and β, respectively.

The parameters of the heterogeneity distributions can be estimated by MLE. The likelihood can be written in terms of the number of purchases (x_i) and the timing of the last purchase ($t_{x,i}$) for each customer. This estimation procedure can be quite tedious from a computational perspective as the likelihood function involves numerous evaluations of the Gaussian hypergeometric function.

Schmittlein et al. (1987) presented some key expressions such as the probability of being active at the end of the calibration period (T_i) and the expected number of future transactions in a given time period for both a randomly chosen customer and a customer with past observed data $(x_i, t_{x,i}, T_i)$.

The Pareto/NBD model allows us to predict also the timing of the last transaction in the calibration period and the timing of the first transaction in the holdout period. Given the individual-level parameters λ_i and μ_i, we derive the equation on the expected timing of the last purchase as

$$\mathbb{E}[t_{x,i}|\lambda_i, \mu_i, T_i] = \frac{1 - e^{-\mu_i T_i}}{\mu_i} - \frac{1 - e^{-(\lambda_i+\mu_i)T_i}}{\lambda_i + \mu_i}, \tag{2.3}$$

see Section 2.7.1 for the associated derivations. By comparing $\mathbb{E}[t_{x,i}|\lambda_i, \mu_i, T_i]$, averaged over the estimated distribution of λ_i and μ_i, to the observed timing of the final purchase, we can assess the model's fit performance.

To measure the model's performance on out-of-sample predictions, we can use the timing of the first purchase in the interval $[T_i, T_i^+]$, where T_i^+ marks the end of the out-of-sample period. A complication here is that a particular customer may not make any purchase in this interval. For example, this may happen if the customer has defected. In turn, this makes it extremely difficult to compare the predictions to realizations. We solve this by instead predicting the minimum of the next purchase timing and T_i^+; for individual i this minimum is denoted by $t_{f,i}$. If the customer has defected, $t_{f,i} = T_i^+$.

In Section 2.7.1, we show that the conditional expectation of $t_{f,i}$ in the Pareto/NBD model equals

$$\begin{aligned}\mathbb{E}[t_{f,i}|x_i, t_{x,i}, T_i, \lambda_i, \mu_i] &= (1 - \mathbb{P}[t_{\Delta,i} > T_i|x_i, t_{x,i}, T_i, \lambda_i, \mu_i])T_i^+ \\ &\quad + \mathbb{P}[t_{\Delta,i} > T_i|x_i, t_{x,i}, T_i, \lambda_i, \mu_i]\,(T_i + \frac{1 - e^{-(\lambda_i+\mu_i)(T_i^+ - T_i)}}{\lambda_i + \mu_i}),\end{aligned} \tag{2.4}$$

where $\mathbb{P}[t_{\Delta,i} > T_i | x_i, t_{x,i}, T_i, \lambda_i, \mu_i]$ gives the probability that individual i is still active at time T_i. This probability can be shown to equal

$$\frac{\lambda_i}{\lambda_i + \mu_i e^{(\lambda_i+\mu_i)(T_i - t_{x,i})}}, \tag{2.5}$$

see Schmittlein et al. (1987). Note that this probability depends on the time between the last (in-sample) purchase and T_i. There is still a chance of defection in this period, but, given the data, a purchase is impossible in that interval.

Sampling of the behavioral parameters for the Pareto/NBD Model

The joint posterior distribution of the behavioral parameters, $\theta_i = (\lambda_i, \mu_i)$, of the Pareto/NBD model is characterized by the likelihood function, the independent gamma priors on these parameters, and the (ML estimates of the) hyperparameters, $\xi = (\alpha, r, \beta, s)$:

$$\begin{aligned} \pi(\theta_i | \mathrm{data}_i, \xi) &= \pi(\lambda_i, \mu_i | r, \alpha, s, \beta, x_i, t_{x,i}, T_i) \\ &\propto f(x_i, t_{x,i}, T_i | \lambda_i, \mu_i) g(\lambda_i | r, \alpha) h(\mu_i | s, \beta) \\ &\propto \frac{\lambda_i^{x_i}}{\lambda_i + \mu_i} (\mu_i e^{-(\lambda_i+\mu_i)t_{x,i}} + \lambda_i e^{-(\lambda_i+\mu_i)T_i}) \frac{\alpha^r}{\Gamma(r)} \lambda^{(r-1)} e^{-\alpha\lambda} \frac{\beta^s}{\Gamma(s)} \mu_i^{(s-1)} e^{-\beta\mu_i} \ . \end{aligned} \tag{2.6}$$

As mentioned before, among the models that rely on MLE, the Pareto/NBD model is the only one that does not have a standard distribution of individual parameters, $\pi(\theta_i | \mathrm{data}_i, \xi)$. A Metropolis-Hastings algorithm (see Hastings (1970)) can be used to sample from this posterior density. Details of this sampling algorithm are presented in Section 2.8.

2.3.3 BG/NBD Model

The BG/NBD model replaces the continuous defection process of the Pareto/NBD model by a discrete process. Customers can now only drop out at the moment of a repeat transaction. This implies that the defection process is explicitly dependent on the purchase process.

Jerath et al. (2011) argue that such a dependency may not be realistic, as heavy buyers eventually get more opportunities to drop out. However, the advantage of this model is that its parameters can be estimated more easily. The individual's purchase process is Poisson with intensity $\lambda_i \sim \Gamma(r, \alpha)$ like in the Pareto/NBD model. The dropout probability for individual i is denoted by p_i and follows a beta distribution with shape parameters a and b. The hyperparameters of the BG/NBD model can be estimated using MLE.

Fader et al. (2005a) present the expression for the expected number of (future) transactions of each customer, conditioned upon the hyperparameters. In Section 2.7.2, we derive the expected timing of the last in-sample transaction and the next out-of-sample transaction. Again, we truncate the next future transaction timing to the end of the out-of-sample period (T_i^+). The expected timing of the last in-sample transaction equals

$$\mathbb{E}(t_{x,i}|T_i, \lambda_i, p_i) = \frac{1}{1-p_i}\left(\frac{1-e^{-\lambda_i p_i T_i}}{\lambda_i p_i} - \frac{1-e^{-\lambda_i T_i}}{\lambda_i}\right), \tag{2.7}$$

and the conditional expectation of the timing of the next transaction equals

$$\begin{aligned}\mathbb{E}(t_{f,i}|x_i, t_{x,i}, T_i, \lambda_i, p_i) = {} & (1 - \mathbb{P}[t_{\Delta,i} > T_i|x_i, t_{x,i}, T_i, \lambda_i, \mu_i])T_i^+ \\ & + \mathbb{P}[t_{\Delta,i} > T_i|x_i, t_{x,i}, T_i, \lambda_i, \mu_i](T_i + \frac{1-e^{-\lambda_i(T_i^+ - T_i)}}{\lambda_i}).\end{aligned} \tag{2.8}$$

For this model, the conditional probability of being active at time T_i equals

$$\mathbb{P}[t_{\Delta,i} > T_i|x_i, t_{x,i}, T_i, \lambda_i, \mu_i] = 1 - \delta_{t_{x,i}>0}\frac{p_i e^{\lambda_i(T_i - t_{x,i})}}{1 - p_i + p_i e^{\lambda_i(T_i - t_{x,i})}},$$

where $\delta_{t_{x,i}>0}$ is a 0/1 indicator, which equals 1 if consumer i made a repeat purchase.

Sampling of the behavioral parameters for the BG/NBD Model

To sample the individual rate parameters of the BG/NBD model, we again make use of ideas from Bayesian statistics. Directly sampling from the joint conditional distribution of λ_i and p_i is not easy. However, we can derive the full conditional distributions of λ_i and p_i. We, therefore, propose to use a Gibbs sampler (Geman and Geman, 1984) which successively draws from the conditional distribution of λ_i given $x_i, t_{x,i}, T_i$ and p_i, and the conditional distribution of p_i given $x_i, t_{x,i}, T_i$ and λ_i. After convergence, this Markov Chain generates draws from the joint conditional distribution. Details of the derivations of both distributions are presented in Section 2.8.2. The conditional density of the purchase rate λ_i is

$$\pi(\lambda_i|x_i, t_{x,i}, T_i, p_i) = \frac{\frac{p_i}{(t_{x,i}+\alpha)^{x_i+r}}}{\frac{p_i}{(t_{x,i}+\alpha)^{x_i+r}} + \frac{1-p_i}{(T_i+\alpha)^{x_i+r}}}\varphi_{x_i+r, t_{x,i}+\alpha}(\lambda_i) + \frac{\frac{1-p_i}{(T_i+\alpha)^{x_i+r}}}{\frac{p_i}{(t_{x,i}+\alpha)^{x_i+r}} + \frac{1-p_i}{(T_i+\alpha)^{x_i+r}}}\varphi_{x_i+r, T_i+\alpha}(\lambda_i), \tag{2.9}$$

where $\varphi_{x,\beta}$ is the density of a gamma distribution with shape parameter x and rate parameter β. The conditional density of the defection probability p_i equals

$$\pi(p_i|x_i, t_{x,i}, T_i, \lambda_i) = \frac{a}{a + (b + x_i - 1)e^{-\lambda_i(T_i - t_{x,i})}} \beta_{a+1,b+x_i-1}(p_i) + \\ \frac{(b + x_i - 1)e^{-\lambda_i(T_i - t_{x,i})}}{a + (b + x_i - 1)e^{-\lambda(T_i - t_{x,i})}} \beta_{a,b+x_i}(p_i), \quad (2.10)$$

where $\beta_{a,b}$ is the density of a beta distribution with parameters a and b. As the distributions are mixtures of gamma or beta distributions, respectively, sampling from these distributions is straightforward.

2.3.4 PDO Model

The most recent BTYD model is the Periodic Death Opportunity (PDO) model. This model is based on the BG/NBD model, but assumes that a customer may only defect after each τ periods of time. The defection process is, therefore, no longer linked to purchase occasions and heavy purchasers do not get more defection opportunities. Jerath et al. (2011) show that the PDO model can be seen as a generalization of the Pareto/NBD and the NBD model. If τ becomes very small, the PDO model approaches the Pareto/NBD model. The PDO model collapses to the NBD model when τ exceeds the observation period, leaving no dropout possibility for customers.

More precisely, the PDO model assumes that the interpurchase time for individual i has an exponential distribution with parameter $\lambda_i \sim \Gamma(r, \alpha)$. Customers may defect with a probability of p_i after each τ periods, where p_i follows a beta distribution with parameters a and b. The PDO model has four hyperparameters for the heterogeneity distributions and the additional *period length* parameter τ. MLE can again be used to estimate the hyperparameters; for more details see Jerath et al. (2011).

The introduction of the τ parameter complicates the prediction of the timing of the last and the next transactions. T_i is likely not a multiple of τ, and we need to deal with the delay between the last opportunity to defect before T_i and, for the computation of the expected first future transaction, the delay between T_i and the first opportunity to defect after T_i. A further complication is the possibility that there is no defection opportunity during $(T_i, T_i^+]$. Details of the derivations are presented in Section 2.7.3. The expected time of the last transaction in the in-sample period is

$$\mathbb{E}(t_{x,i}|T_i,\lambda_i,p_i)=\sum_{n=1}^{N_i}p_i(1-p_i)^{n-1}\left(n\tau-\frac{1-e^{-n\lambda_i\tau}}{\lambda_i}\right)+(1-p_i)^{N_i}\left(T_i-\frac{1-e^{-\lambda_i T_i}}{\lambda_i}\right), \tag{2.11}$$

where N_i equals the number of defection opportunities, that is, $N_i=\lfloor T_i/\tau\rfloor$. The expected time of the first purchase in the out-of-sample period $(T_i,T_i^+]$ is

$$\begin{aligned}&\mathbb{E}(t_{f,i}|x_i,t_{x,i},T_i,\lambda_i,p_i,T_i^+)=\\&\quad(1-p_i^+)T_i^+ + p_i^+\Bigg[(T_i+\frac{1}{\lambda_i})e^{-\lambda_i T_i}-(\bar{T}_i+\frac{1}{\lambda_i})e^{-\lambda_i\bar{T}_i}+\delta_{T_i^+<(N_i+1)\tau}T_i^+e^{-\lambda_i(T_i^+-T_i)}\\&+\delta_{T_i^+\geq(N_i+1)\tau}\left(e^{-\lambda_i((N_i+1)\tau-T_i)}p_iT_i^+ + (1-p_i)\Big((N_i+1)\tau+\mathbb{E}(t^+|\lambda_i,p_i,T_i^+-(N_i+1)\tau)\Big)\right)\Bigg],\end{aligned} \tag{2.12}$$

where $\bar{T}_i$ is the minimum of the first defection opportunity in the out-of-sample period for customer i and T_i^+, that is, $\bar{T}_i=\min((N_i+1)\tau,T_i^+)$. Furthermore, p_i^+ is shorthand notation for the conditional probability that individual i is active at time T_i. This probability is given by

$$p_i^+=\mathbb{P}(t_{\Delta,i}>T_i|x_i,t_{x,i},T_i,\lambda_i,p_i)=\frac{(1-p_i)^{N_i}e^{-\lambda_i T_i}}{p_i e^{-\lambda_i\tau}\sum_{n=m_{x,i}}^{N_i}\left((1-p_i)e^{-\lambda_i\tau}\right)^{n-1}+(1-p_i)^{N_i}e^{-\lambda_i T_i}}$$

where $m_{x,i}$ is the first opportunity to defect *after* $t_{x,i}$, that is, $m_{x,i}=\left\lfloor\frac{t_{x,i}}{\tau}+1\right\rfloor$ and we define $\sum_{n=a}^{b}(\cdot)=0$ whenever $a>b$. Finally, $\mathbb{E}(t^+|\lambda_i,p_i,T_i^+-(N_i+1)\tau)$ is the expected value of the minimum of the time of the first transaction in $(0,T_i^+-(N_i+1)\tau)$ and $(T_i^+-(N_i+1)\tau)$. The expression for this expectation is given in Equation (2.35) of the appendix.

Sampling of the behavioral parameters for the PDO Model

To sample λ_i and p_i, we again propose a Gibbs sampler; see Section 2.8.3 for the details. Conditional on the data and p_i, λ_i follows a mixture of gamma distributions, that is,

$$\pi(\lambda_i|x_i,t_{x,i},T_i,p_i)=\sum_{n=m_{x,i}}^{N_i}\frac{w_{x_i,p_i}^{(n)}}{W_{x_i,t_{x,i},p_i}}\varphi_{x_i+r,\alpha+(n-1)\tau}(\lambda_i)+\frac{w_{x_i,p_i}^{(N_i+1)}}{W_{x_i,t_{x,i},p_i}}\varphi_{x_i+r,\alpha+T_i}(\lambda_i) \tag{2.13}$$

where $W_{x_i,t_{x,i},p_i} = \sum_{n=m_{x,i}}^{N_i+1} w_{x_i,p_i}^{(n)}$, and

$$w_{x_i,p_i}^{(n)} = \begin{cases} p_i \frac{(1-p_i)^{n-1}}{(\alpha+(n-1)\tau)^{x_i+r}} & \text{if } 1 \le n \le N_i \\ \frac{(1-p_i)^{N_i}}{(\alpha+T_i)^{x_i+r}} & \text{if } n = N_i + 1 \ . \end{cases}$$

The conditional distribution of p_i is a mixture of beta distributions, that is,

$$\pi(p_i|x_i, t_{x,i}, T_i, \lambda_i) = \sum_{n=m_{x,i}}^{N_i} \frac{v_\lambda^{(n)}}{V_{t_{x,i},\lambda_i}} \beta_{a+1,b+n-1}(p_i) + \frac{v_{\lambda_i}^{(N_i+1)}}{V_{t_{x,i},\lambda_i}} \beta_{a,b+N_i}(p_i) \tag{2.14}$$

where $V_{t_{x,i},\lambda_i} = \sum_{n=m_{x,i}}^{N_i+1} v_{\lambda_i}^{(n)}$, and

$$v_{\lambda_i}^{(n)} = \begin{cases} B(a+1, b+n-1)e^{-\lambda(T_i-(n-1)\tau)} & \text{if } m_{x,i} \le n \le N_i \\ B(a, b+N_i) & \text{if } n = N_i + 1, \end{cases}$$

where $B(\cdot,\cdot)$ is the beta function. Note that the value $V_{t_{x,i},\lambda_i}$ depends on the data only through $m_{x,i}$.

2.3.5 Hierarchical Bayes Extension of the Pareto/NBD Model

The models presented above do not allow the individual-level parameters to be correlated and they do not take into account customer characteristics. In many cases, individual-level characteristics are available and may be useful in predicting customer behavior. Abe (2009a), therefore, proposes a Hierarchical Bayes [HB] extension of the Pareto/NBD model in which the individual-level parameters follow a bivariate log-normal distribution. The mean of this distribution may depend on customer characteristics.

The disadvantage of this extension is that closed-form expressions for interesting metrics, such as the expected number of purchases, are no longer available. Besides, MLE can no longer be straightforwardly used to obtain parameter estimates. Abe proposes the use of Markov chain Monte Carlo [MCMC] techniques to estimate the (hyper)parameters and to calculate various metrics.

Abe (2009a) makes the same individual-level assumptions as in the Pareto/NBD model, but assumes that $(\log \lambda_i, \log \mu_i) \sim N(w_i\beta, \Gamma)$, where w_i is a $1 \times K$ vector of individual characteristics, including an intercept. In case no covariates are available, the distribution reduces to $N(\beta, \Gamma)$. Γ is not restricted to a diagonal matrix and, therefore, this model allows the individual-level parameters to be correlated.

The joint density of the data and all parameters forms the basis for the inference. This density is given by

$$\pi(\{x_i, t_{x,i}, T_i, \lambda_i, \mu_i\}_{i=1}^N, \beta, \Gamma) = \prod_{i=1}^N \Big(\pi(x_i, t_{x,i}|\lambda_i, \mu_i)\pi(\lambda_i, \mu_i|\beta, \Gamma)\Big)\pi(\beta, \Gamma) \; .$$

Here $\pi(\beta, \Gamma)$ is the prior distribution of the population-level parameters β and Γ. The standard conjugate prior is used, that is, $\beta \sim N(\beta_0, A_o)$ and Γ follows an inverted Wishart distribution with parameters $(\nu_0,\, \Gamma_0)$. As the individual-level behavioral assumptions of the HB model are identical to the Pareto/NBD model, conditional on λ_i and μ_i, all timing related expressions are the same. Draws for the individual-level parameters are a natural by-product of the MCMC sampler.

Abe (2009b) proposes an extension of the HB model by adding the amount of spending. Hereby, the individual parameter vector, θ_i, extends to three dimensions, including the rate of average log-spending of customers, $(\log \lambda_i, \log \mu_i, \log \eta_i)$. We also include this extension in our empirical study. Consequently, we consider four different configurations of the HB model. The first configuration (HB1) represents the HB model without any covariates and without spending. The second configuration (HB2) incorporates only the customer-specific covariates. The third and fourth configurations represent the HB models with the average spending parameter, and without or with covariates, respectively.

Sampling of the hyperparameters and the behavioral parameters for the HB Model

We use MCMC for inference on the hyperparameters and the individual parameters for the HB models. More specifically, we use a Metropolis within Gibbs sampler. The sampler uses the latent variables z_i and $t_{\delta,i}$, where z_i is the binary variable representing whether customer i is active ($z_i = 1$) or inactive ($z_i = 0$) at the end of the calibration period; and if already inactive, $t_{\delta,i}$ is the defection time (see Abe (2009a)). As our sampler differs from the one presented in Abe (2009a), we present the main steps of the sampler:

[0] Set initial value for $\theta_i, i = 1, \ldots, N$.

[1a] Generate $z_i|t_{x,i}, x_i, T_i, \theta_i$ according to the being active probability given in Equation (2.5), for $i = 1, \ldots, N$.

[1b] If $z_i = 0$, generate $t_{\delta,i}|t_{x,i}, x_i, T_i, z_i, \theta_i$ using an exponential distribution truncated to $(t_{x,i}, T_i)$.

[2] Generate $\beta, \Gamma|\{\theta_i\}_{i=1}^N$ using a standard multi-variate normal regression update (see Rossi et al. (2005, Page 34)).

[3] Generate $\theta_i|t_{x,i}, x_i, T_i, z_i, t_{\Delta,i}, \beta, \Gamma$ with a Gaussian random-walk MH algorithm, for $i = 1, \ldots, N$.

The step size in the random-walk MH algorithm is set by applying an adaptive MH method in the burn-in phase (Gilks et al., 1996).

2.4 Data

We compare the performance of the presented models on three datasets. Below, we briefly discuss these three datasets.

The first dataset contains daily transaction data of an online grocery retailer in a Western European country (OG hereafter). We base our analysis on a random set of 1460 customers who started buying from the company in January 2009. We ignore all Sundays as OG does not provide delivery on that day. The available data contains the initial and the repeat purchase information of each customer over a period of 309 days. To estimate the model parameters, we use the transaction data of all customers over the first 154 days, leaving a 155 day holdout period for model validation.

The second dataset is the commonly used CDNOW data. This publicly available dataset covers the transactions data of 2357 customers who made their first transaction in the first quarter of 1997. The data spans a period of 78 weeks from January 1997 through June 1998. We set the calibration and holdout periods to 39 weeks each.

The final dataset comes from a Turkish grocery store. This set is also used by Batislam et al. (2007) and Jerath et al. (2011). It contains the transactions of 5479 customers who made their first purchase between August 2011 and October 2011, covering a period of 91 weeks. To be consistent with the earlier papers, we use the first 78 weeks for calibration and leave 13 weeks for validation purposes. Detailed descriptive statistics of all datasets appear in Table 2.3.

The three datasets have quite different characteristics. Together they span a wide range of purchase and activity patterns. For instance, in the first dataset, the majority of customers are frequent customers, whereas the other two datasets include a large group of incidental buyers. Although the first two datasets both deal with online retailers, the industries in which these retailers operate are different, namely groceries versus CDs. We see a clear difference in the customer's loyalty to the firm; the average frequency of shopping per customer is higher at the OG than at the CD retailer. The fraction of

Table 2.3: Descriptive statistics over the three datasets

	OG	CDNOW	Grocer
Number of customers	1460	2357	5479
Available time frame	309 days	78 weeks	91 weeks
Time split (in-sample/out-of-sample)	154/155	39/39 weeks	78/13 weeks
Available time units	days	weeks/days	weeks
Zero repeaters in estimation period (fraction)	174 (0.12)	1,411 (0.60)	2,221 (0.41)
Zero repeaters in holdout period (fraction)	295 (0.20)	1,673 (0.70)	4,577 (0.84)
Zero repeaters in estimation and holdout periods (fraction)	135 (0.09)	1,218 (0.51)	2,179 (0.40)
Number of purchases in estimation period (all)	16,252	2,457	24,840
Number of purchases in holdout period	12,827	1,882	2,907
Average number of purchases per customer in estimation period (stdev)	11.13 (10.76)	1.04 (2.190)	4.53 (9.17)
Average number of purchases per customer in holdout period (stdev)	8.79 (10.78)	0.798 (2.057)	0.53 (1.72)
Average length of the observation period (T) (stdev)	143.76 (7.39)	32.72 (3.33)	22.81 (26.87)
Average recency as a fraction of T ($(T - t_x)/T$)	0.27	0.79	0.67

customers without a repeat purchase (zero-repeat buyers) is also much smaller for the OG compared to CDNOW. A customer's final observed purchase tends to be close to the end of the sample for the online retailer. This is reflected in the last row of Table 2.3, which gives the average recency normalized by the average observation period.

Customer behavior at the brick-and-mortar grocer is quite different compared to that at the online grcer. Contrary to the general claim in the literature, the customers of the OG are more loyal to the company than those of the grocer chain. The rate of zero-repeat buyers in the grocer's data base is considerably higher, and the average normalized recency is significantly lower than for the OG. In what follows, we relate the performance of the models on three datasets to their characteristics.

2.5 Empirical Findings

We split this section in two parts. First, we discuss the parameter estimates for all models and datasets.[1] Next, we focus on the predictive performance of the models, where we distinguish between (1) expected number of transactions; and (2) expected timing of transactions. We especially focus on the performance of the models in predicting the timing of the last in-sample purchase and the first out-of-sample purchase.

For the online retailer datasets (OG and CDNOW), covariate data on the average number of shopping items per customer is available. This data is used in the HB model configurations HB2 and HB4. As both datasets also have individual-level spending infor-

[1] All calculations are performed using MATLAB R2011b.

mation, the spending extension of the HB models (HB3 and HB4) can be applied as well. We mean-center the covariate (average number of items in the shopping basket) so that the mean of the behavioral parameters, θ_i, given average covariate values will be entirely determined by the intercept. As no covariate nor spending information is available for the third dataset (grocer), only the HB1 model can be used. For all HB models, the MCMC steps were repeated $256,000$ iterations, of which the last $32,000$ were used to infer the posterior distribution of parameters. Convergence was monitored visually and checked with the Geweke test on all datasets (Geweke et al., 1991).

2.5.1 Parameter Estimates

Maximum Likelihood-based models

First we present the parameter estimates that are based on ML estimation; namely for the Pareto/NBD, BG/NBD, and PDO models. Using the estimates, we can gain insight in the degree of heterogeneity in each customer base as well as in some key quantities for a random customer. Table 2.4 reports the estimated hyperparameters for the OG. According to the Pareto/NBD model a random customer makes 0.072 transactions per day while active. Note that this statistic cannot be calculated directly from the data as it intrinsically contains the condition of being active. The shape parameter ($r = 0.958$) indicates a moderate level of heterogeneity in purchase rates across customers (Schmittlein et al., 1993). For this dataset, the PDO model fits best when the period length τ is set to about 20 days. The parameters related to the purchase process in the PDO model are very similar to those in the Pareto/NBD model. The BG/NBD model also gives a very similar result for the purchase rate of an average customer while active (0.071 purchases). The relatively small shape parameter value ($r = 0.897$) indicates slightly more differences in purchase rates across customers within the BG/NBD model.

Table 2.4: Results of the Pareto/NBD, BG/NBD and PDO Maximum Likelihood Estimates - OG

Pareto/NBD		BG/NBD		PDO ($\tau = 20.001$)	
r	0.96	r	0.90	r	0.94
α	13.35	α	12.64	α	13.13
r/α	0.072	r/α	0.071	r/α	0.071
s	0.04	a	0.03	a	0.04
β	38.24	b	3.00	b	2.18
s/β	0.001	a/(a+b)	0.010	a/(a+b)	0.018
log-likelihood	-49,208	log-likelihood	-49,212.3	log-likelihood	-49,201.4

The estimated average defection rate for the Pareto/NBD model is given by $s/\beta = 0.001$. As the shape parameter s is less than 1, the expected lifetime value of a random customer from the cohort diverges to infinity. From another perspective, half of the customers in the cohort defect after $(2^{1/s} - 1)\beta = 383,014,675$ days. This shows that a short-term measure rather than these long lifetime estimations would be more useful for a manager. The probability of a random customer defecting in the next day is only $1 - e^{-s/\beta} = 0.001$. In other words, it is highly unlikely that such a customer will drop out in the near future. However, the very small value of s suggests that there is a very large dispersion in defection rates.

The estimation results for the CDNOW data are given in Table 2.5. We obtain the same parameter estimates as Fader et al. (2005a). We find that an average customer makes around 0.05 transactions per week, while active. The small shape parameter value indicates substantial differences in purchase rates across customers. Similar to the previous dataset, the heterogeneity on defection rates is extremely high on this dataset ($s = 0.606$ in the Pareto/NBD model) and the expected lifetime value of a random customer from the cohort diverges to infinity.

Table 2.5: Results of the Pareto/NBD, BG/NBD and PDO Maximum Likelihood Estimates - CDNOW

Pareto/NBD		BG/NBD		PDO ($\tau = 3.001$)	
r	0.55	r	0.24	r	0.52
α	10.58	α	4.41	α	10.40
r/α	0.052	r/α	0.055	r/α	0.05
s	0.61	a	0.79	a	0.43
β	11.66	b	2.43	b	2.61
s/β	0.052	a/(a+b)	0.246	a/(a+b)	0.142
log-likelihood	-9,595	log-likelihood	-9,582.4	log-likelihood	-9,585.6

When applying the models on the Turkish grocery dataset, we find that while active, an average customer places approximately 0.1 orders per week; see Table 2.6. The population is quite heterogeneous in purchase rates. The heterogeneity is even greater according to the BG/NBD model. For an in-depth discussion on the customer lifetime, we recommend the discussion in Jerath et al. (2011).

MCMC-based models

In order to apply the HB models we first need to set the prior distributions. In many contexts, the prior is set diffuse enough so that it does not affect the posterior. In other

Table 2.6: Results of the Pareto/NBD, BG/NBD and PDO Maximum Likelihood Estimates - grocery retailer

Pareto/NBD		BG/NBD		PDO ($\tau = 1.001$)	
r	0.48	r	0.28	r	0.46
α	4.38	α	2.34	α	4.38
r/α	0.11	r/α	0.12	r/α	0.105
s	0.57	a	0.40	a	0.62
β	17.60	b	2.09	b	22.19
s/β	0.033	a/(a+b)	0.161	a/(a+b)	0.027
log-likelihood	-67,925.8	log-likelihood	-68,008.3	log-likelihood	-67,757.3

words, the prior variance is set to a very large value. For the prior on Γ, we initially use $\nu_0 = J + 3$ and $\Gamma_0 = \nu_0 I$, where J represents the number of behavioral parameters of a customer (see Rossi et al. (2005, Page 30)). This is an extremely spread prior. However, in case limited data per individual is available, such a prior may have a strong impact on the posterior. Indeed, looking at the likelihood function for the HB model given in Equation (2.36), it can be seen that the likelihood for a zero-repeat buyer ($x_i = 0 = t_{x,i}$) tends to 1 as μ approaches ∞ for any value of λ. Therefore, without a proper prior the posterior does not exist. The prior needs to ensure that the posterior density for large values for μ approaches 0 quickly enough. Very diffuse priors fail to deliver this property, leading to (very) unstable estimates.

Among the datasets in our study, the CDNOW dataset is unique in terms of having a very large proportion of zero-repeat buyers. In other words, the data does not provide much information. We, therefore, need to set a relatively informative prior for this dataset. Accordingly, we choose $\nu_0 = J + 30$ and $\Gamma_0 = \nu_0 I$. In this way, extreme estimates are avoided and population-level estimates are reasonable.[2] Still, we have experimented with a diffuse prior on this dataset. A detailed look at the results per individual (not reported) reveals that there are indeed extreme values for some parameters (in a range of 5.10^8). We also observe very different predictions for individuals with a history of zero-repeat transactions, following the reasoning stated above. A further elaboration on the selection of the prior parameters on the CDNOW dataset is given in Section 2.9.

The hyperparameters of the HB models are not directly comparable to the hyperparameters of the other BTYD models, not only because of the different heterogeneity distribution (log-normal distribution versus gamma and beta distributions), but also because the multi-variate structure of the log-normal distribution allows correlation between

[2] With a more diffuse prior, an extremely large number of iterations is needed to obtain accurate estimates of posterior quantities as the posterior variance will be very large.

parameters for a single customer. Table 2.7 gives the median and the mode of the posterior mean of behavioral parameters across customers in each dataset. It is interesting to note that the location of the population distribution in the HB models seems to be different to that for the other models. In the next section, we investigate whether this has an impact on the models' performance.

Table 2.7: Median and mode of the behavioral rates of HB model estimates

		HB1		HB2		HB3		HB4	
		λ	μ	λ	μ	λ	μ	λ	μ
OG	median	0.0474	0.0008	0.0471	0.0008	0.0479	0.0002	0.0479	0.0003
	mode	0.0204	0.0003	0.0233	0.0004	0.0086	0.0001	0.0085	0.0001
CDNOW	median	0.0045	0.0129	0.0072	0.0170	0.0081	0.3834	0.0089	0.5117
	mode	0.0045	0.0132	0.0073	0.0019	0.0080	0.0006	0.0083	0.0004
Grocer	median	0.0469	0.0568	-	-	-	-	-	-
	mode	0.0464	0.0080	-	-	-	-	-	-

2.5.2 Unconditional Predictions

We follow the procedure described in Section 2.3.1 to obtain unconditional predictions. As individuals in the customer database make their first purchases at different times, the time span T varies across customers. Consequently, we obtain different in-sample predictions for different values of T. We calculate the unconditional predictions for each of the T_i values in the database and average over them. These predictions are only based on the population-level parameters, estimated using all the data in the customer base. Hence, they serve as good indicators of the model's ability to fit the overall data pattern. Table 2.8 shows some statistics on the unconditional expectations on the number of transactions and the timing of the last transaction for each model and each dataset. The first row shows the statistics based on the observed values for each dataset.

The mean predictions for the HB models are very different from the other model predictions on CDNOW data.[3] However, the predicted values are much closer to the median and mode of the data. In other words, it seems that the large number of zero-repeat buyers pulls the predictions from the HB models towards smaller values. This is probably due to the shape of the population distribution. As can be seen in Table 2.4, the mode for the population distributions of λ_i and μ_i are at 0. The log-normal distribution does not allow for a mode at 0 without also pulling the mean towards 0 (or having an extreme variance). This explains why the mean predictions for the HB models are pulled

[3]Note that the mean unconditional predictions move even further away with the most diffuse prior. For example, it becomes 0.09 for the HB2 model, see Table 2.15.

towards 0. For the other datasets, the percentage of zero-repeat buyers is not as large, therefore this phenomenon is not observed there.

Table 2.8: Average of unconditional expectations versus observed quantities in calibration period

		Number of transactions			Time of last transaction		
		mean	median	mode	mean	median	mode
OG	True	10.132	6	0	105.421	128	0
	Pareto/NBD	7.926	8.000	8.300	76.786	77.831	78.410
	BG/NBD	6.593	6.647	6.970	57.841	58.571	61.670
	PDO	9.789	9.884	10.360	104.217	105.574	111.540
	HB1	10.573	10.694	11.150	103.157	104.419	110.650
	HB2	10.707	10.826	11.320	106.048	107.289	113.780
	HB3	11.231	11.341	11.290	101.139	102.482	107.830
	HB4	11.139	11.256	11.360	101.662	102.942	104.270
CDNOW	True	1.042	0	0	6.864	0	0
	Pareto/NBD	1.071	1.071	1.100	6.804	6.790	6.860
	BG/NBD	1.058	1.057	1.000	6.913	6.889	7.760
	PDO	1.079	1.078	1.150	6.915	6.900	6.540
	HB1	0.227	0.227	0.220	2.884	2.862	3.090
	HB2	0.245	0.244	0.230	3.020	2.997	2.590
	HB3	0.232	0.231	0.220	2.900	2.880	3.410
	HB4	0.235	0.235	0.220	2.953	2.926	2.690
Grocer	True	4.534	1	0	22.805	7	0
	Pareto/NBD	4.462	4.443	4.320	22.589	22.411	21.850
	BG/NBD	4.240	4.222	4.150	23.951	23.731	23.000
	PDO	4.424	4.403	4.290	22.841	22.667	22.110
	HB1	4.839	4.816	4.700	22.485	22.313	21.910

We also provide some performance measures for the number of in-sample transactions (x) and the time of the last in-sample transaction (t_x) for each model. Table 2.9 shows the in-sample Mean Squared Error (MSE), Mean Absolute Error (MAE) on all predictions and Mean Error on the over- (ME+) and underpredicted (ME−) observations for all models on the three datasets. At a first glance, all models have a similar fit when predicting x. The PDO model performs slightly better with respect to MSE on the CDNOW and the grocery data. The estimated hyperparameters for this model lead to a low probability of extreme values on these datasets. On the other hand, the HB model fits the best in terms of MSE on the OG dataset. In terms of absolute errors in the unconditional predictions of x, the BG/NBD model has the best fit for the OG and the grocer data.

The HB models perform well on the CDNOW dataset in terms of the MAE. The high MSE and the low MAE values for the HB models on CDNOW link back to our earlier discussion. The high number of zero-repeat buyers in this dataset causes the predictions to move towards the mode of the data. Consequently, on this dataset, the mean of the unconditional predictions of the HB models approaches the strong mode of the data. This fact leads to a low MAE for the HB models. All models show an asymmetry in the unconditional prediction error. If the forecast is too high, the error tends to be relatively small.

The Pareto/NBD, BG/NBD and PDO models have a very similar performance when predicting the last purchase time on the CDNOW dataset. The PDO and the HB are the best performing models with respect to the unconditional predictions on this measure for the CDNOW and the OG datasets (considering the MSE and the MAE, respectively). On the grocer dataset, all models have a similar fit on predicting t_x, except the BG/NBD model which fits slightly worse on this metric.

Among the different configurations of HB models, we see that inclusion of covariates generally causes a slight increase in model fit on both measures. On the other hand, adding the spending parameter into the estimation procedure leads to a slight decrease in model fit for the frequency and the timing of in-sample transactions on the OG data.

Table 2.9: In-sample predictive performance for unconditional predictions of the number of transactions (x) and the time of last transaction (t_x)

		x				t_x in weeks			
		MSE	MAE	ME+	ME−	MSE	MAE	ME+	ME−
OG	Pareto/NBD	116.636	7.803	4.847	11.841	90.526	8.926	9.106	8.873
	BG/NBD	124.992	**7.725**	4.096	11.516	131.352	10.809	7.560	11.573
	PDO	111.038	8.123	6.367	10.880	**66.809**	6.774	10.523	5.071
	HB1	**110.832**	8.302	6.923	10.666	67.110	6.852	10.598	5.205
	HB2	110.910	8.335	7.009	10.647	66.822	**6.664**	10.672	4.803
	HB3	111.485	8.473	7.371	10.513	67.495	6.986	10.430	5.505
	HB4	111.323	8.442	7.292	10.559	67.337	6.949	10.466	5.427
CDNOW	Pareto/NBD	4.789	1.282	0.886	2.411	114.655	8.899	6.353	14.758
	BG/NBD	4.788	1.276	0.879	2.377	114.640	8.942	6.462	14.647
	PDO	**4.786**	1.286	0.888	2.446	**114.610**	8.940	6.455	14.683
	HB1	5.455	1.087	0.227	2.370	130.332	**7.547**	2.772	16.282
	HB2	5.426	1.090	0.244	2.352	129.251	7.586	2.895	16.282
	HB3	5.448	**1.088**	0.231	2.365	130.195	7.551	2.787	16.265
	HB4	5.442	1.089	0.235	2.362	129.796	7.567	2.835	16.271
Grocer	Pareto/NBD	83.958	5.454	3.554	11.381	**719.044**	24.024	19.359	31.472
	BG/NBD	84.097	**5.341**	3.342	11.503	720.197	24.341	20.457	30.755
	PDO	**83.949**	5.435	3.517	11.413	719.137	24.082	19.571	31.323
	HB1	84.081	5.650	3.900	11.298	719.229	**24.001**	19.274	31.532

2.5.3 Conditional Predictions

In this section, we consider individual-level predictions conditional on the individual's history. As discussed in Section 2.3.1, for some metrics of interest, obtaining closed-form expression conditioned on an individual's history and hyperparameters can be extremely cumbersome because of the integral in Equation (2.2). We, therefore, first obtain draws for the individual's behavioral parameters from the posterior densities and next calculate the expected value of the metrics of interest by averaging over these draws. For the Pareto/NBD model, we use a Gaussian random-walk MH sampler to obtain draws of individual parameters conditional on the hyperparameters. To satisfy convergence, we repeat the iterations 300,000 times, of which only the last 10,000 iterations were used.[4] For the BG/NBD and PDO models, we use a two-step Gibbs algorithm with 30,000 iterations, of which only the last 8,000 draws are used.

For metrics like the transaction frequency of a customer with history $(x_i, t_{x,i}, T_i)$, closed-form expressions for the Pareto/NBD, BG/NBD and PDO models are available conditional on both hyperparameters and behavioral parameters. This allows us to test our procedure based on the posterior draws on individual's parameters. We compare our simulation-based predictions to the results computed by the closed-form expressions conditioned on hyperparameters given in Schmittlein et al. (1987), Fader et al. (2005a) and Jerath et al. (2011). In all cases, the correlation between the expectations is more than 99.995%.

We consider the number of transactions in the out-of-sample period as well as the timing of the first out-of-sample transaction. More precisely, with the timing of the first out-of-sample transaction, we mean the minimum of the timing of the next transaction and the end of the out-of-sample period. We use MSE, MAE and the correlation between predicted and observed values. As the above measures do not distinguish between over- and underpredictions, we also provide the mean over all positive errors (ME+: overprediction) and the mean over all negative errors (ME−: underprediction).

Predicting future transaction frequency

Table 2.10 summarizes the predictive performance on the number of future transactions. The HB models perform best in terms of the MSE, MAE and correlation measures on the grocer and the OG datasets. Taking into account that the covariate information works well for the OG, the HB2 model performs, consequently, the best among the HB models. For this model, the coefficient of the average number of items in the shopping basket is

[4] We use an extreme number of burn-in iterations, in practice convergence is achieved much earlier.

significant at the 5% level (based on the highest posterior density [HPD] interval). Adding the average spending worsens the out-of-sample predictions on transaction frequency. Therefore, the HB3 and HB4 models do not perform as well.

The good predictive performance of the HB model can be explained by the relaxation of the independence assumption in the heterogeneity distribution. Note that the HB and the Pareto/NBD models share the same individual-level assumptions. To further investigate the dependence, we take a look at the estimated correlations between purchase and defection rates. As emphasized by Abe (2009a), it makes most sense to look at the estimated correlations for the no-covariate configuration of the HB models (HB1 and HB3). Table 2.11 reports the posterior mean correlations for each pair of parameters on each dataset for the HB3 model, together with the highest posterior density regions (Hyndman, 1996). We find a strong and significant negative correlation between purchase and defection rates for the OG data. Accordingly, we see a remarkable improvement on the prediction performance of the HB models on this dataset. We find a significant, but relatively smaller, negative correlation on the grocery data. The HB1 model performs only slightly better than the other models on this data. There is no significant correlation between the purchase and defection rates for the CDNOW dataset, and consequently, the Pareto/NBD model is the best predicting model with its more flexible gamma heterogeneity distribution.

The final two columns in Table 2.10 summarize the model's performance with regard to over- (ME+) and underpredictions (ME−). We find that for the Pareto/NBD model, the magnitude of underpredictions is bigger than that of overpredictions on all datasets. For the other models, the difference between ME+ and ME− depends on the data. The average underprediction is always larger than the average overpredictions on the CDNOW and grocery retailer datasets. It is exactly the other way around for the OG data, where the customers are relatively more loyal to the company. To further elaborate on this, we construct Table 2.12. This table presents summary statistics on the group of observations that are under- or overpredicted. We list the size of the group, mean values of the purchase frequency ($\overline{x}$) and the recency ($\overline{T} - \overline{t_x}$) in the calibration period, observed frequency in the holdout period ($\overline{x^*}$) and predictions ($\overline{\mathbb{E}[x]}$) for both groups. All models overpredict the transaction frequency, x, for the majority of customers in each datasets. In general, the overprediction occurs for those customers with a low transaction frequency and a long recency; and vice versa for the underprediction. In other words, the BTYD models overestimate transaction frequency for incidental buyers and underestimate it for frequent buyers.

Table 2.10: Model's prediction performance on the number of transactions

		Correlation	MSE	MAE	ME+	ME−
OG	Pareto/NBD	0.9207	21.556	3.055	2.344	3.830
	BG/NBD	0.9195	20.840	2.996	3.253	2.340
	PDO	0.9169	21.219	3.047	3.347	2.343
	HB1	0.9243	18.807	2.806	3.008	2.363
	HB2	**0.9250**	**18.543**	**2.779**	2.941	2.419
	HB3	0.9218	20.242	2.942	3.089	2.530
	HB4	0.9221	20.168	2.934	3.075	2.538
CDNOW	Pareto/NBD	**0.6304**	**2.568**	0.754	0.429	1.866
	BG/NBD	0.6248	2.589	0.787	0.456	1.831
	PDO	0.6214	2.709	0.903	0.696	1.737
	HB1	0.6235	2.962	0.717	0.209	2.083
	HB2	0.6127	2.954	0.736	0.253	2.054
	HB3	0.6241	2.743	0.680	0.234	2.090
	HB4	0.6223	2.740	**0.678**	0.236	2.095
Grocer	Pareto/NBD	0.8230	0.954	0.398	0.242	1.615
	BG/NBD	0.8216	0.966	0.416	0.265	1.602
	PDO	0.8189	0.983	0.460	0.317	1.591
	HB1	**0.8238**	**0.951**	**0.394**	0.239	1.600

Note that ME+ and ME− give the average of over- and underpredictions over the groups

We next study the relation between the prediction error and the number of in-sample purchases. The plots in Figure 2.1 show the average predicted number of out-of-sample purchases as a function of the number of in-sample purchases. Figure 2.2 gives the MAE as a function of the number of in-sample purchases. To be able to focus on the main differences between the model classes, we do not show the results for the HB models including spending and/or covariates.

