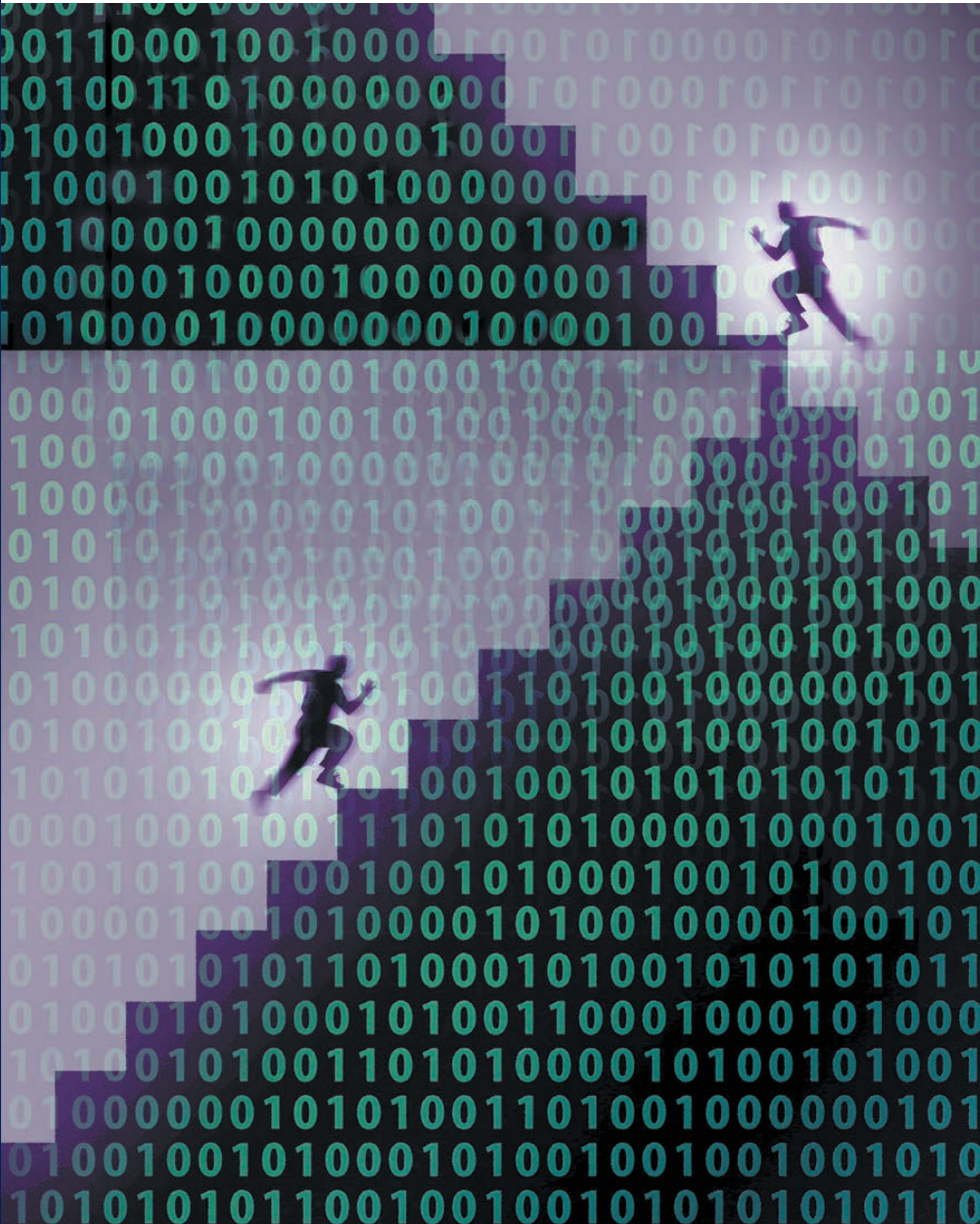
WEI LI

Competition in the Retail Market of Consumer Packaged Goods



Competition in the Retail Market of Consumer Packaged Goods

Competition in the Retail Market of Consumer Packaged Goods

Concurrentie in het Schap

Thesis

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

Prof. dr. F. A. van der Duijn Schouten

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on
Thursday the 14th of January 2021 at 9:30 hrs

by

Li, Wei
born in Hubei, China

Erasmus University Rotterdam

Doctoral Committee

Doctoral dissertation supervisor(s):

Prof.dr. D. Fok
Prof.dr P.H.B.F. Franses

Co-supervisor(s):

Other members:

Prof.dr. R. Paap
Prof.dr. A.C.D. Donkers
Prof.dr. F. Sotgiu

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: www.erim.eur.nl

ERIM Electronic Series Portal: repub.eur.nl/

ERIM PhD Series in Research in Management, 503
ERIM reference number: EPS-2021-503-MKT
ISBN 978-90-5892-597-8

Design: PanArt, www.panart.nl

This publication (cover and interior) is printed by Tuijtel on recycled paper, BalanceSilk®
The ink used is produced from renewable resources and alcohol free fountain solution.
Certifications for the paper and the printing production process: Recycle, EU Ecolabel, FSC®C007225
More info: www.tuijtel.com

FSC
www.fsc.org
RECYCLED
Paper made from recycled material
FSC® C007225

Preface

I would not have been able to complete this dissertation without the help of lots of people. First I would like to thank my supervisors prof. dr. Dennis Fok and prof. dr. Philip Hans Franses, whose knowledge, patience, and motivation guide me through the long journey of my PhD trajectory. Dennis, thank you for your guidance on the direction of my research and valuable comments on modeling, coding, and academic research in general. Philip Hans, thank you for always being there providing helpful feedback on paper writing and positioning. Also I would like to thank dr. Christaan Heij who had provided precious help on the sales forecasting chapter.

I would like to express my immense gratitude to colleagues at the University of Nottingham, in particular Professor Kevin Lee and Professor David Harvey. Without their help and encouragement, it would have been much more difficult, if not impossible, to resume and complete this dissertation.

I would like to thank prof. dr. Albert Wagelmans and prof. Girma Sourafel for being very supportive in my job seeking process. Albert thank you for showing me that an excellent lecturer is not only being good at teaching but also being systematic at organising. Girma thank you for your insightful suggestions on under-graduate dissertation supervision.

I would like to thank all the lecturers who have taught me at Erasmus University Rotterdam, particularly prof. dr. Richard Paap, prof. dr. Patrick Groenen, and dr. Alex Koning, apart from my supervisors. The knowledge I have learnt from all of you is the base for my current and future research and work.

I am very grateful to my close friends dr. Chen Li, dr. Zhihua Li, dr. Vitalie Spinu, dr. Yijing Wang, prof. dr. Chen Zhou, who have made Rotterdam a memorable place full of fun and joy to me. My friends in the UK, dr. Zhihua Li, Haitao Peng, dr. Jing Chen, and dr. Ziming Cai, thank you all for your company and the time spent with you is always so happy.

Last but not least, I would like to express my utmost gratitude to my family.

Mum and dad thank you for always providing a harbour for me when I feel tired. Yang, my dear husband, thank you for being supportive all the way. Bodun, my dear son, your unconditional love definitely motivates me to set a sample for you and being better and stronger!

Contents

Chapter 1

Introduction

Competition in the retail market of consumer packaged goods is an important and fascinating research topic to both practitioners and scholars. It can be studied from various angles, at various aggregation levels, and using various methods. The competition can be between brands, stores, and all the way up to retail chains. The marketing mix actions, for example price promotions, advertising campaigns, new product launches etc., are widely used by managers to increase profits. One can study the impact of marketing mix instruments on consumers or the impact on competitors. The role the marketing mix plays in competition can be approached from the individual level using household panel data or survey data, or from the product level using aggregated sales data. Given a specific topic and available data, researchers can use theoretical model or empirical model or a combination of the both. All these research efforts will help to improve our understanding of competition in the market and eventually contribute to management decision making. This dissertation chooses three specific topics pertaining to competition in the retail market of consumer packaged goods and the chosen research questions are introduced in the following section. To answer the research questions two new models are employed. These are a Hierarchical Bayes Ordered Probit and a Hierarchical Bayes Market Share model, which will be introduced in section 1.2. Our empirical findings will be summarised in section 1.3.

1.1 Research questions

This dissertation investigates three research questions pertaining to competition in the retail market of consumer packaged goods. Chapter 2 looks into how brands react to each other's price promotions in terms of own price promotion planning. Chapter 3 zooms out and studies the competition between retail chains and the role a retailer's private labels play. Chapter 4 comes back to the classic retail forecasting issue and investigates how to include competition information to improve brand sales forecasts.