Table 2.11: 95% Highest Posterior Density Region and mean of correlations between behavioral rates

	$\rho_{\theta_\lambda\theta_\mu}$			$\rho_{\theta_\lambda\theta_\eta}$			$\rho_{\theta_\eta\theta_\mu}$		
	HPDR		mean	HPDR		mean	HPDR		mean
OG	-0.718	-0.297	-0.501*	0.694	0.770	0.732*	-0.765	-0.687	-0.730*
CDNOW	-0.215	0.197	-0.011	0.235	0.421	0.332*	-0.729	-0.675	-0.703*
Grocer	-0.259	-0.115	-0.184*	-	-	-	-	-	-

* Indicates that 0 is not contained in the 95% HPDR (highest posterior density region).

Table 2.12: Statistics on the groups of over- and underpredictions of future transaction frequency

		Overpredicted observations						Underpredicted observations					
		ME+	cus. %	$\overline{x}$	$\overline{(T-t_x)}$	$\overline{x^*}$	$\overline{\mathbb{E}[x]}$	ME−	cus. %	$\overline{x}$	$\overline{(T-t_x)}$	$\overline{x^*}$	$\overline{\mathbb{E}[x]}$
OG	Pareto/NBD	2.344	52	6.593	8.855	3.138	5.482	3.830	48	13.984	3.705	14.934	11.104
	BG/NBD	3.253	72	8.912	6.887	6.072	9.325	2.340	28	13.243	5.119	15.710	13.371
	PDO	3.347	70	8.730	7.030	5.795	9.142	2.343	30	13.412	4.889	15.787	13.444
	HB1	3.008	69	8.909	6.998	5.917	8.925	2.363	31	12.806	5.058	15.061	12.698
	HB2	2.941	69	8.961	7.037	5.993	8.934	2.419	31	12.733	4.949	14.993	12.574
	HB3	3.089	74	8.908	6.599	6.172	9.261	2.530	26	13.560	5.802	16.109	13.580
	HB4	3.075	74	8.944	6.573	6.212	9.287	2.538	26	13.482	5.872	16.047	13.509
CDNOW	Pareto/NBD	0.429	77	0.851	27.303	0.170	0.598	1.866	23	1.695	20.977	2.946	1.079
	BG/NBD	0.456	76	0.813	27.698	0.144	0.600	1.831	24	1.764	20.113	2.859	1.028
	PDO	0.696	80	0.913	26.942	0.216	0.912	1.737	20	1.564	21.567	3.136	1.399
	HB1	0.209	73	0.631	28.748	0.041	0.250	2.083	27	2.116	18.666	2.836	0.753
	HB2	0.253	73	0.639	28.672	0.046	0.299	2.054	27	2.111	18.760	2.853	0.798
	HB3	0.234	76	0.733	27.811	0.108	0.342	2.090	24	1.982	20.343	2.977	0.887
	HB4	0.236	76	0.742	27.744	0.115	0.351	2.095	24	1.968	20.477	2.988	0.893
Grocer	Pareto/NBD	0.242	89	3.516	51.082	0.145	0.387	1.615	11	12.464	17.209	3.533	1.918
	BG/NBD	0.265	89	3.573	51.029	0.155	0.420	1.602	11	12.105	17.298	3.489	1.887
	PDO	0.317	89	3.541	51.002	0.149	0.466	1.591	11	12.411	17.287	3.561	1.970
	HB1	0.239	88	3.404	51.180	0.152	0.391	1.600	12	12.095	17.151	3.450	1.850

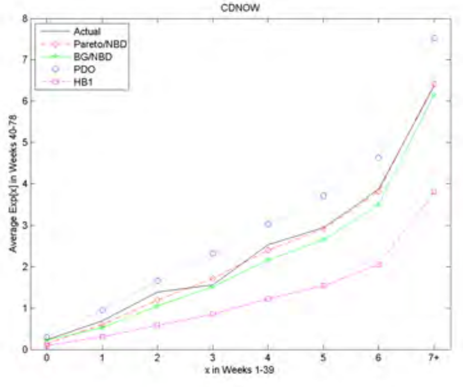
The PDO model tends to yield higher predictions for CDNOW data. This matches our findings in Tables 2.10 and 2.12. On average, the HB1 model yields the lowest predicted transaction numbers. Remarkably, this is not reflected in a poor forecasting performance for this model. In fact, Figure 2.2a shows that the HB1 model predicts very well for all values of the in-sample number of transactions. For the grocer dataset, all models show a very similar prediction pattern. Only the PDO model stands out with its relatively high predictions. Figure 2.2b shows that this leads to higher MAEs. The Pareto/NBD model is different from the other models for the online grocer data. This model has the tendency to underpredict transaction numbers (see also Tables 2.10 and 2.12).

The MAE tends to increase with the number of in-sample transaction numbers for the CDNOW and grocer datasets, contrasting with what is observed for the OG data (see Figure 2.2). The OG dataset stands out with its data center leaning toward frequent buyers. The predictions now result from models pulling values to this center.

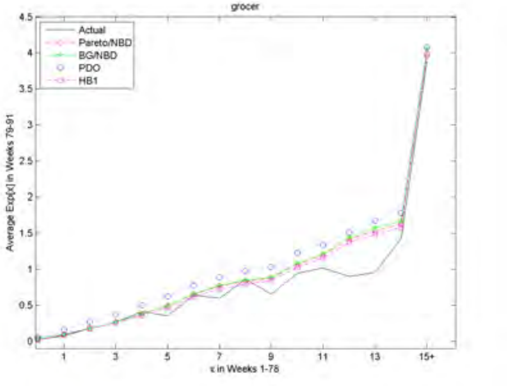
Predicting future transaction timing

Finally, we consider the performance on predicting future transaction timing.[5] Table 2.13 presents an overview of the main results. Interestingly, the PDO model has a good performance on the CDNOW and grocer datasets. This model did not perform particularly well on predicting the number of transactions. Note that the timing of transactions is

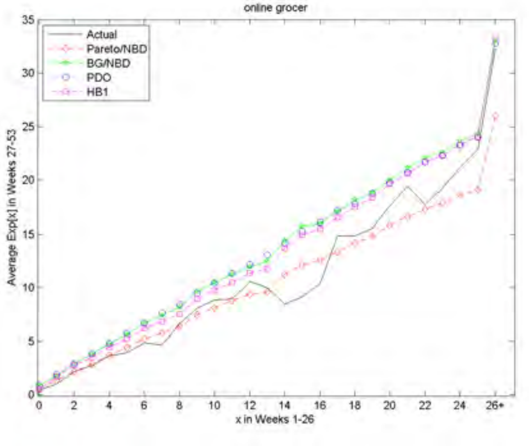
[5]We thank Batislam et al. (2007) and Fader et al. (2005b) for making the out-of-sample timing data available.



(a)

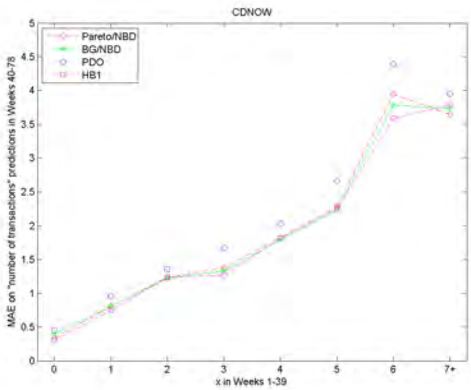


(b)

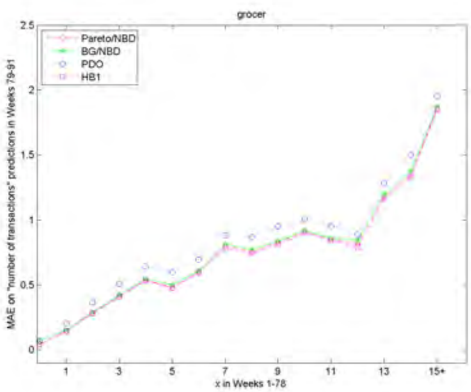


(c)

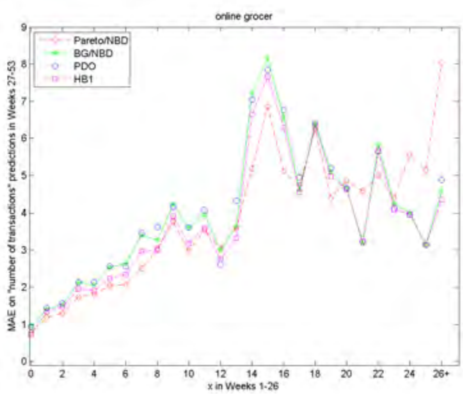
Figure 2.1: Conditional expectation of future transaction numbers on CDNOW, grocer and OG datasets. All plots right-censor the horizontal axis for readability. For CDNOW data, the group having ≥ 7 repeat-purchases corresponds to only 3% of the observations; for the grocer dataset 9% of the observations are in the group ≥ 15; and for the OG 6% are ≥ 26.



(a)



(b)



(c)

Figure 2.2: MAE on the number of future transaction predictions on CDNOW, grocer and OG datasets

strongly influenced by the defection process and that the PDO model specially focuses on this process. Jerath et al. (2011) demonstrate that the PDO model allows the defection process to be somewhere in between the extremes implied by the Pareto/NBD model and the no-defection NBD model. The PDO model performs the worst on the OG data. One reason may be the long (estimated) defection period interval ($\tau = 20.001$ days).

The HB models also perform rather well on the grocer and OG datasets. For both datasets we found a significant correlation between the behavioral parameters. Among the HB models, a remarkable point is the improved performance of the HB3 model when taking into account the average spending amount on CDNOW and OG datasets. This can be explained by the existence of the strong and significant negative correlation between the spending and defection parameters in both datasets (see Table 2.11).

Table 2.13: Model's prediction performance on the timing of next transaction

		Correlation	MSE	MAE	ME+	ME−
OG	Pareto/NBD	0.7296	46.674	4.508	2.649	5.801
	BG/NBD	0.7259	47.173	4.523	2.668	5.792
	PDO	0.6780	50.668	5.116	3.152	7.769
	HB1	**0.7328**	**43.416**	4.223	2.991	5.134
	HB2	0.7254	44.374	4.296	3.068	5.210
	HB3	0.7201	46.594	**4.067**	2.973	4.772
	HB4	0.7204	46.504	4.073	2.983	4.777
CDNOW	Pareto/NBD	0.5789	125.451	**7.372**	17.013	4.027
	BG/NBD	0.5750	125.153	8.122	17.027	5.033
	PDO	**0.5828**	**123.441**	8.517	15.343	6.228
	HB1	0.5486	273.555	15.660	10.062	17.051
	HB2	0.5449	282.423	15.865	9.781	17.352
	HB3	0.5687	270.514	15.408	9.229	16.898
	HB4	0.5689	270.028	15.376	9.214	16.850
Grocer	Pareto/NBD	0.8183	7.684	1.442	4.590	1.182
	BG/NBD	0.8192	7.770	1.542	4.551	1.293
	PDO	**0.8226**	7.976	1.734	4.469	1.514
	HB1	0.8190	**7.602**	**1.426**	4.639	1.171

ME+ and ME− give the average over the groups of overpredictions and underpredictions

In Table 2.14, we investigate for what type of observation the purchase time is over- or underpredicted. We present the size of the over- and underpredicted group, group-specific characteristics in the calibration period, the average observed timing ($\overline{t_f^*}$) in the holdout period and the average predicted time ($\overline{\mathbb{E}[t_f]}$). In line with the previous results,

Table 2.14: Statistics on the groups of over- and underpredictions of future transactions timing

		Overpredicted observations						Underpredicted observations					
		ME+	cus. %	$\overline{x}$	$\overline{(T-t_x)}$	$\overline{t_f^*}$	$\overline{\mathbb{E}[t_f]}$	ME−	cus. %	$\overline{x}$	$\overline{(T-t_x)}$	$\overline{t_f^*}$	$\overline{\mathbb{E}[t_f]}$
OG	Pareto/NBD	2.65	41	11.80	4.47	27.49	30.14	5.80	59	8.97	7.72	37.93	32.13
	BG/NBD	2.67	41	11.79	4.49	27.51	30.18	5.79	59	8.99	7.69	37.85	32.05
	PDO	3.15	57	13.71	3.38	27.39	30.54	7.77	43	5.30	10.45	42.11	34.34
	HB1	2.99	43	11.69	4.61	27.60	30.59	5.13	57	8.97	7.71	38.13	32.99
	HB2	3.07	43	11.62	4.57	27.60	30.67	5.21	57	9.03	7.75	38.15	32.94
	HB3	2.97	39	11.73	4.86	27.59	30.56	4.77	61	9.10	7.38	37.55	32.78
	HB4	2.98	39	11.75	4.85	27.58	30.56	4.78	61	9.09	7.38	37.57	32.79
CDNOW	Pareto/NBD	17.01	26	2.26	18.52	43.49	60.51	4.03	74	0.62	28.42	71.34	66.31
	BG/NBD	17.03	26	2.21	18.80	43.68	60.69	5.03	74	0.64	28.32	71.28	67.25
	PDO	15.34	25	2.27	18.46	43.04	58.39	6.23	75	0.63	28.36	71.25	65.03
	HB1	10.06	20	2.45	17.60	39.80	49.86	17.05	80	0.69	28.10	70.22	53.17
	HB2	9.78	20	2.50	17.40	39.62	49.40	17.35	80	0.69	28.12	70.17	52.82
	HB3	9.23	19	2.38	18.29	39.80	49.03	16.90	81	0.72	27.88	70.05	53.15
	HB4	9.21	19	2.38	18.33	39.74	48.96	16.85	81	0.72	27.85	70.01	53.16
Grocer	Pareto/NBD	4.59	8	7.17	23.32	75.03	79.62	1.18	92	4.32	49.21	82.77	81.59
	BG/NBD	4.55	8	7.10	23.21	75.05	79.60	1.29	92	4.32	49.22	82.77	81.48
	PDO	4.47	7	7.18	22.88	74.81	79.28	1.51	93	4.32	49.18	82.77	81.26
	HB1	4.64	8	7.18	23.30	75.03	79.67	1.17	91	4.18	49.23	82.76	81.59

all BTYD models underpredict the timing of the next purchase for customers who have a low transaction frequency and high recency; and vice versa for the groups of higher predictions.

In Figure 2.3, we show the average predictions as a function of the time of the last in-sample transaction (t_x). Note that the timing predictions are explicitly influenced by t_x (see Equations (2.4), (2.8), and (2.12)). We show the corresponding MAE values in Figure 2.4. Figure 2.3a clearly shows that the HB1 model gives quite different predictions compared to the other models for CDNOW; for HB1 the predictions tend to be smaller. Based on Figure 2.4a we conclude that these predictions are too low. The MAE for the HB1 model is the highest among all models. However, for the recent buyers (high t_x values) the differences between the models are relatively small.

For the grocer dataset, we see that all the models, except the PDO model, have almost identical predictions and performance for the non-recent buyers (see Figures 2.3b and 2.4b). The PDO model has lower predictions and higher MAE for those customers. Again for recent buyers, all models have very similar predictions so that it is difficult to distinguish between the models for this group of observations.

For the OG data, the PDO model also performs relatively poorly for non-recent buyers (see Figures 2.3c and 2.4c). The PDO model tends to underpredict the timing of the first transaction for customers who do not have recent transactions. On this data, the majority of customers are frequent buyers who had recent transactions. For instance, the percentage of customers who have $t_x \leq 10$ weeks is just 15% and therefore the left hand side of the figure does not have a big weight in the overall predictive performance of the models for this dataset. However, for the other datasets, a large part of the dataset have low values of t_x (53% of customers have $t_x \leq 10$ on the grocery dataset and 73% of customers has $t_x \leq 10$ on the CDNOW dataset).

2.6 Discussion

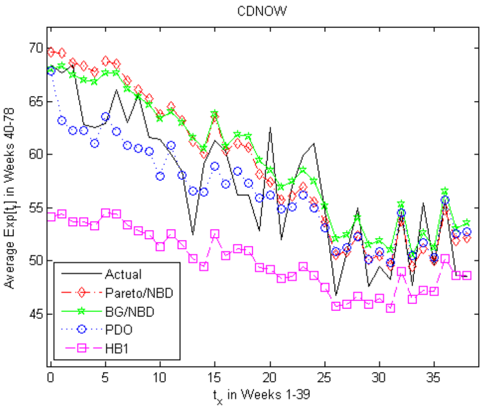
In this paper, our aim is to present a new use of the existing buy-till-you-defect [BTYD] models. In the current literature, the main focus is on predicting the transaction frequency. We argue that prediction of the future transaction timing of an individual is also very relevant. For each of the most popular BTYD models, we develop a method to calculate such predictions.

First of all, these timing predictions are useful to compare the quality of the existing models on an additional metric. Next, timing predictions have a clear managerial purpose. For example, consider an online retailer implementing micro-marketing strategies. The most appropriate time to contact its customers depends on their expected timing of the next purchase. High quality timing predictions may contribute to achieving the full potential of micro-marketing (Zhang and Krishnamurthi, 2004).

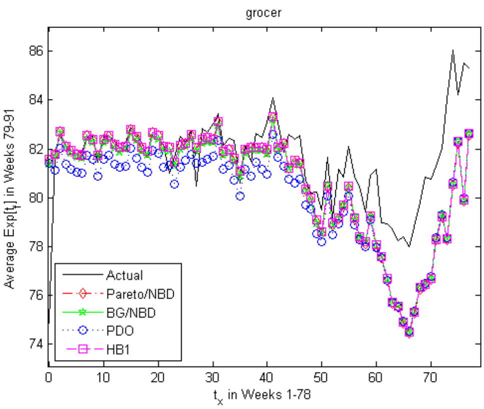
Following the pioneering research by Gupta (1988), there is a growing literature that examines the effectiveness of promotions on whether to buy, '*when*' to buy, and how much to buy (see the summary of relevant literature in Gönül and Hofstede (2006)). We believe that using the BTYD models to predict the timing of transactions provides a new means of answering the 'when' question.

An operations manager may also use predictions on the timing and transaction value as input for Revenue Management. For example, online retailers have limited delivery capacity at a given time. Given the appropriate predictions, operations managers can prioritize valued customers for highly demanded delivery time slots (Talluri and Van Ryzin, 2005). Tereyağoğlu et al. (2012) emphasize the crucial role of having accurate timing predictions to improve revenues. In summary, we believe that the ability to predict the timing of future transactions can be helpful to accelerate research on aforementioned topics in industries that operate in a noncontractual setting.

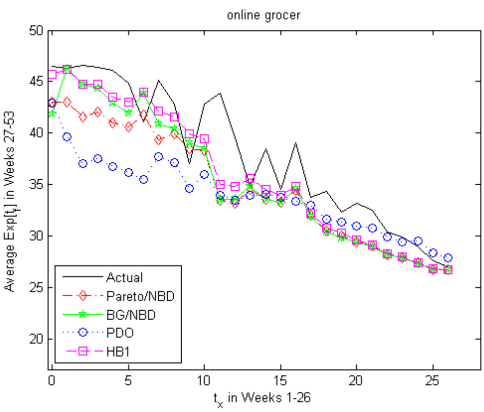
We present a general method and specific formulas that can be used to predict the timing of the next purchase for four of the established BTYD models. Such formulas have not been presented before. We use these methods to compare the predictive performance of all models on three very different datasets. We find that the predictive performance of



(a)

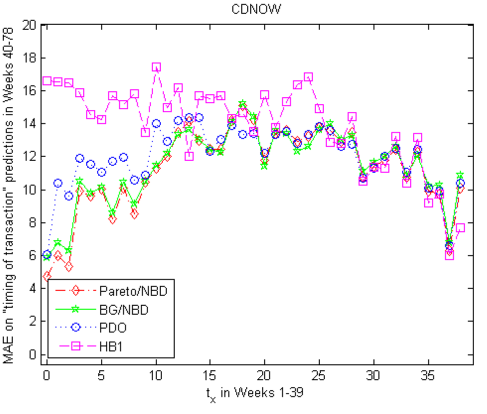


(b)

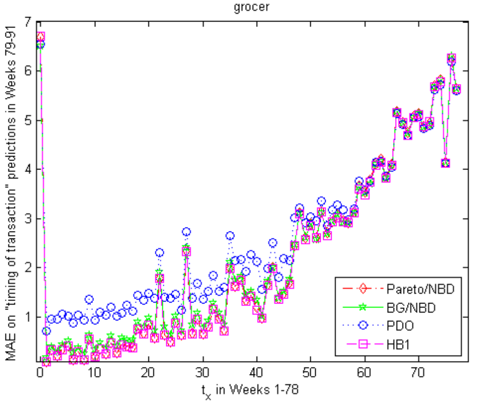


(c)

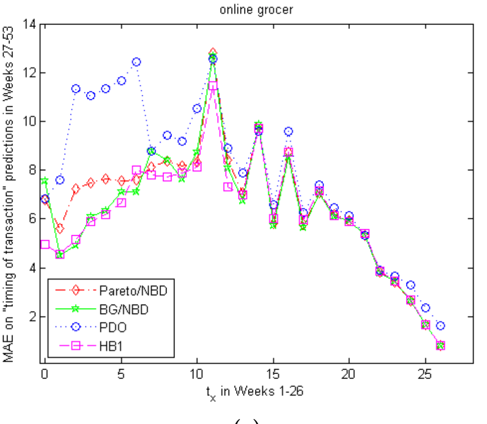
Figure 2.3: Conditional expectation of future transaction timing on CDNOW, grocer and OG datasets



(a)



(b)



(c)

Figure 2.4: MAE of future transaction timing predictions on CDNOW, grocer and OG datasets

the models varies not only with the characteristics of the data, but also with respect to the performance metric.

Managers who aim to forecast their customers' transaction frequency should first examine general characteristics of the customer cohort and then choose the best fitting model. The HB models tend to perform relatively poorly in case data is weak due to many zero-repeat buyers. On the other hand, they do have a clear advantage if there are many repeat buyers and there are significant correlations between the behavioral parameters.

The PDO and HB models perform well on the timing of transaction predictions, again conditional on some data characteristics. Our conclusions on model choice are based on informally relating data characteristics to forecasting performance on just three datasets. There are studies that attempt to formally quantify and validate such relations through classification and regression trees and random forests (Schwartz et al., 2014). Such a formal study is very welcome in this context to arrive at more general recommendations.

By comparing the predictive performance on future frequency versus timing, we found that the BTYD models perform rather poorly on the latter. A closer focus on the defection process may lead to better timing predictions. The ideas of Bueschken and Ma (2012) may be helpful in this context. They provide a new perspective on possible switches between active and inactive states, and allow for both regular and incidental buyers by relaxing the Poisson process assumption on the arrival of transactions.

2.7 Appendix: Timing expressions

In this section, we present the derivations of the expected timing of the last transaction, t_x, in the observation period $[0, T]$ and the expected timing of the next event (either the first purchase or the end of the forecast interval), t_f, conditioned on an individual's parameters. The hyperparameters do not play a role here. In all sections of this appendix we drop the i subscript, representing customer i, for notational simplicity. In the notation we also do not condition on the length of the observational interval T.

2.7.1 Timing of transactions for Pareto/NBD and HB models

The derivations in this section apply to the original Pareto/NBD model and its HB extension. The expressions are the same as both models have the same assumptions on individual behavior. The time of defection, t_Δ, has the probability function[6]

$$\mathbb{P}(\mathrm{d}t_\Delta|\lambda,\mu) = \mu e^{-\mu t_\Delta}\,\mathrm{d}t_\Delta\ . \tag{2.15}$$

Setting $t_\delta = \min(t_\Delta, T)$, we obtain

$$\mathbb{P}(\mathrm{d}t_\delta|\lambda,\mu) = \begin{cases} \mu e^{-\mu t_\delta}\,\mathrm{d}t_\delta & \text{if } 0 \leq t_\delta < T \\ e^{-\mu T}\delta_T(t_\delta)\,\mathrm{d}t_\delta & \text{if } t_\delta = T \\ 0 & \text{otherwise,} \end{cases} \tag{2.16}$$

where $\delta_w(x)$ is the Dirac-delta function at w evaluated at x.[7] Conditioning on the unobserved value t_δ, we find the density of t_x on $(0,T]$ as

$$\mathbb{P}(\mathrm{d}t_x|t_\delta,\lambda,\mu) = \left(\lambda e^{-\lambda(t_\delta - t_x)} + \delta_0(t_x)e^{-\lambda t_\delta}\right)\mathrm{d}t_x, \tag{2.17}$$

where we make use of the memoryless property of the Poisson process. Informally, we can look back in time and do as if the process starts at t_δ. Integrating over t_δ, one obtains

$$\begin{aligned} \mathbb{P}(\mathrm{d}t_x|\lambda,\mu) &= \int_{t_\delta\in[t_x,T]} \mathbb{P}(\mathrm{d}t_x|t_\delta,\lambda,\mu)\,\mathbb{P}(\mathrm{d}t_\delta|\lambda,\mu) \\ &= \begin{cases} \lambda\,\frac{\mu e^{-(\lambda+\mu)t_x} + \lambda e^{-(\lambda+\mu)T}}{\lambda+\mu}\ \mathrm{d}t_x & \text{if } 0 < t_x \leq T \\ \left(\frac{\mu}{\lambda+\mu} + \frac{\lambda e^{-(\lambda+\mu)T}}{\lambda+\mu}\right)\ \delta_0(t_x)\ \mathrm{d}t_x & \text{if } t_x = 0\ . \end{cases} \end{aligned} \tag{2.18}$$

Based on Equation (2.39), the expected value on the time of the last transaction is calculated as follows,

$$\mathbb{E}(t_x|\lambda,\mu) = \int_0^\infty t_x\,\mathbb{P}(\mathrm{d}t_x|\lambda,\mu) = \frac{1-e^{-\mu T}}{\mu} - \frac{1-e^{-(\lambda+\mu)T}}{\lambda+\mu}\ . \tag{2.19}$$

Next, we present the derivations for the predictions of the time of next event from the end of the calibration period conditional on x and t_x: $\mathbb{E}(t_f|x,t_x,\lambda,\mu)$. Let T^+ be some future horizon $T^+ > T$. Consider the first future transaction after T. We define t_f as the

[6] We use a rather formal notation here as our stochastic variables have a mixed discrete/continuous distribution. For practical purposes one can see the part before $\mathrm{d}t_\Delta$ on the right-hand side of (2.15) as the traditional probability density function.

[7] More precisely, $\delta_w()$ is a point mass at w normalized such that for any continuous function g, $\int g(t)\delta_w(t)\ \mathrm{d}t = g(w)$.

time of this occurrence or T^+, whichever is first. We have

$$\mathbb{E}(t_f|x, t_x, \lambda, \mu) = \mathbb{E}(t_f|x, t_x, z = 1, \lambda, \mu)\, p^+ + \mathbb{E}(t_f|x, t_x, z = 0, \lambda, \mu)\, (1 - p^+),$$

where $z = 1$ indicates that a customer is active at time T and

$$p^+ = \mathbb{E}(z|x, t_x, \lambda, \mu) = \frac{\lambda}{\lambda + e^{(\lambda+\mu)(T-t_x)}} \ . \tag{2.20}$$

Consider an active customer; the density of the first timing, t, of a transaction on (T, ∞) is $\lambda e^{-(\lambda+\mu)(t-T)}$ and t has a point mass at infinity of $\frac{\mu}{\lambda+\mu}$ as defection may have been the first event to happen. Therefore, on the interval $(T, T^+]$ the density of t_f given a customer's transaction data and that the customer is active at time T is $\pi_f(t|x, t_x, z = 1, \lambda, \mu) = \lambda e^{-(\lambda+\mu)(t-T)}$. The expectation is computed as,

$$\begin{aligned}
\mathbb{E}(t_f|x, t_x, \lambda, \mu) &= p^+ \int_T^{T^+} t\pi_f(t|x, t_x, z = 1, \lambda, \mu)\, \mathrm{d}t \\
&\quad + p^+ \left(1 - \int_T^{T^+} \pi_f(t|x, t_x, z = 1, \lambda, \mu)\, \mathrm{d}t\right) T^+ + (1 - p^+)T^+ \\
&= T + \frac{\mu e^{(\lambda+\mu)(T-t_x)}}{\lambda + \mu e^{(\lambda+\mu)(T-t_x)}}(T^+ - T) + \frac{\lambda}{\lambda + \mu e^{(\lambda+\mu)(T-t_x)}}\, \frac{1 - e^{-(\lambda+\mu)(T^+-T)}}{\lambda + \mu}
\end{aligned} \tag{2.21}$$

2.7.2 Timing of transactions for BG/NBD model

In the BG/NBD model, the timing of defection, t_Δ, is also the timing of the last transaction and its density is

$$\mathbb{P}(\mathrm{d}t_\Delta|\lambda, p) = \lambda p e^{-\lambda p t_\Delta}\, \mathrm{d}t_\Delta, \tag{2.22}$$

see Fader et al. (2005a). It should be noted that the first purchase at time 0 is special in that a customer cannot defect at time 0. Given that $t_\delta = \min(t_\Delta, T)$:

$$\mathbb{P}(\mathrm{d}t_\delta|\lambda, p) = \begin{cases} \lambda p e^{-\lambda p t_\delta}\, \mathrm{d}t_\delta & \text{if } 0 < t_\delta < T \\ e^{-\lambda p T}\delta_T(t_\delta)\, \mathrm{d}t_\delta & \text{if } t_\delta = T \\ 0 & \text{otherwise .} \end{cases} \tag{2.23}$$

Conditioning on the unobserved value t_δ, we find the density of t_x as

$$\mathbb{P}(\mathrm{d}t_x|t_\delta,\lambda,p) = \begin{cases} \delta_{t_\delta}(t_x)\,\mathrm{d}t_x & \text{if } t_\delta < T \\ \big(\lambda(1-p)e^{-\lambda(1-p)(T-t_x)} + e^{-\lambda(1-p)T}\delta_0(t_x)\big)\,\mathrm{d}t_x & \text{if } t_\delta = T \\ 0 & \text{otherwise .} \end{cases} \tag{2.24}$$

Integrating over t_δ, one obtains the probability

$$\begin{aligned} \mathbb{P}(\mathrm{d}t_x|\lambda,p) &= \int_{t_\delta\in[t_x,T]} \mathbb{P}(\mathrm{d}t_x|t_\delta,\lambda,p)\,\mathbb{P}(\mathrm{d}t_\delta|\lambda,p) \\ &= \big(\lambda p e^{-\lambda p t_x} + (1-p)\lambda e^{-\lambda(T-(1-p)t_x)} + e^{-\lambda T}\delta_0(t_x)\big)\,\mathrm{d}t_x \ . \end{aligned}$$

and, therefore

$$\mathbb{P}(\mathrm{d}t_x|\lambda,p) = \begin{cases} \lambda\left(pe^{-\lambda p t_x} + (1-p)e^{-\lambda T}e^{\lambda(1-p)t_x}\right)\mathrm{d}t_x & \text{if } 0 < t_x \le T \\ e^{-\lambda T}\delta_0(t_x)\,\mathrm{d}t_x & \text{if } t_x = 0 \\ 0 & \text{otherwise .} \end{cases} \tag{2.25}$$

Using Equation (2.25), the expected value of the time of the last transaction in the observation interval $[0,T]$ can be calculated as

$$\mathbb{E}(t_x|\lambda,p) = \int_0^T t_x\lambda\left(pe^{-\lambda p t_x} + (1-p)e^{-\lambda T}e^{\lambda(1-p)t_x}\right)\mathrm{d}t_x = \frac{1}{1-p}\left(\frac{1-e^{-\lambda p T}}{\lambda p} - \frac{1-e^{-\lambda T}}{\lambda}\right) \ . \tag{2.26}$$

For the case $x, t_x > 0$ one easily sees, by referring to the Pareto/NBD result on p^+ in Equation (2.20) under substituting $(1-p)\lambda$ for λ and λp for μ, that

$$p^+ = \mathbb{P}(z=1|x,t_x,\lambda,p) = \begin{cases} \frac{1-p}{1-p+pe^{\lambda(T-t_x)}} & \text{if } x, t_x > 0 \\ 1 & \text{if } x = 0 = t_x \ . \end{cases} \tag{2.27}$$

The density of the first future transaction given the rates, the observed transaction data and the customer being active at T is $\pi_f(t|x,t_x,z=1,\lambda,\mu) = \lambda e^{-\lambda(t-T)}$. Note that an active customer will always make at least one future purchase. The expected value of the first future purchase timing (or T^+) is

$$\begin{aligned}\mathbb{E}(t_f|x,t_x,\lambda,p) &= p^+\lambda\int_{t_f=T}^{t_f=T^+} t_f e^{-\lambda(t_f-T)}\,\mathrm{d}t_f + \left(1-p^+ + p^+e^{-\lambda(T^+-T)}\right)T^+ \\ &= T + (1-p^+)(T^+-T) + p^+\frac{1-e^{-\lambda(T^+-T)}}{\lambda}\ . \end{aligned} \tag{2.28}$$

2.7.3 Timing of transactions for PDO model

In the periodic-defection-model (PDO) (Jerath et al., 2011) the time of defection, t_Δ, has a discrete distribution with support $\{n\tau\}_{n=1,2\cdots}$ which is given as

$$\mathbb{P}(t_\Delta = n\tau|\lambda,p) = p(1-p)^{n-1}, \tag{2.29}$$

where τ can be treated as a known value (estimated using MLE at the customer base level). Let $t_\delta = \min(t_\Delta, T)$ be the time after which no transactions are observed. Given t_δ the distribution of the time, t_x, of the last observed transaction in $[0,T]$ is

$$\mathbb{P}(\mathrm{d}t_x|t_\delta,\lambda,p) = I_{[0,t_\delta]}(t_x)e^{-\lambda(t_\delta-t_x)}\left(\lambda+\delta_0(t_x)\right)\mathrm{d}t_x, \tag{2.30}$$

I_A is the indicator function of the set A. Note the distribution's point mass at 0. One computes

$$\begin{aligned}\mathbb{P}(\mathrm{d}t_x|\lambda,p) &= \int_{t_\delta\in[t_x,T]}\mathbb{P}(\mathrm{d}t_x|t_\delta,\lambda,p)\,\mathbb{P}(\mathrm{d}t_\delta|\lambda,p) \\ &= \left(\sum_{n=m_x}^{N} p(1-p)^{n-1}e^{-\lambda n\tau} + (1-p)^N e^{-\lambda T}\right)\left(\lambda+\delta_0(t_x)\right)e^{\lambda t_x}\,\mathrm{d}t_x, \end{aligned} \tag{2.31}$$

where we use the notations N for $\lfloor T/\tau \rfloor$ and m_x as the time of the first opportunity to defect after or at t_x, expressed as a multiple of τ, that is, $m_x = \left\lfloor \frac{t_x}{\tau}+1\right\rfloor$. Using Equation (2.31) together with the observation that in our case, it holds that

$$\sum_{m=1}^{N}\int_{\{m_x=m\}}\sum_{n=m}^{N}(\,\cdot\,)\,\mathrm{d}t_x = \sum_{n=1}^{N}\sum_{m=1}^{n}\int_{\{m_x=m\}}(\,\cdot\,)\,\mathrm{d}t_x = \sum_{n=1}^{N}\int_{t_x=0}^{t_x=n\tau}(\,\cdot\,)\,\mathrm{d}t_x\,,$$

the expected value for the time of the last observed transaction in the interval $[0, T]$ is found as[8]

$$\mathbb{E}(t_x|\lambda, p) = \sum_{n=1}^{N} p(1-p)^{n-1} \left(n\tau - \frac{1 - e^{-n\lambda\tau}}{\lambda} \right) + (1-p)^N \left(T - \frac{1 - e^{-\lambda T}}{\lambda} \right) . \quad (2.32)$$

Now let us turn to the timing of the first repeat transaction, t_1, where, by convention, we set $t_1 = \infty$ in case there is no repeat transaction after the initial transaction at time 0. More in particular, we study t_1 capped by the observation period's length, $t^+ = \min(t_1, T)$. Then, by analogy to Equations (2.30) and (2.31) we obtain

$$\mathbb{P}(\mathrm{d}t^+|t_\delta, \lambda, p) = \left(I_{[0,t_\delta]}(t^+)\lambda e^{-\lambda t^+} + e^{-\lambda t_\delta}\delta_T(t^+) \right) \mathrm{d}t^+ \quad (2.33)$$

and

$$\begin{aligned}\mathbb{P}(\mathrm{d}t^+|\lambda, p) = &\left(\sum_{n=\lceil t^+/\tau \rceil}^{N} p(1-p)^{n-1}\lambda e^{-\lambda t^+} + (1-p)^N \right) \lambda e^{-\lambda t^+} \, \mathrm{d}t^+ \\ &+ \left(\sum_{n=1}^{N} p(1-p)^{n-1} e^{-n\lambda\tau} + (1-p)^N e^{-\lambda T} \right) \delta_T(t^+) \, \mathrm{d}t^+ . \quad (2.34)\end{aligned}$$

From the density in Equation (2.34), the expected value for the timing of the first transaction becomes

$$\begin{aligned}\mathbb{E}(t^+|\lambda, p, T) = &\sum_{n=1}^{N} p(1-p)^{n-1} \left(\frac{1 - (n\lambda\tau + 1)e^{-n\lambda\tau}}{\lambda} \right) + (1-p)^N \left(\frac{1 - (\lambda T + 1)e^{-\lambda T}}{\lambda} \right) \\ &+ \left(pe^{-\lambda\tau} \frac{1 - \left((1-p)e^{-\lambda\tau}\right)^N}{1 - (1-p)e^{-\lambda\tau}} + (1-p)^N e^{-\lambda T} \right) T\end{aligned}$$

or

$$\begin{aligned}\mathbb{E}(t^+|\lambda, p, T) = 1/\lambda &\left(1 - \sum_{n=1}^{\lfloor T/\tau \rfloor} p(1-p)^{n-1}(n\lambda\tau + 1)e^{-n\lambda\tau} - (1-p)^{\lfloor T/\tau \rfloor}(\lambda T + 1)e^{-\lambda T} \right) \\ &+ \left(pe^{-\lambda\tau} \frac{1 - \left((1-p)e^{-\lambda\tau}\right)^{\lfloor T/\tau \rfloor}}{1 - (1-p)e^{-\lambda\tau}} + (1-p)^{\lfloor T/\tau \rfloor} e^{-\lambda T} \right) T . \quad (2.35)\end{aligned}$$

[8] For reasons of computational efficiency, in cases where N is a large number, the summation in Equation (2.32) may be written as $\frac{\tau}{p}\left(N(1-p)^{(N+1)} - (N+1)(1-p)^N + 1\right) - \frac{1-(1-p)^N}{\lambda} + \frac{pe^{-\lambda\tau}}{\lambda}\frac{\left((1-p)e^{-\lambda\tau}\right)^N - 1}{(1-p)e^{-\lambda\tau} - 1}$.

This expression for the timing of the first transaction in the calibration period is reused for calculating the timing of the first future transaction after T, see Equation (2.12).

2.8 Appendix: Estimation procedure for Pareto/NBD, BG/NBD and PDO models

To calculate the various expectations, we also need draws from the conditional density of the individual-level parameters. Below we discuss how to obtain such draws for the Pareto/NBD, BG/NBD and PDO model.

For the BG/NBD and PDO models, the relevant parameters are the transaction rate, λ, and the probability of defection, p, per defection opportunity. Below, we argue that we can easily draw from the full conditional distributions $\pi(\lambda|x, t_x, p)$ and $\pi(p|x, t_x, \lambda)$. We rely on Gibbs sampling to obtain draws from the joint conditional distribution $\pi(\lambda, p|x, t_x)$.

For the Pareto/NBD model, sampling from the full conditionals is not straightforward. Therefore, we need to develop a different method. We propose to use a random-walk Metropolis-Hastings algorithm to obtain draws from the individual-level posterior distribution.

2.8.1 The Pareto/NBD model

The likelihood function for the Pareto/NBD model is

$$f(x, t_x|\lambda, \mu) = \frac{\lambda^x}{\lambda + \mu}(\mu e^{-(\lambda+\mu)t_x} + \lambda e^{-(\lambda+\mu)T}) \; . \tag{2.36}$$

Given the likelihood function and the independent gamma priors on the defection and purchase rates, the joint posterior distribution of the behavioral parameters can be written as

$$\begin{aligned}\pi(\lambda, \mu|r, \alpha, s, \beta, x, t_x) &\propto f(x, t_x|\lambda, \mu)g(\lambda|r, \alpha)h(\mu|s, \beta) \\ &\propto \frac{\lambda^x}{\lambda + \mu}(\mu e^{-(\lambda+\mu)t_x} + \lambda e^{-(\lambda+\mu)T})\lambda^{(r-1)}e^{-\alpha\lambda}\mu^{(s-1)}e^{-\beta\mu} \; .\end{aligned} \tag{2.37}$$

Note that we consider the hyperparameters (r, α, s, β) to be fixed. The candidate draws in our random-walk Metropolis-Hastings sampler are generated using

$$\begin{aligned}\lambda^c &= \exp(\log\lambda + \varepsilon_\lambda), \qquad &\varepsilon_\lambda \sim N(0, \sigma_\lambda^2) \\ \mu^c &= \exp(\log\mu + \varepsilon_\mu), \qquad &\varepsilon_\mu \sim N(0, \sigma_\mu^2).\end{aligned}$$

In this way we ensure that the parameters always remain positive.

The parameters are now drawn sequentially using the following two-step Gibbs sampler:

1. Start sampling with initial values for λ and μ
2. Update λ
   - Draw the candidate value: λ^c
   - Compute $\alpha = \min\left(1, \pi(\lambda^c, \mu|r, \alpha, s, \beta, x, t_x)/\pi(\lambda, \mu|r, \alpha, s, \beta, x, t_x)\right)$.
   - With probability α, set $\lambda = \lambda^c$
3. Update μ:
   - Draw the candidate value: μ^c
   - Compute $\alpha = \min\left(1, \pi(\lambda, \mu^c|r, \alpha, s, \beta, x, t_x)/\pi(\lambda, \mu|r, \alpha, s, \beta, x, t_x)\right)$.
   - With probability α, set $\mu = \mu^c$
4. Repeat steps 2 and 3.

2.8.2 BG/NBD model

For the conditional posterior distribution of the transaction rate, we have $\pi(\lambda|x, t_x, p) \propto \pi(\lambda, p)\pi(x, t_x|\lambda, p)$ such that

$$\pi(\lambda|x, t_x, p) \propto \lambda^{x+r-1} \times \begin{cases} pe^{-\lambda(t_x+\alpha)} + (1-p)e^{-\lambda(T+\alpha)} & \text{if } 0 < t_x \le T \\ e^{-\lambda(T+\alpha)} & \text{if } x = 0 = t_x \ . \end{cases}$$

We, therefore, have

$$\pi(\lambda|x, t_x, p) = \frac{\frac{p}{(t_x+\alpha)^{x+r}}}{\frac{p}{(t_x+\alpha)^{x+r}} + \frac{1-p}{(T+\alpha)^{x+r}}}\varphi_{x+r,t_x+\alpha}(\lambda) + \frac{\frac{1-p}{(T+\alpha)^{x+r}}}{\frac{p}{(t_x+\alpha)^{x+r}} + \frac{1-p}{(T+\alpha)^{x+r}}}\varphi_{x+r,T+\alpha}(\lambda), \quad (2.38)$$

where $\varphi_{x,\beta}$ is the density of a gamma distribution with shape parameter x and rate parameter β.

Likewise, for the conditional posterior distribution of the defection probability, we have

$$\begin{aligned}\mathbb{P}(\mathrm{d}t_x|\lambda,\mu) &= \int_{t_\delta\in[t_x,T]} \mathbb{P}(\mathrm{d}t_x|t_\delta,\lambda,\mu)\,\mathbb{P}(\mathrm{d}t_\delta|\lambda,\mu) \\ &= \begin{cases} \lambda\,\frac{\mu e^{-(\lambda+\mu)t_x}+\lambda e^{-(\lambda+\mu)T}}{\lambda+\mu}\;\mathrm{d}t_x & \text{if } 0<t_x\le T \\ \left(\frac{\mu}{\lambda+\mu}+\frac{\lambda e^{-(\lambda+\mu)T}}{\lambda+\mu}\right)\,\delta_0(t_x)\;\mathrm{d}t_x & \text{if } t_x=0\ . \end{cases}\end{aligned} \tag{2.39}$$

$$\begin{aligned}\pi(p|x,t_x,\lambda) &\propto \pi(\lambda,p)\pi(x,t_x|\lambda,p) \\ &\propto \begin{cases} p^a(1-p)^{b+x-2}e^{-\lambda t_x}+p^{a-1}(1-p)^{b+x-1}e^{-\lambda T} & \text{if } 0<t_x\le T \\ p^{a-1}(1-p)^{b-1} & \text{if } x=0=t_x \end{cases}\end{aligned}$$

and so

$$\pi(p|x,t_x,\lambda) = \frac{a}{a+(b+x-1)e^{-\lambda(T-t_x)}}\beta_{a+1,b+x-1}(p) + \frac{(b+x-1)e^{-\lambda(T-t_x)}}{a+(b+x-1)e^{-\lambda(T-t_x)}}\beta_{a,b+x}(p) \tag{2.40}$$

where $\beta_{a,b}$ is the density of a beta distribution with parameters a and b.

2.8.3 PDO model

For the conditional posterior distribution of the transaction rate in the PDO model, we get

$$\begin{aligned}\pi(\lambda|x,t_x,p) &\propto \pi(\lambda,p)\pi(x,t_x|\lambda,p) \\ &\propto p\sum_{n=m_x}^{N}\frac{(1-p)^{n-1}}{(\alpha+(n-1)\tau)^x}\varphi_{x+r,\alpha+(n-1)\tau}(\lambda) + \frac{(1-p)^N}{(\alpha+T)^x}\varphi_{x+r,\alpha+T}(\lambda),\end{aligned}$$

so that

$$\pi(\lambda|x,t_x,p) = \sum_{n=m_x}^{N}\frac{w_{x,p}^{(n)}}{W_{x,t_x,p}}\varphi_{x+r,\alpha+(n-1)\tau}(\lambda) + \frac{w_{x,p}^{(N+1)}}{W_{x,t_x,p}}\varphi_{x+r,\alpha+T}(\lambda), \tag{2.41}$$

where

$$w_{x,p}^{(n)} = \begin{cases} p\frac{(1-p)^{n-1}}{(\alpha+(n-1)\tau)^{x+r}} & \text{if } 1\le n\le N \\ \frac{(1-p)^N}{(\alpha+T)^{x+r}} & \text{if } n=N+1, \end{cases}$$

and $W_{x,t_x,p} = \sum_{n=m_x}^{N+1} w_{x,p}^{(n)}$

For the conditional posterior distribution of the defection probability, it holds

$$\pi(p|x,t_x,\lambda) \propto \pi(\lambda,p|x,t_x)$$
$$\propto \pi(\lambda,p)\pi(x,t_x|\lambda,p) \propto p^a \sum_{n=m_x}^{N}(1-p)^{b+n-2}e^{-\lambda(T-(n-1)\tau)} + p^{a-1}(1-p)^{b+N-1}\ .$$

Therefore,

$$\pi(p|x,t_x,\lambda) = \sum_{n=m_x}^{N}\frac{v_\lambda^{(n)}}{V_{t_x,\lambda}}\beta_{a+1,b+n-1}(p) + \frac{v_\lambda^{(N+1)}}{V_{t_x,\lambda}}\beta_{a,b+N}(p), \tag{2.42}$$

is a mixture of beta distributions where

$$v_\lambda^{(n)} = \begin{cases} B(a+1,b+n-1)e^{-\lambda(T-(n-1)\tau)} & \text{if } m_x \le n \le N \\ B(a,b+N) & \text{if } n = N+1\ . \end{cases}$$

and $V_{t_x,\lambda} = \sum_{n=m_x}^{N+1} v_\lambda^{(n)}$ and $B(\cdot,\cdot)$ is the beta function. Note that the value $V_{t_x,\lambda}$ depends on the data only through m_x.

2.9 Appendix: HB estimation with a very diffuse prior on CDNOW dataset

Table 2.15 presents the mean of unconditional expectations for the CDNOW data under a very diffuse prior distribution. Recall that the prior parameters are chosen as $\nu_0 = J+3$ and $\Gamma_0 = \nu_0 I$, where J represents the number of parameters of a customer (see Rossi et al. (2005, Page 30)).

Table 2.15: Average of unconditional expectations in calibration period - under a diffuse prior on CDNOW data

	HB1	HB2	HB3	HB4
Avg. $\mathbb{E}[x]$	0.228	0.096	0.253	0.209
Avg. $\mathbb{E}[t_x]$	2.852	1.110	3.151	2.654

Although a very diffuse prior leads to badly estimated individual-level parameters, this does not necessary lead to bad predictions on the future transaction number and the timing predictions. The main reason for this is that these metrics are bounded. Figure 2.5 and Tables 2.16 to 2.19 show the forecasting performance of the HB models under this very diffuse prior. Hence, it is important to also look at the posterior distributions of the individual-level parameters. As noted earlier, these are very extreme under a diffuse prior for this dataset.

Figure 2.5: Conditional expectation of future transaction frequency and future transaction timing on CDNOW - under a diffuse prior

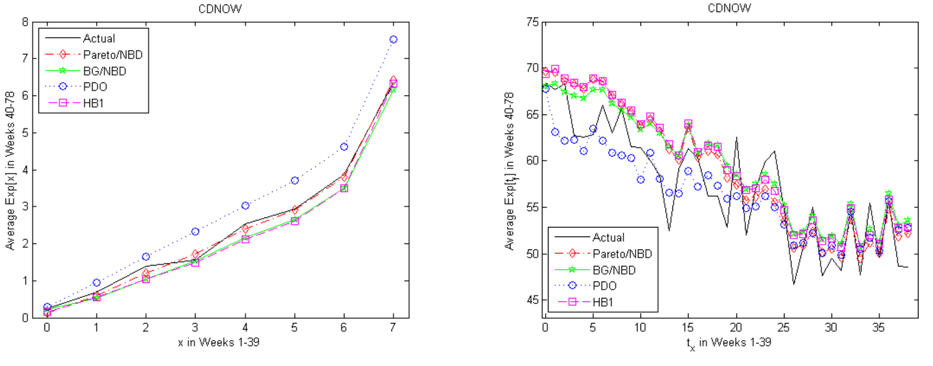


Table 2.16: In-sample predictive performance for unconditional predictions of the expected number of transactions and expected timing of last transaction - under a diffuse prior on CDNOW data

		$\mathbb{E}[x]$				$\mathbb{E}[t_x]$ **-weeks-**			
		MSE	MAE	ME+	ME−	MSE	MAE	ME+	ME−
	HB1	5.454	1.087	0.228	2.369	130.586	7.537	2.747	16.236
CDNOW	HB2	5.689	**1.061**	0.096	2.501	147.785	**7.081**	1.094	16.653
	HB3	5.414	1.092	0.253	2.344	128.279	7.626	3.024	16.172
	HB4	5.486	1.083	0.208	2.388	132.239	7.481	2.556	16.357

Table 2.17: Model's prediction performance on the number of transactions - under a diffuse prior on CDNOW data

		Correlation	MSE	MAE	ME+	ME−
	HB1	0.6245	2.606	0.758	0.413	1.858
CDNOW	HB2	0.6154	2.890	0.748	0.302	1.990
	HB3	0.6185	2.997	0.795	0.523	1.962
	HB4	0.6173	2.744	**0.680**	0.247	2.094

Table 2.18: Highest Posterior Density Region and mean of correlations between behavioral rates - under a diffuse prior on CDNOW data

	$\rho_{\theta_\lambda \theta_\mu}$			$\rho_{\theta_\lambda \theta_\eta}$			$\rho_{\theta_\eta \theta_\mu}$		
	HPDR		mean	HPDR		mean	HPDR		mean
CDNOW	-0.163	0.297	0.078	0.070	0.312	0.188*	-0.868	-0.835	-0.853*

Table 2.19: Model's prediction performance on the time of next transaction - under a diffuse prior on CDNOW data

	Correlation	MSE	MAE	ME+	ME−
HB1	0.5770	126.257	7.502	17.232	-4.052
HB2	0.5538	291.028	16.054	9.423	-17.628
HB3	0.5491	142.779	**6.494**	5.329	-10.314
HB4	0.5665	271.112	15.367	9.053	-16.873

Chapter 3

The Need for Market Segmentation in Buy-Till-You-Defect Models

3.1 Introduction

In recent years, improvements in information technology have enabled firms to record a tremendous amount of data on their customers' transactions. Even small grocery chains record various details associated with each transaction. Simultaneously, recent advances in quantitative techniques, such as Bayesian estimation, bring new opportunities to thoroughly analyze the growing transaction data. Furthermore, companies have started to realize that advanced marketing models can offer detailed customer insights. In a survey of nearly $3,000$ executives, managers and analysts working across more than 30 industries and 100 countries, half of the respondents said that improving information systems and adopting advanced quantitative models are top priorities for their organizations (LaValle et al., 2014). This survey underpins the widespread belief that advanced models offer value to companies. Therefore, managers increasingly adopt these models to enhance their business performance.

In many cases, the use of advanced information systems and marketing models aims at better understanding the customer base and more accurate predictions of customer behavior. Detailed insights in customer behavior and heterogeneity are essential to develop marketing strategies tailored to particular segments or even to specific individuals.

Segmentation and predictive modeling are two must-have tools in today's customer centric landscape. Even though they rely on different sets of techniques that have been studied extensively in marketing, they both support managers to develop customized marketing strategies for each of the target units, namely segments or individual customers.

Marketers have traditionally dealt with customer heterogeneity by segmenting the market (Bhatnagar and Ghose, 2004). Despite the fact that companies are moving toward a marketing era where the only relevant segment is the individual customer, segmentation still offers a lot of value for managers in having an overall understanding of their customer base. Moreover, segmentation forms the very first step toward more advanced one-to-one marketing strategies.

In this paper, we present a customer-level predictive model which also provides an inherent segmentation. This model is relevant for companies operating in a non-contractual setting. In such a setting, customers can stop buying from the company without letting the company know. For instance, the majority of online retailers operates in a non-contractual setting. The unobserved defection of customers adds a big challenge to predicting customer behavior. On the other hand, it needs to be taken into account, if the company wants to generate accurate predictions of individual behavior. Needless to say, more accurate predictions can help to improve returns on marketing actions by better distributing the limited marketing budgets.

Our proposed model is positioned under the so-called Buy-till-you-defect [BTYD] modeling stream. The common modeling approach for these models is to assume stochastic arrival processes (with steady and heterogeneous rates) for each customer's purchase and defection behavior. While a customer is active (the defection has not arrived yet), her transactions arrive according to the assumed arrival process. Usually a Poisson purchase process is assumed as this requires only limited data on a customer's purchase history. On the population level, the heterogeneity over the customer base is modeled by assuming some standard continuous probability distribution.

The current BTYD models have two common weaknesses. The first one is related to their predictive performance. BTYD models provide detailed predictions on the individual's purchase frequency and defection behavior. Especially the predictions on customer defection are key contributions of BTYD models since firms can directly obtain customer lifetime predictions using these models. As we will show in this paper, in many situations these models generate unexpectedly long lifetime predictions for customers. This extreme lifetime prediction problem has also been observed by Wübben and Wangenheim (2008) in their empirical validation study. The failure of BTYD models to deliver what they initially promised lowers the face validity of these models, making it more difficult to get them to be used in practice. A very managerially relevant application of these predictions is calculating the so-called Customer-Lifetime-Value [CLV]. This marketing metric has a central importance for companies. By making good use of CLV, companies can focus on long-term customer satisfaction rather than short-term metrics (Zhang et al., 2014).

The extremely long lifetime predictions indicate that the models could be improved as such predictions are obviously incorrect. From a technical perspective, these extreme predictions may be difficult to explain as it seems to be a counter-intuitive phenomenon for hierarchical models. One may expect that the heterogeneity distribution would shrink outlying customers toward the center of the population. This normally results in fewer extremes. To date, there is not a clear explanation in the literature on the reasons behind the extreme lifetime predictions. Even though there are some models that focus solely on the defection process (Fader et al. (2005a), Jerath et al. (2011)), the lifetime predictions are still not reasonable enough that they can be directly used for managerial decision making.

The second weakness that points toward a potential improvement for BTYD models is their limited descriptive power. More specifically, they lack customer-base level insights that managers can directly act upon. This weakness is at least as important as the former one, since understanding the heterogeneity and identifying behavioral patterns in the customer base is crucial for academic researchers and industry practitioners alike. Researchers and marketing managers are particularly interested in understanding the relationship between various details associated with a transaction and consumer's purchase and defection behavior. For instance, finding a link between consumers' defection rate and their basket size or the paid delivery fee would provide actionable insights. These insights may lead to important managerial implications regarding customized pricing or promotion strategies, issues that are of key interest to companies.

Following the earlier discussion, we pose the following research questions: (1) What are the reasons behind the extreme lifetime prediction problem that limits the adoption of BTYD models; and how can we address this problem? (2) Can a BTYD model also provide insights on segments within the customer base; and is it possible to relate these segments to customer characteristics?

Regarding the first research question, we conduct a detailed simulation study to investigate the reasons behind the extremely long lifetime estimates. Our explanation consists of two parts. First, the data is not very informative on the lifetime of a specific individual. We only observe consumer behavior during a limited time interval and we cannot observe defection directly. Second, the customer base likely contains a number of segments. At least two segments are expected for online retailers: customers who only purchase the service/product a few times, and customers who become frequent buyers. This leads to a multi-modal heterogeneity distribution, which cannot be fitted well using any of the current BTYD models. In fact, under the uni-modal heterogeneity distributions of the existing BTYD models, the variance is forced to be large in order to capture the one-time users as well as the more regular users. The fact that this inflates the customer lifetimes of regular users is not sufficiently penalized through the fit of the model as we only observe the customers for a limited time period. This phenomenon will also lead to biased estimates for individual level parameters. In sum, more attention should be paid to heterogeneity modeling for the BTYD models, especially in cases where multiple customer segments exist.

Based on our findings, we develop a new BTYD model that overcomes the lifetime estimation problem and yields more detailed insights in the customer base. Our model provides an inherent segmentation that segregates customers directly on their purchase and defection parameters. Typically, since defection is not observed, current segmentation models require other covariate data as a proxy for customer defection. In our model, however, we provide a refined segmentation by using predicted behavioral parameters of customers. In other words, the segment membership directly tells us something about customer's purchase frequency and defection. In fact, the shape of our proposed heterogeneity distribution over the behavioral parameters reveals the inherent customer segments. Moreover, our model has the capability of incorporating other available covariate data. This way, one can study whether there exists a substantial statistical relationship between a certain covariate and customer's purchase and defection parameters. By a better understanding of the customer base through the relationship between segments and covariates, the company may also be able to predict the purchase or defection behavior for a new customer based only on the covariate data from her first purchase.

Based on our model building, simulation and empirical studies, our contribution is twofold. First, as our proposed heterogeneity distribution can accommodate multi-modal, heavy-tailed and skewed distributions, we obtain better lifetime predictions than the ones from the benchmark BTYD models, namely the Pareto/NBD model (Schmittlein et al., 1987) and its hierarchical Bayes extension [HB] model (Abe, 2009a). This is especially true for datasets where there exists inherent multimodality. Second, in line with Van Oest and Knox (2011), Reinartz and Kumar (2000), and Schmittlein and Peterson (1994), we show that different customer segments may exhibit different patterns concerning purchase and defection behavior. We also show that other customer shopping characteristics can be linked to this segmentation to gain more insight on the customer base.[1] Using data from an online retailer in a Western European country, we illustrate the added descriptive power of the proposed model.

[1]This extends the results of Van Oest and Knox (2011) who show using a modified BG/NBD model that customer complaints can be indicators of customer defection.

We show that our model not only improves the direct usability of customer lifetime predictions, but also substantially increases the descriptive power of BTYD models with its segmentation scheme. In the literature, predictive models and segmentation have often been used in conjunction with each other. Several studies showed that the accuracy of predictions can be improved by first using certain kinds of segmentation methods (Morwitz and Schmittlein (1992), Chen et al. (2007)), and vice versa (Hwang et al. (2004), Kim et al. (2006)). Our model sets itself apart from these studies by proposing a unifying framework where predictions and segmentation are executed simultaneously and dynamically in a BTYD framework. We believe that the adoption of BTYD models will accelerate as marketing managers obtain not only an actionable segmentation, but also meaningful lifetime predictions for their customers.

In the next two sections we briefly review the BTYD models with a focus on our benchmark Pareto/NBD and HB models. Then we give an initial theoretical analysis of the extreme lifetime prediction problem. In Section 3.4, we present two variants of our proposed Mixture Hierarchical Bayes BTYD model including estimation details. Prediction results from a simulation study showing the contribution of our models are presented in Section 3.5. Section 3.6 presents the results of our empirical study where we compare the newly proposed models to their benchmark models. General conclusions are discussed in Section 3.7.

3.2 BTYD Models

All BTYD models describe the transaction behavior of individuals $i = 1, \ldots, N$ over a time period starting at the first transaction for each individual. As the time of the first purchase of different individuals usually do not coincide, each individual is observed for a different length of time. We measure time relative to the first purchase. Hence, for each customer $t = 0$ corresponds to the time of the first purchase. We denote the total observation time for customer i as T_i.

In BTYD models, customer i remains active for a stochastic and unobserved lifetime which is denoted by $t_{\Delta,i}$. The Pareto/NBD model and the HB model have the same individual level assumptions: The customer makes purchases according to a Poisson process with rate λ_i until the lifetime ends (defection occurs), and her lifetime $t_{\Delta,i}$ has an exponential distribution with rate μ_i. The observed customer data is denoted by the vector $[x_i, t_{x,i}, T_i]$, where x_i represents the number of repeat purchases, and $t_{x,i}$ represents the time of the last observed purchase.[2] Using these distributional assumptions, we obtain[3]

$$\begin{aligned} \text{Prob}(X_i = x|\lambda_i, t_{\Delta,i}, T_i) &= e^{-\lambda_i(t_{\Delta,i}\wedge T_i)}\frac{(\lambda_i(t_{\Delta,i}\wedge T_i))^x}{x!}, \\ \pi(t_{\Delta,i}|\mu_i, T_i) &= \mu_i e^{-\mu_i t_{\Delta,i}}, \end{aligned} \tag{3.1}$$

where $\pi(.)$ denotes a density function. The purchase and the defection rates are assumed to be distributed according to some standard distributions across the population. While Schmittlein et al. (1987) assume two independent gamma distributions for the Pareto/NBD model, Abe (2009a) relaxes the independence assumption by employing a bivariate log-normal distribution in his HB model. This allows for a correlation between purchase and defection rates. In a situation where this correlation is non-zero, the HB model outperforms other BTYD models in terms of forecasting performance (see Chapter 2). In the HB model, it is also possible to incorporate observed customer characteristics. These characteristics for individual i are collected in a $(1 \times R)$ row vector D_i. This vector does not contain a constant. Using the row vector $\theta_i = [\log(\lambda_i), \log(\mu_i)]$ the HB model specifies

$$\theta_i|\beta, \Gamma, \Delta \sim N(\beta + D_i\Delta,\, \Gamma), \tag{3.2}$$

where β is a (1×2) vector of intercepts, Δ is an $(R \times 2)$ matrix of coefficient parameters and Γ denotes a (2×2) variance-covariance matrix.

Both Pareto/NBD and HB models yield extreme lifetime predictions for a substantial group of customers when applied on a dataset from an online grocery retailer from a Western European country.[4] Similarly, Wübben and Wangenheim (2008) obtain exceptionally long lifetime predictions from the Pareto/NBD model on a dataset from an apparel retailer. In our e-grocer data, the HB model performs the best compared to the other BTYD models. This is due to a strong correlation between the purchase and defection parameters in this particular dataset. In the following section where we investigate the reasons behind these estimates, we focus on the superior HB model.

[2]Thanks to the memorylessness property on the inter-arrival time distribution, $[x_i, t_{x,i}, T_i]$ summarizes customer i's full history without loss of information.

[3]The value $(t_{\Delta,i} \wedge T_i)$ is the minimum of $t_{\Delta,i}$ and T_i.

[4]Other established BTYD models such as BG/NBD model (Fader et al., 2005a) and PDO model (Jerath et al., 2011) generate extreme lifetime predictions on this dataset as well.

3.3 An initial investigation of the lifetime prediction problem

To understand whether the extreme lifetime prediction problem stems from an inherent characteristic of the HB model, or from a lack of fit of the model, we conduct an initial simulation study.[5] For this purpose, we generate data exactly matching the assumptions of the model, that is, Poisson arrivals combined with an exponential lifetime for the individuals, and a bi-variate log-normal for the heterogeneity distribution. For now we assume that customer characteristics are not available. The four steps of the data generation process are as follows:

1. Fix the hyper-parameters (β and Γ) to some *known* values:

   We choose the following values, $\beta^*_\lambda = \log(0.08)$ and $\beta^*_\mu = \log(0.04)$.[6] The variance-covariance matrix is chosen to be equal to the identity matrix.

2. Draw behavioral parameters θ^*_i for $i = 1, \dots, N$ according to the heterogeneity distribution:

   Draw $\theta^*_i \sim \pi(\theta_i|\beta^*, \Gamma^*)$ from the multivariate normal distribution. Here we take $N = 1,000$.

3. Draw lifetimes, $t^*_{\Delta,i}$ for $i = 1, \dots, N$ according to the specified lifetime distribution:

   Draw $t^*_{\Delta,i} \sim \pi(t_{\Delta,i}|\theta^*_i)$ from an exponential distribution with rate parameter e^{θ_i} for customer i.

4. Draw the number of repeat transactions x_i and the last purchase time $t_{x,i}$, given an observation period T_i, lifetime $t^*_{\Delta,i}$ and behavioral parameters θ^*_i:

   For $i = 1, \dots, N$, draw $x_i, t_{x,i} \sim \pi(x_i, t_{x,i}|t^*_\Delta, \theta^*_i, T_i)$.[7] We fix the observation period length T_i to 154 days to match the generated data with the real data from the online grocer.

We next apply Markov Chain Monte Carlo [MCMC] simulation to obtain estimates of parameters from the generated data.[8] In this ideal setting we do not find any evidence

[5]All calculations throughout the paper are performed using MATLAB R2011b.

[6]Note that, if no covariate data is used, or in case covariates are mean-centered, β values give the mean of the log behavioral parameters.

[7]See the details of this sampling process in the 5$^{\text{th}}$ step of generating data for MHB model testing (for segmented data) given in Section 3.10.

[8]Details on the MCMC sampler can be found in Abe (2009a) or by simplifying the sampler in Section 3.8

of extreme lifetimes using the HB model. Contrary to common findings on real data, all lifetime predictions are reasonable and they tend to shrink toward the center of the data. Figure 3.1 contrasts the predictions against the true, simulated lifetimes. In the plot on the right hand-side we zoom in on shorter lifetimes where we observe that the HB model can retrieve the true values of the lifetime to a large extent.

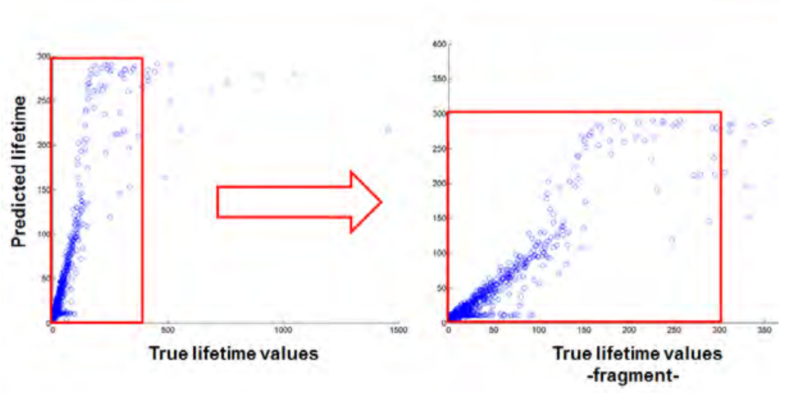


Figure 3.1: Lifetime predictions from the HB model versus true lifetimes on a generated dataset

Based on this simulation study, we conclude that the HB model gives reasonable lifetime predictions if it is applied to a dataset that satisfies all model assumptions. The extreme lifetime predictions that are obtained for real data are, therefore, most likely due to a violation of one of the model assumptions. This conclusion is the very motivation of our paper. We believe that the HB model's fit problem stems from the fact that the log-normal distribution (or the gamma distribution for the Pareto/NBD model) does not accurately capture the true population distribution. The true distribution is likely to be multi-modal, as the population contains various types of customers. The existence of individuals with very short lifetimes leads to a thick right-hand tail of the log defection rate distribution; and due to the symmetry of the normal distribution we also obtain a thick left-hand tail. For the individuals in this part of the distribution, we would erroneously conclude that their defection rate is virtually zero (leading to infinitely long lifetime predictions). All in all, we need to capture the multimodality in the data to avoid drawing wrong conclusions on the customer level.

3.4 Mixture HB BTYD Model

Based on our earlier motivation, we propose to model customer heterogeneity in a way that allows for latent classes where each class corresponds to a different log-normal heterogeneity distribution. We propose two different variants of the Mixture Hierarchical-Bayes BTYD model. In the first variant (hereafter denoted as MHB model), a-priori segment probabilities are independent of customer covariates. In the second (denoted as MHB-C model), we allow covariates to influence the segment probabilities. In the mixture model literature such covariates are called concomitant variables. In principle one would be able to obtain better predictive performance with the MHB-C model that accommodates concomitant variables.

3.4.1 MHB Model

To allow for a multi-modal heterogeneity distribution, we replace the multivariate normal distribution over the log purchase and log defection rates by a mixture of K multivariate normal distributions.[9] One can also view this as a distribution that allows for K segments in the population where there are also within segment differences. The mixture of normals approach provides a great deal of flexibility. First, it may capture a distribution with multiple modes. Next, it could capture a distribution with fat tails if one of the components is a normal component with a large variance. The mixture of normals approach has become quite popular in marketing due to its flexibility and the potential interpretation of each mixture component as representing a 'segment'. Finally, the parameters in these models are relatively easy to estimate (Rossi et al., 2005).

More formally, we write the heterogeneity distribution as

$$\begin{aligned}
\theta_i &= D_i \Delta + \eta_i, \\
\eta_i &\sim \mathrm{N}(\beta_{s_i}, \Gamma_{s_i}), \\
s_i &\sim \mathrm{Multinomial}_K(p),
\end{aligned}$$

where s_i indicates the segment to which customer i belongs. For each segment (or component) we associate a mean vector and a variance-covariance matrix, namely β_k and Γ_k, $k = 1, \ldots, K$. The vector p contains the K segment probabilities where their values sum up to 1.

[9] Data examination shows us that there are generally two major segments in the customer base of grocery e-tailers, namely frequent and incidental buyers. However in the model we present here, we do not fix the number of latent components.

The proposed model is visualized in Figure 3.2.[10] The joint distribution of the observable data and all latent variables and parameters can be decomposed as

$$\begin{aligned}
&\pi(\{(x_i, t_{x,i}), t_{\Delta,i}, z_i, \theta_i, s_i\}_{i=1}^N, \Delta, \{\beta_k, \Gamma_k\}_{k=1}^K, p) \\
&\quad = \prod_{i=1}^N \left[\pi((x_i, t_{x,i})|t_{\Delta,i}, z_i, \theta_i)\,\pi(t_{\Delta,i}|z_i, \theta_i)\,\pi(z_i|\theta_i)\,\pi(\theta_i|\Delta, \beta_{s_i}, \Gamma_{s_i})\,\pi(s_i|p)\right] \times \\
&\qquad\qquad \pi(\Delta)\,\pi(p) \prod_{k=1}^K \left[\pi(\beta_k|\Gamma_k)\,\pi(\Gamma_k)\right]. \quad (3.3)
\end{aligned}$$

The observables are x_i, $t_{x,i}$ and T_i. The variables z_i and $t_{\Delta,i}$ relate to the unobserved defection process. z_i is a latent binary indicator denoting whether customer i is active ($z_i = 1$) or inactive ($z_i = 0$) at the end of the calibration period (T_i). The latent lifetime is given by $t_{\Delta,i}$. The set of values $(x_i, t_{x,i}), (t_{\Delta,i}, z_i), \theta_i, s_i$ are distributed independently across individuals when conditioned on $(\Delta, p, \{\beta_k, \Gamma_k\}_{k=1}^K)$.

As said, D_i is the observable characteristics (covariate) row vector of an individual and does not include an intercept. We follow the advice by Rossi et al. (2005, Page 144) to mean-center all covariates, so that the mean of θ for the average customer is entirely determined by the mixture component means (β_k). Therefore $\mathbb{E}[\theta_i|D_i = \overline{D}, p, \{\beta_k\}_{k=1}^K] = \sum_{k=1}^K p_k\beta_k$.

We choose the standard conditionally conjugate priors to complete the model specification, that is,

$$\begin{aligned}
\text{vec}(\Delta) = \delta &\sim N(\bar{\delta}, \bar{A}_\delta^{-1}), \\
p &\sim \text{Dirichlet}(\alpha), \\
\beta_k|\Gamma_k &\sim \text{N}(\bar{\beta}, \Gamma_k \otimes \bar{A}^{-1}), \\
\Gamma_k &\sim \text{IW}(\bar{\Gamma}, \bar{\nu}).
\end{aligned}$$

IW denotes the Inverse Wishart distribution. A discussion on setting the values of the prior parameters is presented in Section 3.6.

Bayesian inference

The posterior distribution for all parameters and latent variables is not available in closed form. We use MCMC sampling for inference on the parameters and the latent variables

[10] Figure 3.2 helps us to easily identify the direct dependency relationships between neighboring parameters. Note that the joint distribution of the observable data and all latent variables and parameters in Equation (3.3) holds since $(x_i, t_{x,i}, T_i), t_{\Delta,i}, z_i$ are independent of $p, \Delta, \beta_{s_i}, \Gamma_{s_i}$ given θ_i.

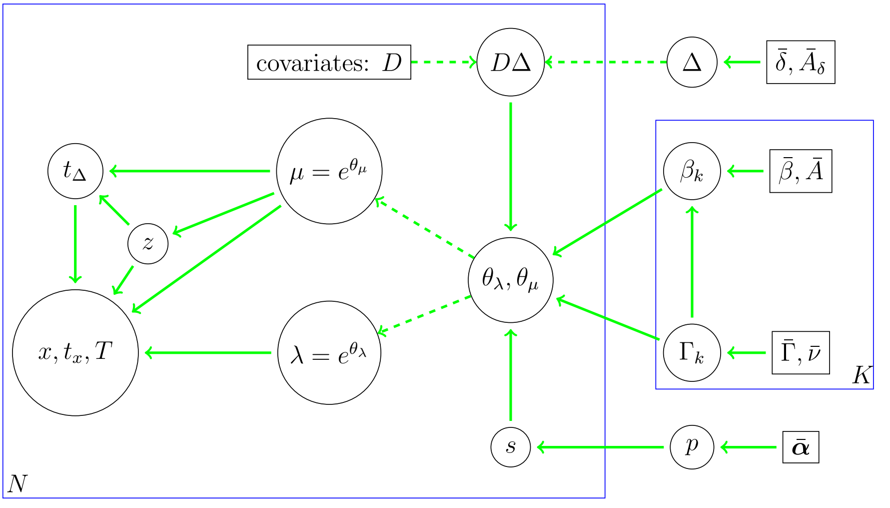


Figure 3.2: MHB model that specifies customer purchase and defection behavior, together with customer heterogeneity. Constant values are enclosed by rectangles. Each variable in the big box is of dimension N, representing each customer. Each value in the smaller box is of dimension K, representing each latent component. The value of the indicator variable $s \in \{1, \cdots, K\}$ picks one out of K components with β_k and Γ_k; $k = 1, \ldots, K$. The covariates, D, are assumed not to include an intercept. The intercept is modeled through β_k. The dashed lines represent deterministic relations.

for the MHB model. More specifically, we use a Metropolis within Gibbs sampler (see Hastings (1970) and Geman and Geman (1984)). The sampler uses the latent variables z_i and $t_{\Delta,i}$. We present the main steps of the sampler below, details of the sampling procedure are given in Section 3.8.
The MCMC sampler for the MHB model is:

[0] Set initial values for $\theta_i, i = 1, \ldots, N$, and repeat the following.

[1a] Generate $z_i | x_i, t_{x,i}, T_i, \theta_i$ according to the being active probability $\frac{\lambda_i}{\lambda_i + \mu_i e^{(\lambda_i+\mu_i)(T_i - t_{x,i})}}$ (as given in Equation (3) in Schmittlein et al. (1987)), for $i = 1, \ldots, N$.

[1b] Generate $t_{\Delta,i} | x_i, t_{x,i}, T_i, z_i, \theta_i$ using an exponential distribution with rate $(\mu_i + \lambda_i)$ truncated to $(t_{x,i}, T_i)$ if $z_i = 0$; and an exponential distribution with rate μ_i truncated to (T_i, ∞) if $z_i = 1$ (see Equation (3.8)).

[2a] Calculate $\tilde{p}_{ik} | \theta_i, D_i, \Delta, \beta_k, \Gamma_k, p_k$, the conditional posterior membership probabilities of customer i for component k using Equation (3.10) in Section 3.8.

[2b] Generate $s_i|\tilde{p}_i$, the indicator variable for the segment to which the customer i belongs by drawing from a multinomial distribution with parameters $\tilde{p}_i = [\tilde{p}_{i1}, \cdots, \tilde{p}_{iK}]$.

[3] Generate $\beta_k|\theta, \Delta, s, \Gamma_k$ and $\Gamma_k|\theta, \Delta, s$ for each latent class k using a multivariate normal regression update (see Rossi et al. (2005, Page 34)). Note that $\pi(\beta_k, \Gamma_k|\theta, \Delta, \{s_i\}_{i=1}^N)$ does not depend on rates θ_i for those customers that do not belong to the component k. Let $\theta^{(k)}$ be the matrix of behavioral parameters for those customers who belong to segment k, that is, $\theta^{(k)} = \{\theta_i\}_{i:s_i=k}$. Then

$$\begin{aligned}\pi(\beta_k, \Gamma_k|\theta, \Delta, \{s\}_{i=1}^N) &= \pi(\beta_k, \Gamma_k|\theta^{(k)}, \Delta) \\ &\propto \pi(\theta^{(k)}, \Delta, \beta_k, \Gamma_k) \\ &= \pi(\theta^{(k)} - D^{(k)}\Delta|\beta_k, \Gamma_k)\, \pi(\beta_k|\Gamma_k)\, \pi(\Gamma_k) \end{aligned} \tag{3.4}$$

[4] Generate $\Delta|\theta, \beta, \Gamma, s$, the regression coefficients over the whole population, using a standard multivariate regression update; $\Delta \sim \pi(\Delta|\theta, \beta, \Gamma, s)$. For this step, the data should be pooled from K components (see Rossi et al. (2005, Page 148)). Details are provided in Section 3.8.

[5] Draw p conditional on $\{s_i\}_{i=1}^N$. This conditional distribution is a Dirichlet, that is, to update on the membership probabilities of the components we use $p|\{s_i\}_{i=1}^N \sim \text{Dir}(\alpha_1 + \sum_{i=1}^N I[s_i = 1], \ldots, \alpha_K + \sum_{i=1}^N I[s_i = K])$, where $I[A]$ denotes an indicator function which equals one if condition A is true, and zero otherwise.

[6] Generate $\theta_i|t_{x,i}, x_i, T_i, z_i, t_{\Delta,i}, \beta_{s_i}, \Gamma_{s_i}$ with a Gaussian random-walk Metropolis Hastings [MH] algorithm, for $i = 1, \ldots, N$. The step size in the random-walk MH algorithm is set by applying an adaptive MH method in the burn-in phase (Gilks et al., 1996).

3.4.2 MHB-C Model with Concomitant Variables

In the previous section, the prior segment probability was equal for all customers. This implies that without purchase histories we cannot distinguish the different types of customers. In this section we extend the MHB model using concomitant variables such that the prior segment probabilities depend on customer characteristics.

We replace the common vector p by an individual specific vector p_i. To relate these probabilities to customer characteristics we build on the multinomial probit [MNP] model. As is common in the MNP model we introduce latent customer specific "utilities" for each segment. These utilities are denoted by u_{ik}, for $i = 1, \dots, N$ and $k = 1, \dots, K$, and they may depend on the concomitant variables C_i as

$$u_{ik} = C_i \omega_k + \varepsilon_{ik}, \tag{3.5}$$

where $\varepsilon_{ik} \sim N(0, 1)$ and C_i contains a constant next to L concomitant variables. Finally we set ω_K to a vector of zeros (with length $(L + 1)$) for identification (Paap and Franses, 2000). Given the utilities, the segment to which a customer belongs is completely determined. The customer is assigned to the segment that has the highest utility, that is,

$$s_i = \text{argmax}_k u_{ik}. \tag{3.6}$$

The MHB-C model is visualized in Figure 3.3. Every relationship in Figure 3.3 is defined in terms of probability distributions (solid arrows) or in a deterministic way (dashed arrows). Note that the probabilities of belonging to a segment depend on the distribution of the utilities. This latter distribution is a function of the MNP model's coefficients $\omega_1 \dots, \omega_K$.

The joint distribution of the data and parameters now becomes,

$$\begin{aligned} &\pi(\{(x_i, t_{x,i}), t_{\Delta,i}, z_i, \theta_i, s_i, u_i\}_{i=1}^N, \Delta, \{\beta_k, \Gamma_k\}_{k=1}^K, \omega) \\ &= \prod_{i=1}^N [\pi((x_i, t_{x,i})|t_{\Delta,i}, z_i, \theta_i)\, \pi(t_{\Delta,i}|z_i, \theta_i)\, \pi(z_i|\theta_i)\, \pi(\theta_i|\Delta, \beta_{s_i}, \Gamma_{s_i})\, I[s_i = \text{argmax}_k u_{ik}]\, \pi(u_i|\omega)] \\ &\qquad \times \pi(\Delta)\, \pi(\omega) \prod_{k=1}^K [\pi(\beta_k|\Gamma_k)\, \pi(\Gamma_k)], \end{aligned} \tag{3.7}$$

where $u_i = (u_{i1}, \dots, u_{iK})$ and $\omega = (\omega_1, \dots, \omega_K)$. Both in Equation (3.3) and Equation (3.7), the dependence of densities on prior parameters has been suppressed.

Bayesian inference

We again use a Metropolis within Gibbs sampler to obtain the posterior conditional densities for each of the parameters. Note that to satisfy the *irreducibility* requirement of the Markov chain the sampler needs to skip the deterministic relationships between parameters. Therefore, we do not sample the segment indicators s_i; these are determined through the utilities u_{ik} as in Equation (3.6).

The resulting sampler is very similar to the one for the previous model. The only difference is in the assignment of customers to different latent components. Therefore,

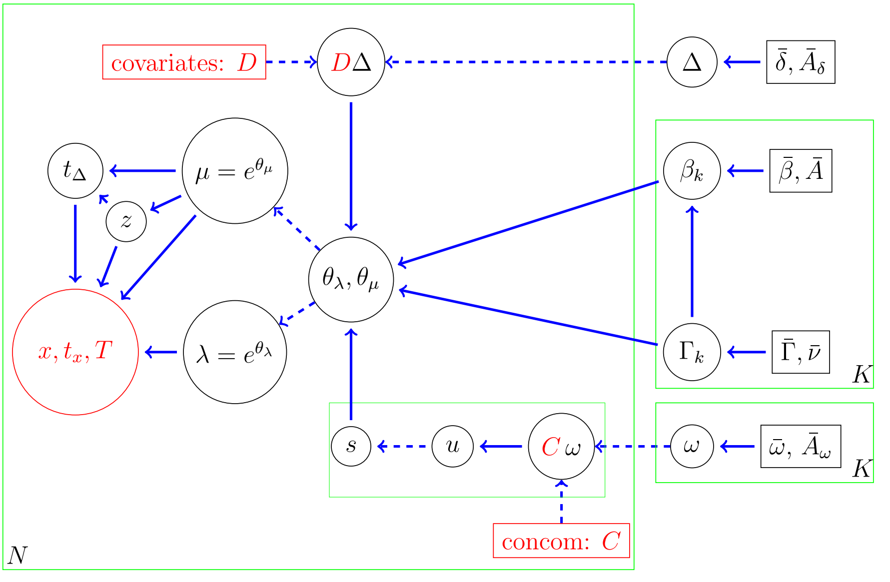


Figure 3.3: MHB-C model with concomintant variables. Constant values are enclosed by rectangles. Each variable in the big box is of dimension N, representing each customer. Each data structure in the smaller boxes on the right hand side of the figure is of dimension K, representing different latent components. The matrices of the inner box are of dimension $(N \times K)$. The dashed line represents a deterministic relation rather than a probabilistic one.

only the second and the third steps of the Gibbs Sampler are different in this sampler. In these steps we update the utility values for each customer and the component-specific probit coefficients ω. The other steps of the sampler are identical to those given under MHB model. The MCMC sampler of the MHB-C model becomes:

[0] Set initial values for $\theta_i, i = 1, \ldots, N$, and repeat the following.

[1a] Generate $z_i | t_{x,i}, x_i, T_i, \theta_i$.

[1b] Generate $t_{\Delta,i} | t_{x,i}, x_i, T_i, z_i, \theta_i$.

[2a] Generate $u_i | C_i, \omega, D_i, \Delta, \theta_i, \beta, \Gamma$, the utility row vector of customer i for the latent segments.

[2b] Update the segment indicators $s_i | u_i$ that assign customers to one of the K components according to the component that has the highest utility value.

[3] Generate $\omega|u$, the latent component specific coefficients using a standard multivariate normal regression update.

[4] Generate $\beta_k|\theta, \Delta, s, \Gamma_k$ and $\Gamma_k|\theta, \Delta, s$ for each latent class k.

[5] Generate $\Delta|\theta, \beta, \Gamma, s$ using a standard multivariate update after pooling data from K components.

[6] Generate $\theta_i|t_{x,i}, x_i, T_i, z_i, t_{\Delta,i}, \beta_k, \Gamma_k$ with a Gaussian random-walk MH algorithm.

The details of the sampling procedures for the nodes ω and u are presented in Section 3.9.

3.5 Model Testing on Generated Data

In order to evaluate the performance of the proposed BTYD models with heterogeneous latent classes, we start by testing them on generated datasets. We generate data based on some *known* parameter values and next see whether we can retrieve those values using the models. This also provides a test to see if our implementation of the MCMC sampler is done properly and converges fast. This approach is especially crucial as some events are unobservable. In our case the segment allocation and the actual lifetime are not observable in a real-life setting. Furthermore, we assess the effects of misspecification, that is, using HB instead of MHB model.

We present the data generation process and some statistics on the generated dataset in Section 3.5.1. Following that, we present the prediction performance of each model under comparison (MHB, MHB-C and HB models). In Section 3.5.3, we give a robustness analysis of the proposed models by testing all models' predictive performance on a generated data with a unimodal heterogeneity distribution.

3.5.1 Data Generation

Considering $N = 1,000$ customers and $K = 2$ latent components, we generate a transaction dataset for $T = 200$ days following three major steps. Details of the data generation, including the exact parameter values, are given in Section 3.10.