Price promotion is the most extensively used among all kinds of marketing actions as it has a large immediate effect on sales. There is a rich literature on the consumer's response to promotions, but there are relatively few papers studying the reactions of brands to the use of this powerful tool. Chapter 2 focuses on the use of one instrument, namely price promotions, and the reactions with the same instrument across many product categories. Retailers use promotional calendars to plan their future promotion activities (Leeflang and Wittink, 2001). Bogomolova et al. (2017) finds that manufacturers take competitors' price promotions as the key determinant in their own promotion planning, and they use intuition and untested assumptions as inputs to their price-promotion decisions. As a result brand managers do not necessarily wait till a competitor's price promotion materialises before planning a reaction, they may anticipate and react when they plan their promotional calendars. The use of multi-category data allows us to look into how the brands' market shares and price levels and the category characteristics influence competitive reactions. Competitive reactions are defined as reactions by one brand (the defender) with one or more instruments to another brand's (the attacker) actions (Horváth et al., 2005). The literature has found asymmetric price and market share impact on the cross-price effect (Blattberg and Wisniewski, 1989a; Sethuraman and Srinivasan, 2002; Horváth and Fok, 2013), which states that price cuts of high-price/high-share brands have a greater impact on low-price/low-share brands than the reverse. As a result, when brand managers plan their competitive reactions, the relative price and/or market share positioning play an important role. Competitive reactions are also likely to be related to market environment, which can be described by product characteristics and the competitive structure in the market, for example the average purchase quantity of a category, the market concentration, and the number of brands in a market. Therefore, on top of competitive reactions, we also study the moderation effects of brand positioning and category characteristics. The chapter is co-authored with prof. dr. Dennis Fok and both authors made significant contributions.

Chapter 3 moves to a highly aggregated level to investigate the competition between retail chains and the role of their private label brands. In the last decades private labels have become more and more prevalent across all fast-moving consumer-goods categories. The increasing importance of private labels offers retailers a way to differentiate themselves from their competitors. We propose a three dimensional measure for the private-label positioning: breadth of the private-label program, price, and assortment size. Using the definition of customer-based brand equity proposed by Keller (1993), we define customer-based *chain* equity as the differential effect of chain knowledge on customer responses to the marketing-mix activities of the chain. To study the differential effect that can be attributed to the private-label program, we first investigate how a retailer's category specific market shares depend on price changes by national brands and private labels. Then we examine the impact of the private-label positioning on the baseline market shares and the price sensitivities. If a certain private-label positioning strategy increases a retailer's baseline share or makes its market share less sensitive to price, we can say that this strategy contributes to customer-based chain equity. This chapter is co-authored with prof. dr. Dennis Fok and prof. dr. Philip Hans Franses. This chapter is mainly written by myself. Prof. dr. Dennis Fok made significant contributions on modeling and structuring and Prof. dr. Philip Hans Franses made valuable contributions on positioning of the paper and wording.

Retail forecasting is relevant to both retailers and manufacturers. Forecasts give an impression of what future sales patterns can look like, and it helps to understand the competition between brands. When there are cross brand effects brand level forecasts are relevant as this can facilitate organising promotions by brand (Fildes et al., 2019). Chapter 4 aims to investigate how much value is added to traditional sales forecasting models in marketing by using modern techniques like factor models, Lasso, elastic net, random forests and boosting methods. Brand sales forecasts are often generated from econometric time series models (Hanssens et al., 2003), where the well-known SCANPRO model (Wittink et al., 1988) is an illustrative example. Such models usually include past sales and own marketing activities (current and past), and frequently also variables concerning past competitor behaviour are included, at least if one knows this competition. As retailers have the most complete information regarding sales and promotions, we take a retailer's point of view and address various ways to include information on competitors. Our key conjecture is that in practice it is often not known which brands are effectively the main competitive brands. One may then resort to a couple of strategies. One option is to simply ignore competition. This makes the model simple to analyse, as there is no need for the sometimes cumbersome collection and preparation of data from competitors. A second strategy is to spend effort in studying which are the most relevant competitive brands. The third strategy, which is the one we will address in chapter 4, is to consider all other possible brands as potential competitors that might be relevant for the forecasts of the own brand sales. This approach is relevant if we do not know beforehand which brands have predictive content, and in this case we can let the data help to decide on this each time we make a forecast. This chapter is co-authored with prof. dr. Dennis Fok and prof. dr. Philip Hans Franses. All three authors have significant contributions to the the chapter.

Chapter 2 and 3 both study how certain effects change cross-sectionally, which motivates the use of hierarchical models. Chapter 4 involves choosing the most powerful predictors from a large variable set, therefore leads to the comparison of various Big Data methods. The methodologies are introduced in the next section.

1.2 Methodologies

All the following three chapters are empirical studies using weekly supermarket scanner data of consumer packaged goods. The Hierarchical Bayes Ordered Probit model and Hierarchical Bayes Market Share model are new to the marketing literature and thus are part of the contributions of the dissertation.

The Hierarchical Ordered Probit model is used in Chapter 2. It consists of two layers: the first layer is an ordered probit model, which investigates the competitive reactions in terms of no price cut, small price cut, and deep price cut in each category; and the second layer is a linear model, which associates these promotion interactions with brand specific prices and market shares and some category specific characteristics.

The Hierarchical Market Share model is employed in Chapter 3. It consists of two layers as well. The first layer is a market share attraction model describing the retailers' market shares in a specific category. The second layer associates chain baseline market shares and price-sensitivities across all categories and chains with private-label positioning variables and some other chain- and/or category-specific characteristics.

In both Chapter 2 and 3 we take a Bayesian approach to simultaneously estimate the parameters from the two layers of the models. The Markov chain Monte Carlo (MCMC) simulation method is used to obtain posterior results.

In Chapter 4 there is a benchmark model that uses only the focal brand's own information, while the other models include competitive sales and marketing activities in various ways. An Average Competitor Model (ACM) summarises all competitive information by averages. Factor-augmented models incorporate all or some competitive information by means of common factors. Lasso and elastic net models shrink the coefficient estimates towards zero by adding a shrinkage penalty to the sum of squared residuals that is to be minimised. Random forests average many tree models obtained from bootstrapped samples. Boosting trees grow many small trees sequentially and then average over all the tree models to deliver forecasts. We use these methods to forecast sales of packaged goods one week ahead and compare their predictive performance.

1.3 Results and implications

Through our empirical analyses, we try to answer the three research questions proposed and find some interesting results.

First, we find that the competitive reactions can be partly explained by several brand specific and category specific characteristics, which implies that brand managers do take into account competition when planning their price promotions. If a brand's market share is 50% or larger (so the attacker's share cannot be greater than the defender's), then its reaction intention increases monotonically with the attacker's market share. However, if the defender's share is smaller than 50%, the reaction intention reaches the maximum when the attacker has the equal share with the defender then falls back with the increase of the attacker's share. Thus a market with two brands that each have half share is the most competitive. Also in categories that represent high budget share and highly concentrated, there are more likely to have price retaliations. While in categories with more brands or more price dispersion, there are less likely to have price competitive reactions.

Second, we did not find significant impact of private-label positioning factors on chain and category specific baseline market shares. However the private-label positioning does influence retail chain's market share price sensitivities. We find that a relatively lower priced private-label programme with larger assortment would be the most effective in weakening the market share price sensitivity to national brands, thus contribute to the chain differentiation.

Last but not least, when it comes to brand sales forecasting, the performance of a benchmark model that only uses own brand information and season dummies can be improved by incorporating competitors information in some way. Among all the methods we test, the Lasso and elastic net are the safest to employ as they are better than the benchmark for most of the brands. The random forest method has better improvement for some of the brands.

Chapter 2

Immediate and Dynamic Competitive Reactions to Price Promotions

2.1 Introduction

Among all kinds of marketing actions, like price adjustments, advertising campaigns, new product launches etc., price promotions (temporary price reduction) is the most extensively used tool. Price promotions have an immediate effect on sales, although often followed by some post-promotional dip. There is a rich literature on the consumer's response to promotions, e.g. Kopalle et al. (1999); Van Heerde et al. (2000); Fok et al. (2006); Horváth and Fok (2013). Yet there are relatively much less papers that study competitive reactions (Reibstein and Wittink, 2005). However, as formulated by Steenkamp et al. (2005) it is the sequence of marketing actions and competitive reactions eventually determines the market structure and the performance of the players in it.

In the literature, competitive reactions are usually defined as reactions by one brand (the defender) with one or more instruments to another brand's (the attacker) actions (see, for example, Horváth et al., 2005). In other words competitors in a market observe or anticipate on each other's actions and the corresponding consumer response. This study focuses on the use of one instrument, namely price promotions, and the reactions with the same instrument across many product categories. The literature that deals with competitive reactions usually study the actual price changes (e.g. Steenkamp et al. 2005; Leeflang and Wittink 2001). However the actual price variation not only comes from price promotions but also from regular price adjustments. And the latter is obviously due to different motivations and have different market responses (Fok et al. 2006). To study the competitive reactions to promotions, regular price changes need to be separated from price promotions. To the best of our knowledge, the promotional competitive reactions across categories has not been examined yet.

Research on competitive reactions uses a variety of approaches. To answer normative questions like how should competitors react or what is the optimal reaction strategy, game-theoretic models have been used extensively (Ailawadi et al., 2005). For instance, Naik et al. (2005) extend the Lanchester model to investigate marketing mix planning in dynamic competitive markets. While studies that focus on empirical generalizations use econometric models to investigate reaction functions, that is, how competitors actually react to each other's attacks and the factors affecting the observed reaction behaviour. For example, Steenkamp et al. (2005) perform a large-scale empirical study using vector-autoregressive models with exogenous variables (VARX models) and they find that competitive reactions tend to be mostly passive (no reaction) and when a reaction occurs, it is usually with the same instrument, and there are few reactions in the long run. Next to these two main streams of approaches, there are some exploratory studies. For example, Montgomery et al. (2005) use interviews to get insight into whether or how managers incorporate competitor behaviour in their decision making. Montgomery and colleagues find that when managers make decisions on the timing of a price change or a product introduction, they incorporate competitor behaviour into their own decision making, but they do not attempt to predict future competitor reactions to their own moves.

Retailers use promotional calendars to plan their future promotion activities (Leeflang and Wittink, 2001). The promotion calendar is a result of negotiation between manufacturers and retailers and is planned over half year ahead (Guyt and Gijsbrechts, 2014). It seems that this leaves no space for the brand managers of manufacturers to react immediately if a competing brand promotes. However, the process of making a promotional calendar incorporates competition already. According to Leeflang and Wittink (2001), the brand managers would consider both past actions by competitors and expectations of future reactions during their promotion planning. A most recent study (Bogomolova et al., 2017) also finds that manufacturers take competitors' price promotions as the key determinant in their own promotion planning, and they use intuition and untested assumptions as inputs to their price-promotion decisions. As a result brand managers do not necessarily wait till a competing brand's price promotion materialises then start to plan a reaction, they may anticipate and react on the calendar. If this is true then the calendar reactions should partly be explained by brands competition relationships. If there is no anticipation in the formulation of promotion calendars, then there should be no patterns exhibited in the calendar reactions.

This chapter aims to provide insights into the calendar price promotion interaction patterns in a retailing context. We have two assumptions here, the first is that brands expect the competition from each other and reflect their expectations on the promotion calendar. The second is that the brand initiating the promotions can block the earliest possible slot on a retailer's calendar, so that the following promotions are reactions. Brands having price promotions planned closely on calendar do have immediate and dynamic cross price promotion effects (Horváth and Fok, 2013). The higher the cross effects, the more likely a brand intends to react competitively. But the calendar reactions may not fully mirror the cross price promotion effects for a variety of reasons. First, not only reaction intentions but also abilities, the resources brand managers can employ, play a role in promotion planning. Second the expectations can simply be wrong. Last but not lease the retailers also have a say on the promotion planning and the retailers' goals are in general different from the manufacturers' (Guyt and Gijsbrechts, 2014; Ailawadi et al., 2009). Hence we expect much less calendar reactions than universally found cross price effects. The data used in this chapter are weekly price promotions of fast-moving consumer goods from 25 categories in a large supermarket chain. The goal of this study is to empirically analyse the immediate and dynamic calendar competitive reactions to price promotions. We explicitly consider the potential asymmetries in these reactions, and the moderating effects of category characteristics on these reactions. Below, we will clarify all important concepts that appear in this research goal.

Immediate and dynamic effects. In the retail context, prices are usually set at a weekly frequency. The term "immediate effect" in this chapter therefore refers to the effect of a current planned price promotion (at week t) on the occurrence of competitive promotions in the next week (at week $t+1$). The term "dynamic effect" refers to the effect of a current planned price promotion (at week t) on the possible promotions in all the following weeks after the next week, that is, week $t+2$, $t+3, \ldots$. The patterns of immediate and dynamic reactions could be very different across brands and across product categories. One brand may react quickly (on calendar) to some brands' promotion attacks while it may react relatively slow to others, or even accommodate (give no reaction to) some attacks.

Symmetry and asymmetry. The competitive reaction between two brands i and j is considered to be symmetric if the (immediate and/or dynamic) reactions of brand j to the promotions of brand i are equal to the reactions the other way around. An analogous concept is the asymmetric price and market share effect on the cross-price effect (Blattberg and Wisniewski 1989a; Sethuraman and Srinivasan 2002; Horváth and Fok 2013), which states that price cuts of high-price/high-share brands have a greater impact on low-price/low-share brands than the reverse. Taking into account these asymmetric price and share effects, brand managers could base their competitive reactions on their relative price and/or market share positioning. We therefore consider prices and market shares as factors that may cause an asymmetry in immediate and/or dynamic competitive reactions.

Category's moderating effect. Competitive reactions are likely to be related to the market environment. The market environment can be described by product characteristics and characteristics of the competitive structure. Important product characteristics may be the average purchase quantity and important descriptors of the competitive environment may be market concentration or the number of brands in the market.

We explicitly focus our analysis on reactions to price promotions. However, we usually only observe the actual prices paid by the consumers. Not all changes in this price level correspond to promotions. The actual price may change because of regular price changes. Furthermore, as is common in this literature, we rely on prices that are aggregated at the brand level. Dependent on the particular type of aggregation the actual (aggregated) price may show some fluctuation that is independent of any promotion. Next it is unlikely that competitors would react to very small changes in the price of a particular brand. Therefore, we transform the actual price into an ordinal price promotion index. We use the algorithm of Fok et al. (2006) to transform the actual price into the regular price and a continuous price promotion index. This price promotion index equals one if the actual price equals the regular price, a price promotion index of .95 indicates a 5% price discount. Such an index is also used by, for example, Macé and Neslin (2004) and Horváth et al. (2005). Regular price changes are not reflected in the price promotion index. To get rid of the small, for our purpose irrelevant, variation in the promotional index due to aggregation issues, the regular price decomposition algorithm, or other minor price variation, we next classify the promotional index in three ordered categories: no discount, shallow discount, and deep discount. The threshold values used for this classification are discussed in Section 2.4. Note that, in practice, when brand managers make promotional decisions, they also tend to make choices between no discount, small discount, and big discount.

As we discretise the price promotions, the competitive reactions we study are also in terms of the same ordinal levels: no reaction, reaction by shallow discount, and reaction by a deep discount. Previous studies do not use such a discretisation. One may argue that we reduce the available information by the discretisation. Of course this is true. However, as argued above the discretisation allows us to focus on the true competitive reactions without having to consider, and model, all other variation in the (promotional) prices. If we would use the continuous promotion index as our main dependent variable, any model will be forced to fit all the small variation which actually may not be the brands' promotional actions.

To explain the promotional reactions of each brand and to explain the differences in competitive reactions across brands and across categories, we use a hierarchical Bayes ordered probit (HB-OP) model. The model consists of two layers: the first layer is an ordered probit model, which specifies the price reaction function; the second associates the reactions with prices and market shares of the brands involved as well as some category characteristics. Following Steenkamp et al. (2005) and Fok et al. (2006), we assume that the category characteristics do not change over time. A Bayesian method is used to estimate the parameters of the two layers simultaneously. This method is more efficient than the two-step procedure that first estimates all individual level models and collects the relevant coefficients (and standard errors), next applies a linear model to these coefficients.

The remainder of this chapter is organised as follows. In Section 2.2, we review the related literature and indicate the added value of this chapter relative to the current literature. We discuss our expectations on the effects of moderating variables on competitive reations in Section 2.3. In Section 2.4, we present the data. Section 2.5 covers the HB-OP model in detail. Next we present the empirical results in Section 2.6. Finally, we draw the conclusions and give some discussions in Section 2.7.

2.2 Literature

This chapter can be related to two research streams. Concerning the methodology it relates to papers based on hierarchical Bayes (HB) models. Concerning the substantive content, it fits a series of papers that study competitive reactions to price promotions.

We first consider the literature on hierarchical Bayes models (HB models) in marketing (e.g. Kim et al. 2002; Talukdar et al. 2002; Van Nierop et al. 2008; Fok et al. 2006; Horváth and Fok 2013). All these papers use the hierarchical Bayesian approach in constructing models, while the topics they deal with and the specific models they employ are very different. Fok et al. (2006) put forth a hierarchical Bayes error correction model to explain the differences in immediate and dynamic effects of promotional prices and regular prices on sales. The first level in their model consists of a vector autoregression that is rewritten in error correction format. The second level correlates the immediate and dynamic effects with various brand-specific and category-specific characteristics. Horváth and Fok (2013) extend this to include cross-effects and competitive reactions. Kim et al. (2002) propose a demand model which is constructed as a HB model to allow for household-specific utility parameters. Talukdar et al. (2002) use the HB methodology to investigate new product diffusion across products and countries and they find that the pooling approach leads to substantial improvements in prediction accuracy. Van Nierop et al. (2008) augment a standard sales model that explains sales by marketing instruments with a second layer that relates the effect parameters to shelf and SKU descriptors.

The current chapter is different from the above literature in the way that it augments an ordered probit model with a second layer that relates immediate and dynamic competitive reactions to brand-specific and category-specific variables. Just like the other hierarchical Bayes models the HB-OP model 'shrinks' the parameters for individual brands, and it allows us to estimate the competitive reactions and the factors that determine them simultaneously.

There is also a literature that explicitly deals with competitive reactions. Leeflang and Wittink (1992) use (log-)linear models to study competitive responses to price and non-price promotions with scanner data on seven brands in one category. They found that competitors are more likely to react with the same than with a different marketing instrument and that reactions have a decreasing probability of occurring as the time lag increases. Leeflang and Wittink (1996) specify the normative criterion that managers should only react to other brands' marketing activities if those activities have nonzero effects on their own market shares. Based on this rule, they find that overreaction occurs more often than underreaction. In another paper, Leeflang and Wittink (2001) try to explain competitive reaction effects using cross- and own-demand elasticities. They find that the greater the cross-brand elasticity, the greater the corresponding reaction elasticity; the greater the own-brand elasticity, the smaller the corresponding reaction elasticity. Next to the works of Leeflang and Wittink, a series of studies (Horváth et al. 2005; Srinivasan et al. 2000; Steenkamp et al. 2005;

Horváth and Fok 2013) use vector-autoregressive models with exogenous variables (VARX) to investigate competitive reactions (using price promotions and/or advertising) and consumer responses (through sales or market shares) simultaneously. Horváth et al. (2005) find that the inclusion/exclusion of competitive reactions has a large impact on the estimated effects of marketing actions on brand sales. Srinivasan et al. (2000) find that a negative price shock results in a strong competitive reaction and this in turn leads to a persistent reduction both in price and promotion prices. On the other hand, Steenkamp et al. (2005) find that the most dominant form of competitive response is no reaction, and when a reaction does occur, it is usually with the same instrument.