1. Allocate customers to components $(s_i^*|\omega^*)$:

Fix the component specific regression coefficient matrix to its *true* value ω^*; generate *true* utilities such as $u^* = C\,\omega^* + \varepsilon$, where $\varepsilon \sim \mathrm{N}(0,1)$; and assign each customer to the component with the the highest utility.[11]

2. Generate customer specific behavioral parameters $\theta_i^*|\beta_{s_i}^*, \Gamma_{s_i}^*$:

   Fix the true hyper-parameter values β^* and Γ^* for each of the components; generate *true* behavioral parameters for each customer by sampling from a MVN distribution such as $\theta_i^* \sim \pi(\theta_i|\beta_k^*, \Gamma_k^*)$.

3. Generate customer lifetime $(t_\Delta^*|\theta^*)$ and transaction data $((x, t_x)|\theta^*, t_\Delta^*, T)$:

   Draw $t_{\Delta,i}^* \sim \pi(t_{\Delta,i}|\theta_i^*)$ from an exponential distribution with the rate parameter of $\theta_{\mu,i}^*$. Given an observation period T and lifetime $t_{\Delta,i}^*$, generate number of transactions and the time of the last purchase based on Poisson purchase arrivals.[12]

The data generation is in line with Section 3.3, apart from the segmentation of customers. We generate one covariate (D) from a standard uniform distribution. As we mean-center all covariate data, it does not affect the mean values of the (component-specific) hyper-parameters. We also generate a concomitant variable (C). In order to keep things simple, for the first half of the data, the concomitant variable is set to 1 and for the other half to -1. Note that randomness is introduced on customers' assignment to components by the utility generation in the first step of generating data.

Table 3.1 shows some descriptive statistics on the generated data. In this dataset we can easily distinguish the two different components, namely Segment 1 with loyal customers and Segment 2 with customers who quickly stop buying. The final two rows show that the concomitant variable cannot perfectly determine the segment allocation.

Table 3.1: Descriptive statistics on the generated data with two components

	All customers	Segment 1	Segment 2
# of customers	1000	528	472
Avg. # of transaction (x)	126.79	238.94	1.34
Std. # of transaction (x)	215.36	247.38	0.80
Avg. last purchase time (t_x)	94.03	171.82	7.01
Std. last purchase time (t_x)	92.62	55.12	20.87
% concomitant (1)	50	68	29
% concomitant (-1)	50	32	71

[11] We fix ω^*, the $((L+1)\times K)$ MNP probit coefficient matrix to $\left[\begin{smallmatrix}0.1 & 0\\ 0.8 & 0\end{smallmatrix}\right]$ where $L = 1$ is the number of concomitant variables.

[12] See the details of sampling process $x, t_x|\theta^*, t_{\Delta,i}^*, T$ in the 5th step given in Section 3.10.

3.5.2 Model Evaluation

In this section we compare the predictive performance of the three models: the HB model proposed by Abe (2009a), the MHB model, and the MHB-C model. We run all the models on the generated data and compare the results on both population and individual levels. For all the hierarchical Bayes models under comparison, the MCMC simulation has run $200,000$ iterations of which the last $40,000$ (with a thinning factor of 10) have been used for posterior inference. Markov chain convergence was monitored using trace plots of posterior draws.

Population level comparison

The MHB and MHB-C models can directly be compared to each other as they both specify two segments. However, the HB model cannot directly be compared with the mixture models on the population level due to a smaller number of parameters. We report the true values of segment specific intercept vectors (β_k^*) as well as the posterior mean predictions from the MHB, MHB-C and HB models in Table 3.2. The values in parentheses give the posterior standard deviation for each parameter. The second and the third rows of Table 3.2 presents the posterior means and standard deviations of the segment specific intercepts (β_k) from the MHB and MHB-C models respectively. These mean β_k values give the population level means of the behavioral parameters (θ vector) for each segment. Based on these two rows, we conclude that both the MHB and MHB-C models perform well in recovering the *true* parameter values presented in the first row. As expected, the mean estimates for the HB model (presented in the last row of the same table) are in between the mixture model's segment specific estimates.

The true values of segment specific variance-covariance matrix Γ_k^*, and the posterior mean of its predictions from the MHB, MHB-C and HB models are presented in Table 3.3. Again as the HB model accommodates only one component, there is only one variance-covariance matrix prediction from this model. The most striking result from these tables is the huge difference in the variance of the log defection rate across the models (see the $\Gamma_{2,2}$ values). This already hints at a potential cause of extreme lifetime predictions. We will further discuss this in the next section.

Individual level comparison

We next compare the model configurations based on their individual level predictions. We focus on the predictions of the purchase and defection rates as well as the predicted lifetime. We measure the predictive performance using the mean absolute error [MAE]

Table 3.2: True (segment specific) intercept vectors (β_k) and their posterior means from MHB, MHB-C and HB models on generated data. As the HB model accommodates one mode, there is only one β prediction from this model. Note that the first element of β is the mean of log purchase rates (θ_λ), and the second is the mean of log defection rates (θ_μ).

	β_1		β_2	
TRUE	0	−6.908	−4.605	−2.996
MHB	−0.033 (0.039)	−6.906 (0.169)	−4.585 (0.232)	−2.972 (0.230)
MHB-C	−0.016 (0.036)	−6.878 (0.147)	−4.687 (0.192)	−2.976 (0.172)
HB	−1.357 (0.095)	−4.248 (0.203)	-	-

Table 3.3: True (segment specific) variance-covariance matrices (Γ_k) and their posterior mean from MHB, MHB-C and HB models on generated data. As HB model accommodates one mode, there is only one Γ estimates from this model.

	Γ_1	Γ_2
TRUE	$\begin{pmatrix} 0.640 & 0 \\ 0 & 0.640 \end{pmatrix}$	$\begin{pmatrix} 0.640 & 0 \\ 0 & 0.640 \end{pmatrix}$
MHB	$\begin{pmatrix} 0.670 & 0.044 \\ 0.044 & 1.250 \end{pmatrix}$	$\begin{pmatrix} 0.837 & 0.113 \\ 0.113 & 0.748 \end{pmatrix}$
MHB-C	$\begin{pmatrix} 0.677 & 0.028 \\ 0.028 & 1.190 \end{pmatrix}$	$\begin{pmatrix} 0.864 & 0.057 \\ 0.057 & 0.787 \end{pmatrix}$
HB	$\begin{pmatrix} 2.76 & -5.28 \\ -5.28 & 19.04 \end{pmatrix}$	-

and the correlation between the predicted and the true values. Table 3.4 summarizes the results.

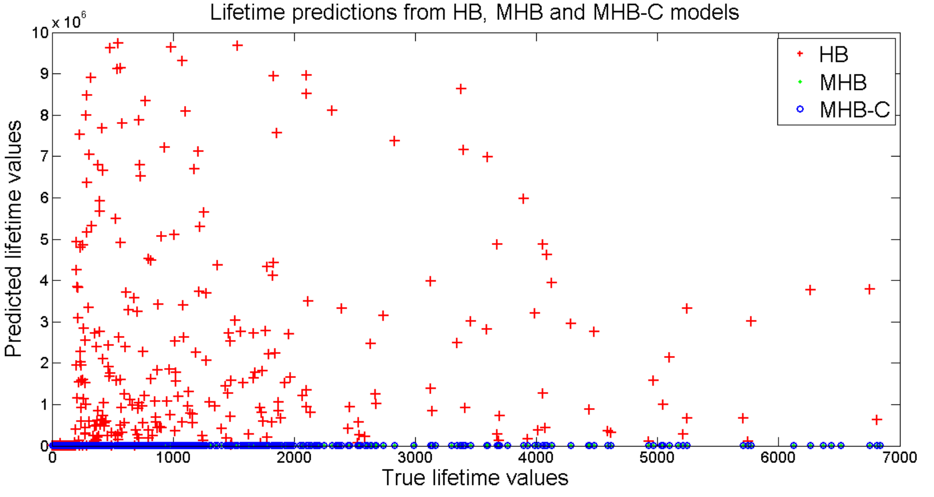
Table 3.4: Comparison of the models on individual metrics (MAE and correlation between true values and predicted means) on generated data

		HB	MHB	MHB-C
Purchase rate (λ)	MAE	0.086	0.045	0.044
	CORR	0.996	0.997	0.997
Defection rate (μ)	MAE	108,658	0.024	0.023
	CORR	0.035	0.547	0.549
Lifetime	MAE	77,381,052	902	852
	CORR	0.026	0.523	0.526

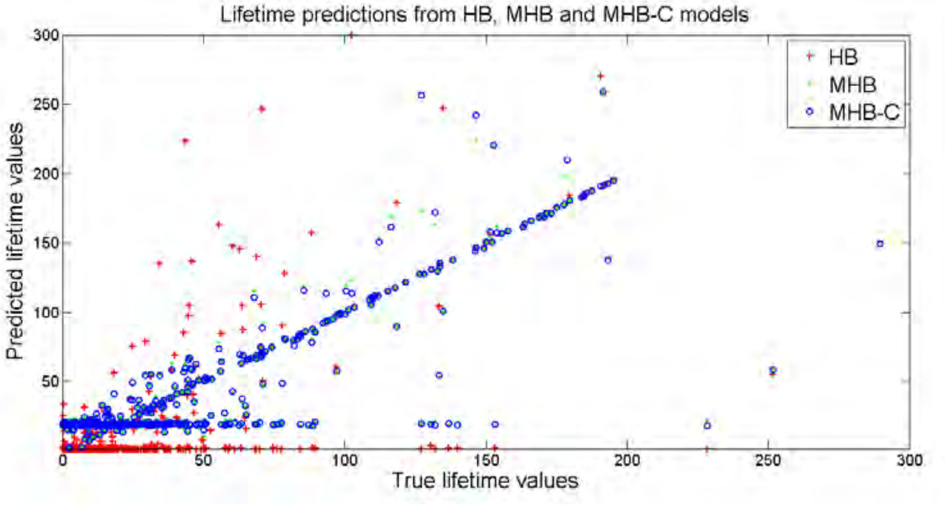
Note that 99.9% of the customers are assigned to their true components for both MHB models.

Table 3.4 shows that all models perform relatively well on predicting the purchase rate λ. Although the MAE for the HB model is about twice as large as that for the MHB and MHB-C models. When it comes to predicting the defection rate μ and the lifetime, there are enormous differences between the HB and the MHB models. Both MHB and MHB-C models predict these measures relatively well, especially considering the fact that we cannot observe the defection. The performance of the HB model clearly demonstrates the earlier mentioned phenomenon of extreme predictions. The predictive performance on the lifetime is illustrated in Figure 3.4a where it is very easy to observe the extremely long lifetime predictions for the HB model. Figure 3.4b gives a small fragment of Figure 3.4a where the axes are limited to the 0 to 300 range. The lifetime predictions based on the HB model hardly show a relation with the true values.

The conclusion from these experiments is quite clear. The MHB and MHB-C models perform well on data representing multiple customer segments. Assuming a unimodal heterogeneity distribution as is done in the HB model can lead to very poor predictive performance on defection and lifetime. In fact the performance is so poor that we observe very extreme lifetime predictions, and hardly any relation with the actual lifetime. This confirms our reasoning that such extreme predictions in earlier applications of BTYD models are due to multimodality. We will further investigate this on real data in Section 3.6.



(a) Scatter plot showing extreme lifetime predictions from HB model. Note the difference in scale on the axes.



(b) A small fragment of the upper scatter plot - axes limited to 300.

Figure 3.4: Scatter plots showing the difference in customer lifetime predictions between HB and MHB models on generated data.

3.5.3 Robustness Analysis on MHB and MHB-C Models - Testing on unimodal data

We also study the performance of the MHB and MHB-C models relative to the HB model in case the customer base has a unimodal heterogeneity distribution. For this purpose, we have generated new data.[13] Table 3.5 shows some statistics on this data.

Table 3.5: Descriptive statistics on the (uni-modal) generated data

# of customers	1000
Avg. # of purchases	5.613
Std. # of purchases	8.965
# of customers with no repeat purchase	367
Avg. last purchase time (t_x)	26.085
Max. last purchase time (t_x)	153.92
Observation time (T)	154

Tables 3.6 and 3.7 present the posterior means of the population level parameters from the three models together with the true parameter values. Based on these tables, we conclude that if the proposed MHB and MHB-C models are applied to a dataset where the heterogeneity distribution is unimodal, the estimates are not deteriorated. All customers are simply assigned to one component, leaving the other empty. As a result the predictive performance of the MHB models is only slightly worse than that of the HB model, see Table 3.8. This loss in predictive performance can entirely be attributed to the fact that MHB and MHB-C models contain more parameters.

Table 3.6: True values and posterior means of β using MHB, MHB-C and the HB models. As the second component from MHB models becomes empty, β_2 values are not reported.

	β	
TRUE	−2.526	−3.219
MHB (β_1)	−2.420	−3.357
	(0.064)	(0.076)
MHB-C (β_1)	−2.483	−3.293
	(0.064)	(0.067)
HB	−1.357	−4.248
	(0.106)	(0.115)

[13] We fix $\beta_\lambda = \log(0.08)$ and $\beta_\mu = \log(0.04)$. The variance covariance matrix Γ is chosen to be equal to the identity matrix.

Table 3.7: True values and posterior means of Γ using MHB, MHB-C and HB models. As the second component from MHB models becomes empty, Γ_2 values are not reported.

	Γ
TRUE	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$
MHB (Γ_1)	$\begin{pmatrix} 1.040 & 0.052 \\ 0.052 & 0.991 \end{pmatrix}$
MHB-C (Γ_1)	$\begin{pmatrix} 0.996 & -0.017 \\ -0.017 & 0.990 \end{pmatrix}$
HB	$\begin{pmatrix} 1.095 & -0.043 \\ -0.043 & 0.947 \end{pmatrix}$

Table 3.8: Comparison of models on individual metrics (MAE and correlation between true values and predicted means) on generated data

		HB	MHB*	MHB-C**
Purchase rate (λ)	MAE	0.061	0.073	0.062
	CORR	0.849	0.798	0.848
Defection rate (μ)	MAE	0.040	0.040	0.042
	CORR	0.376	0.372	0.343
Lifetime	MAE	17.165	17.293	17.448
	CORR	0.783	0.782	0.769

* 99.7% of the customers is assigned to Component 1.
** 100% of the customers is assigned to Component 1.

3.6 Empirical Study

In this section, we test our MHB-C model on real-life data.[14] We first present the explanatory contribution of the MHB-C model by revealing the segments in the customer base as well as by showing how these segments differ from each other. Next, we compare the predictive performance of the MHB-C model against benchmark models. In this section, we consider both the Pareto/NBD model and the HB model. To provide a fair judgment on the performance of the models in consideration, we focus on out-of-sample predictive power.

The dataset we consider contains daily transaction data of an online grocery retailer (called OG hereafter) in a Western European country. We base our analysis on a random set of 1460 customers who started buying from the company in January 2009. We ignore all Sundays as OG does not provide delivery on that day. The data contains the initial and the repeat purchase information of each customer over a period of 309 days. To estimate the model parameters, we use the transaction data of all customers over the first 154 days, leaving a 155 day holdout period for model validation. The transaction data contains information on the number of shopping items, the Euro values of the shopping basket and the delivery fee, the number of discounted items in the basket and also the percentage discount rate of each basket. Table 3.9 presents some descriptive statistics. According to this table an average customer purchases 11 times in the calibration period. However, this number drops to 9 in the validation period mostly because of customers who have left the company by then. On average, the first transaction of customers contains a basket made up of 64 items of which 6 come with a discount. The average initial basket is worth 126 Euros after discount and the delivery fee is 7 Euros.

We use the number of items in the basket together with the basket value and the delivery fee from the initial purchase as explanatory factors in our MHB-C model. These variables are used as covariate and as concomitant variables. We standardize the covariate vector so that the β_k vector represents the average values of the log of the purchase and defection rate for the k^{th} component. Moreover, we applied a log transformation on the number of items in the initial shopping basket as this variable is highly skewed.[15]

There are two points that one needs to pay attention when applying the MHB model. The first concerns the number of segments, i.e. latent components, in the customer base.

[14] We do not include the MHB model in this section for two reasons. First of all, this model is dominated by the MHB-C model due to lack of ability to explain how the segments differ from each other. Second, in order to provide a concise overview of the predictive results from the models in comparison, we include only the MHB-C model together with the benchmark Pareto/NBD and HB models.

[15] Our computational experiments revealed that a highly skewed covariate might cause very unstable estimations.

Table 3.9: Descriptive statistics for the OG dataset

# of customers	1460
Available time frame	309 days
Time split (in-sample/out-of-sample)	154/155
Zero repeaters in estimation period (%)	174 (12%)
Zero repeaters in holdout period (%)	295 (20%)
Zero repeaters in estimation and holdout periods (%)	135 (9%)
# of purchases in estimation period (all)	16,252
# of purchases in holdout period	12,827
Avg. # purchases per customer in estimation period (std.)	11.13 (10.76)
Avg. # purchases per customer in holdout period (std.)	8.79 (10.78)
Avg. observation time T (std.)	143.76 (7.39)
Avg. recency rate $((T - t_x)/T)$	0.27
Avg. # of items in the first purchase (std.)	64.34 (40.67)
Avg. # of discounted items in the first purchase (std.)	5.93 (8.14)
Avg. basket value after discount -in €- (std.)	125.73 (71.51)
Avg. discount rate of the basket (%)	4.08%
Avg. delivery fee of the first purchase -in €- (std.)	6.97 (1.37)

To set the number of mixture components, we run the MHB-C model with different numbers of latent components and choose the optimum one based on the number of customers assigned to each component (Frühwirth-Schnatter, 2006). If additional segments become too small, we stop adding segments. We do not use likelihood-based measures as obtaining the marginal likelihood is computationally very challenging, even in the basic BTYD model. As an alternative, one may choose the number of segments based on out-of-sample predictive performance. However, in our case we would then have to split our data in three parts, to leave one part for a fair comparison against the alternative HB model. Although there is a growing literature on Bayesian analysis of mixtures when the number of components are unknown (Richardson and Green (1997), Stephens (2000), Hurn et al. (2003), Dellaportas and Papageorgiou (2006), Nobile and Fearnside (2007)), we leave this issue for further research.

Secondly, in order to apply the MHB model, we need to set the prior parameters. In many Bayesian applications, the prior is chosen to be uninformative by setting a very large variance so that the prior will not affect the posterior. However, for the MHB-C model, setting a very diffuse prior on the Γ_k has a major impact on the posterior distribution of behavioral parameters as well as on the group membership parameters. We, therefore, set $\nu_0 = J + 30$ and $\Gamma_0 = \nu_0 I$, where $J = 2$ represents the number of behavioral parameters for an individual customer (see Rossi et al. (2005, Page 150)). We have carried out a simulation study where we set different prior degrees of freedom. The results confirm that setting a too diffuse prior leads to unstable estimates. Setting the prior degrees of freedom to $J + 30$ seems to be informative enough to obtain stable results without the prior influencing the posterior results too much.

To obtain posterior results, we apply our Metropolis within Gibbs sampler as presented in Section 3.4.2. The MCMC steps are repeated for $400,000$ iterations of which the last $40,000$ were used to infer the posterior distribution of parameters. Convergence was monitored visually and checked with the Geweke test (Geweke et al. 1991). For each of the hyper-parameters, the Geweke convergence diagnostic concludes that the two non-overlapping parts of the Markov chain[16] are from the same posterior distribution.

For our dataset from OG, we end up with two segments, with a customer share of 41% and 59%. When we increase the number of components to three, one of the component covers only 4 customers, while the others contain the rest in a balanced share. Similarly for the four-component case, the two additional components together cover only 1% of the whole customer base. A detailed discussion of the results from MHB-C model with

[16] We chose the two non-overlapping parts of the Markov chain as the first 0.1 proportion of the chain just after the burn-in iterations and the last 0.5 proportion of the chain.

three or four segments is presented in Section 3.11. One noteworthy conclusion is that the MHB-C model with two latent components gives better out-of-sample predictions than the ones with three or four latent components on this dataset. In general one may also expect to find two major segments: the frequent buyers and those who try the service only a couple of times and quit very early.

We first investigate the differences between the two identified segments. To this end we first allocate each individual to one of the segments based on the posterior segment membership probabilities. Next we take a look at descriptive statistics of the resulting two groups. Table 3.10 shows these statistics. The first component (41%) clearly contains customers who buy more frequently (on average 19.3 times) and more recently from the company. The difference between the end of the observation period and the last purchase time is evidently much higher for the second component (59.93 vs. 7.30 days as 'average recency' as Table 3.10 shows). Conversely, the customers in the second component ordered only a couple of times (on average 3.75 times) and these orders took place a long time ago. Next to the differences between segments on shopping frequency (x) and recency ($T - t_x$), we gain further insight on the additional variables. We see clear differences between segments on characteristics of the first purchase, that is, the average number of shopping items, average basket value, average delivery fee and the average number of discounted items. It seems that the frequent buyers on average have smaller initial shopping baskets both in value and in number of items, and pay higher amounts for the delivery of their first purchase. We can, therefore, conclude that these customers are less price sensitive as they do not mind to pay a high delivery fee. The lower average discount rate on their baskets reveals the same fact as well. On the other hand, there is a major group of customers who uses the service provided by OG to buy once in a while in bigger quantities. These customers tend to pay less in delivery fees and they seek more discount. On this particular dataset, we clearly see two distinct segments in the customer base with different willingness to pay on home delivered groceries. All in all, besides providing predictions on purchase frequency and customer lifetime like the other BTYD models do, our proposed MHB-C model further provides an inherent segmentation where we can distinguish segments also on additional variables. Below, we elaborate on the difference between segments by considering the posterior results for the regression coefficients (ω) appearing in the segment membership MNP model.

The MHB-C model allows us to make inference on the differences across the segments based on the concomitant variables. We have included three concomitant (and covariate) variables, namely the log number of items, basket value and delivery fee from the initial purchases of customers. Table 3.11 shows the posterior mean and 95% highest posterior

Table 3.10: Descriptive statistics on the two segments obtained from MHB-C model

	Segment 1	Segment 2
# of customers	599	861
% of customers	41.03	58.97
Avg. observation time T	147.33	141.27
Avg. last purchase time t_x	140.04	81.34
Avg. recency $(T - t_x)$	7.30	59.93
Avg. # of purchases x	19.31	3.75
Avg. # of items in the basket	59.75	67.54
Avg. basket value (in €)	106.06	139.41
Avg. delivery fee (in €)	7.19	6.81
Avg. # of discounted items	5.03	6.56
Avg. discount rate of basket (%)	0.03	0.05

density region (HPDR) for the coefficients ω in the MNP choice model. Based on the highest posterior density region from the posterior draws on ω, we conclude that components substantially differ from each other on all of the concomitant variables included. Table 3.11 confirms the conclusions from Table 3.10 such as that Segment 1 is less likely than Segment 2 (at the average value of the concomitant variables) through the negative intercept (−0.435), and the customers from the first component buy in smaller amounts and pay higher fees.

Table 3.11: Posterior mean and 95% highest posterior density region on ω

	Mean ω_1	HPDR	
Intercept	−0.435*	−0.812	−0.129
Log # of items	1.002*	0.285	1.621
Basket value	−0.014*	−0.021	−0.007
Delivery fee	0.190*	0.067	0.346

* Indicates that 0 is not contained in the 95% HPDR.

Recall that we restrict ω_2 (referring the second segment) to zeros vector. Therefore, the coefficients in this part of the model are relative to Segment 2.

Table 3.12 and Table 3.13 present the posterior means of the segment specific means and variances of the log purchase and log defection rates. These tables again support our previous findings. The posterior mean on log purchase rate is higher for the first component than that of the second component (−2.221 vs. 3.811) which says that customers in the first component buy more frequently. The result on the log defection rate is also intuitive as the customers in Segment 1 are more loyal and have longer lifetimes.

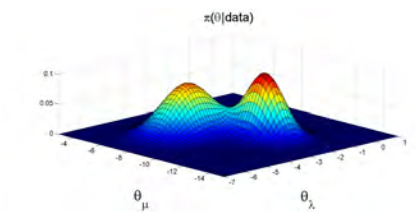
Table 3.12: Segment-specific posterior mean (and standard deviation) of the log purchase and log defection rates for the two-component MHB-C model

	β	
MHB Component 1	−2.221	−10.419
	(0.055)	(0.917)
MHB Component 2	−3.811	−7.272
	(0.093)	(0.308)

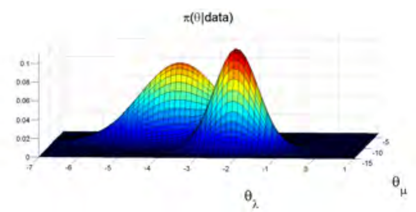
We next consider the shape of the heterogeneity distribution. We visualize the posterior distribution with the plots in Figure 3.5. These plots are created by using the segment sizes, the mean values of β_k and Γ_k and the "`gmdistribution`" function in MATLAB. The multimodality on the heterogeneity distribution is very clear from these figures.

Table 3.13: Posterior mean variance-covariance within segments (Γ_k) for the two-component MHB-C model

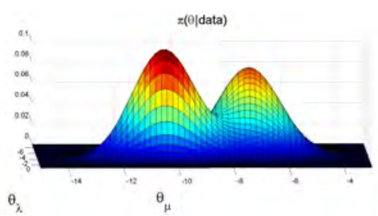
	Γ_1	Γ_2
MHB	$\begin{pmatrix} 0.299 & 0.017 \\ 0.017 & 1.275 \end{pmatrix}$	$\begin{pmatrix} 1.004 & 0.029 \\ 0.029 & 1.260 \end{pmatrix}$



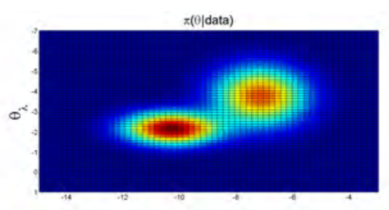
(a) Bivariate Gaussian mixture heterogeneity distribution



(b) From θ_λ perspective



(c) From θ_μ perspective



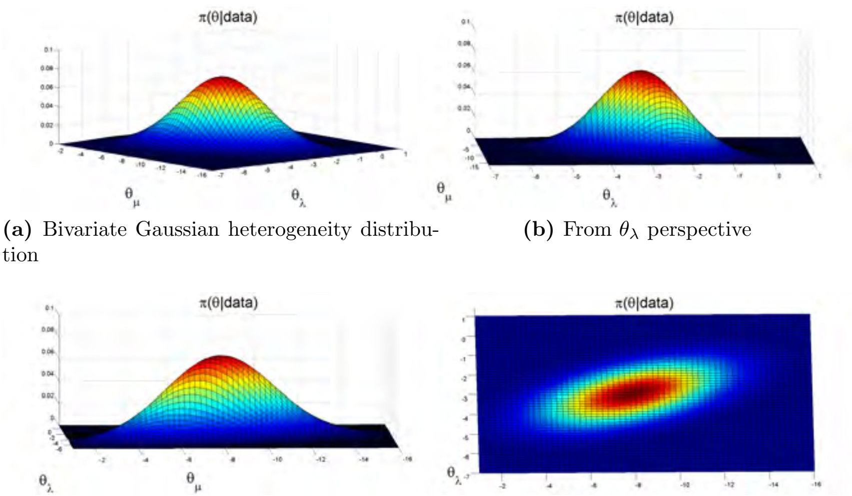
(d) Contour plot of the heterogeneity distribution

Figure 3.5: The shape of the posterior heterogeneity distribution (bivariate Gaussian mixture distribution)

It is also interesting to compare the heterogeneity distribution from the MHB-C model against the one from the HB model. We, therefore, present the posterior means of the hyper-parameters β and Γ in Table 3.14 for the HB model[17] and show the shape of the heterogeneity distribution over the whole population in Figure 3.6. As the HB model tries to fit a unimodal distribution, we see higher variance on the heterogeneity distribution, especially on the log defection parameter which ultimately causes extreme lifetime predictions. The heterogeneity distribution of the HB model masks the bi-modal structure over the behavioral parameters' distribution.

Table 3.14: Posterior mean of the intercept vector β and the variance-covariance matrix Γ for the HB model

	HB	
β	−3.062	−8.083
	(0.036)	(0.929)
Γ	$\begin{pmatrix} 1.016 & -1.339 \\ -1.339 & 6.369 \end{pmatrix}$	



(a) Bivariate Gaussian heterogeneity distribution

(b) From θ_λ perspective

(c) From θ_μ perspective

(d) Contour plot of the heterogeneity distribution

Figure 3.6: The shape of the posterior heterogeneity distribution (bivariate Gaussian distribution) for OG

[17]All the MCMC settings are the same for the HB and MHB-C models.

Next we closely look at the correlation between the log defection and log purchase rates within each segment.[18] The HB model has been shown to outperform earlier BTYD models in the case where there is a correlation between the log purchase and log defection rates (see Chapter 2). Table 3.15 shows that for the HB model we obtain a significant correlation (−0.596). This fact can also easily be observed on Figure 3.6d. For the MHB-C model, we do not find evidence for correlation between behavioral parameters within each segment even though one can observe such correlation on the overall customer base (see Figure 3.5d). Apparently the correlation has now been taken up in the segment structure.

Table 3.15: Posterior mean and 95% highest posterior density region of correlations between log purchase and log defection rates

	$\rho_{\theta_\lambda\theta_\mu}$		
	Mean	95% HPDR	
HB	−0.596*	−0.789	−0.364
MHB-C Segment 1	−0.013	−0.303	0.285
MHB-C Segment 2	0.001	−0.172	0.176

* Indicates that 0 is not contained in the 95% HPDR.

Finally, we move on to the predictive performance. We also want to compare the performance against the Pareto/NBD model. Pareto/NBD model parameters are estimated differently than for the MHB-C and HB models. The hyper-parameters of this model are estimated by Maximum Likelihood Estimation [MLE]. In order to estimate the behavioral rates for every individual, we use a Metropolis-Hastings within Gibbs sampler as discussed in Chapter 2. To provide a fair comparison, we do not incorporate any covariates for the HB and MHB-C models as the Pareto/NBD model cannot accommodate such additional information. Table 3.16 presents statistics on the out-of-sample predictions of the number of transaction as well as lifetime predictions for the MHB-C, HB and Pareto/NBD models. For the predicted number of transactions we can measure the predictive performance. We use MSE, MAE and the correlation between predicted means and observed values. For the predicted lifetime value, we cannot evaluate the performance as the actual lifetime cannot be observed. Instead, we present the mean and median prediction in days.

Table 3.16 shows that the hierarchical Bayes models (HB and MHB-C) outperform the standard Pareto/NBD model. This finding matches the results in earlier papers and the fact that we found a significant correlation between behavioral parameters (see

[18] As emphasized by Abe (2009a), it makes most sense to look at the estimated correlations without any covariates for the HB and MHB models. Therefore, Table 3.15 reports the posterior mean correlations between the behavioral parameters for a model without covariates.

Table 3.15). The HB and MHB-C models perform very similarly on the out-of sample number of transaction predictions. However, the HB model tends to perform slightly better in predicting the number of purchases on all three measures.

Table 3.16: Out-of-sample predictions from the Pareto/NBD, HB and MHB-C models

MODELS	# of purchases			lifetime	
	CORR	MSE	MAE	Mean	Median
MHB-C	0.922	19.172	2.866	7.23E+3	2.15E+3
HB	0.924	18.581	2.774	8.17E+45	4.80E+3
Pareto/NBD	0.921	21.556	3.005	5.30E+130	4.11E+9

When it comes to lifetime metric, there is a clear difference among the models' predictions. The Pareto/NBD model[19] and the HB model both produce extremely long mean lifetime predictions. Whereas the mean lifetime prediction from the MHB-C model is around 20 years. We also check the median posterior results on lifetime predictions as they result in less extreme values. The median lifetime for the Pareto/NBD model is still extremely long. For the HB model it is 16 years, meanwhile the results from the MHB-C model is 7 years. Based on these results we can say that the lifetime predictions obtained from the MHB model can be used as a customer loyalty index for managerial decision making. This is in sharp contrast to the results from the other models.

3.7 Conclusions

The contribution of this chapter is twofold. First, we propose a new BTYD model that addresses the extreme lifetime prediction problem of current BTYD models. If current BTYD models are applied on datasets where the true heterogeneity distribution is multi-modal, one is very likely to obtain extreme lifetime predictions. The main reason for this is that the assumed heterogeneity distribution very poorly fits reality. As a result the variance in the distribution is inflated and extreme lifetime predictions are generated. In other words, if there are several segments in the customer base, the standard BTYD models should not be used. We have substantiated this claim through a simulation experiment as well as through a real-life application. Using a mixture of normals as the heterogeneity distribution yields better predictive results on both lifetime and number of

[19]The hyper-parameter estimations of the Pareto/NBD model on defection rate are $s = 0.04$ and $\beta = 38.24$ (shape and scale parameters of the gamma heterogeneity distribution). The estimated average defection rate for the Pareto/NBD model is given by $s/\beta = 0.001$. As the shape parameter s is less than 1, analytically the expected lifetime value of a random customer from the cohort diverges to infinity.

transactions compared to two major benchmark models, namely the Pareto/NBD model and the HB model.

Second, our MHB-C model increases the descriptive power of BTYD models. While the existing literature on these models has focused primarily on prediction accuracy, this study provides detailed customer base level insights within a segmentation framework. We endorse the claim by Cooil et al. (2008) that segmentation through latent classes is an important method not only for predictive but also for descriptive studies.

Especially our second contribution may be very relevant in practice. If firms are able to predict the segment to which a customer belongs, they can allocate their limited marketing resources in a more efficient way. Based on the predicted segment membership, the customer can be assigned a particular treatment. In other words, effective segmentation allows a company to determine which customers they should try to serve and how to best position their products and services for each segment. Our model also provides information to managers on customers without prior purchase history. For instance, if a transaction from a new customer to OG contains small basket size and if this new customer pays relatively high delivery fee, it is more likely that she will continue buying from OG than another new customer who orders in a bigger quantity and pays less in delivery fee. We believe that our MHB-C model provides a solid methodology to empirically investigate what kind of customer characteristics relate to the lifetime or shopping frequency of customers.

As a future extension, the MHB-C model can be further developed to endogenize the number of segments. The current version of the model does not treat the number of latent components as a model parameter. However, there is a growing literature on finding the number of latent components within the parameter estimation process. The reversible jump Markov Chain Monte Carlo (RJMCMC) method may be useful here, see Richardson and Green (1997); Stephens (2000); Nobile and Fearnside (2007) and Dellaportas and Papageorgiou (2006). The model-specific set-up of this method, however, requires further investigation as incorporating RJMCMC in the proposed complex model is not straightforward. Alternatively one may build on the Dirichlet Process Prior as in Rasmussen (1999), Ishwaran and James (2002) and McAuliffe et al. (2006).

We also advocate further testing of this model on other datasets. The lifetime estimates resulting from BTYD models have not been used a lot in the past. The main reason for this is the poor performance of those estimates. We believe that this situation has been improved with our proposed model. We, therefore, hope to see more applications of these models to predict customer lifetime and to calculate CLV.

3.8 Appendix: MCMC Sampling steps for the MHB model

1. **Nodes z and t_Δ.**

In this subsection the focus is on data and parameters of a single customer. We suppress in our notation the conditioning on T_i which is assumed throughout the subsection. In our MCMC sampler, we draw $t_{\Delta,i}$ and z_i according to the following

$$\begin{aligned}\pi(t_{\Delta,i}, z_i|x_i, t_{x,i}, \lambda_i, \mu_i, \varpi) &= \pi(t_{\Delta,i}|z_i, x_i, t_{x,i}, \lambda_i, \mu_i)\pi(z_i|x_i, t_{x,i}, \lambda_i, \mu_i)\\ &= \pi(t_{\Delta,i}|z_i, t_{x,i}, \lambda_i, \mu_i)\pi(z_i|t_{x,i}, \lambda_i, \mu_i)\end{aligned}$$

where ϖ signals parameters other than written explicitly. The right hand side shows that the conditional probability does not depend on the ϖ parameters. $t_{\Delta,i}$ is the defection time. However, we will derive the conditional distribution of $t_{\Delta,i}, z_i$ using $\pi(t_{\Delta,i}, z_i|x_i, t_{x,i}, \lambda_i, \mu_i, \varpi) = \pi(t_{\Delta,i}|x_i, t_{x,i}, \lambda_i, \mu_i)\pi(z_i|t_{\Delta,i}, x_i, t_{x,i}, \lambda_i, \mu_i)$. For the distribution of the time of defection, $t_{\Delta,i}$, of a customer conditioned on the data $(x_i, t_{x,i})$ and parameters (λ_i, μ_i) of that customer, we have

$$\pi(t_{\Delta,i}|x_i, t_{x,i}, \lambda_i, \mu_i) \propto \pi(t_{\Delta,i}, x_i, t_{x,i}|\lambda_i, \mu_i) = \pi(x_i, t_{x,i}|t_{\Delta,i}, \lambda_i, \mu_i)\,\pi(t_{\Delta,i}|\lambda_i, \mu_i)$$

and

$$\pi(x_i, t_{x,i}|t_{\Delta,i}, \lambda_i, \mu_i) = \pi(x_i|t_{x,i}, t_{\Delta,i}, \lambda_i, \mu_i)\,\pi(t_{x,i}|t_{\Delta,i}, \lambda_i, \mu_i) \propto \pi(t_{x,i}|t_{\Delta,i}, \lambda_i, \mu_i),$$

where $\pi(x_i|t_{x,i}, t_{\Delta,i}, \lambda_i, \mu_i)$ is a constant as far as dependence on $t_{\Delta,i}$ is concerned. So, $\pi(t_{\Delta,i}|x_i, t_{x,i}, \lambda_i, \mu_i) \propto \pi(t_{x,i}|t_{\Delta,i}, \lambda_i, \mu_i)\,\pi(t_{\Delta,i}|\lambda_i, \mu_i) \propto I_{[t_{x,i},\infty)}(t_{\Delta,i})\,e^{-\lambda_i(t_{\Delta,i}\wedge T_i)}\,e^{-\mu_i t_{\Delta,i}}$ and

$$\pi(t_{\Delta,i}|x_i, t_{x,i}, \lambda_i, \mu_i) = \frac{I_{[t_{x,i},\infty)}(t_{\Delta,i})\,e^{-\lambda_i(t_{\Delta,i}\wedge T_i)}\,e^{-\mu_i t_{\Delta,i}}}{C(x_i, t_{x,i}, \lambda_i, \mu_i)} \tag{3.8}$$

with the constant $C(x_i, t_{x,i}, \lambda_i, \mu_i)$ determined as

$$C(x_i, t_{x,i}, \lambda_i, \mu_i) = \int_{t_{x,i}}^{\infty} e^{-\lambda_i(t_{\Delta,i}\wedge T_i)}\,e^{-\mu_i t_{\Delta,i}}\,\mathrm{d}t_{\Delta,i} = \frac{e^{-(\lambda_i+\mu_i)t_{x,i}} - e^{-(\lambda_i+\mu_i)T_i}}{\lambda_i+\mu_i} + \frac{e^{-(\lambda_i+\mu_i)T_i}}{\mu_i}\,.$$

Once we have the conditional distribution of $t_{\Delta,i}$ we can easily find the (discrete) distribution of the binary variable z_i indicating whether the customer is active at T_i (corresponding to $z_i = 1$) or not (corresponding to $z_i = 0$). The value of z_i is determined as $z_i = I_{[T_i,\infty)}(t_{\Delta,i})$. Unconditional on $t_{\Delta,i}$, we have

$$\begin{aligned}\text{Prob}(z_i = 1|x_i, t_{x,i}, \lambda_i, \mu_i) &= \frac{\int_{T_i}^{\infty} e^{-\lambda_i T_i}\, e^{-\mu_i t_{\Delta,i}}\, \mathrm{d}t_{\Delta,i}}{C(x_i, t_{x_i,i}, \lambda_i, \mu_i)} = \frac{\frac{e^{-(\lambda_i+\mu_i)T}}{\mu_i}}{\frac{e^{-(\lambda_i+\mu_i)t_{x,i}} - e^{-(\lambda_i+\mu_i)T_i}}{\lambda_i+\mu_i} + \frac{e^{-(\lambda_i+\mu_i)T_i}}{\mu_i}} \\ &= \frac{1}{\frac{\mu_i}{\lambda_i+\mu_i}\left(e^{(\lambda_i+\mu_i)(T_i - t_{x_i,i})} - 1\right) + 1}\,. \end{aligned} \tag{3.9}$$

See Abe (2009a) and Schmittlein et al. (1987) for Equation (3.9). The distribution $\pi(t_{\Delta,i}|z_i, t_{x,i}, \lambda_i, \mu_i)$ is now the distribution given in Equation (3.8) truncated to the interval $(t_{x,i}, T_i)$ if $z_i = 0$, and to the interval (T_i, ∞) if $z_i = 1$.

2. **Node** s.

   Draw indicator variables for latent class membership, for each customer i; $s_i \sim \pi(s_i|\theta_i, \Delta, \beta_{s_i}, \Gamma_{s_i}, p_k) \propto \pi(\theta_i - D_i\Delta|\beta_k, \Gamma_k)\, p_k$. This is done in two steps:

   (a) Calculate the conditional membership probabilities for each customer and each component as
   $$\widetilde{p}_{ik} = \frac{p_k\, \varphi(\theta_i - D_i\Delta|\beta_k, \Gamma_k)}{\sum_{\ell=1}^{K} p_\ell\, \varphi(\theta_i - D_i\Delta|\beta_\ell, \Gamma_\ell)}, \tag{3.10}$$
   where $\varphi(\cdot)$ is the multivariate normal density.

   (b) Draw the indicator variables of customer i from the multinomial distribution with the parameters of membership probabilities to each groups: $s_i \sim \text{Multinomial}_K(\widetilde{p}_i)$ where $\widetilde{p}_i = [\widetilde{p}_{i1}, \ldots, \widetilde{p}_{iK}]$.

3. **Nodes** $\boldsymbol{\beta}$ **and** $\boldsymbol{\Gamma}$.

   Draw hyper-parameters for each latent class k; $(\beta_k, \Gamma_k) \sim \pi(\beta_k, \Gamma_k|\theta, \Delta, s)$. Note that the value of the quantity $\pi(\beta_k, \Gamma_k|\theta, \Delta, s)$ does not depend on rates θ for those customers that do not belong to the class indicated by s. Let $\theta^{(k)}$ be the rates for those customers for which the class indicator variable has value k: $\theta^{(k)} = \{\theta_i\}_{i:s_i=k}$. Then, according to Equation (3.4) on Page 72,

   $$\pi(\beta_k, \Gamma_k|\theta, \Delta, s) \propto \pi(\theta^{(k)} - D^{(k)}\Delta|\beta_k, \Gamma_k)\, \pi(\beta_k|\Gamma_k)\, \pi(\Gamma_k).$$

   This comes down to the linear regression update:

   (a) **Node** $\boldsymbol{\beta}$.

The conditionally conjugate prior for the intercept (or mean) of each class is given as

$$\beta_k|\Gamma_k \sim \mathrm{N}(\bar{\beta}, \bar{\Gamma} \otimes \bar{A}^{-1})$$

where $\bar{\beta}$ stands for the location parameter, and $\bar{A}$ stands for the shape parameter determining the tightness of the prior.

The posterior density for β_k is sampled from a normal distribution with a mean $\widetilde{\beta}_k$ where $\widetilde{\beta}_k = (\iota'\iota + \bar{A})^{-1}(\iota'(\theta^{(k)} - D^{(k)}\Delta) + \bar{A}\bar{\beta}$ and a variance of $\Gamma_k \otimes (\iota'\iota + \bar{A})^{-1}$. ι is a vector of ones with the size of the number of customers in the k^{th} component.

(b) **Node $\boldsymbol{\Gamma}$.**

The conjugate prior on the covariance structure of each latent class is

$$\Gamma_k \sim \mathrm{IW}(\bar{\Gamma}, \bar{\nu}),$$

where $\bar{\Gamma}$ gives the location parameter, $\bar{\nu}$ gives the degrees of freedom.

The posterior density for Γ_k is sampled from the inverse Wishart distribution with the scale matrix of $\bar{\Gamma}_k + ((\theta^{(k)} - D^{(k)}\Delta) - \iota\widetilde{\beta})'((\theta^{(k)} - D^{(k)}\Delta) - \iota\widetilde{\beta}) + (\widetilde{\beta} - \bar{\beta})'\bar{A}(\widetilde{\beta} - \bar{\beta})$ and the degrees of freedom $\bar{\nu} + \iota'\iota$.