Our study differs from the above papers in three important aspects. First we focus on price promotions instead of considering all price variation. A change in actual price may come from two different pricing decisions: price discounting and regular price changes (Fok et al., 2006). Naturally the regular price change is not a promotional action and thus needs to be separated from discounting. Next the price, or promotional index, may contain other variation next to true promotions. For example, due to some data aggregation. Second, our estimation methodology is more efficient as we do not rely on a two step procedure. Third, we consider factors that explain the competitive reactions across brands in many different categories. The latter is also done by Steenkamp et al. (2005). However, in our work we consider a larger set of moderating variables. For example, we consider brand prices, category budget share, number of brands and price dispersion in a category. We also consider a possibly asymmetric effect of the distance in market share *and* price between two brands. All the moderation variables we consider will be thoroughly discussed in the following paragraphs.

Leeflang and Wittink (2001) have shown a positive correlation between cross-price effects and competitive reaction effects. The asymmetry in cross-price effects has already been shown in Blattberg and Wisniewski (1989a). So we expect a similar asymmetric pattern in competitive reaction effects. This expectation is also based on the fact that managers take into account consumer's responses when they make decisions on competitive reactions (Leeflang and Wittink, 2001). Following this reasoning, the factors that influence the cross- and own-price effects should also (partly) explain cross-brand competitive reactions. More specifically, we consider market shares, brand prices, and several category characteristics as potential factors that may affect the intensity of competitive reactions. Below, we will discuss these factors one by one.

2.3 Moderating factors

In this section we discuss our expectations on a number of factors that are likely to affect competitive reaction. We categorize the factors into being market share related, price related, and product category characteristic related.

Market share factors

We expect that the reaction intensity between two brands depends on their market shares. According to Steenkamp et al. (2005), the promotions of a high-share brand are more easily noticed and are therefore considered to be more threatening. High-share brands are thus more likely to receive reactions from other brands. Sethuraman and Srinivasan (2002) analytically and empirically show that the cross-price effect of high-share brands on the market share of low-share brands is smaller than the reverse. The intuitive explanation is that a low-share brand has more loyal consumers as they target a niche market. Therefore if the purpose of a reaction is to maintain a brand's current share, a low-share brand is less likely to react to a high-share brand's attack than the reverse. In addition, brands with a low market share often have less resources to respond to price promotion attacks by high-share brands. That being said, the competition reactions received by high-share brands are mostly from brands that fall into the same high-share group. This expectation is in line with the finding of Horváth and Fok (2013) that large brands have bigger cross promotion effects on each other than small brands.

Steenkamp et al. (2005) also show that the share difference between the attacker and the defender moderates the competitive reactions. But they don't disentangle the neighborhood effect (of the share distance) from the asymmetric effect (of whether the attacker's share is greater the defender). However, we expect this neighborhood effect to be asymmetric, that is, the market share distance may be most important if the attacker has the highest market share. This asymmetric neighborhood effect has not been considered in the literature on competitive reactions. We expect that the greater the share asymmetry in favor of the attacker, the smaller the probability of a reaction. Furthermore, a two-brand market with equal shares is considered to be more competitive than a market where two brands' shares are far apart (Raju, 1992). Thus the share distance between the attacker and the defender is expected to have a negative effect on the reaction likelihood.

Price factors

We are also interested in the effect of prices on competitive reaction. Many existing studies associate cross-price effects with price levels. Sethuraman et al. (1999) and Horváth and Fok (2013) empirically demonstrate that brands that are priced close to each other, have larger cross-price effects than brands that are priced far apart. The argument is that brands within the same price tier can more easily be substituted (Russel, 1992) and they compete for the same group of consumers. Therefore, we expect the reaction intensity to be higher for brands that are close in price. Many studies (e.g. Blattberg and Wisniewski 1989a, Sethuraman et al. 1999, Horváth and Fok, 2013) have found that a price promotion by a higher-priced brand affects the market share or sales of a lower-priced brand more so than the reverse, which is known as asymmetric price effect. Based on the positive correlation between competitive reactions and cross-price elasticities that was found by Leeflang and Wittink (2001), we expect a lower-priced brand to be more likely to react to the attacks from a higher-priced brand than the reverse. Analogous to the effect of market share of the attacker, we also expect when the higher-priced is the attacker, its promotions are more likely to be noticed and perceived to be a threat, thus more likely to encounter reactions.

Category characteristics

Budget share. Brands in categories that represent a large share of the consumer's shopping budget are expected to be more likely to react to competitor's promotions. This expectation is based on Macé and Neslin (2004) who indicate that products that take a larger share of the budget are associated with more stockpiling and a larger post-promotional dip when they are on sale. Fok et al. (2006) also empirically show that categories with a high budget share tend to have a larger immediate promotion price elasticity. Because consumers are sure to spend a large part of their budget on products of such a category, they are more willing to stockpile. Thus in such a category, when one brand is on sale, consumers would like to give up buying other brands and even other categories to stockpile the discounted product. The direct cross-brand price promotional effect in such categories is relatively large compared to categories of lower budget share, consequently managers are expected to be more likely to react.

Number of brands in a category. The competitive intensity in a market is partly captured by the number of brands in the category. However, the expected sign of the effect of the number of brands on competitive reactions is not clear. On the one hand, a large number of brands implies that each brand has a number of close substitutes available within the product category. Due to the presence of a large number of very similar brands, brand switching effects will dominate sales changes (Raju, 1992). The brand switching effects may trigger more competitive reactions between brands. On the other hand, Sethuraman et al. (1999) find stronger cross-price effect when there are less competing brands. The reason could be that in a category with less brands, each brand has a relatively high market share and brand managers have higher margins, so they are more likely to react.

Market Concentration. Market concentration captures another dimension of competitive intensity in a product category. A high market concentration implies that a small set of brands is dominant in the category, while a low market concentration implies that the market shares of the brands are close to each other and no one is dominant. If one brand or a set of brands is dominant because of its unique positioning in a category, these brands are less exposed to brand switching effect, thus the competition in such a category is less intense (Raju, 1992). Steenkamp et al. (2005) also find that in a concentrated market manufacturers tend to perform less price cuts to maintain their high margins. Hence a high market concentration implies less intensive competition, suggesting a negative moderating effect for competitive reactions. On the other hand, Horváth and Fok (2013) find that in highly concentrated markets there are greater cross-price effets, therefore more likely to lead to price retaliation. Overall we keep the expectation of the sign of the moderating effect of market concentration open.

Product differentiation. The product differentiation measures how different the products in a category are. It is expected to have a negative moderating effect on competitive reactions. Differentiation makes brands less exposed to competitor's actions (Narasimhan et al., 1996). The empirical results of Fok et al. (2006) also show that categories with a large price dispersion have smaller price promotion effects. If brand A and brand B are very different, they attract consumers with different preferences. Even when brand A has a price cut, the consumers who prefer brand B are not very likely to change their preferences and choose brand A. So managers are less likely to react to other brands' promotion in a market with very differentiated products.

2.4 Data

The data are also studied in Srinivasan et al. (2004) and Fok et al. (2006). The dataset covers 25 product categories[1] from a large supermarket chain in the Chicago area, Dominick's Finer Foods. It contains weekly data over the period from September 1989 until May 1997. The data are aggregated from SKU level to brand level using static weights. See Fok et al. (2006) for more details. Each category contains four brands: brands 1-3 represent the three major brands in the category and 4 is the "rest". We treat brand 4 as a real brand as is commonly done (see for example, Fok et al., 2006).

The changes in a brand's actual price series come from different sources: price promotions, regular price changes, and the aggregation process from SKU to brand level. Given our research questions, we focus on promotional price changes. Reactions to regular price changes may be completely different and they are therefore beyond the scope of this chapter. The original data contains only the actual price. Fok et al. (2006) smooth the actual price series using cubic splines with asymmetric weights to decompose the actual price series into regular and promotional prices. The size of a promotion is then measured by a promotion index, which equals the ratio of the actual price and the regular price. So a lower promotion index means a deeper discount. As shown in Figure 2.1, price goes up and down sometimes due to regular price adjustments, but the promotion index can never be larger than one by definition.