4. **Node $\boldsymbol{\Delta}$.**

The regression coefficient matrix (without an intercept) over the customer base has the following conjugate prior

$$\mathrm{vec}(\Delta) = \delta \sim \mathrm{N}(\bar{\delta}, \bar{A}_\delta^{-1}) \ .$$

The posterior density for $\mathrm{vec}(\Delta)$ is again a normal distribution with mean $(X'X + \bar{A}_\delta)^{-1}(X'y + \bar{A}_\delta\bar{\delta})$ and variance $((X'X) + \bar{A}_\delta)^{-1}$ where

$$X'X = \sum_k \Gamma_k^{-1} \otimes D'^{(k)} D^{(k)}$$

$$X'y = \mathrm{vec}\left(\sum_k D'^{(k)}(\theta^{(k)} - \iota\beta_k)\Gamma'^{-1}_k\right)$$

Details of Δ sampling:

As this model does not distinguish the slope among different components, the regression coefficients are drawn over the whole population; $\Delta \sim \pi(\Delta|\theta, \beta, \Gamma, s)$. In these expressions we consider data for all customers.

At this stage we use the mean β and variance-covariance matrix Γ of each component, parameter values θ for each customer. Besides, we have the information on covariates D and the prior distribution on regression coefficients $\delta = \text{vec}(\Delta)$ which is given as $N(\bar{\delta}, \bar{A}_{\delta}^{-1})$.

We create a linear regression model that covers customer data in all segments. In order to be able to pool data from K components, we collect the multivariate regression models across the components. To do so, we standardize all equations.

- Customer data should be shifted by the intercept of the component that she belongs.
- For component k, we have

$$\theta^{(k)} - \iota\beta_k = D^{(k)}\Delta + \varepsilon^{(k)}$$
$$\text{vec}(\theta^{(k)} - \iota\beta_k) = \text{vec}(D^{(k)}\Delta) + \text{vec}(\varepsilon^{(k)}),$$

given that $\text{vec}(\varepsilon^{(k)}) \sim N(0, \Gamma_k \otimes I)$ and using the property of $\text{vec}(ABC) = (C' \otimes A))\text{vec}(B)$, we obtain

$$\text{vec}(\theta^{(k)} - \iota\beta_k) = (I \otimes D^{(k)})\text{vec}(\Delta) + \text{vec}(\varepsilon^{(k)}) \tag{3.11}$$

Next we standardize the error for the MVR model of each component.

$$(M_k'^{-1} \otimes I)\text{vec}(\theta^{(k)} - \iota\beta_k) = (M_k'^{-1} \otimes I)(I \otimes D^{(k)})\text{vec}(\Delta) + U_k$$
$$(M_k'^{-1} \otimes I)\text{vec}(\theta^{(k)} - \iota\beta_k) = (M_k'^{-1} \otimes D^{(k)})\text{vec}(\Delta) + U_k, \tag{3.12}$$

where $M_k'M_k = \Gamma_k$ and U_k represents errors with a unit covariance structure.

- In Equation (3.12), if we write the expressions as $y_k = (M_k'^{-1} \otimes I)\text{vec}(\theta^{(k)} - \iota\beta_k)$, $X_k = (M_k'^{-1} \otimes D^{(k)})$, $\delta = \text{vec}(\Delta)$, and then we have the regression model $y_k = X_k\delta + U_k$. After stacking all the regression models from the mixture components, we deal with a standard normal regression update, where errors are independent and of unit size. The vectors y_k are stacked into y and matrices X_k are stacked into X. Δ can therefore be sampled from a normal distribution with mean $(X'X + \bar{A}_\delta)^{-1}(X'y + \bar{A}_\delta\bar{\delta})$ and variance $((X'X) + \bar{A}_\delta)^{-1}$. Note that the matrices M_k are not explicitly used in this sampling process.

The moments mentioned can be calculated efficiently as follows:

$$X'X = \sum_k {\Gamma_k}^{-1} \otimes D'^{(k)} D^{(k)}$$

$$X'y = \text{vec}\left(\sum_k D'^{(k)} (\theta^{(k)} - \iota\beta_k) {\Gamma'}_k^{-1} \right)$$

5. **Node $\boldsymbol{p}$.**

   Draw $p \sim \pi(p|s)$. Dirichlet update: $\upsilon \sim \text{Dir}(\bar{\boldsymbol{\alpha}} + \#)$. Here $\#_k = |\{i|s_i = k\}|$. We set $\bar{\boldsymbol{\alpha}}$ as 1.

6. **Node θ.**

   Draw, for each customer i, a new value for $\theta_i \sim \pi(\theta_i|x_i, t_{x,i}, y_i, z_i, \Delta, \beta_k, \Gamma_k, s_i)$. Note that

$$\pi(\theta_i|x_i, t_{x,i}, y_i, z_i, \Delta, \beta_k, \Gamma_k, s_i) \propto \pi(x_i, t_{x,i}, y_i, z_i, \theta_i\Delta, \beta_k, \Gamma_k, s_i)$$

   and this is proportional to $\pi(x_i, t_{x,i}, y_i, z_i|\theta_i)\pi(\theta_i|\beta_k + D_i\Delta, \Gamma_k)$. Sampling of θ_i requires the Metropolis Hastings algorithm. We use a Gaussian random walk algorithm for generating candidate values. The step size in the random-walk MH algorithm is set by applying an adaptive MH method in the burn-in phase (Gilks et al., 1996).

3.9 Appendix: MCMC Sampling steps for the MHB-C model

1. **Node ω.**

   The conjugate prior on the latent component-specific regression coefficients is $\omega_k \sim \text{N}(\bar{\omega}, \bar{A}_\omega^{-1})$. ω_k is dimension of $((L + 1) \times 1)$ where L is the number of concomitant variables. It describes the effect of concomitant variables on each of the latent classes. The draws from the posterior distribution can be obtained by a standard regression update process on the following model.

$$u_{ik} = C_i\,\omega_k + \varepsilon_{ik}$$

where $\varepsilon_{ik} \sim \mathrm{N}(0, \mathbb{I}_{\mathbb{K}})$, $\mathbb{I}_{\mathbb{K}}$ is the identity matrix of dimension K. The normal regression update on the component specific regression coefficients give

$$(\omega_k|u_k) \sim \mathrm{N}((C'C + \bar{A}_\omega)^{-1}(C'u_k + \bar{A}_\omega\,\bar{\omega}), (C'C + \bar{A}_\omega)^{-1}),$$

for $k = 1, \ldots, K-1$.

Note that for identification, we restrict $\omega_K = 0$.

2. **Node $\boldsymbol{u}$.**

   In order to assign each customer to a latent component, we use latent utility variable u. The selector function $\varsigma(u)$ determines which component each customer is assigned to, that is,

$$\varsigma(u_i) = k,\ \text{if } u_{ik} > u_{ij} \text{ for all } j \neq k,$$

   where $u_{ik} = C_i\,\omega_k + \varepsilon_{ik}$ is the utility of customer i being assigned to the latent component k. C_i is the row vector of component-invariant behavioral characteristics (concomitant variables) of customer i (together with an intercept), ω_k is the component specific regression coefficients, and ε_{ik} is the stochastic error term.

   The probability of customer i being a member of component k is equal to

$$\begin{aligned}
\mathrm{Prob}(s_{ik} = 1) &= \mathrm{Prob}(u_{ik} \geq u_{ij},\ \text{for all } j \text{ in } (K-1) \text{ components}) \\
&= \mathrm{Prob}(u_{ij} - u_{ik} \leq 0,\ \ \text{all } j \neq k) \\
&= \mathrm{Prob}(\varepsilon_{ij} - \varepsilon_{ik} \leq C_i\,(\omega_k - \omega_j),\ \ \text{all } j \neq k) \\
&= \mathrm{Prob}(\widetilde{\varepsilon}_{ikj} \leq C_i\,\widetilde{\omega}_{kj},\ \ \text{all } j \neq k)
\end{aligned}$$

   where $\widetilde{\varepsilon}_{ikj} = \varepsilon_{ij} - \varepsilon_{ik}$ and $\widetilde{\omega}_{kj} = (\omega_k - \omega_j)$.

   To allocate customer i to latent components, we need to sample from

$$u_i \sim \pi(u_i|\cdots, \theta, \beta, \Gamma, \omega, \cdots) \propto \pi(\theta_i|\Delta, \beta_{\varsigma(u_i)}, \Gamma_{\varsigma(u_i)})\,\pi(u_i|C_i\,\omega)\ . \tag{3.13}$$

Note that $\pi(u_i|C_i\,\omega)$ in Equation (3.13) is a multivariate normal density. So, Equation (3.13) expresses that we need to sample from a multivariate normal density with different multiplicative constants $(\pi(\theta_i|\Delta, \beta_{\varsigma(u_i)}, \Gamma_{\varsigma(u_i)}))$ in different domains. This is difficult to efficiently accomplish due to the very high rejection frequencies. We, therefore, use the insight from McCulloch and Rossi (1994) and specify a Gibbs sampler by breaking each draw of u_i into a sequence of K univariate truncated normal draws by cycling through the u_i vector (one-at-a-time sampling or one dimensional sampling).

We need to take into account the discrete jumps that may happen through $\varsigma(u)$ as this results in new parameter values on β, Γ and ω. We separately investigate each component of Equation (3.13).

- $\pi(\theta|\Delta, \beta_{\varsigma(u)}, \Gamma_{\varsigma(u)})$: The dependency here is interceded through $\varsigma(u)$. Recall that

  $$\varsigma(u_i) = k, \text{ if } u_{ik} > u_{ij} \text{ for all } j \neq k (\text{ or if } s_{ik} = 1)\ .$$

  Dropping the customer index i momentarily so that $s_k = s_{ik}$ and $u_k = u_{ik}$ in the following, we have $\pi(s_k|u) = \pi(s_k|u_k, u_{-k})$, that is,

  $$\pi(s_k|u_k, u_{-k}) = I(s_k = 1)I(u_k > u_o) + I(s_k \neq 1)I(u_k < u_o)$$

  where $o = \text{argmax}\{u_l|l \neq k\}$. We, therefore, get

  $$\pi(\theta|\Delta, \beta_{\varsigma(u)}, \Gamma_{\varsigma(u)}) = I(u_k > u_o)\pi(\theta|\Delta, \beta_k, \Gamma_k) + I(u_k < u_o)\pi(\theta|\Delta, \beta_o, \Gamma_o).$$

- $\pi(u|C\,\omega)$: Utilities have a multivariate Normal distribution, that is,

  $$\pi(u|C\,\omega) \propto e^{-1/2(u-\bar{u})'(u-\bar{u})},$$

  where $\bar{u} = C\omega$.

So the conditional density of utilities can be written as

$$\pi(u_k|\theta, \boldsymbol{\beta}, \boldsymbol{\Gamma}, u_{-k}, \omega) \propto (I(u_k > u_o)\,|\Gamma_k|^{-1/2} e^{-1/2(\theta-(\beta_k+D\Delta))(\Gamma_k)^{-1}(\theta-(\beta_k+D\Delta))'} \\ + I(u_k < u_o)\,|\Gamma_o|^{-1/2} e^{-1/2(\theta-(\beta_o+D\Delta))(\Gamma_o)^{-1}(\theta-(\beta_o+D\Delta))'})\; e^{-1/2(u_k-\bar{u})^2}\;. \quad (3.14)$$

Expression (3.14) is a combination of two truncated normal densities, see Figure 3.7. We write Ω_r as the scaling factor of the truncated normal distribution on the right,

$$\Omega_r = |\Gamma_k|^{-1/2} e^{-1/2(\theta-(\beta_k+D\Delta))(\Gamma_k)^{-1}(\theta-(\beta_k+D\Delta))'}$$

where $u_k < u_o(\max(u_{-k}) = u_o)$; and Ω_l as the scaling factor of the other truncated normal distribution

$$\Omega_l = |\Gamma_o|^{-1/2} e^{-1/2(\theta-(\beta_o+D\Delta))(\Gamma_o)^{-1}(\theta-(\beta_o+D\Delta))'}$$

where $u_k > u_o$.

Then,

$$\pi(u_k|\theta, \boldsymbol{\beta}, \boldsymbol{\Gamma}, u_{-k}, \omega) \propto (I(u_k > u_o)\Omega_r + I(u_k < u_o)\Omega_l)\, e^{-1/2(u_k-\bar{u})^2}\;. \quad (3.15)$$

The normalization constant is easily computed. Let ϕ be the density function of the standard normal distribution.Then, the final version for the sampling distribution is

$$\pi(u_k|\theta, \boldsymbol{\beta}, \boldsymbol{\Gamma}, u_{-k}, \omega) = \frac{\Omega_r I(u_k > u_o) + \Omega_l I(u_k < u_o)}{(1-\Phi(u_o-\bar{u}_k))\Omega_r + \Phi(u_o-\bar{u}_k)\Omega_l}\phi(u_k-\bar{u}_k)\;. \quad (3.16)$$

The sampling is now done by applying the following to all utility components:

- Sample $U \sim$ Uniform$[0,1]$ to determine which truncated normal distribution to sample from.
- If $U < \frac{\Omega_l\Phi(u_o-\bar{u}_k)}{\Omega_l\Phi(u_o-\bar{u}_k)+\Omega_r(1-\Phi(u_o-\bar{u}_k))}$, then truncate to the right and sample from the left side of the truncated normal distribution. In particular, set

$$u_k^{\text{new}} = \Phi^{-1}(\Phi(u_o-\bar{u}_k)U') + \bar{u}_k$$

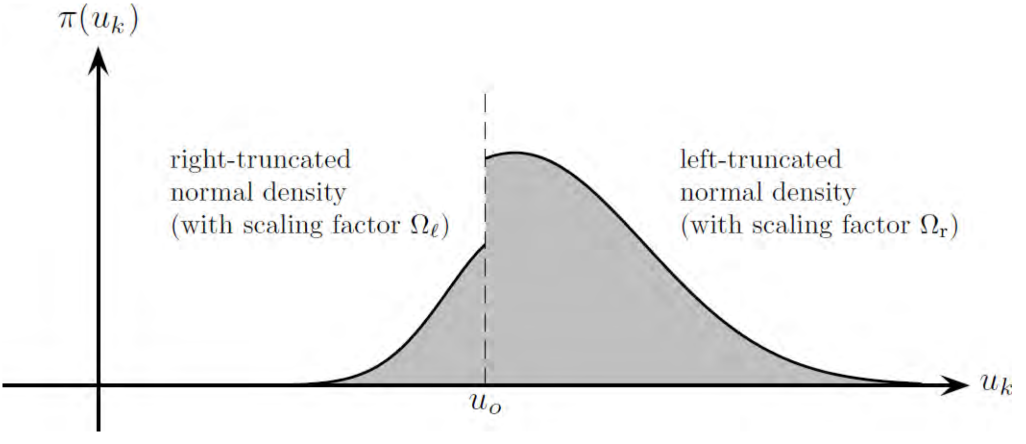


Figure 3.7: The sampling density for the utilities.

where $U' \sim \text{Uniform}[0,1]$.

- If $U > \frac{\Omega_r(1-\Phi(u_o-\bar{u}_k))}{\Omega_l\Phi(u_o-\bar{u}_k)+\Omega_r(1-\Phi(u_o-\bar{u}_k))}$, then truncate to the left and sample from the right side of the truncated normal distribution. In particular, set

$$u_k^{\text{new}} = \Phi^{-1}\left((1-\Phi(u_o-\bar{u}_k))U' + \Phi(u_o-\bar{u}_k)\right) + \bar{u}_k \ .$$

3.10 Appendix: Data generation for MHB model testing

Consider $N = 1,000$ customers and $K = 2$ latent components. We generate a single covariate data, D $(N \times 1)$, for all customers from a standard uniform distribution. We create another customer characteristics matrix including an intercept and a concomitant variable, C $(N \times L)$ where $L = 2$. In order to keep it simple, for the first half of the population the concomitant variable is set to 1 and for the other half it is set to -1. The transaction data of customers are generated in five steps:

1. Fix the component specific regression coefficient matrix, ω^* $(L \times K)$ to $\left[\begin{smallmatrix} 0.1 & 0 \\ 0.8 & 0 \end{smallmatrix}\right]$. Using the concomitant matrix together with the ω^* matrix, we generate utilities, u^*, using the normally distributed error term.[20] More specifically, we use the following utility generation form: $u^* = C\,\omega^* + \varepsilon$, where $\varepsilon \sim \text{N}(0, \mathbb{I})$. Note that the used parameter values are chosen to balance the random and deterministic components of utilities. Given the *true* utility values u^*, customers are assigned to each component,

$$s_i^* = k, \text{ if } u_{ik}^* > u_{ij}^* \text{ for all } j \neq k.$$

   Based on this procedure, we add randomness on assigning customers to their *true* components. In our sample 52.8% of the customers is assigned to segment 1.

2. Fix the hyper-parameter values β^* and Γ^* for each of the components: We aim to generate a customer dataset that has $K = 2$ distinct groups or in other words that has a bi-modal heterogeneity distribution over the customer base. As the covariate data, D, is standardized, the β vector represents the average values of parameters of interest (log of purchase and defection parameters) for each component. Our main concern is on distinguishing between the components. We, therefore, use a rather different set of parameters for each component. We set $\beta_1^* = [\log(1), \log(1/1000)]$

[20]The proposed model employs an MNP sub-model to assign customers to latent components.

and $\beta_2^* = [\log(1/100), \log(1/20)]$. The (2×2) covariance matrices Γ_k^* are chosen to be equal to $0.64 \times \mathbb{I}$ for each of the components.

3. Generate behavioral parameters $\theta_i^* \sim \pi(\theta_i | \beta_{s_i}^*, \Gamma_{s_i}^*)$ for each of the customers: Conditional on the membership to one of the two components, customer's behavioral parameters are generated from normal distributions independently given the associated hyper-parameters.

4. Generate lifetime $t_{\Delta,i}^*$ for each of the customers: For $i = 1, \ldots, N$, draw $t_{\Delta,i}^* \sim \pi(t_{\Delta,i} | \theta_i^*)$. As the lifetime is distributed according to an exponential distribution with the rate parameter of e^{θ_μ}, this step is straightforward.

5. Generate repeat transaction frequency x_i and the last transaction time in calibration period $t_{x,i}$ for each of the customers: For $i = 1, \ldots, N$, draw $x_i \sim \pi(x_i | t_{\Delta,i}^*, \theta_i^*)$. Transaction data basically contains two elements: transaction number x_i and the time of the last transaction $t_{x,i}$. Note that the time of the first order t_0 and the total observation time T are fixed ($t_0 = 0, T = 200$) and they are common across the customers. The sampling scheme of transaction data $(x_i, t_{x,i})$, given the defection time $t_{\Delta,i}$ and the parameters θ_i is the following:[21]

   Let $(V_l)_{l=1,2,\ldots}$ be iid exponentially distributed with mean $1/\lambda$. Put $E_x = \sum_{l=1}^{x} V_l$. Then E_x has an Erlang-x distribution: the sum of x independent exponential distributions with average $1/\lambda$. Write $\hat{t}_\Delta = \min(t_\Delta, T)$ where $\hat{t}_\Delta$ is the effective time of defection. Now, for $x \geq 1$, we can compute

$$\begin{aligned}\pi(x, t_x | t_\Delta, \theta) &= \pi(E_x = t_x, V_{x+1} + E_x > \hat{t}_\Delta) = \pi(E_x = t_x)\, \pi(V_{x+1} > \hat{t}_\Delta - t_x | E_x = t_x) \\ &= \pi(E_x = t_x)\, \pi(V_{x+1} > \hat{t}_\Delta - t_x) = \frac{\lambda^x t_x^{x-1}}{(x-1)!} e^{-\lambda t_x}\, e^{-\lambda(\hat{t}_\Delta - t_x)} = \frac{\lambda^x t_x^{x-1}}{(x-1)!}\, e^{-\lambda \hat{t}_\Delta}\end{aligned}$$

   Performing the integral of t_x over the interval $(0, \hat{t}_\Delta)$ leads to[22]

$$\text{Prob}(x, t_x \leq t | t_\Delta, \theta) = \frac{\lambda^x t^x}{x!} e^{-\lambda \hat{t}_\Delta}$$

[21] We drop the i index in the following derivations for the sake of simplicity on notation.

[22] And in turn to

$$\text{Prob}(x = 0 | \theta) = \int_0^\infty e^{-\lambda \hat{t}_\Delta} \mu e^{-\mu t_\Delta} \mathrm{d}t_\Delta = \frac{\mu}{\lambda + \mu} + \frac{\lambda}{\lambda + \mu} e^{-(\lambda+\mu)T} \,.$$

for $t < \hat{t}_\Delta$ and $x \neq 0$. Clearly, $\text{Prob}(x = 0, t_x \leq t | t_\Delta, \theta) = e^{\lambda \hat{t}_\Delta}$, and for $t < T$

$$\begin{aligned} F(t) &\equiv \text{Prob}(t_x \leq t | t_\Delta, \theta) \\ &= \begin{cases} 0 & \text{if } t < 0 \\ \sum_{x=0}^{\infty} \frac{\lambda^x t^x}{x!} e^{-\lambda \hat{t}_\Delta} & \text{if } 0 \leq t < \hat{t}_\Delta \\ 1 & \text{if } t \geq \hat{t}_\Delta \end{cases} \\ &= \begin{cases} 0 & \text{if } t < 0 \\ e^{-\lambda(\hat{t}_\Delta - t)} & \text{if } 0 \leq t < \hat{t}_\Delta \\ 1 & \text{if } t \geq \hat{t}_\Delta \end{cases} \end{aligned}$$

and for $s \in [0, 1]$,

$$F^{-1}(s) = \begin{cases} 0 & \text{if } s \leq e^{-\lambda \hat{t}_\Delta} \\ \hat{t}_\Delta + \ln(s)/\lambda & \text{if } s > e^{-\lambda \hat{t}_\Delta} \end{cases}$$

All this leads to the following sampling scheme for recency-frequency (RF) data.

(a) Draw $t_\Delta \sim \text{EXP}(\mu)$.

(b) Draw $U \sim \text{U}[0, 1]$. Put

$$t_x = \begin{cases} 0 & \text{if } U \leq e^{-\lambda \hat{t}_\Delta} \\ \hat{t}_\Delta + \ln(U)/\lambda & \text{if } U \geq e^{-\lambda \hat{t}_\Delta} \end{cases}$$

(c) Put

$$x = \begin{cases} 0 & \text{if } t_x = 0 \\ 1 + \text{POISSON}(\lambda t_x) & \text{if } t_x > 0 \end{cases}$$

3.11 Appendix: Setting the number of components for MHB-C model

Table 3.17 shows the out-of-sample prediction accuracy of the MHB-C model for different numbers of components. The MHB-C model with 2 components performs best in predicting the number of purchases in the validation period. As discussed earlier, our main criterion of determining the optimum number of components is the number of members within each group (Frühwirth-Schnatter, 2006). Based on this criterion, we decide that the optimum number of components is 2 with a general customer share of 59% and 41%

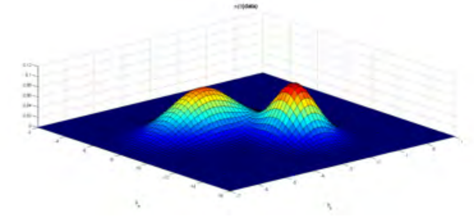
for the two segments. When the number of components increases to 3, one of the component covers only 4 customers (0.2%) of the customer base while the others contain the rest of it in a balanced share. For the 4 component case, two additional components cover only around 1% of the customers.

We did not use the Bayesian counterparts of likelihood based model comparison methods, i.e. the marginal likelihood comparison, because of the lack of the closed-form solution to the marginal likelihood. Schwarz criterion is not used either, because it is not evident that the regularity conditions for deriving Schwarz's criterion through asymptotic expansions actually hold (Frühwirth-Schnatter, 2006).

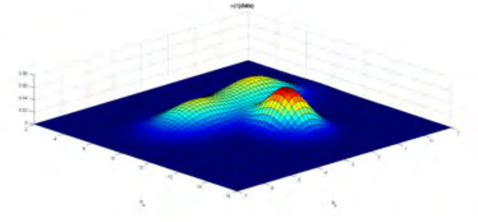
Table 3.17: The out-of-sample prediction performance of the MHB-C model with different number of components on OG data

MHB-C Model	# of purchase			# of customers (%) in each component			
	Correlation	MSE	MAE	Comp1	Comp2	Comp3	Comp4
2-Component	0.9208	19.556	2.851	599 (41%)	861 (59%)	-	-
3-Component	0.9207	19.654	2.860	601 (41%)	855 (59%)	4 (0.2%)	-
4-Component	0.9200	19.738	2.857	602 (41%)	839 (58%)	15 (1%)	4 (0.2%)

Figure 3.8 shows the heterogeneity distribution for the OG data using the MHB-C model with 3 or 4 components. The plot given in Figure 3.8a is not different that the MHB-C model with 2 components where there are only two peaks, i.e. the additional component does not capture a different (heterogeneous) characteristic. However, when 4 components are forced on the MHB-C model, we observe three peaks on OG data (see Figure 3.8b). Despite this additional peak in the 4 component model, which may capture different characteristics of the heterogeneity distribution, this model clearly deteriorates out-of-sample estimation results. Note that, this model performs the worst in out-of-sample predictions. We therefore opt for the 2 component model in this paper.



(a) Bivariate Gaussian mixture heterogeneity distribution from MHB-C model with 3 components



(b) Bivariate Gaussian mixture heterogeneity distribution from MHB-C model with 4 components. Note the vertical scale.

Figure 3.8: The shape of the posterior heterogeneity distributions (bivariate Gaussian mixture distribution) over the online retailer's customer base when the MHB-C model is run with 3 and 4 components.

Chapter 4

An Empirical Investigation of Demand for Online Services: Evidence from Online Grocery Shopping and Delivery Fees

4.1 Introduction

It is by now clear that the internet has drastically changed retailing and final consumer's purchasing behavior in the last two decades. From the firms' perspective, online retailing has allowed companies to price-discriminate consumers in ways that were unimaginable before, increasing profits and consumer surplus simultaneously. From the consumers' perspective, online retailing has allowed consumers with little spare time to purchase goods and services without leaving the comfort of their homes avoiding unnecessary trips or phone calls.

Clear examples of online shopping trends are the growing complexity of airline pricing (see Klein and Loebbecke (2003), Mohammed (2005), Belobaba (2002), Robinson (2002), Barnhart et al. (2003), Smith et al. (2001)) or the rapid increase of online retailing that has taken place in the last few years (Lewis (2006), Laudon and Traver (2007), Baier and Stüber (2010)). For example, in the US alone e-commerce grew 13% (while offline retail barely grew 1%) in the first quarter of 2013, and it is expected to raise its sales up to \$370BN by 2017 with the help of tablets and smart phones.[1] This rapid change is leading firms to think strategically on how to manage their revenue sources and hence asking the

[1]This forecast is based on a report published by Forrester Research "US Online Retail Forecast, 2012 To 2017" by Mulpuru et al. (2013).

question of what particular parts of the services and goods provided are valued the most by their customers, and how customer heterogeneity plays a role when customers value different goods.

Although the literature is replete with classic examples of how to price two goods that are complements (such as admission tickets and rides in amusement parks, blades and razors, show tickets and concessions, or video games and consoles as primary/access goods and their secondary/complematary goods), the impact of internet on retailing has also brought a dilemma on how to manage different revenue channels through internet platforms. When the manufacturer of smart phones maximizes profits, she must decide on the optimal pricing strategy of both hardware and software. Similarly, when shopping platforms maximize profits they must decide on the price of items as well as their delivery fees.

In this chapter, we investigate the optimal pricing strategies of the online operations of a grocery retailer. This online retailer derives its revenues and profits from two different sources: shipping fees and grocery sales. When maximizing revenues, the retailer may consider whether to sell groceries at a discount and make up for its profits with high shipping fees. Alternatively, it may offer cheap (or even free!) delivery and charge higher prices for groceries. The optimal strategy depends on how the demand for groceries is correlated with the demand for home delivery. The intuition is that customers' demand intensity for the groceries provides a meter of how much the customer is willing to pay for the delivery service. If increases in delivery service demand is associated with decreases in grocery demand per delivery, this would indicate a positive correlation between grocery demand and willingness to pay for the delivery service. In case of consumers with high willingness to pay for home delivery associated with a high demand for groceries, firms ought to charge high prices for groceries and low delivery fees (Gil and Hartmann, 2009).

In order to provide optimal pricing strategies for online grocery retailers, we build our theory on the well-grounded two-part tariff literature. A two-part tariff exists when a fixed payment is made for the access good before any secondary good purchases are allowed. Since the well-known paper from Oi (1971) where the optimal pricing policies are presented under a two-part tariff, the literature has concentrated on different dimensions of the two-part tariffs such as consumer heterogeneity or budget constraints (see Ng and Weisser (1974), Littlechild (1975), Schmalensee (1981), Rosen and Rosenfield (1997)). The profitability of two-part tariffs relative to a single uniform pricing has been steadily studied over the years (as a recent example see Iyengar et al. (2011)). However, very little attention has been paid on a repeat purchase setting as the literature has focused on modeling the buyer behavior where they have been restricted to visit the firm and pay the fixed fee at most one time. To our knowledge, only Phillips and Battalio (1983) and Yang et al. (2005) allow for repeat buying under a two-part tariff. While the latter investigates whether free shipping is profitable for firms, the former focuses on the substitution effect between visit frequencies and consumption per visit as we aim to concentrate on further. In their paper, however, Phillips and Battalio (1983) do not consider consumer heterogeneity and focus on a single consumer case. We expand on the theoretical framework of Schmalensee (1981) by considering customer heterogeneity and allowing for repeat number of visits. We make theoretical predictions on how the total number of primary good sales, the total amount of secondary good sales and the average secondary good consumption per primary good change in the primary and secondary good prices. This is the very essence of our theoretical contribution.

We adapt the model to the institutional setting of an online grocery store and derive testable implications regarding variation in the price of deliveries (access good) and the price of groceries (secondary good). We take these predictions to the data using a unique dataset detailing transaction information from an online grocery retailer in a Western European country following the empirical work of Gil and Hartmann (2009).[2] Our data is the result of an extraction of all transactions between 2008 and 2009 of a random sample of customers of this online grocery store. This firm (OG hereafter) structures its online operations into eight different time slots in any given day from Monday to Friday, only offering five morning slots on Saturdays, and no delivery on Sundays. The resulting dataset has a total of $953,107$ transactions from $29,988$ customers located in 44 different cities in this country that made purchases between January 2008 and December 2009.

We verify our theoretical predictions by replicating them on a real dataset and find that there is a positive relationship between the demand for home delivery services and groceries. This is basically consistent with a two-part pricing policy that will charge high margins for delivery services and offer discounts (or not charge extra mark-ups) for groceries. Next, we conduct an in-dept analysis on our data and find that there are two groups of customers with very different willingness to pay and price sensitivities. We use this fact to improve the optimal pricing strategy of OG by combining second and third-degree price discrimination schemes, and consequently to propose a discriminating two-part tariff.

We can summarize our empirical findings as follows. First, delivery time slots with bigger number of transactions also have larger basket sizes per transaction. This fact is true across time slots and cities as well as within time slots and cities. Second, higher delivery fees are associated with fewer transactions. Third, we find that there is a positive

[2] Due to a data confidentiality agreement, we cannot reveal the identity of the retailer.

association between delivery fees and basket sizes across time slots. This result does not hold within time slots because price-sensitive customers increase their average basket size in high-demand periods.

In a final attempt to reconcile the observed set of prices and optimal prices, we also investigate the correlation between operational profits obtained through delivery services and operational profits through the sale of groceries. Our results show that operational profits are mainly driven by the number of transactions and average basket sizes, whereas the number of transactions are driven by delivery fees. Using our estimates of the relation between delivery fees, number of transactions and profit margins, we show that OG was underpricing delivery across time slots for households and overcharging B2B customers.

We are not the first to empirically study this topic and so we build upon a number of papers that have explored a wide variety of sectors such as cellphone pricing (Miravete and Röller, 2004), sports pricing (Marburger (1997); and Fort (2004)), or concession pricing (Gil and Hartmann, 2009).[3] If anything, to the best of our knowledge the closest papers to ours are Lewis et al. (2006) and Chintagunta et al. (2012) in that they also explore pricing and consumer behavior in online grocery shopping. While the latter measures the relative importance of transactions costs in consumer choice between online and offline grocery shopping, the former uses an ordered probability model to study the non-linear impact of shipping fees on size and incidence of orders. Our study differs from these in that we extend the theory of Schmalensee (1981) and others[4] to derive testable implications that we take to the data. We not only show that correlations between delivery fees, basket sizes and number of transactions are consistent with an optimal two-part pricing scheme, but also estimate optimal prices that discriminate across consumer types.

Our paper also provides clear managerial implications for firms that manage a portfolio of products with interrelated demands as well as firms that may be able to screen customers that differ in their willingness to pay. Our results suggest that OG would benefit from exploiting price discrimination between B2B and household customers because they show significant differences in their sensitivities to delivery fees.

The chapter's organization is as follows. In the following section, we present our theoretical framework departing from Schmalensee (1981) and provide testable implications. Section 4.3 describes the data and the institutional details around online grocery shopping in this particular Western European country. In Section 4.4, we introduce our empirical methodology and show results. Section 4.5 presents results from "diff-in-diffs" estima-

[3]Examples in the popular press also covered the hotel industry (Landsburg, 2006) and the airline industry (Saporito (2011) and Rane (2013)).

[4]Other representative papers on the same topic are Oi (1971) and Rosen and Rosenfield (1997).

tion, while in Section 4.6 we discuss the managerial implications of our findings. Finally, Section 4.7 concludes our study.

4.2 Theory

In this section, we extend the theoretical work of Schmalensee (1981) on access service pricing to repeat purchase occasions where consumers adjust their number of visits to the firm and the amount of secondary good purchase per visit. Then we discuss our theoretical predictions. In our model, we assume away the income effects which can be considered reasonable for the online grocery retail environment where the delivery fees are relatively small compared to basket values.[5] This implies that the demand of the secondary good is independent of primary good price. In order to derive cleaner predictions, we assume the firm offers one representative secondary good at price p.

We allow the demand of secondary goods per visit to be a function of the secondary good price p, expected number of visits n in a given period, and the consumer's type θ. We assume a continuous distribution of consumer types θ with $0 \leq \theta \leq 1$. Let

$$\begin{aligned} S(p,n,\theta) &= \text{surplus of consumer type } \theta \text{ for n primary goods,} \\ q(p,n,\theta) &= \text{secondary good demand of consumer type } \theta \text{ per primary good if n primary} \\ & \text{goods are purchased.} \end{aligned}$$

The consumer surplus $S(p,n,\theta)$ increases in θ for all $n \geq 0$, and decreases in p; and can be calculated as $S(p,n,\theta) = n\int_p^\infty q(t,n,\theta)dt$.

A type θ consumer will purchase n primary goods if and only if $S(p,n,\theta) \geq nx$ where x is the unit price of the primary good. For each i, there exists a marginal consumer type θ^i defined implicitly by[6]

$$S(p,i,\theta^i) = ix, \ \ i = 0,1,...,n.$$

If $\theta \in [\theta^i, \theta^{i+1})$, then type θ consumer makes i primary good purchases with $q(p,i,\theta)$ secondary good purchases per primary good purchase. Notice that $0 = \theta^0 \leq \theta^1 \leq ... \leq$

[5] For online grocery retailing, there is generally a lower limit of basket size in order to receive the groceries, as we also have in our dataset. Moreover, assuming that secondary good demand is not affected by the price of the primary good is common in discrete-choice demand literature (Gil and Hartmann, 2009).

[6] Note that the secondary good demand of the marginal consumer is no longer independent of the primary good price. Therefore, no income effects assumption (meaning that changes in primary good price do not affect the demand of the secondary good demand) is valid for the inframarginal consumer and access prices play a role on the overall customer base of the company.

$\theta^i \leq ...\theta^M \leq \theta^{M+1} = 1$ for $i > 1$ where M is the maximum possible number of primary good purchases.[7]

The demand of secondary good per primary good purchase, $q(p, n, \theta)$, is not monotonically increasing in θ due to sudden jumps in the repeat purchase frequency n. This implies that $q(p, n, \theta)$ decreases at the switching points from one level of repeat purchase to the next level. Within each repeat purchase frequency, $q(p, n, \theta)$ is increasing in θ. Finally, $q(p, n, \theta)$ is decreasing in p.

If $m(\theta)$ is the density function of consumer types, total market demand for the primary good N and and the secondary good Q are given by

$$\begin{aligned} N(x,p) &= \sum_{i=1}^{M} i \int_{\theta^i(x,p)}^{\theta^{i+1}(x,p)} m(\theta)d\theta \\ Q(x,p) &= \sum_{i=1}^{M} i \int_{\theta^i(x,p)}^{\theta^{i+1}(x,p)} q(p,i,\theta)m(\theta)d\theta. \end{aligned}$$

Differentiation of the marginal consumer θ^i with respect to x and p results in[8]

$$\begin{aligned} \theta^i_x &= \frac{\partial\theta^i(x,p)}{\partial x} = \frac{i}{S_\theta} > 0 \text{ where } S_\theta = \partial S(p,i,\theta^i)/\partial\theta \\ \theta^i_p &= \frac{\partial\theta^i(x,p)}{\partial p} = q(p,i,\theta^i)\theta^i_x > 0. \end{aligned}$$

The demand function of the marginal consumer who makes i purchases, $q(p, i, \theta^i)$, is increasing in the delivery fee x as $q_x(p, i, \theta^i(x, p)) = q_\theta(p, i, \theta^i)\theta^i_x \geq 0$. The basket size of the marginal consumer with respect to an increase in the prices of the secondary goods depends on how fast the demand curve changes as a function of taste parameter θ as well as the price p $(q_p(p, i, \theta^i(x, p)) = q_p(p, i, \theta^i) + q_\theta(p, i, \theta^i)\theta^i_p)$. The demand $q(p, i, \theta^i)$ increases in θ but decreases in p. As p increases, θ^i that describes the i^{th} marginal consumer increases; therefore, the demand of the marginal consumer, $q(p, i, \theta^i)$, may go up or down as p increases.

[7]The reason behind taking M as a finite number is not only the technical details on avoiding that the analytic properties of summations play a role, but also the realistic setting that has an upper limit on the number of purchases.

[8]Note that we use subscript as a shorthand notation for the partial derivation, i.e. $q_p = \partial q/\partial p$.

Now we have a look at how the total market demand for the primary and secondary goods changes in their prices.

$$
\begin{aligned}
N_p(x,p) &= \frac{\partial N(x,p)}{\partial p} = -\sum_{i=1}^{M} \frac{1}{i} m(\theta^i) q(p,i,\theta^i)\theta_x^i \\
N_x(x,p) &= \frac{\partial N(x,p)}{\partial x} = -\sum_{i=1}^{M} m(\theta^i)\theta_x^i \\
Q_p(x,p) &= \frac{\partial Q(x,p)}{\partial p} = \sum_{i=1}^{M} i\Big[\int_{\theta^i(x,p)}^{\theta^{i+1}(x,p)} q_p(p,i,\theta)m(\theta)d\theta + q(p,i,\theta^{i+1})m(\theta^{i+1})\theta_p^{i+1} \\
&\qquad - q(p,i,\theta^i)m(\theta^i)\theta_p^i\Big] \\
Q_x(x,p) &= \frac{\partial Q(x,p)}{\partial x} = \sum_{i=1}^{M} i[q(p,i,\theta^{i+1})m(\theta^{i+1})\theta_x^{i+1} - q(p,i,\theta^i)m(\theta^i)\theta_x^i]
\end{aligned}
$$

Notice that the demand for the primary good decreases as the prices of primary and secondary goods increase: N_x and N_p are non-positive for any distribution of consumer tastes $m(\theta)$ and any normal demand function $q(p,i,\theta)$.[9] However, the direction of the change of aggregate demand for the secondary good as prices increase is more complex. Let us assume that θ is uniformly distributed over the population. The aggregate secondary good demand increases with the primary good prices, Q_x is non-negative, if $q_\theta(p,i,\theta^i) \leq q_\theta(p,j,\theta^j)$ for $i > j$ and may decrease otherwise. In words, this sufficient condition on $Q_x \geq 0$ indicates that, with increasing frequency of repeat purchases, the basket size of the marginal consumer changes less in θ. Rewriting Q_x as

$$
Q_x(x,p) = \sum_{i=1}^{M} \left[(i-1)q(p,i-1,\theta^i) - iq(p,i,\theta^i)\right] \theta_x^i m(\theta^i) + Mq(p,M,\theta^{M+1})\theta_x^{M+1} m(\theta^{M+1})
$$

helps us to elaborate more on the condition for $Q_x \geq 0$. The second part of the above equation takes value 0 as $\theta^{M+1} = 1$. The full expression's sign, therefore, comes down to the sign of $(i-1)q(p,i-1,\theta^i) - iq(p,i,\theta^i)$ which compares the total consumption of the marginal consumer at the purchase frequency of i to her total consumption at $i-1$ repeat purchase level. Accordingly, if the marginal consumer at the repeat purchase level i decreases her purchase frequency and her total consumption does not decrease, then

[9] Normal demand function (q) stands for the demand function of a normal good that satisfies the condition of $\partial q/\partial I > 0$ where I stands for consumer's budget. In words, normal goods are any goods for which demand increases when income increases, and falls when income decreases as opposed to inferior goods' demand.

$Q_x \geq 0$. This is plausible as the price sensitive consumer spends less in the delivery fee due to buying less frequently and she does not have to reduce her grocery purchase.

On the other hand, keeping the price of the primary good x constant, one expects that the aggregate demand on the secondary goods decreases as p increases, in other words Q_p is non-positive.

Proposition 1 *If $m(\theta)$ is uniformly distributed and $q_\theta(p,j,\theta^j) \geq q_\theta(p,i,\theta^i)$ for $i > j$; then average basket size is increasing in the delivery fee.*

Proposition 1 states that as the delivery fee increases the average basket size of the population increases. This occurs if the total secondary good (groceries) demand of a consumer who is marginal at i primary good purchase (deliveries) does not decrease in decreasing i.

Next we study the optimal primary and secondary good prices with repeat purchase instances. Firm's profit function is given by

$$\pi(x,p) = (x-f)N(x,p) + (p-c)Q(x,p) \tag{4.1}$$

where f and c are the costs of providing each unit of primary and secondary good respectively. Differentiating (4.1) with respect to x and p, we obtain

$$\begin{aligned} \pi_x &= N(x,p) + (x-f)N_x(x,p) + (p-c)Q_x(x,p) = 0 \\ \pi_p &= Q(x,p) + (x-f)N_p(x,p) + (p-c)Q_p(x,p) = 0 \end{aligned}$$

Eliminating the term $(x-f)$ and solving both equalities, we obtain

$$\begin{aligned} p-c &= -\frac{Q(x,p)N_x(x,p)}{N_x(x,p)Q_p(x,p) - N_p(x,p)Q_x(x,p)} \frac{(Q(x,p)/N(x,.p))N_x(x,p) - N_p(x,p)}{(Q(x,p)/N(x,.p))N_x(x,p)} \\ &= -\frac{Q(x,p)N_x(x,p)}{N_x(x,p)Q_p(x,p) - N_p(x,p)Q_x(x,p)} \frac{\sum_{i=1}^{\infty}\left(q(p,i,\theta_i) - \frac{Q(x,p)}{N(x,p)}\right)m(\theta_i)\theta_i^x}{(Q(x,p)/N(x,.p))N_x(x,p)} \end{aligned}$$

Proposition 2 *If $q_\theta(p,j,\theta^j) \geq q_\theta(p,i,\theta^i)$ for $i > j$ and $Q_p \leq 0$, at the optimum $(p-c)$ has the sign of*

$$\sum_{i=1}^{\infty}\left(\frac{Q(x,p)}{N(x,p)} - q(p,i,\theta^i)\right)m(\theta^i)\theta_i^x.$$

Notice that if the number of purchases is limited to one, then Proposition 2 reduces to Proposition 5 of Schmalensee (1981) that says $(p-c)$ has the sign of $\left(\frac{Q(x,p)}{N(x,p)} - q(p,1,\theta^1)\right)$

where $q(p,1,\theta^1)$ is the consumption of marginal consumer and Q/N is the average consumption. Similarly, we show here that firms should charge a premium on secondary goods if the secondary good consumption of the average consumer is higher than that of the marginal consumer, $\sum_{i=1}^{M}\left(\frac{Q(x,p)}{N(x,p)} - q(p,i,\theta^i)\right)m(\theta^i)\theta^i_x > 0$. One sufficient condition for this to hold is $\frac{Q(x,p)}{N(x,p)} > q(p,1,\theta^1)$ since $q(p,i,\theta^i)$ is decreasing in i. Alternatively, if the expected demand of marginal consumers (demand of marginal consumers weighted by the type distribution) is higher than the average basket size, then the firm should offer a discount on the secondary goods, otherwise it should charge a premium on those.