To focus on true price promotions, we study three competitive reaction levels: "no discount", "shallow discount", and "deep discount".We convert the continuous promotion index into a three-level ordinal promotion index, see Figure 2.2. The two cutoffs are set at 0.95 and 0.9 and are the same across all brands in all categories[2]. The cutoff of 0.95 is widely used to recode promotional price into zero-one promotion dummies (e.g. Fok et al. (2006) and Raju (1992)). Indeed, when the promotion index is larger than 0.95, the difference between actual and regular prices is really small, and it may not even come from a real discount but from the aggregation or decomposition processes. The cutoff 0.9 is to distinguish between shallow and deep discounts. We choose 0.9 such that the incidence of deep discount is the lowest for

[1] The 25 categories are: bottled juice, cereals, cheese, cookies, crackers, canned soup, dish detergent, "front-end" candies (i.e., those displayed by the checkout registers), frozen diners, frozen juice, fabric softener, laundry detergents, oatmeal, paper towels, refrigerated juice, soft drinks, shampoos, snack crackers, toothbrushes, canned tuna, toothpaste, and bathroom tissue.

[2] For the shallow discount cutoff, we also tried alternative thresholds (0.94 and 0.97). It made no dramatic change in the histogram of price-promotion index. We also tried a 0.8 for the deep discount cutoff, but there are some brands who have never done such "big" promotions. To ensure each brand has three levels in promotion, we choose 0.9 for the cutoff.

Figure 2.1: Price Series and Promotion Index of a Brand in Fabric Softeners

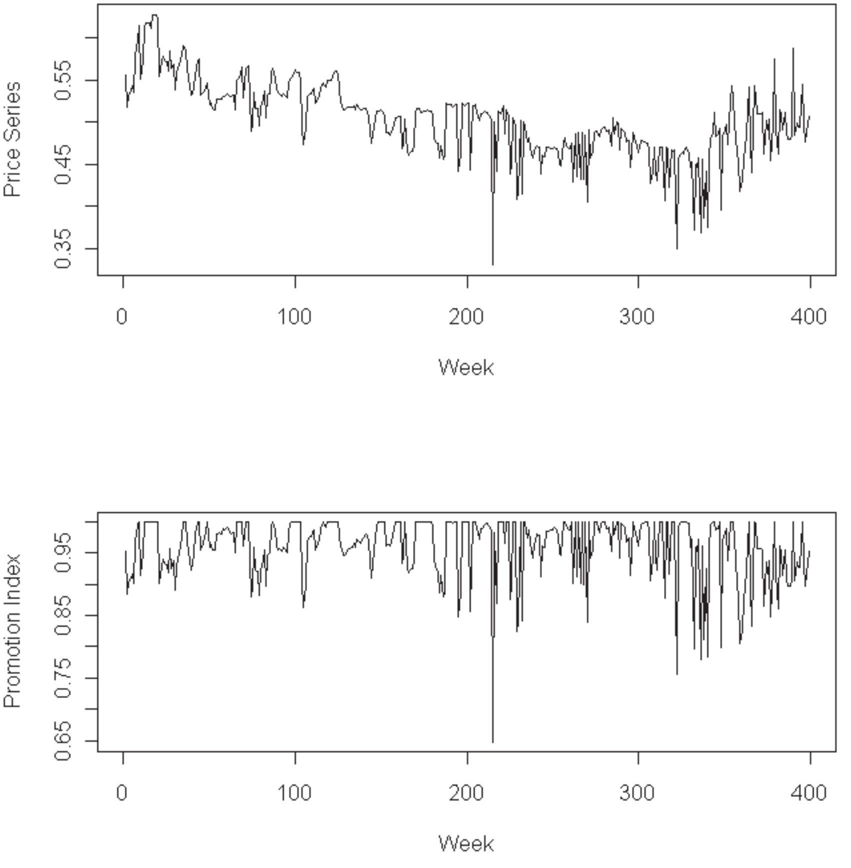
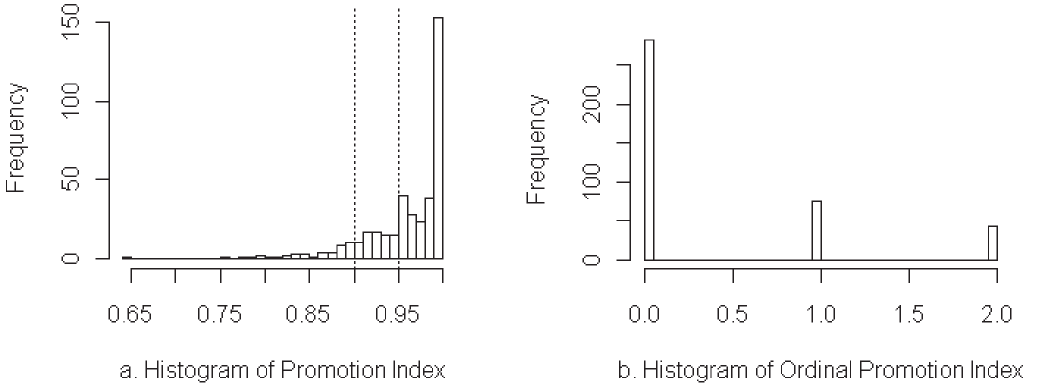


Figure 2.2: Recoding of Promotion Index into Three Promotional Levels



most brands. Because naturally a deep discount does not happen very often and is a special (strong) promotion. The ordinal promotion variable captures the most important variation in the promotion index. It is not affected by the small variations that may not come from promotions.

2.5 Hierarchical Bayes Ordered Probit model

The HB-OP model we employ consists of two layers: the first layer is an ordered probit model, which investigates the immediate and dynamic competitive interactions in each category; and the second is a linear model, which associates these promotion interactions with brand specific prices and market shares and some category specific characteristics. The following two subsections will address these two layers one by one.

2.5.1 The first layer of HB-OP model: immediate and dynamic competitive reactions

The first layer of the HB-OP model explains the ordinal price promotion index (denoted by PPI) for each brand. The variable PPI takes the values: no discount (0), shallow discount (1), and deep discount (2). This layer of the model assumes that managers make discount decisions based on some unobservable discount intention (denoted by y^*), the higher this intention, the more likely that managers will take (big) actions. The ordered probit model is specified for each of the four brands in all 25 categories, so there are 100 of such models in total. In a category (denoted by subscript c), each brand acts as both an attacker and a defender. The price-promotion index of the defender (denoted by subscript j) is treated as the dependent variable and the (lagged) price-promotion indexes of the attackers (denoted by subscript i) are used as explanatory variables. We include the lagged price promotion indexes as explanatory variables to study the immediate and dynamic calendar reactions as discussed in the introduction section. However the ordered price-promotion index cannot be used as explanatory variables directly, so we code it into a price promotion dummy variable PPD, which equals to 1 if there is either a shallow or a deep discount, otherwise it is 0.[3] We include K lagged terms of these promotion dummies. We also include K lagged sales volumes (denoted by S) of the defender. The lagged own sales and own promotion dummies together capture the defender's own promotion planning strategy, on which we will not focus.

To control for possible seasonality and holiday effects in promotions, we include seasonal dummies (denoted by SD) and holiday dummies (denoted by HD) as explanatory variables. There are 13 seasonal dummies, each covering four consecutive weeks, and 10 dummies for special holidays (Lent, Easter, Memorial Day, July 4th, Labor Day, Thanksgiving, the week following Thanksgiving, Halloween, Christmas and Superbowl). Based on individual ordered probit models, we select those seasonal dummies and holiday dummies with significant effects (at 0.05 level) and include them in the HB-OP model.

In sum, the explanatory variables we used in this layer consist of lagged price promotion dummies of all brands (including the attackers and the defender itself) in category c, lagged sales volumes of the defender, seasonal dummies, and holiday dummies, that is,

[3] We also experimented with another dummy called $DeepPPD$, which equals to 1 if there is a deep discount, otherwise is zero, and include both PPD and $DeepPPD$ as explanatory variables in the ordered probit model in our pre-analysis. The results show that the proportion of significant competitive effects decreased compared with the model not including $DeepPPD$. Given the fact that we want to explain the variation in competitive reactions, a lower proportion of significant competitive reactions would make it harder to be explained so we exclude the variable $DeepPPD$.

$$\begin{aligned} y^*_{jc,t} &= \alpha_{0jc} + \sum_{k=1}^{K} \alpha_{kjc} PPD_{jc,t-k} + \sum_{k=1}^{K} \sum_{i \neq j} \beta_{kijc} PPD_{ic,t-k} \\ &\quad + \sum_{k=1}^{K} \lambda_{kjc} S_{jc,t-k} + \sum_{s=2}^{13} \nu_{sjc} SD_{s,t} + \sum_{h=1}^{10} \mu_{hjc} HD_{h,t} + \varepsilon_{jc,t}, \end{aligned} \tag{2.1}$$

$$PPI_{jc,t} = \begin{cases} 0 & if\ -\infty < y^*_{jc,t} \leq \gamma_{1jc} \\ 1 & if\ \gamma_{1jc} < y^*_{jc,t} \leq \gamma_{2jc} \\ 2 & if\ \gamma_{2jc} < y^*_{jc,t} \leq +\infty \end{cases}, \tag{2.2}$$

where the subscript t denotes a particular week and the error term $\varepsilon_{jc,t} \sim N(0,1)$. In (2.2) γ_{1jc} and γ_{2jc} are unobserved thresholds that transfer the intention $y^*_{jc,t}$ into actual decisions $PPI_{jc,t}$, where $\gamma_{1jc} < \gamma_{2jc}$. Here we set threshold γ_{1jc} to zero so that the constant term α_{0jc} and the other threshold γ_{2jc} are both identified. In (2.1), $PPD_{jc,t-k}$ and $PPD_{ic,t-k}$ denote k week lagged price promotion dummies of the defender j and the attacker i ($i \neq j$) in category c. We set the number of lags (K) based on a pre-analysis using the ordered probit model for every brand in each category. The Schwarz Information Criterion (SIC) shows that $K = 2$ is the best choice for most brands[4]. So we choose to set $K = 2$ in the HB-OP model. Finally, $S_{jc,t-k}$ denotes the k week lagged sales volume of the defender.

Coefficients α_{1jc} and α_{2jc} measure the own effect of brand j's prior promotions on present decisions. These two α parameters capture the fact that a brand's current action is partly based on its prior actions. A positive α means that if there is no promotion during the last two weeks, then the likelihood of no promotion now is high. The other way around, a negative α means that if there is no promotion during the last two weeks the likelihood of a present promotion is high. The two α parameters can have opposite signs. Coefficient β_{1ijc} measures the effect of brand i's attacks on brand j's decisions, we call this the *immediate effect*. While coefficient β_{2ijc} measures the effect of brand i's attacks two weeks ago on brand j's current actions, we call this the *dynamic effect*. Actually the "total" dynamic effect of brand i's actions at week $t-2$ on brand j's decision-making at week t is more than β_{2ijc}, because the immediate reaction effect is partly carried over to week t through the parameter α_{1jc}. For the discussion in this chapter we will ignore the carry over effect and focus on the "true" competitive reaction effect through β_{1ijc} and β_{2ijc}. Positive β_{1ijc} or β_{2ijc}

[4]Only for brand 1 in frozen dinners category, SIC indicates that $K = 3$ is the best.

would be the evidence of competitive reactions.

2.5.2 The second layer of the HB-OP model: explaining competitive reactions

In the second layer of the HB-OP model we stack all immediate reaction effects β_{1ijc} (each brand has 3 competitors in the same category, so in total we have $3 \times 4 \times 25 = 300$ reaction effects) and explain the variation in it by brand and category characteristics. We do the same thing with the dynamic reaction effect β_{2ijc}. The model thus examines the cross-sectional determinants of immediate and dynamic competitive reactions. Specifically, it correlates the reaction effects to the attacker's price and market share, the distance and asymmetry relative to the attacker, and the share and price of the defender, additionally there are some category characteristic variables. This part of the model reads,

$$\begin{aligned}
\beta_{1ijc} &= \theta_0 + \theta_1 P_{ic} + \theta_2 |P_{ic} - P_{jc}| + \theta_3 |P_{ic} - P_{jc}| I(P_{ic} > P_{jc}) \\
&\quad + \theta_4 MS_{ic} + \theta_5 |MS_{ic} - MS_{jc}| + \theta_6 |MS_{ic} - MS_{jc}| I(MS_{ic} > MS_{jc}) \\
&\quad + Z_c' \theta_7 + \eta_{1ijc} && (2.3) \\
\beta_{2ijc} &= \delta_0 + \delta_1 P_{ic} + \delta_2 |P_{ic} - P_{jc}| + \delta_3 |P_{ic} - P_{jc}| I(P_{ic} > P_{jc}) \\
&\quad + \delta_4 MS_{ic} + \delta_5 |MS_{ic} - MS_{jc}| + \delta_6 |MS_{ic} - MS_{jc}| I(MS_{ic} > MS_{jc}) && (2.4) \\
&\quad + Z_c' \delta_7 + \eta_{2ijc} \, ,
\end{aligned}$$

where again subscripts i, j, and c denote attacker, defender, and category respectively. P_{ic} denotes the time-average regular price of attacker i in category c. $|P_{ic} - P_{jc}|$ denotes the absolute value of the price distance. $I(P_{ic} > P_{jc})$ is an indicator function, which equals to one if the average regular price of brand i is larger than that of brand j and zero otherwise[5]. $|P_{ic} - P_{jc}| I(P_{ic} > P_{jc})$ is included to measure the price asymmetry between the attacker and the defender. For the time-average market share (MS_{ic}) we have similar components. Z_c is a M-dimensional vector containing the category characteristics we discussed in Section 2.3. Finally the error terms $(\eta_{1ijc}, \eta_{2ijc})'$ follows a multivariate normal distribution with mean zero and covariance matrix Σ.

In general, positive coefficients in (2.3) and (2.4) mean a strengthening moderating effect on competitive reactions, that is an increase in the likelihood of a competitive

[5]There are no $P_{ic} = P_{jc}$ cases in the data.

reaction.

2.5.3 Measurement of the category characteristics.

The budget share of a category is measured by the average total expenditures (Fok et al., 2006) (scaled down by 1,000,000) in the category over time, that is

$$BS_c = \frac{1}{T_c} \sum_{t=1}^{T_c} \sum_{i=1}^{4} S_{ict} P_{ict} \ , \tag{2.5}$$

where T_c is the total number of weeks (observations) for category c.

In the original data there is no information regarding the number of brands in each category, so we proxy it with the market share of the remainders other than the top three leading brands. It is measured by the time average share of "brand" 4, that is

$$MS_{4c} = \frac{1}{T_c} \sum_{t=1}^{T_c} (\frac{S_{4ct}}{\sum_{i=1}^{4} S_{ict}})' \ . \tag{2.6}$$

Market concentration in category c is measured by $\sum_{i=1}^{4} MS_{ic} log MS_{ic}$ (Raju, 1992), which is large for highly concentrated markets.

To calculate the price dispersion in category c (PD_c), the average regular price in category c, P_c, is first calculated as

$$P_c = \frac{1}{4T_c} \sum_{t=1}^{T_c} \sum_{i=1}^{4} P_{ict} \ . \tag{2.7}$$

Next we compare the highest regular price with the lowest regular price in a category (Fok et al., 2006)

$$PD_c = \frac{\sum_{t=1}^{T_c} [max_i(P_{ict}) - min_i(P_{ict})]}{T_c P_c} \ . \tag{2.8}$$

2.5.4 Estimation methodology

We use a Bayesian approach to simultaneously estimate the parameters in (2.1), (2.2), (2.3) and (2.4). We use the Markov chain Monte Carlo (MCMC) simulation method to obtain posterior results. The Gibbs sampling method (Geman and Geman, 1984) with data augmentation (Tanner and Wong, 1987) is applied. Thus the latent variables y^*_{jct}, β_{1ijc} and β_{2ijc} are sampled alongside the model parameters. The Bayesian analysis is based on uninformative priors for the model parameters, except for imposing an inverted Wishart prior on the Σ parameter to improve convergence of the MCMC sampler, see Appendix A for more details on the MCMC methodology.

2.6 Results

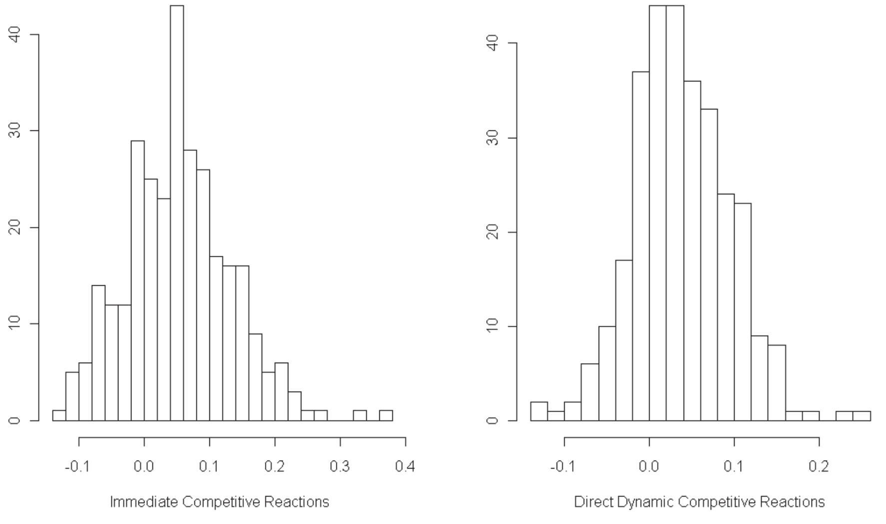
Our posterior results are based on 200,000 draws, of which we use the first 100,000 as burn-in. To remove auto-correlation in the chain, we save every 20th draw (so obtain 5000 draws) to compute posterior results. The estimation procedure is implemented in R and we use package Bayesian Output Analysis (BOA) (Smith, 2007a) for convergence tests. The Geweke test and the Heidelberger and Welch stationarity tests both show that the Markov chain has converged.

2.6.1 Immediate and dynamic reactions

The distribution of the posterior means of the immediate and dynamic reactions to price promotion attacks is presented in Figure 2.3. These histograms show the posterior means of the competitive effects across all brands and categories (so 300 posterior means in each histogram). As can be seen in the figure, the dispersion in the dynamic reactions is slightly smaller than that in the immediate reactions.