Proposition 3 *If $q_\theta(p,j,\theta^j) \geq q_\theta(p,i,\theta^i)$ for $i > j$ and $Q_p \leq 0$, at the optimum $(x-f) \geq 0$.*

Table 4.1 summarizes our theoretical predictions regarding the relationship among the delivery fee (x), grocery price p and the number of transactions, total basket size, average basket size.

Table 4.1: Summary of Theoretical Predictions

	Delivery fee (x)	Unit Price (p)
Total number of transactions (N)	-	-
Total Basket Size (Q)	+*	-
Average Size of the Basket(Q/N)	+*	+/-
Marginal consumer demand ($q(p,i,\theta^i)$)	+	+/-

+ indicates increasing; - indicates decreasing; +/- indicates may increase or decrease. * holds if $q_\theta(p,j,\theta^j) \geq q_\theta(p,i,\theta^i)$ for $i > j$.

In the following sections, we first provide information on our dataset and then validate our theoretical predictions on this particular dataset using simple regression techniques.

4.3 Data

Our data comes from an online grocery store in a Western European country. This online grocery store is the internet channel of the leading brick and mortar grocery chain in the country in terms of market share, employing more than 200,000 people. The online retailer offers approximately 10,000 stock keeping units [SKU], including fresh groceries such as meat, milk, and fruit. Customers of this company choose a convenient delivery time slot that they need to pay an additional time-specific delivery fee before they continue with their grocery shopping. The company offers its attended home delivery service in all major urban areas in the country; 44 different cities that vary widely in size, and roughly 65% of the country's households can access this service.

The transaction data that we use in this study is from 2008 and 2009. In this period, the online store had more than 1,920,000 transactions from approximately 200,000 different customers. We select 29,988 customers and all of their transactions during years 2008 and 2009 to a total of 953,107 transactions using two different criteria. First, we randomly select 10,000 customers that order at least once during both the first three months of 2008 and the last three months of 2009. We choose these selection criterion so that we are able to capture behavior from those customers that are the most loyal to the company. Second, we randomly select 10,000 customers among all who purchased online in each of January, February and March of 2009 and then merge both of these datasets avoiding customer and transaction repetition.[10]

Among all transactions in our data, 72% of the orders were from regular household customers which form 81% of the sampled customer base. The remaining 28% of the transactions comes from small businesses without professional catering service such as child-care centers, senior centers, law firms and IT firms. As this type of business customers have different characteristics than the regular households such as higher order volumes and frequent orders, we create a dummy variable B2B in our data that distinguishes customers between businesses and households.

The online grocer delivers orders in six days of the week (Monday to Saturday) and in eight two-hour time slots a day. Upon login, the customer reserves a two-hour delivery slot. In order to plan the delivery routes more effectively, these slots are overlapping with each other, such as 8 AM to 10 AM, 9 AM to 11 AM, 10 AM to 12 PM, 11 AM to 1 PM, 12 PM to 2 PM, 4 PM to 6 PM, 5 PM to 7 PM, and 6 PM to 8 PM. All time slots are available daily, except for Saturday that is missing the three slots after 4 PM and Sunday that does not offer service. The online retailer uses differentiated delivery fees to steer demand. This helps to improve the capacity utilization of the delivery service by balancing the demand across week days as well as within a day. In the same way that the delivery fee ranges from 4.95 to 11.95 based on the popularity of the time slot, customers display a wide range of demand intensity in their online purchase behavior. Figure 4.1 shows the distribution of transactions per customer in our sample.

Our transaction data contains information on several dimensions such as the number of items in the shopping basket, Euro value of the basket, the number of items with price discounted, the number of items per category such as frozen, cold, inedible types, the

[10]It is important to note that we append datasets sampled using two distinct criteria and drop the repeats and end up with a total number of 953,107 transactions. We combined these criteria to make sure our final dataset includes a fair amount of loyal customers that had purchased from the grocer at the beginning and the end of our sample period, as well as other customers that only purchased groceries randomly in the middle of this period.

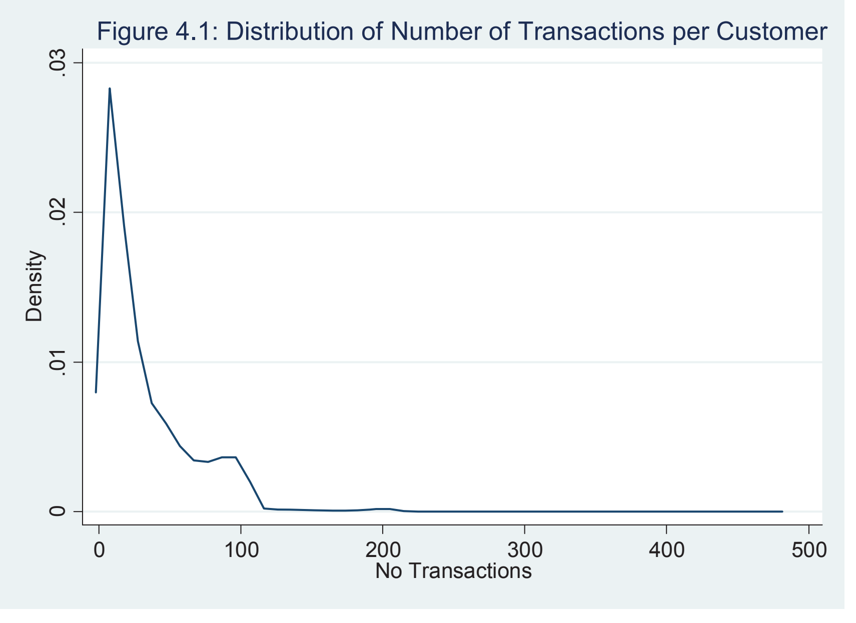


Figure 4.1: Distribution of Number of Transactions per Customer
Density
0
.01
.02
.03
0
100
200
300
400
500
No Transactions

basket's profit, the delivery fee, and whether the customer is a business or a household. Table 4.2 provides summary statistics for the whole dataset at the transaction level.

Table 4.2 shows that the average transaction contains a basket made up of 68 items of which 6 come with a discount. This average basket is worth 131 Euros before discount and 126 Euros after discount. On average, 62% of items fall under the generic definition of Category 1 while cool products represent 34% of the basket. The remaining 6% is divided into deep-freeze items, inedible and Category 2 type of goods (mainly crates of beverage).

Table 4.2: Summary Statistics

Variable	Obs.	Mean	Std. dev.	Min	Max
No. of items in the bundle	953107	68.43	40.82	1	1838
Value of the bundle (before discount)	953107	131.24	69.40	60	2545.65
Value of the bundle (after discount)	953107	126.43	67.46	37.25	2545.65
No. of items with a discounted price (if any)	953107	6.22	8.76	0	260
% of items with a discounted price	953107	9.95%	13.35%	0	100%
No. of items from the product group 1	953107	42.57	30.01	0	1650
No. of items from the product group 2	953107	0.60	1.55	0	80
No. of items from the cooled product group	953107	22.98	17.80	0	396
No. of items from the deep-freezed product group	953107	1.82	3.20	0	213
No. of items from the inedible product group	953107	0.45	1.40	0	150
% of items from the product group 1	953107	61.55%	16.40%	0%	100%
% of items from the product group 2	953107	1.35%	5.74%	0%	100%
% of items from the Cooled product group	953107	33.53%	16.66%	0%	100%
% of items from the deep-freezed product group	953107	2.73%	4.21%	0%	100%
% of items from the unedible product group	953107	0.84%	2.82%	0%	100%
Delivery fee per transaction	953107	7.26	1.41	4.95	11.95
% of discount amount of the bundle	953107	3.57%	4.89%	0%	53.86%
B2B dummy variable	953107	0.286	0.45	0	1
Marginal dummy variable	953107	0.498	0.50	0	1

Note: This table provides summary statistics of all variables used in our empirical analysis.

In our dataset we classify customers into two different categories: B2B vs. B2C, as well as marginal and inframarginal customers. The former division comes from information readily available in the dataset that specifies whether a customer is a business or a regular household. The latter definition is driven by our observation that most customers only purchase on a handful amount of time slot and day combinations. We define those purchasing (placing orders) 11 times or less during our sample period in a given time slot and day combination as marginal consumer and those placing orders more often as inframarginal consumer.[11] In our empirical study, we verify that this classification helps us to identify those who are more price sensitive compared to the others. Table 4.2 shows that almost 29% of transactions come from B2B and that close to 50% of transactions are from marginal consumers. Table 4.3 tabulates the interrelation between B2B and marginal customers, and shows that businesses are more likely to be inframarginal than households are. In other words, households are going to be more sensitive in the margin to changes in price offerings as well as relative improvements to their outside option.

Table 4.3: Cross-Tabulation of B2B Vs. Marginal Customers

	Inframarginal	Marginal	Total
B2C	308,705	371,693	680,398
B2B	169,989	102,720	272,709
Total	478,694	474,413	953,107

Note: This table cross-tabulates number of transactions for whether the customer is a firm (B2B) or a household (B2C), as well as marginal (shows up less than 11 times in its time slot) or inframarginal (more than 11 times).

Table 4.4 combines Tables 4.2 and 4.3 and provides separate summary statistics for businesses and households as well as marginal and inframarginal customers. When comparing B2B to B2C and marginal to inframarginal, we show that these groups of customers are different from each other in all variables. In particular, while businesses and inframarginal customers purchase baskets with less discounted number of items and less percentage discount in their baskets, and pay higher delivery fees; shopping baskets of business and marginal customers are larger in value. Hence, rather than directly associating B2B customers with the inframarginal customers, one should be careful on the higher dimensionality in the data.

Finally, Table 4.5 breaks down the sample into the 45 time slot-day combinations for which the customers can order their online deliveries. This table provides averages for

[11] We chose 11 as the dividing number because that is the median value of the number of transactions per time slot, day, and city per customer.

Table [illegible] Summary Statistics by Consumer Type: B2B Vs. B2C & Marginal Vs. Inframarginal Consumer

Variable	B2B - B2C			Marginal - Inframarginal		
	B2B (1)	B2C (0)	Diff (0) - (1)	Marginal (1)	Inframarginal (0)	Diff (0) - (1)
No. of items in the bundle	83.82	62.26	-21.55***	65.28	71.55	6.27***
	(60.02)	(27.53)		(37.75)	(43.43)	
Value of the bundle (before discount)	143.00	126.53	-16.47***	133.27	129.22	-4.05***
	(90.30)	(58.33)		(71.18)	(67.54)	
Value of the bundle (after discount)	-141.17	120.52	-20.65***	128.03	124.84	-3.18***
	(89.16)	(55.37)		(68.61)	(66.25)	
No. of items with a discounted price (if any)	2.93	7.53	4.60***	6.46	5.97	-0.49***
	(6.11)	(9.31)		(9.26)	(8.23)	
% of items with a discounted price	0.04	0.12	0.09***	0.10	0.09	-0.01***
	(0.08)	(0.14)		(0.14)	(0.12)	
No. of items from the product group 1	51.06	39.17	-11.88***	43.40	41.75	-1.65***
	(43.94)	(21.16)		(30.11)	(29.90)	
No. of items from the product group 2	0.60	0.61	0.01***	0.80	0.41	-0.39***
	(1.69)	(1.49)		(1.90)	(1.05)	
No. of items from the cooled product group	30.98	19.78	-11.20***	18.66	27.26	8.60***
	(26.97)	(10.79)		(14.77)	(19.44)	
No. of items from the deep-freezed product group	1.00	2.16	1.16***	1.95	1.70	-0.24***
	(3.78)	(2.87)		(3.29)	(3.10)	
No. of items from the inedible product group	0.19	0.55	0.36***	0.47	0.43	-0.04***
	(1.02)	(1.52)		(1.53)	(1.26)	
% of items from the product group 1	0.61	0.62	0.01***	0.66	0.58	-0.08***
	(0.22)	(0.14)		(0.17)	(0.14)	
% of items from the product group 2	0.013	0.014	0.001***	0.02	0.01	-0.01***
	(0.05)	(0.06)		(0.07)	(0.03)	
% of items from the Cooled product group	0.37	0.32	-0.05***	0.29	0.38	0.10***
	(0.22)	(0.13)		(0.17)	(0.15)	
% of items from the deep-freezed product group	0.01	0.03	0.02***	0.030	0.025	-0.005***
	(0.04)	(0.04)		(0.05)	(0.04)	
% of items from the inedible product group	0.003	0.011	0.008***	0.009	0.008	-0.001***
	(0.02)	(0.03)		(0.03)	(0.03)	
Delivery fee per transaction	7.69	7.09	-0.60***	7.10	7.42	0.32***
	(1.61)	(1.28)		(1.28)	(1.52)	
% of discount amount of the bundle	0.01	0.05	0.03***	0.04	0.03	-0.003***
	(0.02)	(0.05)		(0.05)	(0.05)	
No. of observations	272,709	680,398		474,413	478,694	

Note: This table provides summary statistics by consumer type reporting averages per variable (top number) and standard deviations (bottom number in brackets).
The first three columns separate transactions by whether consumers are businesses (B2B) or households (B2C). The third column reports differences between both groups.
The second set of three columns separate transactions by whether consumers are marginal or inframarginal. We define consumers as marginal if the number of times they purchase goods in a given time slot, day and city is below the median of the distribution (11 times or less during the two-year period of our data). Inframarginal consumers are those purchasing more often than 11 times during our sample period. All differences are statistically significant at the 1% level, and so we signal this with ***.

four variables that characterize the relevance of each time slot in each window. These variables are the basket value in Euros (upper left corner), the delivery fee in Euros (upper right corner), the number of items per basket (lower left corner), and the percentage of transactions occurring in each time slot (lower right corner). As delivery service is not available on Saturday afternoon and Sundays, we have no summary statistics for these variables at those time slots.

Among other things, Table 4.5 shows that afternoon slots are more popular than morning slots in any given day as well as the fact that slots in Monday morning are more popular than the same time slots in any other day (except for Saturday). For the most part, first and second morning slots account for the largest baskets in value and number of items. Finally, this table also depicts the seemingly random[12] (although not quite) variation in delivery fees across time slots and days. Table 4.5 provides averages across 105 weeks in 2008 and 2009 and shows that delivery fees are especially higher on Monday, Friday and Saturday morning slots as well as afternoon slots across the board.[13]

We also compute averages per time slot and day of the week for all other variables in our dataset. We show in Figures 4.2 to 4.4 the empirical relation in our sample between pairs of variables taking as observation time slot and day combination. Figures 4.2A, 4.2B and 4.2C show no clear-cut relationship between number of transactions and delivery fees, average basket value and delivery fee, or average basket value and number of transactions. Only Figure 4.2B depicts slight evidence on the fact that average basket value increases in delivery fees. Figure 4.3A shows a positive correlation between delivery fee and percentage profit per basket but Figures 4.3B and 4.3C find no relationship between profits and average basket size or number of transactions. Finally, Figure 4.4 investigates the composition of baskets and finds that larger baskets are likely to have lower shares of deep-freeze items, group 2 items and inedible items, but higher shares of cold items.

These figures call for deeper empirical work and are useful to justify the introduction of variables that may avoid potential spurious correlations, and control for basket and customer heterogeneity. Once presented the data, we introduce the empirical methodology and results in the following section.

[12]Even though the delivery fees takes exact values from 4.95 to 11.95 Euros, due to the fact that some day-time slots changed fees over time, average values are not exact.

[13]Delivery fees did not change at all for most slots during our sample period of time. Only a few slots changed pricing at the end of our sample in December of 2009.

Table 4.5: Bundle Size, Delivery Fee, and Number of Items per Time Slot and Day of the Week

TIME SLOT\WEEK DAY	MONDAY	TUESDAY	WEDNESDAY	THURSDAY	FRIDAY	SATURDAY
8:00 AM 10:00 AM	€ 171.08 € 9.03 112.05 1.86%	€ 147.67 € 7.00 87.52 1.50%	€ 140.25 € 6.95 79.07 1.32%	€ 140.11 € 6.95 76.19 1.22%	€ 140.28 € 7.94 73.68 1.71%	€ 133.39 € 8.96 66.37 2.23%
9:00 AM 11:00 AM	€ 141.89 € 10.03 86.15 2.59%	€ 141.93 € 8.00 79.36 1.85%	€ 137.12 € 7.95 73.66 1.66%	€ 135.79 € 7.95 72.13 1.73%	€ 146.68 € 8.94 77.03 2.05%	€ 135.29 € 8.95 68.42 2.27%
10:00 AM 12:00 PM	€ 127.64 € 10.03 74.87 2.51%	€ 128.43 € 6.08 71.52 2.14%	€ 127.74 € 5.99 67.78 1.99%	€ 131.21 € 6.00 68.28 1.89%	€ 137.24 € 7.94 70.85 2.12%	€ 128.69 € 7.97 64.06 2.35%
11:00 AM 13:00 PM	€ 125.23 € 9.01 71.21 1.30%	€ 127.26 € 6.06 64.20 1.03%	€ 128.79 € 5.99 62.24 1.00%	€ 130.09 € 5.99 60.76 0.98%	€ 135.85 € 7.95 65.99 0.68%	€ 126.25 € 7.96 60.30 0.72%
12:00 PM 2:00 PM	€ 137.87 € 7.98 81.26 1.80%	€ 131.89 € 5.01 69.65 1.43%	€ 130.84 € 4.95 71.92 1.80%	€ 127.19 € 4.95 65.40 1.63%	€ 133.77 € 6.95 69.07 1.59%	€ 125.70 € 6.97 61.99 2.33%
4:00 PM 6:00 PM	€ 125.35 € 5.99 64.00 3.56%	€ 123.39 € 4.95 64.17 3.67%	€ 122.39 € 4.95 61.21 3.88%	€ 129.44 € 6.94 63.89 3.02%	€ 132.83 € 7.97 66.32 4.02%	
6:00 PM 8:00 PM	€ 123.90 € 6.95 60.96 3.40%	€ 126.89 € 6.94 62.62 3.05%	€ 123.75 € 6.14 60.49 3.39%	€ 132.76 € 7.94 66.46 2.74%	€ 130.85 € 8.94 65.53 3.51%	
7:00 PM 9:00 PM	€ 122.42 € 6.95 60.71 2.85%	€ 124.68 € 6.94 62.09 2.79%	€ 122.62 € 6.13 61.46 2.95%	€ 130.95 € 7.94 65.31 2.55%	€ 131.34 € 8.94 66.33 3.33%	

Note: This table provides summary statistics per time slot of delivery and day of the week. Each window contains four numbers: average bundle value is in the upper left corner, average delivery fee in the upper right corner, average number of items per bundle in the lower left corner, and the percentage of observations over total number of observations in the lower right corner.
There is no delivery service on Saturday afternoon and Sundays.

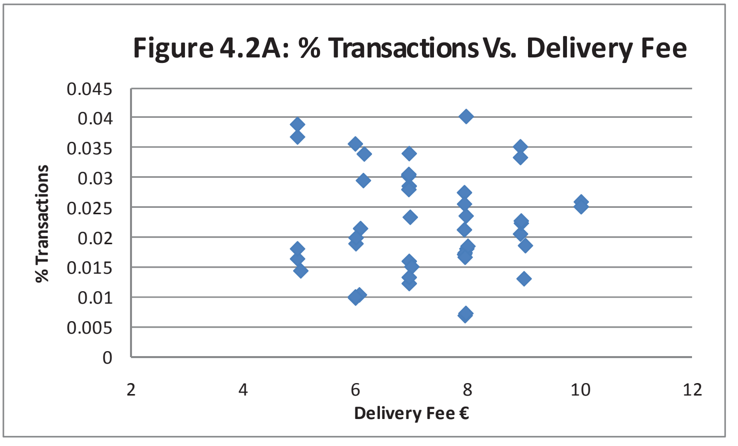


Figure 4.2A: % Transactions Vs. Delivery Fee

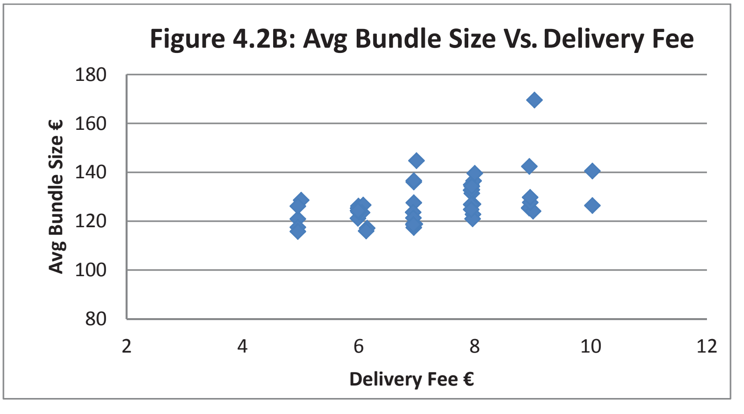


Figure 4.2B: Avg Bundle Size Vs. Delivery Fee

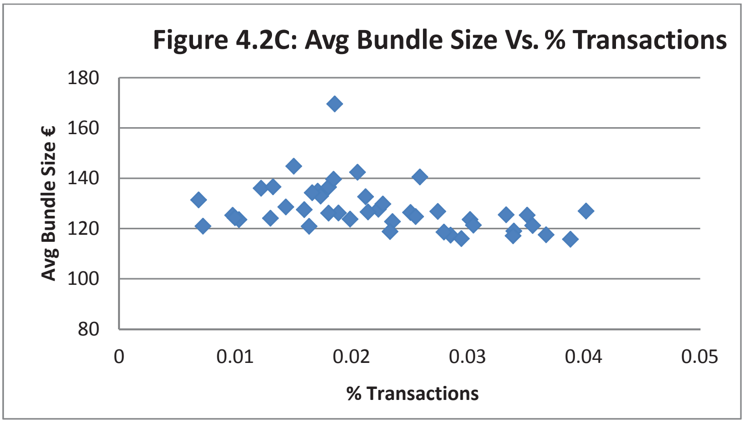


Figure 4.2C: Avg Bundle Size Vs. % Transactions

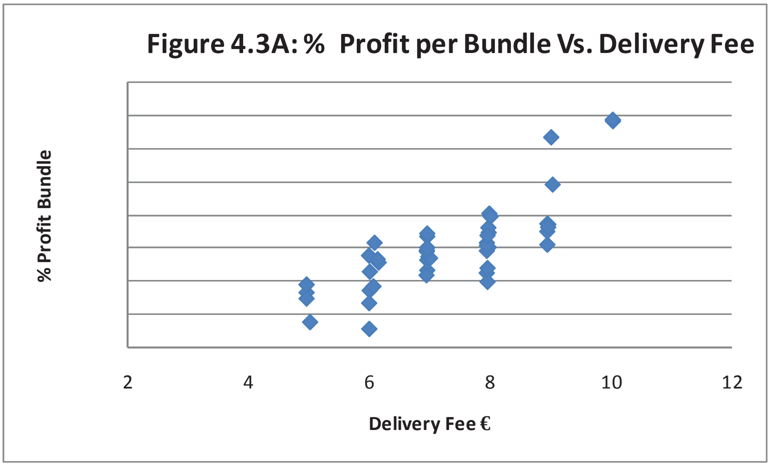


Figure 4.3A: % Profit per Bundle Vs. Delivery Fee
% Profit Bundle
2
4
6
8
10
12
Delivery Fee €

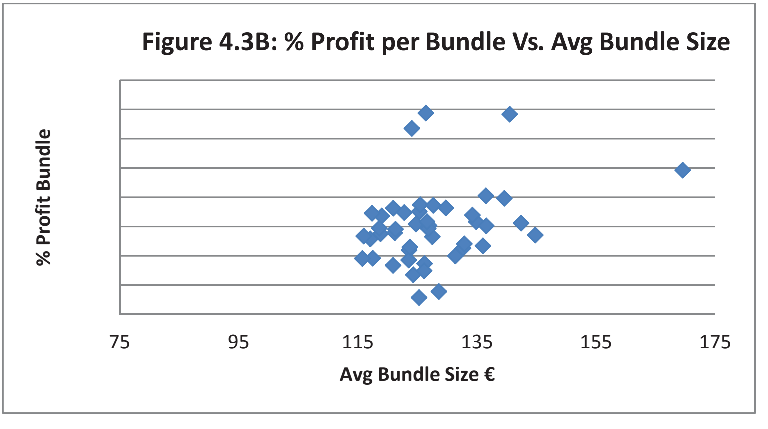


Figure 4.3B: % Profit per Bundle Vs. Avg Bundle Size
% Profit Bundle
75
95
115
135
155
175
Avg Bundle Size €

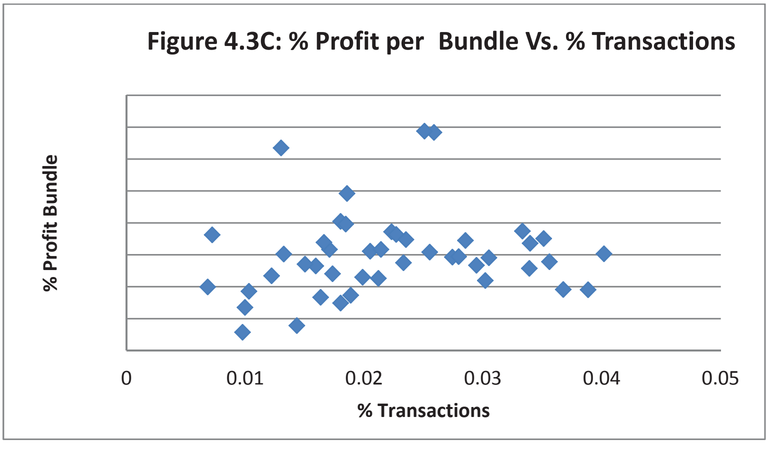
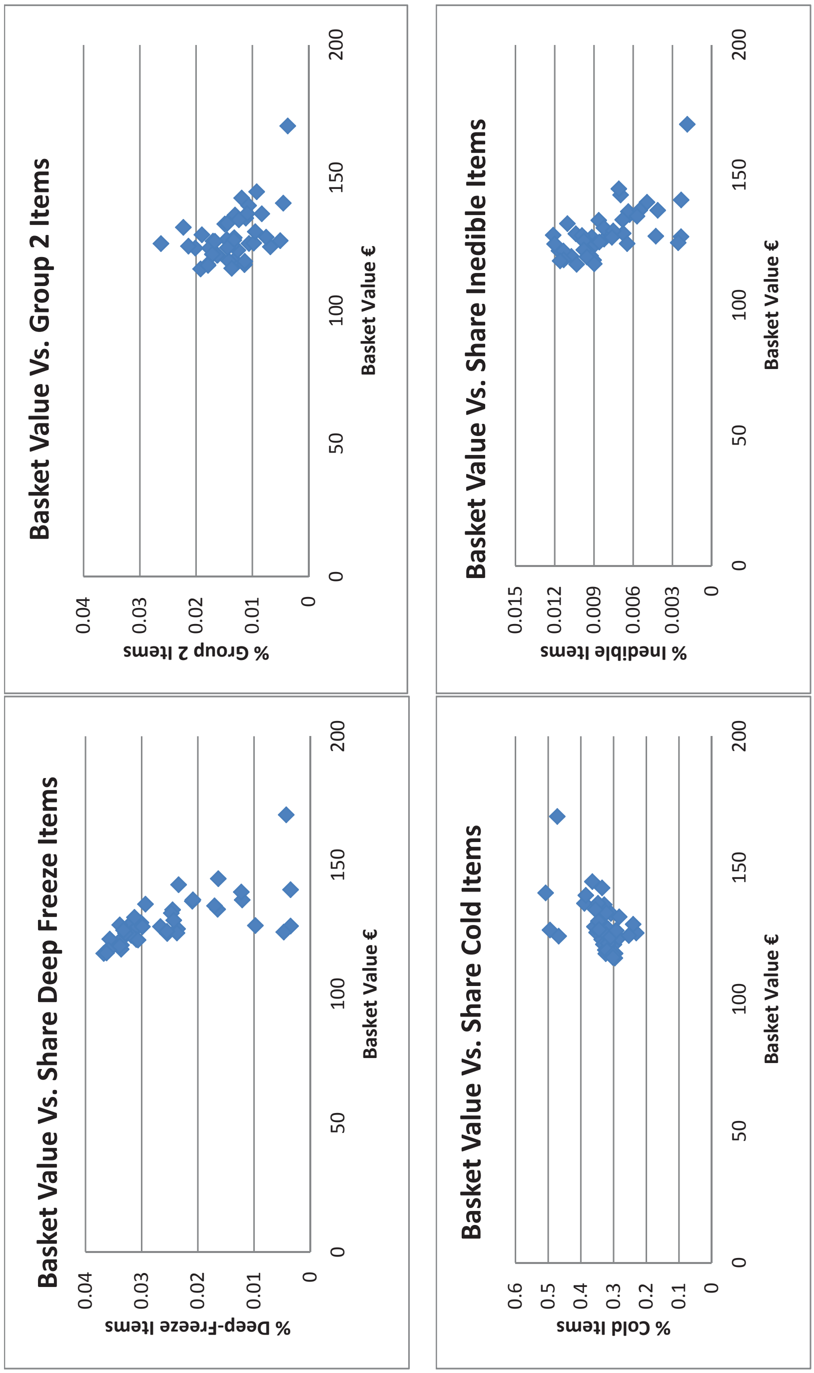


Figure 4.3C: % Profit per Bundle Vs. % Transactions
% Profit Bundle
0
0.01
0.02
0.03
0.04
0.05
% Transactions

Figure 4.4 Basket Value Vs. Basket Item Composition



4.4 Empirical Methodology and Results

To uncover joint demand distribution of primary good and secondary good from data, we employ an approach that is similar to Gil and Hartmann (2009). The main difference from Gil and Hartmann (2009) is in our dataset consumers adjust their number of visits to the firm and the amount of secondary good purchases per visit. We first present our empirical methodology and then the results from the empirical analysis in the following subsections.

4.4.1 Methodology

Our empirical methodology mainly consists two parts. First, we are going to reveal the difference between marginal and inframarginal consumers' secondary good demand. Second, following our theoretical exploration, we are going to empirically examine the relationship between three pairs of variables (number of transactions, basket value and delivery fee) in our dataset following our theoretical predictions in Table 4.1.

First, we are going to explore the correlation between the average basket value and the number of transactions. This empirical exploration is very crucial as it provides an understanding on how the secondary good (groceries) demand and willingness to pay for the primary good (home delivery service) are related to each other. The idea of using the consumer's intensity of demand for the secondary good as a meter of how much the consumer is willing to pay for the primary good is known as metering. Consequently, we unfold how marginal and inframarginal consumers differ in their secondary good demand. This central comparison helps us to come up with an optimal uniform two-part pricing policy over a heterogeneous customer base. The intuition behind this analysis is that increases in primary good demand typically involve more low willingness-to-pay customers, such that increases (decreases) in secondary good demand per buyer would indicate a negative (positive) correlation between secondary good demand and willingness to pay for the primary good. In order to reveal the correlation between the average basket value and the number of transactions, we exploit here the methodology in Gil and Hartmann (2009) and use the following regression specification,

$$\log(Total\_Sales_{tdcw}) = \alpha + \beta \log(No\_Transactions_{tdcw}) + \gamma X_{tdcw} + u_{tdcw}$$

such that $Total\_Sales_{tdcw}$ are the total sales of the online grocer in time slot t, week day d, city c and week w, $No\_Transactions$ is the number of transactions, X_{tdcw} are variables that control for basket composition as well as time slot and week fixed effects. We use log of $Total\_Sales$ instead of the average basket size to avoid potential problems of mechanical negative correlation between the average basket size and the number of transactions, and test whether β is greater than, equal to, or less than one.

Second, we use transaction level data to investigate the relationship between basket value and the delivery fee. For this purpose we run OLS regressions such that,

$$Value\_Basket_{ctrw} = \alpha + \beta Delivery\_Fee_{ctrw} + \gamma X_{ctrw} + u_{ctrw}$$

where the observation unit here is a transaction tr unique to a customer c and a week w. This test links the theory on how the secondary good consumption (grocery sales) changes in primary good price (delivery fee) to the empirical level.

Finally, we run OLS regressions to check whether increases in delivery fees deter online grocery shopping such that,

$$No\_Transactions_{tdcw} = \alpha + \beta Delivery\_Fee_{tdcw} + \gamma X_{tdcw} + u_{tdcw}$$

where the dependent variable is the number of transactions that took place for our set of consumers within a time slot, week day, city and week. According to our theoretical predictions, we would expect to see a decrease on the number of transactions in increasing delivery fee.

Our theory section also yields predictions on the impact of item prices on the number of transactions, basket size and average basket size. As our data is detailed at the transaction (basket) level, we will only be able to point out the correlation between the percentage value of the discount and our three variables of study.

Finally, in order to have a thorough understanding on our data, we repeat all the empirical tests on the level of business vs. household customers as well as marginal vs. inframarginal customers, and draw comparisons on both classification. By this means we can also explore the validity of our theoretical predictions on the positive relation between delivery fee and the demand of the marginal consumer. Note, however, that the marginal-inframarginal classification slightly differs between the empirical part and the theoretical part. While the marginal consumer has been defined at each level of repeat purchase in the theoretical part, we empirically capture the price sensitivity in consumer's behavior by their number of repeat purchases in our transaction data. We next proceed to show our results.

4.4.2 Results

First, we show results regarding the relationship between the average basket size and the number of transactions. According to results in Table 4.6, the logarithm of sales is positively correlated with the logarithm of the number of transactions. Although this is not surprising, we are interested in whether the coefficient value is above one as a coefficient larger than one means that the average basket size increases with the number of transactions. We tested the coefficients in columns 1 to 4 as we include extra controls as well as fixed effects and find that all are statistically significant and higher than one. This is consistent with marginal consumers making larger purchases than inframarginal consumers and with our pricing scenario that increases margins in the primary good or access fees and/or lowers prices of the secondary goods.

Table 4.7 repeats the analysis taking into account the potentially different behavior of businesses and households, as well as marginal and inframarginal consumers. Columns 1 to 6 investigate behavior of marginal consumers (relative to inframarginal consumers) and find that marginal consumers purchase higher basket values. Even when we consider only marginal consumers, the more marginal consumers purchase in a given week, time slot, day and city the higher the average basket size (coefficient larger than one). Columns 7 to 12 explore behavior of business customers (relative to households) and find that business customers purchase higher value baskets. Once we control for basket composition, business customers seem to purchase lower value baskets. This indicates that households and business customers purchase very different types of baskets, so that when adjusting for composition the initial results flip. In addition to this, both B2B and households increase their average basket value with the number of transactions which also indicates that both groups contain marginal consumers. Results in Table 4.7 are consistent with those of Table 4.6 in that marginal consumers have higher values for the secondary good.

In our second part of the analysis, we explore the relationship between basket value and delivery fee. Table 4.8 exhibits a significant positive correlation between basket size and delivery fee from columns 1 to 5. These columns exhibit time slot/day/city and customer id fixed effects. These basically show that those slots with higher delivery fees are also attracting the most valuable customers. Contrary to this, column 6 include customer id fixed effects and basket heterogeneity controls but finds a negative correlation between delivery fees and basket size. Columns 7 and 8 combine week fixed effects with other fixed effect and provide no statistically significant result. These disappointing results from columns 6 to 8 can possibly be explained by the fact that only a few time slots changed delivery fees and these did so right before the Christmas season of 2009 when sales are ready slow. Later in the paper we explore this particular event more carefully

Table 4.6: Total Sales Value and Number of Transactions

	(1)	(2)	(3)	(4)
Dep Var:	**ln(sales)**			
ln(No. Transactions)	1.026***	1.041***	1.045***	1.042***
	(0.003)	(0.002)	(0.002)	(0.002)
No. Items		0.008***	0.008***	0.008***
		(0.001)	(0.001)	(0.001)
Share Discounted Items		0.122***	0.095***	0.165***
		(0.026)	(0.021)	(0.027)
Percentage Discount		-0.391***	-0.338***	-0.460***
		(0.065)	(0.052)	(0.072)
Share Deep-Freeze Items		-0.813***	-0.728***	-0.782***
		(0.249)	(0.246)	(0.284)
Share Group 1 Items		-0.938***	-0.855***	-0.775***
		(0.245)	(0.245)	(0.281)
Share Group 2 Items		-0.248	-0.132	-0.054
		(0.250)	(0.250)	(0.285)
Share Cool Items		-1.179***	-1.090***	-1.019***
		(0.245)	(0.245)	(0.280)
Share Inedible		0.566**	0.588**	0.682**
		(0.274)	(0.270)	(0.307)
Constant	4.763***	5.169***	5.054***	4.970***
	(0.007)	(0.246)	(0.246)	(0.282)
Week FE	No	No	No	Yes
Time Slot/Day/ City FE	No	No	Yes	Yes
Observations	139,056	139,056	139,056	139,056
R-squared	0.92	0.97	0.98	0.97

Note: This table presents OLS specifications that regress ln(total sales) per time slot, week day and city on the ln(number of transactions). A coefficient larger than one implies that the average transaction increases with the number of transactions, and it is easy to show that all coefficients are statistically larger than 1.
Robust standard errors in parentheses clustered at the time slot, week day and city level.
*** p<0.01, ** p<0.05, * p<0.1.

Table 4.7: Total Sales Value and Number of Transactions: Marginal Vs. Inframarginal, and B2B Vs. B2C.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dep Var:	**ln(sales)**											
ln(No. Transactions)	1.034***	1.046***	1.043***	1.047***	1.041***	1.043***	1.030***	1.041***	1.042***	1.063***	1.041***	1.063***
	(0.003)	(0.002)	(0.002)	(0.001)	(0.002)	(0.001)	(0.003)	(0.003)	(0.002)	(0.004)	(0.002)	(0.004)
Marginal?	0.015***	0.040***										
	(0.005)	(0.003)										
B2B?							0.141***	-0.024***				
							(0.007)	(0.006)				
No. Items		0.008***	0.008***	0.009***	0.008***	0.008***		0.007***	0.012***	0.007***	0.012***	0.007***
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share Discounted Items		0.133***	0.044	0.151***	0.053	0.170***		0.168***	0.119***	0.524***	0.156***	0.529***
		(0.019)	(0.035)	(0.019)	(0.038)	(0.020)		(0.028)	(0.026)	(0.051)	(0.028)	(0.052)
Percentage Discount		-0.376***	-0.152**	-0.435***	-0.082	-0.427***		-0.578***	-0.479***	-1.079***	-0.459***	-1.070***
		(0.047)	(0.072)	(0.048)	(0.083)	(0.051)		(0.069)	(0.063)	(0.130)	(0.071)	(0.133)
Share Deep-Freeze Items		-0.745***	-0.417	-0.997***	-0.337	-0.941***		-1.120***	-0.769***	-1.388**	-0.676**	-1.060***
		(0.203)	(0.287)	(0.216)	(0.311)	(0.262)		(0.285)	(0.231)	(0.600)	(0.258)	(0.614)
Share Group 1 Items		-0.821***	-0.544*	-1.054***	-0.434	-0.943***		-1.099***	-0.723***	-1.492**	-0.568**	-1.153*
		(0.201)	(0.283)	(0.216)	(0.309)	(0.261)		(0.274)	(0.230)	(0.594)	(0.258)	(0.610)
Share Group 2 Items		-0.126	0.375	-0.392*	0.479	-0.284		-0.427	0.171	-0.849	0.320	-0.512
		(0.203)	(0.296)	(0.219)	(0.321)	(0.265)		(0.271)	(0.222)	(0.576)	(0.252)	(0.592)
Share Cool Items		-1.018***	-0.639**	-1.340***	-0.525*	-1.234***		-1.341***	-0.854***	-1.742***	-0.698***	-1.403**
		(0.201)	(0.284)	(0.216)	(0.310)	(0.261)		(0.275)	(0.231)	(0.592)	(0.259)	(0.608)
Share Inedible		0.837***	1.046***	0.434*	1.163***	0.522*		0.142	1.272***	-0.613	1.405***	-0.267
		(0.224)	(0.320)	(0.233)	(0.340)	(0.277)		(0.296)	(0.270)	(0.609)	(0.286)	(0.626)
Constant	4.741***	5.018***	4.699***	5.288***	4.557***	5.146***	4.713***	5.394***	4.703***	5.816***	4.536***	5.438***
	(0.007)	(0.202)	(0.286)	(0.218)	(0.311)	(0.263)	(0.005)	(0.276)	(0.231)	(0.595)	(0.260)	(0.614)
Marginal?	All	All	No	Yes	No	Yes	-	-	-	-	-	-
B2B?	-	-	-	-	-	-	All	All	No	Yes	No	Yes
Week FE	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Time Slot/Day/ City FE	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
Observations	234,648	234,648	119,748	114,900	119,748	114,900	196,941	196,941	118,028	78,913	118,028	78,913
R-squared	0.88	0.95	0.97	0.95	0.97	0.95	0.89	0.96	0.98	0.93	0.98	0.93

Note: This table runs same OLS regressions as Table 5 taking into account whether consumers are marginal or businesses. We define marginal as those that purchase in a given time slot, day and city less than the median (11 times in 2 years). Robust standard errors in parentheses clustered at the slot, day and city level. *** p<0.01, ** p<0.05, * p<0.1.

using diff-in-diff methodology. This table also shows that higher discounts are associated with larger baskets once we include customer fixed effects.

Table 4.9 separates purchasing behavior between business and households, and marginal and inframarginal consumers. The first set of regressions in Table 4.9 examines whether business customers (versus households) are more or less sensitive to delivery fees. Columns 4 and 6 show striking results on the fact that basket sizes from business customers are not sensitive to changes in delivery fees. On the other hand, results from columns 3 and 5 show mixed results about the relation between delivery fees and basket size purchases in household customers. If anything, column 5 shows a strong positive and statistically significant correlation between these two variables after including week and customer-specific fixed effects. This initial results on the very different behavior of household and business customers already encourage us to seek ways of price discriminate between groups rather than imposing a uniform two-part tariff over the whole customer base. Clearly the minor group of business customers in OG's database has different behavioral characteristics that our theoretical model does not cover.

The second half of Table 4.9 (columns 7 to 12) explores the differences in behavior between marginal and inframarginal customers. Once again marginal customers purchase much larger sizes at the same time that we find an overall positive correlation between delivery fees and basket sizes. The rest of columns show no significant correlation between delivery fees and basket size once we break the sample into marginal and inframarginal consumers and include week and customer-specific fixed effects. Table 4.9 also shows that household customers increase their basket sizes when item prices are more heavily discounted.

Third and finally, we check whether higher delivery fees are associated with a lower number of online transactions. Table 4.10 shows that there is indeed a negative relationship between delivery fees and the number of online purchases but that this one only shows up as statistically significant when including slot time/day/city and week fixed effects. This result again could be driven by the change in delivery fee before the Christmas season of 2009 and therefore grants further exploration. Note that discounted item prices seem to have no statistically significant relation with the number of transactions according to columns 3 and 4 once time slot fixed effects are included.

Table 4.11 examines differences in purchasing behavior between business and household customers.[14] Columns 1 to 4 show no overall relationship between the number of transactions and delivery fees. If anything, we find that business customers order less

[14]We do not explore differences in behavior between marginal and inframarginal consumers because we use the number of transactions to define that classification.