Most of the posterior means are positive, although many are close to zero. The mean immediate effect (0.053) and the mean dynamic effect (0.038) over all brands are also positive. However when using the 90% highest posterior density (HPD) intervals, only 32 immediate reaction coefficients out of 300 are significantly positive and for dynamic reactions the proportion is 11 to 300. There are no significantly negative competitive reactions. This finding is consistent with our expectation that calendar price retaliation does exist but not happen universally. According to our discussion in the introduction section, this could partly due to the fact that the price promotions are usually planned in advance based on brand managers' experiences and anticipations, which do not always reflect the competition reality. Overall, the immediate effect tends to be a bit larger than the dynamic effect. The posterior probability that the magnitude of the immediate effect is larger than that of the dynamic effect is 0.541. This implies that if possible brands tend to react sooner than later in promotion planning. Finally, the posterior correlation between the unexplained parts of the immediate and the dynamic reactions is 0.230 (computed from the estimation results for Σ, the covariance matrix of the error terms in the 2nd layer). This correlation reflects some consistencies in a brand's reaction strategy

Figure 2.3: Histogram of Posterior Means of Competitive Reactions for all 100 Brands



over time. It also partly explains the similarity in the distribution of the immediate and the dynamic effects.

2.6.2 Moderating factors of immediate and dynamic competitive reactions

The second layer of our HB-OP model studies the effect of moderating factors on immediate and dynamic competitive reactions. We calculate the posterior mean of the percentage of explained variance of the reaction effects. For the immediate effect, we explain 32.9% of the variance, and for the dynamic effect, we explain 22.0%. This implies that there are some patterns in the calendar reactions that do reflect the promotion competition between brands. Table 2.1 presents the posterior means and posterior standard deviations of the parameters in the second layer of our model (i.e. equation (2.3) and (2.4)). These results describe the moderating factors on the competitive reactions.

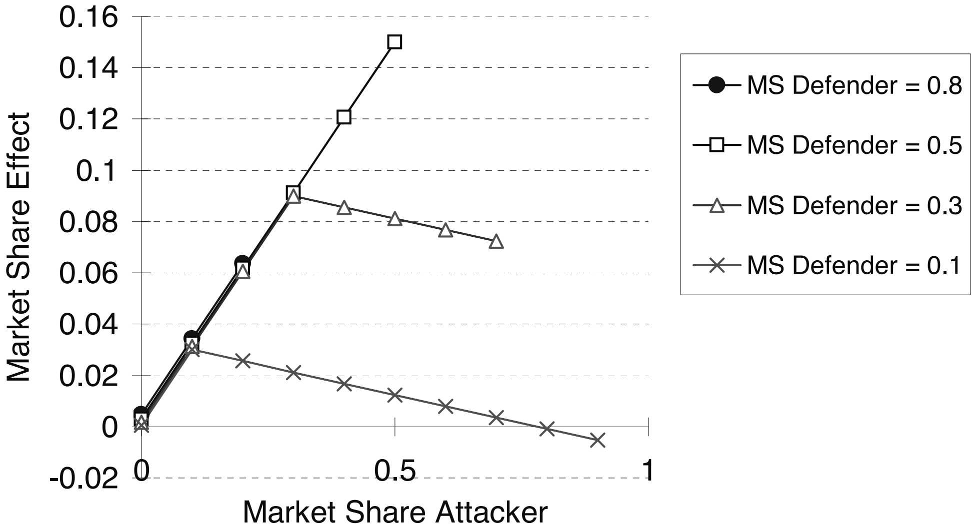
From the estimates for immediate reactions, we find that the larger the market share of the attacker, the more likely the defender will react. This is consistent with our expectation and Steenkamp et al. (2005), which indicate that a more powerful brand initiating a price promotion should expect more aggressive reactions by other

Table 2.1: Posterior Means of the Effects of Brand and Category Characteristics on Immediate and Dynamic Reactions

Characteristic		Immediate Reactions			Dynamic Reactions		
	Intercept	.280	(.128)	**	.132	(.126)	
Brand Share	Market Share of the Attacker	.300	(.160)	*	-.002	(.161)	
	Distance in Share	.006	(.092)		-.060	(.091)	
	Share Asymmetry	-.350	(.184)	*	.014	(.183)	
Brand Price	Price of the Attacker	.003	(.005)		.005	(.005)	
	Distance in Price	-.009	(.018)		-.034	(.018)	*
	Price Asymmetry	-.001	(.020)		.012	(.020)	
Category Characteristics	Budget Share	.139	(.056)	**	.013	(.055)	
	Share of Brand 4	-.167	(.078)	**	.008	(.076)	
	Market Concentration	.150	(.090)	*	.045	(.089)	
	Price Dispersion	-.131	(.077)	*	-.0740	(.076)	

Notes: Posterior standard deviations are in brackets.
* and ** = zero not contained in 90% and 95% HPD interval respectively.

Figure 2.4: Moderating Effect of Market Share on Immediate Reactions



brands in the category when controlling for all other factors. Next we find that if the attacker's share is greater than the defender, the likelihood that the defender reacts will decrease with the increase of the attacker's share. The total market share effect is graphically presented in Figure 2.4, where the horizontal axis represents the market share of attacker, the vertical axis represents the total market share effect (the sum of the effects of the three market share terms in (2.3)), and each line represents a certain market share of the defender.

As can be seen in Figure 2.4 that if the defender's market share is 0.5 or larger (so the attacker's share cannot be greater than the defender's share), then the total market share effect increases monotonically with the attacker's market share. However, if the defender's share is smaller than 0.5, the total market share effect reaches the maximum when the attacker has the equal share with the defender then falls back with the increase of the attacker's share. Thus a market with two brands that each have half share is the most competitive.

All three brand price factors have no significant moderating effect on immediate competitive reactions, which reveals that managers may not take these price factors into account when they make immediate reaction decisions.

All four category characteristic variables have a significant moderating effect on immediate reactions. First, consistent with our expectation, in categories that rep-

Table 2.2: Market Shares and Prices of the brands in Category Beer and Oat Meal

Category	Brand	1	2	3	4
Beer	Market Share (%)	44.6	9.6	3.0	42.8
	Price ($/10oz)	4.73	4.84	4.77	4.75
Oat Meal	Market Share (%)	68.8	11.7	11.7	7.7
	Price ($/10oz)	0.21	0.23	0.14	0.20

resent a larger budget share, brands are more likely to react to each other's attacks compared with brands in categories that account for a smaller budget share. Second, with the increase of number of brands in a category, the competitive reactions become less intense. This captures the fact that brands have a low ability to react in a market with many players, i.e. brand managers operate under a lower margin so are less able to react. Third, in highly concentrated categories there are more intense price retaliations, which implies a more competitive market. This implies that the higher cross price effects in a more concentrated market are captured by brand managers and build up their reaction intention. Moreover the dominant brands in such a market do have more resources to implement price retaliations. Finally, just as we have expected, in a category with differentiated products, the likelihood of the occurrence of competitive reactions is lower than in a category with relatively similar products.

For the moderating factors of dynamic competitive reactions, as can be seen in Table 2.1, among all ten moderating factors, only one factor, namely distance in price, has a marginally significant effect on dynamic competitive reactions. The negative moderating effect of distance in price on dynamic reactions implies that two brands who are close in price are more likely to compete with each other than two brands who are far apart in price levels. One reason for the few number of significant effects could be that the ratio of significant moderating effects out of all dynamic effects is very low (lower than the ratio for immediate effects), and the magnitude of the dynamic effect is too small. So there may be insufficient variation in it to examine the corresponding moderating factors. Another possible reason is that the dynamic effect itself is not affected by most of the moderating factors we discussed in this study. In any case this finding calls for further research.

2.6.3 Posterior promotion probabilities

To illustrate the effect sizes of the moderating variables, we calculate the posterior promotion probabilities for some brands (brand 1 and 2 in the beer category and

Table 2.3: Posterior Mean Promotion Probabilities of Brand 1 in Category Beer

Brand 1's Promotion Probabilities under Different Scenarios	Promoting brand None	Brand 2	Brand 3	Brand 4
Scenario 1 (Baseline)	.442	-	-	-
Scenario 2	-	.507	.474	.479
Scenario 3	-	.494	.464	.451
Scenario 4	-	.802	.779	.770
Scenario 5 (Expected Probabilities)	-	.650	.613	.604

Table 2.4: Posterior Promotion Probabilities of Brand 2 in Category Beer

Brand 2's Promotion Probabilities under Different Scenarios	Promoting brand None	Brand 1	Brand 3	Brand 4
Scenario 1 (Baseline)	.385	-	-	-
Scenario 2	-	.413	.435	.411
Scenario 3	-	.438	.424	.413
Scenario 4	-	.480	.466	.454
Scenario 5 (Expected Probabilities)	-	.454	.442	.430

Table 2.5: Posterior Promotion Probabilities of Brand 1 in Category Oat Meal

Brand 1's Promotion Probabilities under Different Scenarios	Promoting brand None	Brand 2	Brand 3	Brand 4
Scenario 1 (Baseline)	.157	-	-	-
Scenario 2	-	.185	.195	.176
Scenario 3	-	.168	.161	.178
Scenario 4	-	.360	.351	.331
Scenario 5 (Expected Probabilities)	-	.204	.198	.205

brand 1 in the oat meal category) under different promotional scenarios in a non-holiday week in the baseline season. The market shares and price levels of all the brands in the two categories are presented in Table 2.2. We study five scenarios in total. In each scenario we look at the promotion probability at time t. The scenarios differ on the weeks prior to t.