Table 4.8: Basket Value and Delivery Fee

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep Var:	**Basket Value**							
Delivery Fee	3.506***	5.428***	1.451***	1.278***	1.851***	-0.090*	-0.243	-0.068
	(0.287)	(0.426)	(0.117)	(0.130)	(0.219)	(0.051)	(0.240)	(0.051)
No. Items				1.366***	1.398***	1.537***	1.398***	1.534***
				(0.017)	(0.017)	(0.016)	(0.017)	(0.016)
Share Discounted Items				28.631***	27.433***	-6.100***	28.622***	-4.545***
				(2.000)	(1.917)	(1.083)	(1.970)	(1.111)
Percentage Discount				-65.678***	-65.297***	19.721***	-62.963***	28.721***
				(4.847)	(4.497)	(2.409)	(4.706)	(2.478)
Share Deep-Freeze Items				-3.251	-9.242	-20.284**	3.645	-3.155
				(13.833)	(13.449)	(8.785)	(14.430)	(9.329)
Share Group 1 Items				7.899	2.889	-29.897***	19.242	-10.187
				(13.384)	(12.925)	(8.781)	(14.013)	(9.340)
Share Group 2 Items				118.730***	116.035***	48.307***	132.241***	67.367***
				(14.396)	(13.914)	(9.851)	(14.896)	(10.269)
Share Cool Items				-29.167**	-29.291**	-42.633***	-12.843	-22.750**
				(13.492)	(13.057)	(8.746)	(14.138)	(9.299)
Share Inedible				270.267***	264.317***	139.578***	279.647***	158.597***
				(17.434)	(17.047)	(10.562)	(17.876)	(11.019)
Constant	100.965***	87.006***	115.892***	24.282*	21.410	53.251***	17.149	30.074***
	(1.962)	(3.074)	(0.851)	(13.654)	(13.169)	(8.840)	(14.266)	(9.366)
Time Slot/Day/City FE	No	Yes	No	No	Yes	No	Yes	No
Customer ID FE	No	No	Yes	No	No	Yes	No	Yes
Week FE	No	No	No	No	No	No	Yes	Yes
Observations	953,107	953,107	953,107	953,107	953,107	953,107	953,107	953,107
R-squared	0.01	0.05	0.62	0.68	0.69	0.88	0.70	0.88

Note: This table presents OLS regressions of basket value on the delivery fee. The observation level in these specifications is the transaction. Robust standard errors in parentheses cluster at the consumer level. *** p<0.01, ** p<0.05, * p<0.1

Table 4.9: Basket Value and Delivery Fee: B2B and Marginal Consumers.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dep Var:	**Basket Value**											
Delivery Fee	2.312***	1.523***	-0.038	-0.044	0.113***	-0.147	3.681***	1.384***	0.190	-0.142***	0.063	-0.026
	(0.263)	(0.131)	(0.039)	(0.116)	(0.038)	(0.120)	(0.338)	(0.176)	(0.140)	(0.052)	(0.143)	(0.053)
B2B?	19.258***	-7.199***										
	(1.302)	(0.633)										
Marginal?							4.363***	7.791***				
							(0.751)	(0.326)				
No. Items		1.382***	1.701***	1.432***	1.697***	1.430***		1.373***	1.476***	1.597***	1.475***	1.592***
		(0.017)	(0.007)	(0.024)	(0.007)	(0.024)		(0.013)	(0.025)	(0.013)	(0.025)	(0.013)
Share Discounted Items		27.207***	-11.786***	32.405***	-9.742***	32.107***		28.957***	-11.513***	-2.031	-10.846***	-0.095
		(2.020)	(0.936)	(4.384)	(0.964)	(4.448)		(1.610)	(1.494)	(1.269)	(1.563)	(1.291)
Percentage Discount		-78.810***	32.514***	-86.029***	44.746***	-80.461***		-66.545***	27.009***	15.411***	36.939***	24.063***
		(4.675)	(2.273)	(10.756)	(2.359)	(10.907)		(4.206)	(3.436)	(2.895)	(3.630)	(2.964)
Share Deep-Freeze Items		-25.297*	-26.216***	21.830	-2.260	31.045		-11.681	-29.529**	-12.735	-13.773	6.314
		(13.807)	(9.008)	(31.509)	(9.490)	(32.192)		(15.396)	(13.060)	(12.332)	(13.272)	(13.741)
Share Group 1 Items		0.962	-35.238***	9.988	-8.161	19.972		1.912	-40.244***	-22.115*	-23.484*	0.607
		(13.413)	(8.972)	(31.819)	(9.465)	(32.506)		(15.392)	(13.088)	(12.243)	(13.297)	(13.651)
Share Group 2 Items		114.135***	45.439***	91.069***	71.783***	100.417***		110.922***	61.445***	50.503***	77.421***	72.502***
		(14.447)	(10.764)	(32.304)	(11.021)	(32.953)		(15.722)	(16.115)	(13.123)	(16.176)	(14.386)
Share Cool Items		-34.849***	-39.976***	-8.376	-12.890	1.752		-28.771*	-46.933***	-39.475***	-29.479**	-17.233
		(13.516)	(8.982)	(31.284)	(9.464)	(32.012)		(15.329)	(12.901)	(12.247)	(13.107)	(13.657)
Share Inedible		253.287***	150.932***	159.940***	177.045***	169.906***		267.590***	131.514***	148.008***	148.777***	169.306***
		(17.333)	(10.449)	(36.197)	(10.907)	(36.739)		(16.748)	(15.409)	(14.223)	(15.545)	(15.440)
Constant	104.125***	31.090**	48.232***	16.520	17.079*	3.946	97.518***	23.090	58.366***	48.317***	38.774***	21.763
	(1.879)	(13.680)	(8.995)	(31.555)	(9.479)	(32.256)	(2.229)	(15.662)	(13.083)	(12.306)	(13.250)	(13.707)
Week FE	No	No	No	No	Yes	Yes	No	No	No	No	Yes	Yes
Customer ID FE	No	No	Yes	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes
B2B?	All	All	No	Yes	No	Yes	-	-	-	-	-	-
Marginal?	-	-	-	-	-	-	All	All	No	Yes	No	Yes
Observations	953,107	953,107	680,398	272,709	680,398	272,709	953,107	953,107	478,694	474,413	478,694	474,413
R-squared	0.02	0.68	0.87	0.88	0.88	0.88	0.01	0.68	0.90	0.86	0.90	0.86

Note: This table explores the relationship of basket value and delivery fee at the transaction level, breaking sample between business and households as well as marginal and inframarginal consumers. Robust standard errors in parentheses clustered at the customer level. *** p<0.01, ** p<0.05, * p<0.1.

Table 4.10: Number of Transactions on Delivery Fee

	(1)	(2)	(3)	(4)
Dep Var:	**No. Transactions**			
Delivery Fee	-0.254	-0.192	-0.306***	-0.349***
	(0.160)	(0.149)	(0.115)	(0.128)
No. Items		-0.018***	-0.002***	-0.002***
		(0.003)	(0.001)	(0.001)
Share Discounted Items		1.757**	0.252	0.152
		(0.752)	(0.183)	(0.180)
Percentage Discount		-3.860*	-0.413	-0.212
		(2.227)	(0.495)	(0.496)
Share Deep-Freeze Items		-24.236***	-19.295***	4.994
		(8.635)	(3.686)	(3.833)
Share Group 1 Items		-15.814*	-19.760***	4.397
		(8.139)	(3.664)	(3.814)
Share Group 2 Items		-14.120*	-19.685***	4.715
		(8.411)	(3.673)	(3.832)
Share Cool Items		-17.166**	-20.572***	3.419
		(8.098)	(3.668)	(3.816)
Share Inedible		-19.715**	-22.409***	2.883
		(9.021)	(3.732)	(3.863)
Constant	8.715***	25.993***	29.277***	5.039
	(1.198)	(8.098)	(3.793)	(3.947)
Week FE	No	No	No	Yes
Time Slot/Day/ City FE	No	No	Yes	Yes
Observations	139,056	139,056	139,056	139,056
R-squared	0.00	0.01	0.86	0.87

Note: This table presents OLS regressions of the number of transactions per time slot, week day and city on the delivery fee.
Robust standard errors in parentheses clustered at the time slot, day and city level.
*** p<0.01, ** p<0.05, * p<0.1

frequently than households (on aggregate terms) and the usual negative coefficient (although not statistically significant) after controlling for time slot and week fixed effects. Columns 5 to 12 separate the sample into business and household customers. The difference in purchasing behavior is clear as household customers purchase less often when delivery fees are higher and business customers seem to order more frequently in time slots with higher delivery fees. Even though this latter result is clearly the outcome of endogeneity, it is clear that household customers are more sensitive to delivery fee prices than business customers are.

Up to this moment, we have mainly exploited variation in delivery fees for different time slots. As we explained above, in our sample we only have one instance when delivery fees changed. This episode occurred in week 102 in December 2009 right before the online grocery sales enter an expected and seasonal decrease in sales. Therefore, it is not surprising that we observe a negative correlation between delivery fees and basket size. To investigate this episode further, in the next section we provide the result of a diff-in-diff estimator around this episode as robustness check.

4.5 Differences in Differences

We mainly observe one change in delivery fees during the sample period. This change occurred in week 102 (out of 105) during the month of December of 2009. The delivery fee increased in only 18 out of the 45 time slots allowing us to examine the impact of a change in delivery on the number of transactions as well as the average basket size taking as a control group those time slots that did not change delivery fee and observing how both groups changed before and after week 102. This strategy provides a cleaner test than the cross-sectional analysis above, but it does not come free of problems such as the problem of customers moving to other time slots where there has been no increase on fees. Moreover, due to the holiday season, sales in December are lower than those in November (and earlier months). We focus on weeks around the fee change from week 99 to week 105 (the last week in our sample) and divide each weekly realization by its 2008 weekly equivalent realizations such that

$$\frac{Var_{it}}{Var_{it-52}} = \alpha_0 + \alpha_1 After_t + \alpha_2 Treated_i + \alpha_3 After_t * Treated_i + \gamma_i + u_{it}$$

where the dependent variable is the ratio of dependent variables of interest with its realization a year before ($\frac{Var_{it}}{Var_{it-52}}$ in our analysis and Var are delivery fee, number of transactions and basket value), $After_t$ is a dummy variable that takes value 1 if week is 102 or higher,

Table 4.11: Number of Transactions on Delivery Fees: B2B Vs. B2C.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dep Var:	**No. Transactions**											
Delivery Fee	0.146	0.108	-0.106	-0.094	-0.271**	0.611***	-0.228*	0.523***	-0.873***	0.313***	-1.084***	0.411***
	(0.110)	(0.105)	(0.109)	(0.116)	(0.134)	(0.170)	(0.133)	(0.159)	(0.164)	(0.108)	(0.189)	(0.116)
B2B?	-2.329***	-2.430***	-2.871***	-2.871***								
	(0.248)	(0.263)	(0.301)	(0.303)								
No. Items		-0.006***	-0.002*	-0.002**			-0.013***	-0.006***	0.001***	-0.001**	0.001**	-0.001***
		(0.001)	(0.001)	(0.001)			(0.003)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Share Discounted Items		0.691*	-0.197	-0.408			-0.226	0.992***	-0.108	0.205*	-0.046	0.119
		(0.398)	(0.289)	(0.293)			(0.536)	(0.328)	(0.139)	(0.121)	(0.138)	(0.108)
Percentage Discount		-3.389***	0.008	0.686			-0.574	-12.094***	-0.232	-0.588*	0.351	-0.793***
		(1.037)	(0.732)	(0.839)			(1.303)	(1.529)	(0.372)	(0.346)	(0.363)	(0.304)
Share Deep-Freeze Items		-20.126***	-19.325***	-4.065			-16.617**	-20.677*	-13.139***	-8.239***	8.156**	-0.115
		(5.892)	(3.465)	(4.014)			(6.709)	(10.666)	(3.050)	(2.884)	(3.373)	(3.094)
Share Group 1 Items		-11.225*	-11.099***	4.057			-9.257	-9.477	-13.498***	-7.412**	7.873**	0.460
		(5.781)	(3.334)	(3.943)			(6.568)	(10.389)	(3.034)	(2.878)	(3.358)	(3.096)
Share Group 2 Items		-10.692*	-13.548***	1.670			-8.206	-9.954	-13.057***	-7.545***	8.336**	0.354
		(5.853)	(3.420)	(4.010)			(6.767)	(10.419)	(3.037)	(2.876)	(3.368)	(3.087)
Share Cool Items		-10.058*	-9.744***	5.361			-12.692*	-7.674	-13.881***	-7.822***	7.289**	-0.055
		(5.759)	(3.341)	(3.953)			(6.529)	(10.313)	(3.035)	(2.881)	(3.361)	(3.092)
Share Inedible		-14.504**	-14.893***	0.880			-16.729**	-10.915	-15.213***	-7.592***	7.105**	0.502
		(6.158)	(3.450)	(4.018)			(7.196)	(10.499)	(3.087)	(2.933)	(3.395)	(3.154)
Constant	4.715***	16.634***	17.829***	2.475	7.707***	-1.006	18.972***	9.305	25.575***	8.823***	4.767	0.930
	(0.850)	(5.724)	(3.400)	(4.005)	(1.032)	(1.124)	(6.501)	(10.321)	(3.304)	(3.016)	(3.567)	(3.385)
Time Slot, Day, City FE	No	No	Yes	Yes	No	No	No	No	Yes	Yes	Yes	Yes
Week FE	No	No	No	Yes	No	No	No	No	No	No	Yes	Yes
B2B?	All	All	All	All	No	Yes	No	Yes	No	Yes	No	Yes
Observations	196,941	196,941	196,941	196,941	118,028	78,913	118,028	78,913	118,028	78,913	118,028	78,913
R-squared	0.03	0.04	0.56	0.57	0.00	0.03	0.01	0.04	0.85	0.89	0.86	0.89

Note: This table presents OLS regressions of number of transactions on delivery fees breaking sample into B2B and B2C.
Robust standard errors in parentheses clustered at the time slot, day and city level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

$Treated_i$ is a dummy variable that takes value 1 if the time slot experiences an increased in delivery fee in week 102, and we include an interaction of these two variables as well as a time slot fixed effect.[15]

Table 4.12 shows that those slots increasing their delivery fees did so by 16 percentage points (columns 1 and 2). This increase in price was associated with a decrease of 8% in the number of transactions and an average basket value of 13%, after controlling for time slot, day and city. These results are somewhat puzzling in the sense that the decrease on the number of transactions is not statistically significant and the grocery sales have decreased. We are, however, aware of the fact that during the time period subject to "diff-in-diffs" analysis the sales are already slow, and also the fact that "diff-in-diffs" do not account for the migration of customers from one slot to another. Therefore, we investigate this phenomenon more in dept by separating household and business customer within our analysis.

Table 4.12: Difference in Difference After Delivery Fee Increase

	(1)	(2)	(3)	(4)	(5)	(6)
Dep Var:	**Ratio Delivery Fee**		**Ratio No Transactions**		**Ratio Basket Value**	
After Week 102?	-0.0021***		0.0217	0.0325	0.0680*	0.0822*
	(0.0010)		(0.0290)	(0.0350)	(0.0380)	(0.0460)
Increased Fee?	-0.0080***		0.0373		0.0410	
	(0.0010)		(0.0300)		(0.0400)	
After*Increased Fee?	0.1668***	0.1621***	-0.0685	-0.0817	-0.1126**	-0.1388**
	(0.0020)	(0.0020)	(0.0440)	(0.0520)	(0.0570)	(0.0680)
Constant	1.0093***	1.0061***	1.0167***	1.0286***	1.0852***	1.0990***
	(0.0010)	(0.0001)	(0.0180)	(0.0100)	(0.0230)	(0.0130)
Time Slot, Day, City FE	No	Yes	No	Yes	No	Yes
Observations	6,753	6,753	6,753	6,753	6,753	6,753
R-squared	0.75	0.98	0.00	0.34	0.00	0.33

Note: This table provides DiD estimates of the effect of an increase in delivery fee in certain time slots and no change in others. The fee change occurred in week 102 and consequently dependent and independent variables are ratios of weekly realizations between weeks 99 to 105 divided by realizations of weeks 47 to 53 respectively.
Robust standard errors in parentheses clustered by time slot, week day and city level.
*** p<0.01, ** p<0.05, * p<0.1.

Table 4.13 replicates the results in Table 4.12 breaking the sample into business customers and households. Note that slots more popular with business customers increased delivery fees an average of 17% against 15% for those of household customers. Household customers decreased their number of transactions by 10% which was associated with a

[15] Not showed here, we have also included ratios of basket value composition variables and those did not qualitatively change the results. We lose a lot of observations since anytime a variable takes value zero in the denominator the observation gets dropped.

decrease in basket value of around 10%. Note that the decrease amount on the grocery sales is a lot less than those of business customers, and not even statistically significant. To evaluate the profitability of the increase in delivery fees, we would also need to evaluate what percentage of these lower 10% basket value is net profit.

On the other end, the results for business customers are interesting as transactions went down around (statistically insignificant) 13% and decreased their basket value around 30% after controlling for time slot, day and city specific fixed effects. Based on these findings we can say that the firm would be losing a lot of money if the number of transactions did not statistically change and existing customers reduced the size of their basket purchases by 30% due to an increase in delivery fees of around 17%. In our data we are aware of the fact that household customers are more flexible in their time slots that they like to receive their groceries than the business customers. Even though a business customer order twice more frequently than a household (on average terms), the average number of different time slots she orders is the same with the other, and even less number of different days. These results verify our ideas of exploiting a third-degree price discrimination scheme, namely pricing primary good differently between households and business customers, in combination with a two-part tariff scheme. Because the loss of the marginal consumers from the business group has severe results on the grocery sales.

Next, we empirically explore the relation between net profits, delivery fees, number of transactions and basket value. We pursue a final empirical exercise that allows us to speak directly about the impact of delivery fee pricing on profits. In order to do so, we produce OLS regressions of total profit within a time slot, day, city and week on delivery fee, number of transactions and revenue controlling for average basket characteristics as well as time slot, day and city specific fixed effects and week fixed effects. In addition, we reproduce the difference-in-difference methodology in the previous section taking the ratio of profits as dependent variable.

Results in Table 4.14 show that once revenue and number of transactions are controlled for the delivery fee has no effect on profit from sales, even though the delivery fee is negatively associated with profits when we do not control for number of transactions or sales. Finally, Table 4.15 shows results of the "diff-in-diffs" estimation. This table shows that those time slots that increased prices saw a decrease in profits and that such decrease came mostly from business customers (in absolute size).

In the next section, we provide some managerial implications regarding our findings.

Table 4.13: Difference in Difference After Delivery Fee Increase: B2B Vs. B2C.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Dep Var:	Ratio Delivery Fee		Ratio No Transactions		Ratio Basket Value		Ratio Delivery Fee		Ratio No Transactions		Ratio Basket Value	
After Week 102?	-0.0008		0.2097***	0.2509***	0.3916***	0.4472***	-0.0024***		-0.0439	-0.0024	-0.018	0.0225
	(0.0010)		(0.0530)	(0.0660)	(0.0840)	(0.1140)	(0.0010)		(0.0290)	(0.0370)	(0.0390)	(0.0510)
Increased Fee?	-0.0021		0.1240***		0.1020*		-0.0096***		-0.0023		-0.0045	
	(0.0010)		(0.0450)		(0.0610)		(0.0010)		(0.0320)		(0.0400)	
After*Increased Fee?	0.1734***	0.1738***	-0.0933	-0.1378	-0.2505**	-0.3052**	0.1601***	0.1506***	-0.1331***	-0.1082*	-0.1246**	-0.1099
	(0.0020)	(0.0020)	(0.0730)	(0.0890)	(0.1050)	(0.1380)	(0.0010)	(0.0010)	(0.0460)	(0.0570)	(0.0630)	(0.0800)
Constant	1.0039***	1.0026***	0.8625***	0.9085***	0.9851***	1.0179***	1.0099***	1.0070***	0.9781***	0.9581***	1.0457***	1.0267***
	(0.0010)	(0.0001)	(0.0280)	(0.0170)	(0.0370)	(0.0260)	(0.0010)	(0.0001)	(0.0190)	(0.0110)	(0.0240)	(0.0150)
Time Slot, Day, City FE	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes
Sample	B2B	B2B	B2B	B2B	B2B	B2B	B2C	B2C	B2C	B2C	B2C	B2C
Observations	3,785	3,785	3,785	3,785	3,785	3,785	5,888	5,888	5,888	5,888	5,888	5,888
R-squared	0.88	0.977	0.009	0.435	0.01	0.386	0.704	0.988	0.005	0.349	0.002	0.313

Note: This table provides DiD estimates of the effect of an increase in delivery fee in certain time slots and no change in others. The fee change occurred in week 102 and consequently dependent and independent variables are ratios of weekly realizations between weeks 99 to 105 divided by realizations of weeks 47 to 53 respectively.
Robust standard errors in parentheses clustered by time slot, week day and city level. *** p<0.01, ** p<0.05, * p<0.1.

Table 4.14: Determinants of Profit

	(1)	(2)	(3)	(4)
Dep Var:	**ln(profit)**			
ln(Delivery Fee)	0.004	-0.410***		
	(0.012)	(0.086)		
ln(No. Transactions)	0.069***		1.046***	
	(0.005)		(0.002)	
ln(Sales)	0.938***			0.998***
	(0.005)			(0.001)
No. Items	0.0006***	0.0077***	0.0085***	0.0001**
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Share Discounted Items	-0.008	0.137**	0.114***	-0.015
	(0.015)	(0.054)	(0.027)	(0.015)
Percentage Discount	-0.347***	-0.543***	-0.649***	-0.327***
	(0.039)	(0.136)	(0.071)	(0.039)
Share Deep-Freeze Items	0.369*	-0.116	-0.268	0.410**
	(0.188)	(1.012)	(0.320)	(0.189)
Share Group 1 Items	0.174	-0.539	-0.502	0.217
	(0.185)	(1.004)	(0.317)	(0.185)
Share Group 2 Items	-0.138	-0.189	-0.140	-0.139
	(0.189)	(1.009)	(0.324)	(0.189)
Share Cool Items	0.634***	-0.559	-0.262	0.689***
	(0.185)	(1.005)	(0.318)	(0.185)
Share Inedible	-0.215	0.142	0.434	-0.258
	(0.203)	(1.028)	(0.349)	(0.203)
Constant	-1.167***	5.878***	3.431***	-1.439***
	(0.186)	(1.016)	(0.319)	(0.185)
Time Slot, Day, City FE	Yes	Yes	Yes	Yes
Week FE	Yes	Yes	Yes	Yes
Observations	139,054	139,054	139,054	139,054
R-squared	0.99	0.75	0.97	0.99

Note: This table presents OLS regressions of total weekly profit from grocery sales on the delivery fee, the number of transactions, and sales. All specifications contain week and city, time slot, and day fixed effects.
Robust standard errors in parentheses, clustered at the time slot, day and city level.
*** p<0.01, ** p<0.05, * p<0.1

Table 4.15: Difference in Difference After Delivery Fee Increase: Effect on Profits

	(1)	(2)	(3)	(4)	(5)	(6)
Dep Var:	**Ratio Profits**					
After Week 102?	0.0806*	0.0983*	0.4660***	0.5182***	0.0032	0.0544
	(0.043)	(0.055)	(0.121)	(0.167)	(0.044)	(0.060)
Increased Fee?	0.0588		0.0983		-0.0153	
	(0.044)		(0.072)		(0.040)	
After*Increased Fee?	-0.1237*	-0.1693**	-0.2648*	-0.3224	-0.1372**	-0.1288
	(0.063)	(0.079)	(0.147)	(0.198)	(0.067)	(0.086)
Constant	1.1145***	1.1365***	1.0772***	1.1101***	1.0600***	1.0341***
	(0.024)	(0.015)	(0.041)	(0.038)	(0.024)	(0.018)
Time Slot, Day, City FE	No	Yes	No	Yes	No	Yes
Sample	All	All	B2B	B2B	B2C	B2C
Observations	6,753	6,753	3,785	3,785	5,888	5,888
R-squared	0.001	0.32	0.01	0.34	0.002	0.31

Note: This table provides DiD estimates of the effect of an increase in delivery fee in certain time slots and no change in others. The fee change occurred in week 102 and consequently dependent and independent variables are ratios of weekly realizations between weeks 99 to 105 divided by realizations of weeks 47 to 53 respectively.
Robust standard errors in parentheses clustered by time slot, week day and city level.
*** p<0.01, ** p<0.05, * p<0.1.

4.6 Managerial Implications

Our paper empirically and theoretically explores revenue management of an online grocery store across different revenue sources, mainly delivery fees and grocery sales. This exercise has very direct and concrete managerial implications as we can compare the model predictions with the empirical results and the managerial decision of this online grocer at face value. We are not aware of the motivations behind each one of the firm's pricing decisions and strategies, so we cannot make any judgments regarding their managerial success. If anything, we can argue whether their strategies are consistent with profit maximizing behavior at large and suggest potential means for improvement.

Our findings are consistent with a story such that the online grocer faces demand from two very different type of customers. There is a clear difference in price sensitivity and behavior between household and business customers. On the one hand, households tend to increase their basket size at more expensive time slots. On the other hand, the loss of the marginal business consumers has a huge negative impact on the grocery sales and accordingly the profits of OG. These differences make it optimal for firms to seek ways to price discriminate and perhaps offer lower delivery fees to business customers while higher fees to households.

Finally, we use estimates from our difference in differences regressions to obtain the profit-maximizing fees of this online grocery store under a third-degree price discrimination scheme. Since we observe different price sensitivity for B2B and households, we estimate different optimal fees for these two types of customers while constraining all time slots to charge the same prices.[16] Based on our estimation on optimal delivery fees for B2B and household customers, we find that B2B should not be charged (free delivery) while the average fee paid by household customers should increase from 7.26 to 9.69 Euros. The reason behind this differential treatment is that even though B2B customers are less price sensitive, their basket sizes are much larger and therefore OG would be losing a larger amount of profits per order.[17]

[16] We obtain elasticities with the diff-in-diff estimation of Table 4.13 and 4.15 such that we recover the equations $\ln(VP) = \alpha' + \beta \ln(N)$ and $\ln(N) = \gamma' + \theta \ln(F)$ where VP is variable profit, N is the number of transactions and F is the delivery fee. Having said this, then the firm maximizes total profit for each type of customer separately (B2B versus households) such that $\Pi = VP(N) + N^{*}F$ subject to $N = \gamma + \theta F$.

[17] Interestingly enough, the uniform pricing policy over different time slots is consistent with the current practices of this company. Moreover, similar to our findings, OG provides discounts on delivery fees as the size of the order increases.

4.7 Conclusions

In this paper, we investigate the optimality of pricing of the online operations of a grocery retailer. In essence, this online retailer derives its revenues and profits from two different sources. The retailer must consider whether to sell groceries at discount and make up for all the profits with high shipping fees or offer cheap delivery and charge higher prices for groceries. The former or latter strategy will be optimal depending on how the demand for groceries is correlated with the demand for online ordering and home delivery.

After presenting the theoretical foundations of our two-part pricing scheme under a repeat buying setting, we conduct an empirical study where we test our theoretical predictions on the data. We estimate the correlation between basket size, number of transactions and delivery fees using detailed transaction information from an online grocery retailer in a Western European country. Our first set of empirical analysis verifies our theoretical findings and shows a positive correlation between number of transactions and basket sizes as well as a positive association between delivery fees and basket sizes, and a negative correlation between delivery fees and the number of transactions. Next to these results, our empirical investigation also shows that our data has two very different types of customers with different willingness to pay and sensitivity to delivery fees. Combining our theoretical predictions with the empirical results, our findings suggest that a pricing policy that will charge high margins for delivery services to households and free delivery for the business customers would be more profitable for our focal company. Therefore, our results suggest that online grocers should follow such pricing policies when observing the same correlations and heterogeneity structure in their sales data.

We believe the use of two-part tariff or other more complex non-linear pricing schemes in combination with third-degree price discrimination schemes will allow online grocers to extract more consumer surplus. Certain extension points characterize this research. First of all, capturing the marginal and inframarginal consumer behavior in the empirical analysis part in a more detailed way could provide more insights for companies. This would also help the empirical part of this study to better align with the theoretical part that characterizes the repeat buying setting in a more stylized manner. Secondly, the theoretical model could be extended in a way that accommodates heterogenous primary good prices. Although our focal company quit heterogenous prices on the delivery service, it is still an interesting future research direction. Here is another research area where future research in this field should concentrate its efforts and where managerial implications will benefit the most: Rather than using readily available data that characterize consumers, finding some third-degree price discrimination with the help of price menus such that consumers reveal their inherent type. A structural modeling approach is also very welcome in the context of deciding upon an optimal pricing strategy under a two-part pricing scheme with repeat purchase instances.

4.8 Appendix

Proof of Proposition 1:

The average basket size is given by $\frac{Q(p,x)}{N(p,x)}$. Taking the derivative with respect to the delivery fee x, we obtain $\frac{Q_x(x,p)N(p,x)-N_x(p,x)Q(p,x)}{N(p,x)^2}$. Since the denominator is always positive, we will focus on the numerator. By $m(\theta)$ uniform and $q_\theta(p,i,\theta) \leq jq_\theta(p,j,\theta)$ for $i<j$, we can show that $Q_x(x,p) \geq 0$ as follows

$$\begin{aligned} Q_x(x,p) &= \sum_{i=1}^{M} i\left[q(p,i,\theta^{i+1})\frac{i+1}{\int_p^\infty (i+1)q_\theta(t,i+1,\theta^{i+1})dt} - q(p,i,\theta^i)\frac{i}{\int_p^\infty iq_\theta(t,i,\theta^i)dt}\right] \\ &= \sum_{i=1}^{M} i\left[q(p,i,\theta^{i+1})\frac{1}{\int_p^\infty q_\theta(t,i+1,\theta^{i+1})dt} - q(p,i,\theta^i)\frac{1}{\int_p^\infty q_\theta(t,i,\theta^i)dt}\right] \end{aligned}$$

Since $q(p,i,\theta)$ is increasing in θ and $q_\theta(p,i,\theta^i)$ is decreasing in i, we obtain $\frac{q(p,i,\theta^{i+1})}{\int_p^\infty q_\theta(t,i+1,\theta)dt} \geq \frac{q(p,i,\theta^i)}{\int_p^\infty q_\theta(t,i,\theta)dt}$.

Proof of Proposition 2:

$$p-c = -\frac{Q(x,p)N_x(x,p)}{N_x(x,p)Q_p(x,p)-N_p(x,p)Q_x(x,p)}\frac{\sum_{i=1}^{\infty}\left(\frac{q(p,i,\theta_i)}{i}-\frac{Q(x,p)}{N(x,p)}\right)m(\theta_i)\theta_i^x}{(Q(x,p)/N(x,.p))N_x(x,p)}$$

The numerator of the first term together with the negative sign is non-negative. The denominator is also non-negative due to the fact the fact that $N_x \leq 0,\ Q_p \leq 0,\ N_p \leq 0,\ Q_x \geq 0$. Therefore, at the optimal signs of $(p-c)$ and the numerator of the second term are opposite of each other.

Proof of Proposition 3:

Recall that

$$\begin{aligned} \pi_x &= N(x,p) + (x-f)N_x(x,p) + (p-c)Q_x(x,p) = 0 \\ \pi_p &= Q(x,p) + (x-f)N_p(x,p) + (p-c)Q_p(x,p) = 0. \end{aligned}$$

Eliminating the term $(p-c)$ and solving both equalities, we obtain

$$x - f = \frac{N(x,p)Q_p(x,p) - Q(x,p)Q_x(x,p)}{N_p(x,p)Q_x(x,p) - N_x(x,p)Q_p(x,p)}.$$

Since both the numerator and the denominator are negative by $N_x \leq 0,\ Q_p \leq 0,\ N_p \leq 0,\ Q_x \geq 0$, the fraction is positive and the result follows.

Chapter 5

Summary and Conclusions

For over 50 years managers have been exhorted to "stay close to customers" to understand purchase behavior. Especially today's personalized marketing concepts such as direct marketing, one-to-one marketing and customer relationship management emphasize the critical importance of such an understanding for firms' success. However, understanding customers one by one is a difficult task not only because of the growing customer bases with millions of registered customers, but also because of unobservable aspects of customer behavior such as defection or price sensitivities.

In this dissertation, we use mathematical and econometric modeling to contribute to the scientific process of understanding and predicting customer behavior. To address the problem of making predictions on the individual level in large customer bases, we employ a hierarchical Bayesian approach and model customer heterogeneity. To understand unobservable customer behavior and sensitivities, we exploit ideas from probabilistic modeling and two-part pricing literature.

In Chapter 2, we extend the so called Buy-Till-You-Defect models to predict the timing of the purchases of every customer. Such detailed predictions help not only marketing managers but also operations managers in their decision-making processes. To our knowledge, we are the first to provide individual level purchase-timing predictions while taking into account also the unobserved defection behavior of customers. We provide analytical derivations on the expected timing of next purchases for each of the most established BTYD models. We also present a methodology to compute individual predictions among the four established BTYD models which differ in their estimation procedures.

A second contribution of Chapter 2 is a rigorous validation and comparison study of the BTYD models. Such a validation and comparison is needed in the field due to two main reasons. First, as BTYD models rely on different estimation methodologies, such as MCMC simulation or MLE based techniques, they do not directly provide predictions on all of the available metrics. This makes the model comparison difficult. Therefore, there is a lack of an extensive comparison study in the field. We deal with this problem by presenting a methodology that allows one to compute any individual-level metric for each of the models. To our knowledge, we are the first to bring all the following models together: Pareto/NBD, BG/NBD, HB, and PDO models. Second, the BTYD models are usually compared only on two metrics, namely the transaction frequency and the customer lifetime. However, as the latter metric is not observable, the only theoretically valid measure that is available to compare the BTYD models' predictive performance is the transaction frequency. Although the existing models are quite different in terms of their specification, they produce similar predictions on this measure. In other words, this measure is not sensitive to differences among the models. Our timing predictions using the BTYD models helps us to overcome this problem and provide more insights on the relative predictive performance of these models. We show that while the Pareto/NBD model and its Hierarchical Bayes [HB] extension perform the best in predicting transaction frequency, the PDO and HB models predict transaction timing more accurately. Furthermore, we find that differences in a model's predictive performance across datasets can be explained by the correlation between behavioral parameters and the proportion of customers without repeat purchases.

In Chapter 3, we show that managers can also obtain a customer segmentation by applying our proposed BTYD models. Effective segmentation that takes different dimensions of customer behavior into account is vital to understand customer heterogeneity. We show that customer segments obtained within a hierarchical mixture modeling framework also helps to improve individual level predictions. More specifically, we address the extreme lifetime prediction problem that limits the adoption of current BTYD models. We provide an explanation on why customers have extremely long lifetime predictions on certain datasets using these models. According to this, a uni-modal heterogeneity distribution that hides different segment structures in data also creates extreme lifetime predictions. In sum, the new BTYD models that we propose in this chapter not only provide customer segmentation that reveals unobserved characteristics of customer behavior such as their defection or price sensitivity, but also improve lifetime predictions on the individual level.

Both Chapters 2 and 3 contribute to the discussion on whether BTYD models would find their way into managerial practice by extending their output and improve their predictive performance. We acknowledge that BTYD models, compared to simple managerial heuristics, require more time and effort to be implemented in a business setting. Therefore, they should offer better results than managerial heuristics do in order to be commonly adopted by companies. We hope that the introduction of individual level timing predictions and the managerial guidelines on model choice presented in Chapter 2, together with improved customer lifetime predictions and an additional customer segmentation scheme presented in Chapter 3 would improve the diffusion of BTYD models.

A company's marketing actions such as promotion and pricing policies should be aligned with the heterogeneous responses and sensitivities in the customer base. In Chapter 4, we propose a discriminating two-part pricing policy based on the customer behavior and heterogeneity insights that we have developed in Chapter 3. Chapter 4 contributes to both theoretical and empirical foundations of the two-part pricing literature. We extend the two-part pricing theory by considering both customer heterogeneity and repeat-buying behavior at the same time. Moreover, we carry our theoretical predictions to its empirical implementation. This chapter, therefore, contributes to empirical validation of two-part pricing schemes by developing a test that can be applied to transaction data where customers repeatedly buy two complementary products. Our conclusion is that firms may increase their revenue and profit by implementing alternative and simpler pricing strategies that combine second and third degree price discrimination schemes.

References

M. Abe. Counting your customers one by one: A Hierarchical Bayes extension to the Pareto/NBD model. *Marketing Science*, 28(3):541–553, 2009a.

M. Abe. Customer lifetime value and RFM data: Accounting your customers: One by one. *Working paper. Available from: http://repository.dl.itc.u-tokyo.ac.jp/dspace/handle/2261/25752. Accessed: 01.01.2010.*, 2009b.

N. Agatz. *Demand Management in E-fulfillment.* PhD thesis, Erasmus University, 2009.

D. Baier and E. Stüber. Acceptance of recommendations to buy in online retailing. *Journal of Retailing and Consumer Services*, 17(3):173–180, 2010.

C. Barnhart, P. Belobaba, and A. R. Odoni. Applications of Operations Research in the air transport industry. *Transportation science*, 37(4):368–391, 2003.

E. Batislam, M. Denizel, and A. Filiztekin. Empirical validation and comparison of models for customer base analysis. *International Journal of Research in Marketing*, 24 (3):201–209, 2007.

P. Bell and J. Chen. Cutting costs or enhancing revenues? An example of a multi-product firm with impatient customers illustrates an important choice facing operational researchers. *Journal of the Operational Research Society*, 57(4):443–449, 2006.

P. P. Belobaba. Back to the future? Directions for Revenue Management. *Journal of Revenue and Pricing Management*, 1(1):87–89, 2002.

A. Bhatnagar and S. Ghose. A latent class segmentation analysis of e-shoppers. *Journal of Business Research*, 57(7):758–767, 2004.

J. Bueschken and S. Ma. When are your customers active and is their buying regular or random? An Erlang mixture state-switching model for customer scoring. *Working paper, available at SSRN 2006410. Accessed: 01.08.2012.*, 2012. URL `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2006410`.

Y. Chen, G. Zhang, D. Hu, and C. Fu. Customer segmentation based on survival character. *Journal of Intelligent Manufacturing*, 18(4):513–517, 2007.

P. K. Chintagunta, J. Chu, and J. Cebollada. Quantifying transaction costs in online/off-line grocery channel choice. *Marketing Science*, 31(1):96–114, 2012.

C. Cirillo and P. Hetrakul. Mixed logit and latent class model for railway Revenue Management. Technical report, Mid-Atlantic Universities Transportation Center, 2011.

B. Cooil, L. Aksoy, and T. Keiningham. Approaches to customer segmentation. *Journal of Relationship Marketing*, 6(3-4):9–39, 2008.

T. H. Davenport. Competing on Analytics. *Harvard Business Review*, 84(1):98, 2006.

P. Dellaportas and I. Papageorgiou. Multivariate mixtures of normals with unknown number of components. *Statistics and Computing*, 16(1):57–68, 2006.

earth.columbia.edu. First world happiness report launched at the United Nations. `http://www.earth.columbia.edu/articles/view/2960`, 2013. Accessed: 01.07.2014.

A. Ehrenberg. *Repeat-buying: Facts, theory and applications*. Griffin New London, 1988.

P. Fader, B. Hardie, and K. Lee. "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Science*, pages 275–284, 2005a.

P. Fader, B. Hardie, and K. Lee. RFM and CLV: Using iso-value curves for customer base analysis. *Journal of Marketing Research*, pages 415–430, 2005b.

M. Fisher. Strengthening the empirical base of Operations Management. *Manufacturing & Service Operations Management*, 9(4):368–382, 2007.

R. Fort. Inelastic sports pricing. *Managerial and Decision Economics*, 25(2):87–94, 2004.

S. Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models: Modeling and Applications to Random Processes*. Springer, 2006.

S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, (6):721–741, 1984.

J. Geweke et al. *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments*. Federal Reserve Bank of Minneapolis, Research Department, 1991.

R. Gil and W. R. Hartmann. Empirical analysis of metering price discrimination: Evidence from concession sales at movie theaters. *Marketing Science*, 28(6):1046–1062, 2009.

W. Gilks, S. Richardson, and D. Spiegelhalter. *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC, 1996.

F. Gönül and F. Hofstede. How to compute optimal catalog mailing decisions. *Marketing Science*, pages 65–74, 2006.

S. Gupta. Impact of sales promotions on when, what, and how much to buy. *Journal of Marketing Research*, pages 342–355, 1988.

S. Gupta, D. Hanssens, B. Hardie, W. Kahn, V. Kumar, N. Lin, N. Ravishanker, and S. Sriram. Modeling customer lifetime value. *Journal of Service Research*, 9(2):139–155, 2006.

W. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, pages 97–109, 1970.

M. Hurn, A. Justel, and C. P. Robert. Estimating mixtures of regressions. *Journal of Computational and Graphical Statistics*, 12(1):55–79, 2003.

H. Hwang, T. Jung, and E. Suh. An LTV model and customer segmentation based on customer value: A case study on the wireless telecommunication industry. *Expert systems with applications*, 26(2):181–188, 2004.

R. J. Hyndman. Computing and graphing highest density regions. *The American Statistician*, 50(2):pp. 120–126, 1996.

H. Ishwaran and L. F. James. Approximate Dirichlet process computing in finite normal mixtures. *Journal of Computational and Graphical Statistics*, 11(3), 2002.

R. Iyengar, K. Jedidi, S. Essegaier, and P. J. Danaher. The impact of tariff structure on customer retention, usage, and profitability of access services. *Marketing Science*, 30 (5):820–836, 2011.

K. Jerath, P. Fader, and B. Hardie. New perspectives on customer 'death' using a generalization of the Pareto/NBD model. *Marketing Science*, 30(5):866–880, 2011.

S. Y. Kim, T. S. Jung, E. H. Suh, and H. S. Hwang. Customer segmentation and strategy development based on customer lifetime value: A case study. *Expert Systems with Applications*, 31(1):101–107, 2006.

S. Klein and C. Loebbecke. Emerging pricing strategies on the web: Lessons from the airline industry. *Electronic Markets*, 13(1):46–58, 2003.

E. Korkmaz, R. Kuik, and D. Fok. "Counting your customers": When will they buy next? An empirical validation of probabilistic customer base analysis models based on purchase timing. *ERIM Report Series Research in Management*, 2013.

E. Korkmaz, D. Fok, and R. Kuik. The need for market segmentation in buy-till-you-defect models. *ERIM Report Series Reference No. ERS-2014-006-LIS*, 2014.

S. Landsburg. How much should hotel web access cost? *Slate (February 1)*, 2006. URL `http://www.slate.com/id/2135226/`.

K. C. Laudon and C. G. Traver. *E-commerce*. Pearson Prentice Hall, 2007.

S. LaValle, E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz. Big data, analytics and the path from insights to value. *MIT Sloan Management Review*, 21, 2014.

M. Lewis. The effect of shipping fees on customer acquisition, customer retention, and purchase quantities. *Journal of Retailing*, 82(1):13–23, 2006.

M. Lewis, V. Singh, and S. Fay. An empirical study of the impact of nonlinear shipping and handling fees on purchase incidence and expenditure decisions. *Marketing Science*, 25(1):51–64, 2006.

S. C. Littlechild. Two-part tariffs and consumption externalities. *The Bell Journal of Economics*, pages 661–670, 1975.

D. R. Marburger. Optimal ticket pricing for performance goods. *Managerial and Decision Economics*, 18(5):375–381, 1997.

J. D. McAuliffe, D. M. Blei, and M. I. Jordan. Nonparametric empirical Bayes for the Dirichlet process mixture model. *Statistics and Computing*, 16(1):5–14, 2006.

R. McCulloch and P. E. Rossi. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics*, 64(1):207–240, 1994.

E. J. Miravete and L.-H. Röller. Estimating price-cost markups under nonlinear pricing competition. *Journal of the European Economic Association*, 2(2-3):526–535, 2004.

R. Mohammed. *The art of pricing: How to find the hidden profits to grow your business*. Random House Digital, Inc., 2005.

V. Morwitz and D. Schmittlein. Using segmentation to improve sales forecasts based on purchase intent: Which "intenders" actually buy? *Journal of Marketing Research*, pages 391–405, 1992.

S. Mulpuru, C. Johnson, and D. Roberge. Us online retail forecast, 2012 to 2017. *Forrester Research (2013 February)*, 2013. URL `http://www.forrester.com/Forrester+Research+Online+Retail+Forecast+2012+To+2017+US/fulltext/-/E-RES90661`.

Y. K. Ng and M. Weisser. Optimal pricing with a budget constraint–the case of the two-part tariff. *The Review of Economic Studies*, 41(3):337–345, 1974.

A. Nobile and A. T. Fearnside. Bayesian finite mixtures with an unknown number of components: The allocation sampler. *Statistics and Computing*, 17(2):147–162, 2007.

W. Y. Oi. A Disneyland dilemma: Two-part tariffs for a Mickey Mouse monopoly. *The Quarterly Journal of Economics*, 85(1):77–96, 1971.

R. Paap and P. H. Franses. A dynamic multinomial probit model for brand choice with different long-run and short-run effects of marketing-mix variables. *Journal of Applied Econometrics*, 15(6):717–744, 2000.

O. R. Phillips and R. C. Battalio. Two-part tariffs and monopoly profits when visits are variable. *The Bell Journal of Economics*, pages 601–604, 1983.

C. E. Rasmussen. The infinite Gaussian mixture model. In *NIPS*, volume 12, pages 554–560, 1999.

W. Reinartz and V. Kumar. On the profitability of long-life customers in a noncontractual setting: An empirical investigation and implications for marketing. *The Journal of Marketing*, pages 17–35, 2000.

W. Reinartz and V. Kumar. The impact of customer relationship characteristics on profitable lifetime duration. *Journal of Marketing*, pages 77–99, 2003.

S. Richardson and P. J. Green. On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: series B (statistical methodology)*, 59(4):731–792, 1997.

S. Robinson. Computer scientists find unexpected depths in airfare search problem. *Journal for Society for Industrial and Applied Mathematics*, July/August, 2002.

S. Rosen and A. M. Rosenfield. Ticket pricing. *JL & Econ.*, 40:351, 1997.

P. Rossi, G. Allenby, and R. McCulloch. *Bayesian statistics and marketing*. John Wiley and Sons, Ltd, 2005.

R. Schmalensee. Monopolistic two-part pricing arrangements. *The Bell Journal of Economics*, pages 445–466, 1981.

D. Schmittlein and R. Peterson. Customer base analysis: An industrial purchase process application. *Marketing Science*, pages 41–67, 1994.

D. Schmittlein, D. Morrison, and R. Colombo. Counting your customers: Who are they and what will they do next? *Management Science*, pages 1–24, 1987.

D. Schmittlein, L. G. Cooper, and D. G. Morrison. Truth in concentration in the land of (80/20) laws. *Marketing Science*, 12(2):pp. 167–183, 1993.

E. Schwartz, E. Bradlow, and P. Fader. Model selection using database characteristics: Developing a classification tree for longitudinal incidence data. *Marketing Science*, 33 (2):188–205, 2014.

B. C. Smith, D. P. Günther, B. V. Rao, and R. M. Ratlife. E-commerce and Operations Research in airline planning, marketing, and distribution. *Interfaces*, 31(2):37–55, 2001.

M. Stephens. Bayesian analysis of mixture models with an unknown number of components-an alternative to reversible jump methods. *Annals of Statistics*, pages 40–74, 2000.

K. Talluri and G. Van Ryzin. *The theory and practice of Revenue Management*, volume 68. Springer Verlag, 2005.

N. Tereyağoğlu, P. Fader, and S. Veeraraghavan. Filling seats at a theater: Estimating the impact of posted prices and dynamic discounts. *Working paper. Accessed: 01.10.2012.*, 2012. URL `http://opim.wharton.upenn.edu/~senthilv/papers/Filling_seats.pdf`.

tuik.gov.tr. Turkey urban happiness index, 2013. `http://www.tuik.gov.tr/PreHaberBultenleri.do?id=18507`, 2013. Accessed: 14.04.2014.

R. Van Oest and G. Knox. Extending the BG/NBD: A simple model of purchases and complaints. *International Journal of Research in Marketing*, 28(1):30–37, 2011.

G. J. van Ryzin. Future of Revenue Management: Models of demand. *Journal of Revenue & Pricing Management*, 4(2):204–210, 2005.

C. Wu and H. Chen. Counting your customers: Compounding customer's in-store decisions, interpurchase time and repurchasing behavior. *European Journal of Operational Research*, 127(1):109–119, 2000.

M. Wübben and F. Wangenheim. Instant customer base analysis: Managerial heuristics often 'get it right'. *Journal of Marketing*, 72(3):82–93, 2008.

Y. Yang, S. Essegaier, and D. R. Bell. Free shipping and repeat buying on the internet: Theory and evidence. *Manuscript, Wharton School, University of Pennsylvania*, 2005.

J. Zhang and L. Krishnamurthi. Customizing promotions in online stores. *Marketing Science*, pages 561–578, 2004.

Y. Zhang, E. Bradlow, and D. Small. Capturing clumpiness when valuing customers: From RFM to RFMC. *Marketing Science, to appear*, 2014.

Nederlandse Samenvatting (Summary in Dutch)

Al meer dan 50 jaar worden managers aangespoord om dicht bij de klant te blijven om het aankoopgedrag van klanten te begrijpen. Vooral de huidige gepersonaliseerde marketing concepten zoals direct marketing, één-op-één marketing en customer relationship management benadrukken het cruciale belang van dergelijk begrip voor het succes van het bedrijf. Echter, één voor één klanten begrijpen is een moeilijke taak, niet alleen vanwege de groeiende klantenbestanden met miljoenen geregistreerde klanten, maar ook omwille van het niet-waarneembare gedrag van klanten, zoals defectie of hun prijs gevoeligheden.

In dit proefschrift gebruiken we wiskundige en econometrische modellen om bij te dragen aan het wetenschappelijke proces van het begrijpen van klanten. Om het probleem van het maken van voorspellingen op het individuele niveau in grote klantenbestanden aan te pakken, gebruiken we een hiërarchische Bayesiaanse benadering en modelleren we klant-heterogeniteit. Om het niet-waarneembare gedrag van klanten en hun gevoeligheden te begrijpen, putten we ideeën uit probabilistische modellering en tweedelige-prijsstellings literatuur.

In hoofdstuk 2, breiden we de zogeheten Buy-Till-You-Defect modellen uit om de timing van de aankopen van elke klant te voorspellen. Dergelijke gedetailleerde voorspellingen helpen niet alleen marketing managers maar ook operationele managers in hun besluitvorming. Voor zover wij weten, zijn wij de eerste om aankoop timing voorspellingen op individueel niveau aan te bieden, terzelvertijd rekening houdend met niet geobserveerde klant-defectie. Wij leveren analytische afleidingen met betrekking tot het verwachte tijdstip van volgende aankopen voor elk van de gevestigde BTYD modellen. We presenteren ook een methodologie om individuele voorspellingen te berekenen voor de vier gevestigde BTYD modellen die verschillen in hun schattingsprocedure.

Een tweede bijdrage van hoofdstuk 2 is een gedegen validatie en vergelijkende studie van de BTYD modellen. Een dergelijke validatie en vergelijking is nodig in het vakgebied omwille van twee belangrijke redenen. Ten eerste, omdat BTYD modellen die zich baseren op verschillende schattingsmethodieken, zoals MCMC simulatie of MLE gebaseerde technieken, niet rechtstreeks voorspellingen verstrekken over alle beschikbare metrieken. Dit maakt modellen vergelijken lastig. Daarom is er een gebrek aan een uitgebreide vergelijkingstudie in het vakgebied. We pakken dit probleem aan door een methodologie te presenteren die toelaat om eender welk individu-niveau metriek te berekenen voor elk van de modellen. Voor zover wij weten, zijn wij de eerste die alle volgende modellen samenbrengen: Pareto/NBD, BG/NBD, HB, en PDO modellen. Ten tweede, de BTYD modellen worden meestal alleen vergeleken op twee metrieken, namelijk de transactie frequentie en de customer lifetime. Aangezien deze laatste metriek niet waarneembaar is, is de enige geldige theoretische metriek die beschikbaar is om voorspellende prestaties van de BTYD modellen te vergelijken, de transactie frequentie. Hoewel de bestaande modellen vrij verschillend zijn in hun specificaties, produceren zij soortgelijke voorspellingen op deze maatstaf. Met andere woorden, deze maatstaf is niet gevoelig voor de verschillen tussen de modellen. Onze timing voorspellingen met behulp van de BTYD modellen stellen ons in staat dit probleem te verhelpen en zorgen voor meer inzicht in de relatieve voorspellende prestaties van deze modellen.

In hoofdstuk 3, laten we zien dat managers ook een klantsegmentatie kunnen verkrijgen door het toepassen van BTYD modellen. Effectieve segmentatie die rekening houdt met verschillende dimensies in het gedrag van klanten, is van vitaal belang om de klantheterogeniteit te begrijpen. We laten zien dat klantsegmenten, verkregen binnen een hirarchisch mix model raamwerk, ook helpen om voorspellingen te verbeteren op het individuele niveau. Kortom, het model dat wij voorstellen in dit hoofdstuk, levert niet alleen klantsegmentatie die niet waargenomen kenmerken van het gedrag van klanten zoals hun defectie of prijsgevoeligheid blootlegt, maar verbetert ook voorspellingen op het individuele niveau.

Een bedrijf z'n marketing acties zoals promotie en prijsbeleid moeten worden afgestemd op de heterogene reacties en gevoeligheden in het klantenbestand. In hoofdstuk 4, stellen we een discriminerend tweedelige prijsbeleid voor op basis van het klantengedrag en de heterogeniteits inzichten die we hebben ontwikkeld in het vorige hoofdstuk. Hoofdstuk 4 draagt bij aan zowel theoretische als empirische grondslagen van de tweedelige prijsstelling literatuur. We breiden de tweedelige prijsstelling theorie uit door tegelijkertijd te kijken naar zowel de klant heterogeniteit als het herhaal-koopgedrag. Bovendien brengen we onze theoretische voorspellingen naar hun empirische implementatie. Onze conclusie is dat bedrijven hun inkomsten en winsten zouden kunnen verhogen door het implementeren van alternatieve en eenvoudigere prijsstellings-strategieën die tweede- en derde-graads prijsdiscriminatie schema's combineren.

About the author

Evşen Korkmaz received a BSc in Industrial Engineering (2005) and an MSc in Industrial Engineering with a specialization in Operations Research (2008) from Istanbul Technical University with distinction. In 2009, she started her PhD research at the Rotterdam School of Management, Erasmus University Rotterdam. Her main research interest lies on the interface of Marketing Modeling and Operations Research. Her research has been presented at several international conferences such as INFORMS Annual Meeting, Production and Operations Management Conference and ISMS Marketing Science Conferences.

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT (ERIM)

ERIM PH.D. SERIES RESEARCH IN MANAGEMENT

The ERIM PhD Series contains PhD dissertations in the field of Research in Management defended at Erasmus University Rotterdam and supervised by senior researchers affiliated to the Erasmus Research Institute of Management (ERIM). All dissertations in the ERIM PhD Series are available in full text through the ERIM Electronic Series Portal: http://hdl.handle.net/1765/1
ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics at the Erasmus University Rotterdam (EUR).

DISSERTATIONS LAST FIVE YEARS

Acciaro, M., *Bundling Strategies in Global Supply Chains*, Promotor(s): Prof.dr. H.E. Haralambides, EPS-2010-197-LIS, http://hdl.handle.net/1765/19742

Akpinar, E., *Consumer Information Sharing; Understanding Psychological Drivers of Social Transmission*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2013-297-MKT, http://hdl.handle.net/1765/50140

Alexiev, A., *Exploratory Innovation: The Role of Organizational and Top Management Team Social Capital*, Promotor(s): Prof.dr. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-208-STR, http://hdl.handle.net/1765/20632

Akin Ates, M., *Purchasing and Supply Management at the Purchase Category Level: Strategy, Structure, and Performance*, Promotor: Prof.dr. J.Y.F. Wynstra, EPS-2014-300-LIS, http://hdl.handle.net/1765/1

Almeida, R.J.de, *Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty,* Promotor Prof.dr.ir. Uzay Kaymak, EPS-2014-310-LIS, http://hdl.net/1765/1

Bannouh, K., *Measuring and Forecasting Financial Market Volatility using High-Frequency Data*, Promotor: Prof.dr.D.J.C. van Dijk, EPS-2013-273-F&A, http://hdl.handle.net/1765/38240

Benning, T.M., *A Consumer Perspective on Flexibility in Health Care: Priority Access Pricing and Customized Care*, Promotor: Prof.dr.ir. B.G.C. Dellaert, EPS-2011-241-MKT, http://hdl.handle.net/1765/23670

Ben-Menahem, S.M., *Strategic Timing and Proactiveness of Organizations*, Promotor(s): Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch, EPS-2013-278-S&E, http://hdl.handle.net/1765/ 39128

Berg, W.E. van den, *Understanding Salesforce Behavior Using Genetic Association Studies,* Promotor: Prof.dr. W.J.M.I. Verbeke, EPS-2014-311-MKT, http://hdl.handle.net/1765/ 51440

Betancourt, N.E., *Typical Atypicality: Formal and Informal Institutional Conformity, Deviance, and Dynamics*, Promotor: Prof.dr. B. Krug, EPS-2012-262-ORG, http://hdl.handle.net/1765/32345

Binken, J.L.G., *System Markets: Indirect Network Effects in Action, or Inaction*, Promotor: Prof.dr. S. Stremersch, EPS-2010-213-MKT, http://hdl.handle.net/1765/21186

Blitz, D.C., *Benchmarking Benchmarks*, Promotor(s): Prof.dr. A.G.Z. Kemna & Prof.dr. W.F.C. Verschoor, EPS-2011-225-F&A, http://hdl.handle.net/1765/226244

Boons, M., *Working Together Alone in the Online Crowd: The Effects of Social Motivations and Individual Knowledge Backgrounds on the Participation and Performance of Members of Online Crowdsourcing Platforms*, Promotor: Prof.dr. H.G. Barkema, EPS-2014-306-S&E, http://hdl.net/1765/50711

Borst, W.A.M., *Understanding Crowdsourcing: Effects of Motivation and Rewards on Participation and Performance in Voluntary Online Activities*, Promotor(s): Prof.dr.ir. J.C.M. van den Ende & Prof.dr.ir. H.W.G.M. van Heck, EPS-2010-221-LIS, http://hdl.handle.net/1765/ 21914

Budiono, D.P., *The Analysis of Mutual Fund Performance: Evidence from U.S. Equity Mutual Funds*, Promotor: Prof.dr. M.J.C.M. Verbeek, EPS-2010-185-F&A, http://hdl.handle.net/1765/18126

Burger, M.J., *Structure and Cooptition in Urban Networks*, Promotor(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.R. Commandeur, EPS-2011-243-ORG, http://hdl.handle.net/1765/26178

Byington, E., *Exploring Coworker Relationships: Antecedents and Dimensions of Interpersonal Fit, Coworker Satisfaction, and Relational Models*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2013-292-ORG, http://hdl.handle.net/1765/41508

Camacho, N.M., *Health and Marketing; Essays on Physician and Patient Decision-making*, Promotor: Prof.dr. S. Stremersch, EPS-2011-237-MKT, http://hdl.handle.net/1765/23604

Cankurtaran, P. *Essays On Accelerated Product Development,* Promotor: Prof.dr.ir. G.H. van Bruggen, EPS-2014-317-MKT, http://hdl.handle.net/1765/1

Caron, E.A.M., *Explanation of Exceptional Values in Multi-dimensional Business Databases*, Promotor(s): Prof.dr.ir. H.A.M. Daniels & Prof.dr. G.W.J. Hendrikse, EPS-2013-296-LIS, http://hdl.handle.net/1765/50005

Carvalho, L., *Knowledge Locations in Cities; Emergence and Development Dynamics*, Promotor: Prof.dr. L. van den Berg, EPS-2013-274-S&E, http://hdl.handle.net/1765/ 38449

Carvalho de Mesquita Ferreira, L., *Attention Mosaics: Studies of Organizational Attention*, Promotor(s): Prof.dr. P.M.A.R. Heugens & Prof.dr. J. van Oosterhout, EPS-2010-205-ORG, http://hdl.handle.net/1765/19882

Cox, R.H.G.M., *To Own, To Finance, and to Insure; Residential Real Estate Revealed*, Promotor: Prof.dr. D. Brounen, EPS-2013-290-F&A, http://hdl.handle.net/1765/40964

Defilippi Angeldonis, E.F., *Access Regulation for Naturally Monopolistic Port Terminals: Lessons from Regulated Network Industries*, Promotor: Prof.dr. H.E. Haralambides, EPS-2010-204-LIS, http://hdl.handle.net/1765/19881

Deichmann, D., *Idea Management: Perspectives from Leadership, Learning, and Network Theory*, Promotor: Prof.dr.ir. J.C.M. van den Ende, EPS-2012-255-ORG, http://hdl.handle.net/1765/ 31174

Desmet, P.T.M., *In Money we Trust? Trust Repair and the Psychology of Financial Compensations*, Promotor: Prof.dr. D. De Cremer & Prof.dr. E. van Dijk, EPS-2011-232-ORG, http://hdl.handle.net/1765/23268

Dietvorst, R.C., *Neural Mechanisms Underlying Social Intelligence and Their Relationship with the Performance of Sales Managers*, Promotor: Prof.dr. W.J.M.I. Verbeke, EPS-2010-215-MKT, http://hdl.handle.net/1765/21188

Dollevoet, T.A.B., *Delay Management and Dispatching in Railways*, Promotor: Prof.dr. A.P.M. Wagelmans, EPS-2013-272-LIS, http://hdl.handle.net/1765/38241

Doorn, S. van, *Managing Entrepreneurial Orientation*, Promotor(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2012-258-STR, http://hdl.handle.net/1765/32166

Douwens-Zonneveld, M.G., *Animal Spirits and Extreme Confidence: No Guts, No Glory*, Promotor: Prof.dr. W.F.C. Verschoor, EPS-2012-257-F&A, http://hdl.handle.net/1765/31914

Duca, E., *The Impact of Investor Demand on Security Offerings*, Promotor: Prof.dr. A. de Jong, EPS-2011-240-F&A, http://hdl.handle.net/1765/26041

Duursema, H., *Strategic Leadership; Moving Beyond the Leader-follower Dyad*, Promotor: Prof.dr. R.J.M. van Tulder, EPS-2013-279-ORG, http://hdl.handle.net/1765/ 39129

Eck, N.J. van, *Methodological Advances in Bibliometric Mapping of Science*, Promotor: Prof.dr.ir. R. Dekker, EPS-2011-247-LIS, http://hdl.handle.net/1765/26509

Essen, M. van, *An Institution-Based View of Ownership*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. G.M.H. Mertens, EPS-2011-226-ORG, http://hdl.handle.net/1765/22643

Feng, L., *Motivation, Coordination and Cognition in Cooperatives*, Promotor: Prof.dr. G.W.J. Hendrikse, EPS-2010-220-ORG, http://hdl.handle.net/1765/21680

Fourné, S. P. L. Managing Organizational Tensions: A Multi-level Perspective on Exploration, Exploitation, and Ambidexterity, Promotor(s): Prof.dr. J.J.P. Jansen, Prof.dr. S.J. Magala & dr. T.J.M.Mom, EPS-2014-318-S&E, http://hdl.handle.net/1765/21680

Gharehgozli, A.H., *Developing New Methods for Efficient Container Stacking Operations*, Promotor: Prof.dr.ir. M.B.M. de Koster, EPS-2012-269-LIS, http://hdl.handle.net/1765/ 37779

Gils, S. van, *Morality in Interactions: On the Display of Moral Behavior by Leaders and Employees*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2012-270-ORG, http://hdl.handle.net/1765/ 38028

Ginkel-Bieshaar, M.N.G. van, *The Impact of Abstract versus Concrete Product Communications on Consumer Decision-making Processes*, Promotor: Prof.dr.ir. B.G.C. Dellaert, EPS-2012-256-MKT, http://hdl.handle.net/1765/31913

Gkougkousi, X., *Empirical Studies in Financial Accounting*, Promotor(s): Prof.dr. G.M.H. Mertens & Prof.dr. E. Peek, EPS-2012-264-F&A, http://hdl.handle.net/1765/37170

Hakimi, N.A, *Leader Empowering Behaviour: The Leader's Perspective: Understanding the Motivation behind Leader Empowering Behaviour*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2010-184-ORG, http://hdl.handle.net/1765/17701

Hensmans, M., *A Republican Settlement Theory of the Firm: Applied to Retail Banks in England and the Netherlands (1830-2007)*, Promotor(s): Prof.dr. A. Jolink & Prof.dr. S.J. Magala, EPS-2010-193-ORG, http://hdl.handle.net/1765/19494

Hernandez Mireles, C., *Marketing Modeling for New Products*, Promotor: Prof.dr. P.H. Franses, EPS-2010-202-MKT, http://hdl.handle.net/1765/19878

Heyde Fernandes, D. von der, *The Functions and Dysfunctions of Reminders*, Promotor: Prof.dr. S.M.J. van Osselaer, EPS-2013-295-MKT, http://hdl.handle.net/1765/41514

Heyden, M.L.M., *Essays on Upper Echelons & Strategic Renewal: A Multilevel Contingency Approach*, Promotor(s): Prof.dr. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2012-259-STR, http://hdl.handle.net/1765/32167

Hoever, I.J., *Diversity and Creativity: In Search of Synergy*, Promotor(s): Prof.dr. D.L. van Knippenberg, EPS-2012-267-ORG, http://hdl.handle.net/1765/37392

Hoogendoorn, B., *Social Entrepreneurship in the Modern Economy: Warm Glow, Cold Feet*, Promotor(s): Prof.dr. H.P.G. Pennings & Prof.dr. A.R. Thurik, EPS-2011-246-STR, http://hdl.handle.net/1765/26447

Hoogervorst, N., *On The Psychology of Displaying Ethical Leadership: A Behavioral Ethics Approach*, Promotor(s): Prof.dr. D. De Cremer & Dr. M. van Dijke, EPS-2011-244-ORG, http://hdl.handle.net/1765/26228

Huang, X., *An Analysis of Occupational Pension Provision: From Evaluation to Redesign*, Promotor(s): Prof.dr. M.J.C.M. Verbeek & Prof.dr. R.J. Mahieu, EPS-2010-196-F&A, http://hdl.handle.net/1765/19674

Hytönen, K.A. *Context Effects in Valuation, Judgment and Choice*, Promotor(s): Prof.dr.ir. A. Smidts, EPS-2011-252-MKT, http://hdl.handle.net/1765/30668

Jaarsveld, W.L. van, *Maintenance Centered Service Parts Inventory Control*, Promotor(s): Prof.dr.ir. R. Dekker, EPS-2013-288-LIS, http://hdl.handle.net/1765/ 39933

Jalil, M.N., *Customer Information Driven After Sales Service Management: Lessons from Spare Parts Logistics*, Promotor(s): Prof.dr. L.G. Kroon, EPS-2011-222-LIS, http://hdl.handle.net/1765/22156

Kagie, M., *Advances in Online Shopping Interfaces: Product Catalog Maps and Recommender Systems*, Promotor(s): Prof.dr. P.J.F. Groenen, EPS-2010-195-MKT, http://hdl.handle.net/1765/19532

Kappe, E.R., *The Effectiveness of Pharmaceutical Marketing*, Promotor(s): Prof.dr. S. Stremersch, EPS-2011-239-MKT, http://hdl.handle.net/1765/23610

Karreman, B., *Financial Services and Emerging* Markets, Promotor(s): Prof.dr. G.A. van der Knaap & Prof.dr. H.P.G. Pennings, EPS-2011-223-ORG, http://hdl.handle.net/1765/ 22280

Kil, J.C.M., *Acquisitions Through a Behavioral and Real Options Lens*, Promotor(s): Prof.dr. H.T.J. Smit, EPS-2013-298-F&A, http://hdl.handle.net/1765/50142

Klooster, E. van't, *Travel to Learn: The Influence of Cultural Distance on Competence Development in Educational Travel,* Promotors: Prof.dr. F.M. Go & Prof.dr. P.J. van Baalen, EPS-2014-312-MKT, http://hdl.handle.net/1765/151460

Koendjbiharie, S.R., *The Information-Based View on Business Network Performance Revealing the Performance of Interorganizational Networks,* Promotors: Prof.dr.ir. H.W.G.M. van Heck & Prof.mr.dr. P.H.M. Vervest, EPS-2014-315-LIS, http://hdl.handle.net/1765/1

Konter, D.J., *Crossing borders with HRM: An inquiry of the influence of contextual differences in the adaption and effectiveness of HRM*, Promotor: Prof.dr. J. Paauwe, EPS-2014-305-ORG, http://hdl.handle.net/1765/1

Korkmaz, E. *Understanding Heterogeneity in Hidden Drivers of Customer Purchase Behavior,* Promotors: Prof.dr. S.L. van de Velde & dr. R.Kuik, EPS-2014-316-LIS, http://hdl.handle.net/1765/1

Lam, K.Y., *Reliability and Rankings*, Promotor(s): Prof.dr. P.H.B.F. Franses, EPS-2011-230-MKT, http://hdl.handle.net/1765/22977

Lander, M.W., *Profits or Professionalism? On Designing Professional Service Firms*, Promotor(s): Prof.dr. J. van Oosterhout & Prof.dr. P.P.M.A.R. Heugens, EPS-2012-253-ORG, http://hdl.handle.net/1765/30682

Langhe, B. de, *Contingencies: Learning Numerical and Emotional Associations in an Uncertain World*, Promotor(s): Prof.dr.ir. B. Wierenga & Prof.dr. S.M.J. van Osselaer, EPS-2011-236-MKT, http://hdl.handle.net/1765/23504

Larco Martinelli, J.A., *Incorporating Worker-Specific Factors in Operations Management Models*, Promotor(s): Prof.dr.ir. J. Dul & Prof.dr. M.B.M. de Koster, EPS-2010-217-LIS, http://hdl.handle.net/1765/21527

Leunissen, J.M., *All Apologies: On the Willingness of Perpetrators to Apoligize*, Promotor: Prof.dr. D. De Cremer, EPS-2014-301-ORG, http://hdl.handle.net/1765/1

Liang, Q., *Governance, CEO Indentity, and Quality Provision of Farmer Cooperatives*, Promotor: Prof.dr. G.W.J. Hendrikse, EPS-2013-281-ORG, http://hdl.handle.net/1765/1

Liket, K.C., *Why 'Doing Good' is not Good Enough: Essays on Social Impact Measurement*, Promotor: Prof.dr. H.R. Commandeur, EPS-2014-307-S&E, http://hdl.handle.net/1765/51130

Loos, M.J.H.M. van der, *Molecular Genetics and Hormones; New Frontiers in Entrepreneurship Research*, Promotor(s): Prof.dr. A.R. Thurik, Prof.dr. P.J.F. Groenen & Prof.dr. A. Hofman, EPS-2013-287-S&E, http://hdl.handle.net/1765/ 40081

Lovric, M., *Behavioral Finance and Agent-Based Artificial Markets*, Promotor(s): Prof.dr. J. Spronk & Prof.dr.ir. U. Kaymak, EPS-2011-229-F&A, http://hdl.handle.net/1765/ 22814

Lu, Y., *Data-Driven Decision Making in Auction Markets, Promotors:* Prof.dr.ir.H.W.G.M. van Heck & Prof.dr.W.Ketter, EPS-2014-314-LIS, http://hdl.handle.net/1765/1

Markwat, T.D., *Extreme Dependence in Asset Markets Around the Globe*, Promotor: Prof.dr. D.J.C. van Dijk, EPS-2011-227-F&A, http://hdl.handle.net/1765/22744

Mees, H., *Changing Fortunes: How China's Boom Caused the Financial Crisis*, Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2012-266-MKT, http://hdl.handle.net/1765/34930

Meuer, J., *Configurations of Inter-Firm Relations in Management Innovation: A Study in China's Biopharmaceutical Industry*, Promotor: Prof.dr. B. Krug, EPS-2011-228-ORG, http://hdl.handle.net/1765/22745

Mihalache, O.R., *Stimulating Firm Innovativeness: Probing the Interrelations between Managerial and Organizational Determinants*, Promotor(s): Prof.dr. J.J.P. Jansen, Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2012-260-S&E, http://hdl.handle.net/1765/32343

Milea, V., *New Analytics for Financial Decision Support*, Promotor: Prof.dr.ir. U. Kaymak, EPS-2013-275-LIS, http://hdl.handle.net/1765/ 38673

Naumovska, I. *Socially Situated Financial Markets:a Neo-Behavioral Perspective on Firms, Investors and Practices,* Promoter(s) Prof.dr. P.P.M.A.R. Heugens & Prof.dr. A.de Jong, EPS-2014-319-S&E, http://hdl.handle.net/1765/ 1

Nielsen, L.K., *Rolling Stock Rescheduling in Passenger Railways: Applications in Short-term Planning and in Disruption Management*, Promotor: Prof.dr. L.G. Kroon, EPS-2011-224-LIS, http://hdl.handle.net/1765/22444

Nijdam, M.H., *Leader Firms: The Value of Companies for the Competitiveness of the Rotterdam Seaport Cluster*, Promotor(s): Prof.dr. R.J.M. van Tulder, EPS-2010-216-ORG, http://hdl.handle.net/1765/21405

Noordegraaf-Eelens, L.H.J., *Contested Communication: A Critical Analysis of Central Bank Speech*, Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2010-209-MKT, http://hdl.handle.net/1765/21061

Nuijten, A.L.P., *Deaf Effect for Risk Warnings: A Causal Examination applied to Information Systems Projects*, Promotor: Prof.dr. G. van der Pijl & Prof.dr. H. Commandeur & Prof.dr. M. Keil, EPS-2012-263-S&E, http://hdl.handle.net/1765/34928

Oosterhout, M., van, *Business Agility and Information Technology in Service Organizations*, Promotor: Prof,dr.ir. H.W.G.M. van Heck, EPS-2010-198-LIS, http://hdl.handle.net/1765/19805

Osadchiy, S.E., *The Dynamics of Formal Organization: Essays on Bureaucracy and Formal Rules*, Promotor: Prof.dr. P.P.M.A.R. Heugens, EPS-2011-231-ORG, http://hdl.handle.net/1765/23250

Otgaar, A.H.J., *Industrial Tourism: Where the Public Meets the Private*, Promotor: Prof.dr. L. van den Berg, EPS-2010-219-ORG, http://hdl.handle.net/1765/21585

Ozdemir, M.N., *Project-level Governance, Monetary Incentives and Performance in Strategic R&D Alliances*, Promotor: Prof.dr.ir. J.C.M. van den Ende, EPS-2011-235-LIS, http://hdl.handle.net/1765/23550

Peers, Y., *Econometric Advances in Diffusion Models*, Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2011-251-MKT, http://hdl.handle.net/1765/ 30586

Pince, C., *Advances in Inventory Management: Dynamic Models*, Promotor: Prof.dr.ir. R. Dekker, EPS-2010-199-LIS, http://hdl.handle.net/1765/19867

Porck, J.P., *No Team is an Island*, Promotor: Prof.dr. P.J.F. Groenen & Prof.dr. D.L. van Knippenberg, EPS-2013-299-ORG, http://hdl.handle.net/1765/50141

Porras Prado, M., *The Long and Short Side of Real Estate, Real Estate Stocks, and Equity*, Promotor: Prof.dr. M.J.C.M. Verbeek, EPS-2012-254-F&A, http://hdl.handle.net/1765/30848

Potthoff, D., *Railway Crew Rescheduling: Novel Approaches and Extensions*, Promotor(s): Prof.dr. A.P.M. Wagelmans & Prof.dr. L.G. Kroon, EPS-2010-210-LIS, http://hdl.handle.net/1765/21084

Poruthiyil, P.V., *Steering Through: How Organizations Negotiate Permanent Uncertainty and Unresolvable Choices*, Promotor(s): Prof.dr. P.P.M.A.R. Heugens & Prof.dr. S. Magala, EPS-2011-245-ORG, http://hdl.handle.net/1765/26392

Pourakbar, M. *End-of-Life Inventory Decisions of Service Parts*, Promotor: Prof.dr.ir. R. Dekker, EPS-2011-249-LIS, http://hdl.handle.net/1765/30584

Pronker, E.S., *Innovation Paradox in Vaccine Target Selection*, Promotor(s): Prof.dr. H.R. Commandeur & Prof.dr. H.J.H.M. Claassen, EPS-2013-282-S&E, http://hdl.handle.net/1765/39654

Retel Helmrich, M.J., *Green Lot-Sizing*, Promotor: Prof.dr. A.P.M. Wagelmans, EPS-2013-291-LIS, http://hdl.handle.net/1765/41330

Rietveld, C.A., *Essays on the Intersection of Economics and Biology*, Promotor(s): Prof.dr. P.J.F. Groenen, Prof.dr. A. Hofman & Prof.dr. A.R. Thurik, EPS-2014-320-S&E, http://hdl.handle.net/1765/1

Rijsenbilt, J.A., *CEO Narcissism; Measurement and Impact*, Promotor: Prof.dr. A.G.Z. Kemna & Prof.dr. H.R. Commandeur, EPS-2011-238-STR, http://hdl.handle.net/1765/ 23554

Roelofsen, E.M., *The Role of Analyst Conference Calls in Capital Markets*, Promotor(s): Prof.dr. G.M.H. Mertens & Prof.dr. L.G. van der Tas RA, EPS-2010-190-F&A, http://hdl.handle.net/1765/18013

Roza, M.W., *The Relationship between Offshoring Strategies and Firm Performance: Impact of Innovation, Absorptive Capacity and Firm Size*, Promotor(s): Prof.dr. H.W. Volberda & Prof.dr.ing. F.A.J. van den Bosch, EPS-2011-214-STR, http://hdl.handle.net/1765/22155

Rubbaniy, G., *Investment Behavior of Institutional Investors*, Promotor: Prof.dr. W.F.C. Verschoor, EPS-2013-284-F&A, http://hdl.handle.net/1765/ 40068

Schellekens, G.A.C., *Language Abstraction in Word of Mouth*, Promotor: Prof.dr.ir. A. Smidts, EPS-2010-218-MKT, http://hdl.handle.net/1765/21580

Shahzad, K., *Credit Rating Agencies, Financial Regulations and the Capital Markets*, Promotor: Prof.dr. G.M.H. Mertens, EPS-2013-283-F&A, http://hdl.handle.net/1765/39655

Sotgiu, F., *Not All Promotions are Made Equal: From the Effects of a Price War to Cross-chain Cannibalization*, Promotor(s): Prof.dr. M.G. Dekimpe & Prof.dr.ir. B. Wierenga, EPS-2010-203-MKT, http://hdl.handle.net/1765/19714

Sousa, M., *Servant Leadership to the Test: New Perspectives and Insight,* Promotors: Prof.dr. D. van Knippenberg & Dr. D. van Dierendonck, EPS-2014-313-ORG, http://hdl.net/1765/1

Spliet, R., *Vehicle Routing with Uncertain Demand*, Promotor: Prof.dr.ir. R. Dekker, EPS-2013-293-LIS, http://hdl.handle.net/1765/41513

Srour, F.J., *Dissecting Drayage: An Examination of Structure, Information, and Control in Drayage Operations*, Promotor: Prof.dr. S.L. van de Velde, EPS-2010-186-LIS, http://hdl.handle.net/1765/18231

Staadt, J.L., *Leading Public Housing Organisation in a Problematic Situation: A Critical Soft Systems Methodology Approach*, Promotor: Prof.dr. S.J. Magala, EPS-2014-308-ORG, http://hdl.handle.net/1765/50712

Stallen, M., *Social Context Effects on Decision-Making; A Neurobiological Approach*, Promotor: Prof.dr.ir. A. Smidts, EPS-2013-285-MKT, http://hdl.handle.net/1765/ 39931

Tarakci, M., *Behavioral Strategy; Strategic Consensus, Power and Networks*, Promotor(s): Prof.dr. P.J.F. Groenen & Prof.dr. D.L. van Knippenberg, EPS-2013-280-ORG, http://hdl.handle.net/1765/ 39130

Teixeira de Vasconcelos, M., *Agency Costs, Firm Value, and Corporate Investment*, Promotor: Prof.dr. P.G.J. Roosenboom, EPS-2012-265-F&A, http://hdl.handle.net/1765/37265

Tempelaar, M.P., *Organizing for Ambidexterity: Studies on the Pursuit of Exploration and Exploitation through Differentiation, Integration, Contextual and Individual Attributes*, Promotor(s): Prof.dr.ing. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-191-STR, http://hdl.handle.net/1765/18457

Tiwari, V., *Transition Process and Performance in IT Outsourcing: Evidence from a Field Study and Laboratory Experiments*, Promotor(s): Prof.dr.ir. H.W.G.M. van Heck & Prof.dr. P.H.M. Vervest, EPS-2010-201-LIS, http://hdl.handle.net/1765/19868

Tröster, C., *Nationality Heterogeneity and Interpersonal Relationships at Work*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2011-233-ORG, http://hdl.handle.net/1765/23298

Tsekouras, D., *No Pain No Gain: The Beneficial Role of Consumer Effort in Decision Making*, Promotor: Prof.dr.ir. B.G.C. Dellaert, EPS-2012-268-MKT, http://hdl.handle.net/1765/ 37542

Tzioti, S., *Let Me Give You a Piece of Advice: Empirical Papers about Advice Taking in Marketing*, Promotor(s): Prof.dr. S.M.J. van Osselaer & Prof.dr.ir. B. Wierenga, EPS-2010-211-MKT, hdl.handle.net/1765/21149

Vaccaro, I.G., *Management Innovation: Studies on the Role of Internal Change Agents*, Promotor(s): Prof.dr. F.A.J. van den Bosch & Prof.dr. H.W. Volberda, EPS-2010-212-STR, hdl.handle.net/1765/21150

Vagias, D., *Liquidity, Investors and International Capital Markets*, Promotor: Prof.dr. M.A. van Dijk, EPS-2013-294-F&A, http://hdl.handle.net/1765/41511

Verheijen, H.J.J., *Vendor-Buyer Coordination in Supply Chains*, Promotor: Prof.dr.ir. J.A.E.E. van Nunen, EPS-2010-194-LIS, http://hdl.handle.net/1765/19594

Venus, M., *Demystifying Visionary Leadership; In Search of the Essence of Effective Vision Communication*, Promotor: Prof.dr. D.L. van Knippenberg, EPS-2013-289-ORG, http://hdl.handle.net/1765/ 40079

Visser, V., *Leader Affect and Leader Effectiveness; How Leader Affective Displays Influence Follower Outcomes*, Promotor: Prof.dr. D. van Knippenberg, EPS-2013-286-ORG, http://hdl.handle.net/1765/40076

Vlam, A.J., *Customer First? The Relationship between Advisors and Consumers of Financial Products*, Promotor: Prof.dr. Ph.H.B.F. Franses, EPS-2011-250-MKT, http://hdl.handle.net/1765/30585

Waard, E.J. de, *Engaging Environmental Turbulence: Organizational Determinants for Repetitive Quick and Adequate Responses*, Promotor(s): Prof.dr. H.W. Volberda & Prof.dr. J. Soeters, EPS-2010-189-STR, http://hdl.handle.net/1765/18012

Waltman, L., *Computational and Game-Theoretic Approaches for Modeling Bounded Rationality*, Promotor(s): Prof.dr.ir. R. Dekker & Prof.dr.ir. U. Kaymak, EPS-2011-248-LIS, http://hdl.handle.net/1765/26564

Wang, Y., *Information Content of Mutual Fund Portfolio Disclosure*, Promotor: Prof.dr. M.J.C.M. Verbeek, EPS-2011-242-F&A, http://hdl.handle.net/1765/26066

Wang, Y., *Corporate Reputation Management; Reaching Out to Find Stakeholders*, Promotor: Prof.dr. C.B.M. van Riel, EPS-2013-271-ORG, http://hdl.handle.net/1765/ 38675

Weenen, T.C., *On the Origin and Development of the Medical Nutrition Industry*, Promotors: Prof.dr. H.R. Commandeur & Prof.dr. H.J.H.M. Claassen, EPS-2014-309-S&E, http://hdl.handle.net/1765/51134

Wolfswinkel, M., *Corporate Governance, Firm Risk and Shareholder Value of Dutch Firms*, Promotor: Prof.dr. A. de Jong, EPS-2013-277-F&A, http://hdl.handle.net/1765/ 39127

Xu, Y., *Empirical Essays on the Stock Returns, Risk Management, and Liquidity Creation of Banks*, Promotor: Prof.dr. M.J.C.M. Verbeek, EPS-2010-188-F&A, http://hdl.handle.net/1765/18125

Zaerpour, N., *Efficient Management of Compact Storage Systems*, Promotor: Prof.dr. M.B.M. de Koster, EPS-2013-276-LIS, http://hdl.handle.net/1765/38766

Zhang, D., *Essays in Executive Compensation*, Promotor: Prof.dr. I. Dittmann, EPS-2012-261-F&A, http://hdl.handle.net/1765/32344

Zhang, X., *Scheduling with Time Lags*, Promotor: Prof.dr. S.L. van de Velde, EPS-2010-206-LIS, http://hdl.handle.net/1765/19928

Zhou, H., *Knowledge, Entrepreneurship and Performance: Evidence from Country-level and Firm-level Studies*, Promotor(s): Prof.dr. A.R. Thurik & Prof.dr. L.M. Uhlaner, EPS-2010-207-ORG, http://hdl.handle.net/1765/20634

Zwan, P.W. van der, *The Entrepreneurial Process: An International Analysis of Entry and Exit*, Promotor(s): Prof.dr. A.R. Thurik & Prof.dr. P.J.F. Groenen, EPS-2011-234-ORG, http://hdl.handle.net/1765/23422

Erasmus Research Institute of Management - ERiM

BRIDGING MODELS AND BUSINESS

UNDERSTANDING HETEROGENEITY IN HIDDEN DRIVERS OF CUSTOMER PURCHASE BEHAVIOR

Recent years have seen many advances in quantitative models in the marketing literature. Even though these advances enable model building for a better understanding of customer purchase behavior and customer heterogeneity such that firms develop optimal targeting and pricing strategies, it has been observed that not many of the advanced models have found their way into business practice.

This thesis aims to bridge the gap between advanced models and their business applications by systematically extending the use of models. We first focus on probabilistic customer base analysis models that deal with understanding customer heterogeneity and predicting customer behavior. These models specify a customer's transaction and defection processes under a non-contractual setting. Through this study, we show that the timing of the next purchase for each customer can be predicted using these models. We also extend them by modeling customer heterogeneity in a more flexible and insightful way. As a result, managers can obtain a refined segmentation. Based on the customer heterogeneity insights, we then focus on pricing strategies for online retailers who derive their revenues from delivery fees and sales. In order to come up with optimal pricing strategies for delivery fees, we use ideas from the two-part tariff literature.

Given the time and costs associated with implementing advanced models/theories in managerial practice, the marketing executives need to be convinced by clearly demonstrating the contributions of such models. Our study serves as a step toward bridging advanced models and business practice by empirically demonstrating their extended contributions.

ERiM

The Erasmus Research Institute of Management (ERIM) is the Research School (Onderzoekschool) in the field of management of the Erasmus University Rotterdam. The founding participants of ERIM are the Rotterdam School of Management (RSM), and the Erasmus School of Economics (ESE). ERIM was founded in 1999 and is officially accredited by the Royal Netherlands Academy of Arts and Sciences (KNAW). The research undertaken by ERIM is focused on the management of the firm in its environment, its intra- and interfirm relations, and its business processes in their interdependent connections.

The objective of ERIM is to carry out first rate research in management, and to offer an advanced doctoral programme in Research in Management. Within ERIM, over three hundred senior researchers and PhD candidates are active in the different research programmes. From a variety of academic backgrounds and expertises, the ERIM community is united in striving for excellence and working at the forefront of creating new business knowledge.

ERIM PhD Series
Research in Management

Erasmus Research Institute of Management - ERiM
Rotterdam School of Management (RSM)
Erasmus School of Economics (ESE)
Erasmus University Rotterdam (EUR)
P.O. Box 1738, 3000 DR Rotterdam,
The Netherlands

Tel. +31 10 408 11 82
Fax +31 10 408 96 40
E-mail info@erim.eur.nl
Internet www.erim.eur.nl

Design & layout: B&T Ontwerp en advies (www.b-en-t.nl) Print: Haveka (www.haveka.nl)