- Scenario 1 (Baseline): no own or competitive promotions in the past two weeks.
- Scenario 2: a competitor promotes at $t-1$ and no prior own promotion.
- Scenario 3: a competitor promotes at $t-2$ and no prior own promotion.
- Scenario 4: a competitor promotes at $t-2$ and there is own promotion at $t-1$.
- Scenario 5: a competitor promotes at $t-2$, no assumption for own promotion at $t-1$ (uses expected promotion probabilities).

The posterior probabilities for the first four scenarios are relatively easy to obtain given a set of parameter draws from the posterior distribution. For each draw the promotional probability for scenario 5 can be obtained from the corresponding probabilities under scenarios 2, 3, and 4. Denoting the l-th draw of the promotional probability under scenario 2, 3 and 4 by $p_2^{(l)}$, $p_3^{(l)}$, and $p_4^{(l)}$ respectively, we obtain the l-th draw for scenario 5 as $p_2^{(l)}p_4^{(l)} + (1-p_2^{(l)})p_3^{(l)}$.

As shown in Table 2.3, 2.4, and 2.5, in general, after a competitor promotes, a brand is more likely to promote given no prior own promotion. The likelihood of a promotion decreases a little bit one week later, but is still higher than the baseline level. While if there is an own promotion in the previous week, the brand's promotion likelihood will increase considerably. Comparing the two categories, promotions are more likely in the beer category compared to oat meal. Below we consider the reactions in more detail.

We first look at the immediate reactions. Table 2.3 shows that the focal brand, brand 1, is more likely to react to the attacks from brand 2 than from brand 3 ($0.507 > 0.474$). Brand 2's attack increases brand 1's promotion probability by 6.5 percentage points, from 0.442 to 0.507, twice as large as the effect size of Brand 3's promotion. This can be explained partly by the fact that the market share of brand 2 (9.6%) is larger than that of brand 3 (3.0%) and that the both attackers' shares are smaller than the defender's share (44.6%). In other words this is a reflection of the attacker's share effect. The contrast with the results in Table 2.3 and 2.4 gives a good example of the market share asymmetry effect, which states that the effect of market shares on reactions also depends on whether the attacker's share is larger than the defender's. Table 2.4 shows that brand 3's promotion increases the focal brand's (brand 2) promotion probability by 5 percentage points, from 0.385 to 0.435, almost twice as the effect size of brand 1's promotion. Although the market share of brand 1 is larger brand 3, when the share of the attacker, brand 1 here, is much larger than that of the defender, the reaction likelihood decreases with the absolute share difference, ceteris paribus.

The effect of price distances on dynamic reactions can be clearly seen in the oat meal category. In Table 2.2 we can see that the price of brand 2 ($0.23/10oz) is closer to brand 1 ($0.21/10oz) than brand 3's price ($0.14/10oz). Table 2.5 shows that brand 2's promotion increases brand 1's expected dynamic reaction probability by about 5 percentage points, from 0.157 to 0.204, which is larger than the effect of brand 3's promotions (from 0.157 to 0.198). Thus we expect that dynamic reactions of the defender to brand 2 are more likely to happen than to brand 3, ceteris paribus.

2.7 Conclusions

In this study, we proposed a HB-OP model to study the immediate and dynamic calendar competitive reactions to price promotions. The first layer of the model examines the immediate and dynamic calendar reactions; the second layer relates the immediate and dynamic calendar reactions to brand and category level characteristics. Our main contributions to the literature on competitive reactions are: (1) we focus on ordinal price promotions instead of all price variation, this variable ignores changes in regular price and is less affected by measurement error, aggregation, and decomposition issues; (2) the HB approach allows us estimate the parameters of the two layers simultaneously, which is more efficient than a two-step approach; (3) we allow more flexibilities in the specification of the moderating variables on competitive reactions.

We applied the model to weekly promotions for 100 brands from 25 product categories and estimate the model parameters using MCMC. Our results show that the calendar reaction to price promotion attacks does exist but to a limited extent. There are important competitive reactions in the market. In general the defenders are more likely to react sooner than later on calendar.

We find that the immediate calendar competitive reactions can be explained by several brand specific and category specific characteristics, which implies that brand managers do take into account competition when plan their price promotions. The moderating effect of the attacker's market share and a category's budget share are consistent with the findings in prior literature. Our new findings are that share distance mainly matters when the attacker's share is larger than the defender's, and that the number of brands and product differentiation both negatively moderate the competition in a category. Also there are more price retaliations in highly concentrated markets.

For the dynamic calendar competitive reactions, we only find one significant moderating factor, that is price distance. Consistent with our expectation, the closer two brands' prices are, the more likely they react to each other's attacks. The few number of effects may due to the small variation in the dynamic effects or the difficulty in finding appropriate moderating factors.

For future research, there are several interesting topics. First, when making promotional decisions, brands first decide whether they will promote and next they will decide the extent of promotion. So one could try to use a Tobit model to fit the promotional data: the first part (a binary Logit/probit model) explains whether there is a promotion, and the second part (a linear model) explains the extent of promotion. Second, we do not take into account the relation between manufacturers and retailers. One could investigate this relation by studying the competitive interactions between store brands and national brands. Third, one could study the non-calendar competitive reactions across brands and categories when appropriate data sets become available . Finally, we focus on price promotions in this chapter while it would be interesting to study multi-competitive reactions to non-price promotions and advertising.

Chapter 3

Retailer's Chain Equity and Private-Label Positioning

3.1 Introduction

In the last decades private labels have become more and more prevalent across all fast-moving consumer-goods categories. Many retail stores have introduced private-label brands and in many product categories these brands offer a variety of products. In almost all categories and retail stores we see that one or two private labels compete with the national brands. As a result, private-label products continue to rise globally with a value share growth rate of 16.7% (Nielsen, 2019). The private labels accounted for more than 30% of grocery sales in Europe and is catching up in the US during the last recession and have reached a share of around 20% (Seenivasan et al., 2016). The increasing importance of private labels offers retailers a way to differentiate from their competitors. In this chapter we focus on this aspect of private labels, that is, the impact of private labels on the competition between retailers by using a two-level market share model to study how a retailer's private label positioning influence its chain equity.

Using the definition of customer-based brand equity proposed by Keller (1993), we define customer-based *chain* equity as the differential effect of chain knowledge on customer responses to the marketing-mix activities of the chain. As the number of chains operate in one region is relatively small while there are many categories in each chain, and a retailer may position its private labels differently across categories, we will study category specific chain equity. To study the differential effect that can be attributed to the private-label program, we first investigate how retailer's category-specific market share depends on price changes by national brands and private labels. Next, we examine the impact of the private-label positioning on the baseline market share and the price sensitivities, both are category-specific. If a certain private-label positioning strategy increases a retailer's baseline share or makes its market share less sensitive to price, we can conclude that this strategy contributes to customer-based chain equity.

There are two main reasons for retailers to introduce and support the growth of private labels. The first one is profitability. Retailers usually have higher retail margins on private labels than on national brands. Moreover, they use private labels as a negotiation leverage with national-brand manufacturers (Ailawadi et al., 2008). Such a strategy brings direct profit gains to retailers. Empirical studies that support this notion include Ailawadi and Harlam (2004) and Pauwels and Srinivasan (2004). The second reason for retailers to carry private labels is differentiation. A well-positioned private label can differentiate a retailer from its competitors because private labels are exclusively available in the own stores. Indeed, Corstjens and Lal (2000) argued that alternative retailer strategies for differentiation, such as, increased service, extended opening hours, lower prices, and broader assortments, are less effective as competitors can easily copy these strategies. This is different for private labels. Private-label programs will in fact create an even larger differentiation when more retailers introduce private labels. The positioning strategies adopted by retailers will likely be different, and this amplifies the differentiating effect of the private labels. At the household level, Seenivasan et al. (2016) find a monotonic, positive impact of private label loyalty on consumer store loyalty, and Ailawadi et al. (2008) and Koschate-Fischer et al. (2014) find this impact is an inverted U shape. While at the aggregate level, to what extent the differentiation created by the private labels contributes to the customer-based chain equity for the retailer[1], remains open to discussion. In this chapter we aim to contribute to this issue by analysing a large dataset on retailer market shares and information on private labels, across categories and retailers.

We measure a retailer's private-label positioning from the following three aspects. First, we consider the breadth of the private label program, which is defined as the number of categories in which a retailer carries a private label. Second, we consider the relative price of the private label, that is, how much lower is the private label's price relative to the national brands in the same category. This measure reflects the

[1] In this chapter we use the term chain and retailer interchangeably. In our dataset a chain or retailer has more than one store.

price positioning of the private label. Finally, we use the assortment size, that is, the amount of product variants a private label has in a specific category.

We choose to study consumer response at an aggregate level using the retailer's market shares. As compared with household-level data, store-level data are less vulnerable to sample representativeness issues (Bucklin and Gupta, 1999). They can also be obtained across several retailers (Sriram et al., 2007), which does not hold for household-level data because households usually do not visit all retailers. We opt for market shares instead of sales as market shares nullify shocks and market developments that affect all retailers simultaneously. Furthermore, market shares allow for a clear focus on competitive structures. This makes market shares very suitable as the unit of analysis. The data we use are extracted from the IRI Academic Data Set (Bronnenberg et al., 2008). More specifically, we use weekly transaction data from four main grocery chains across 21 categories in a US metropolitan area (one of the 50 IRI markets) covering 261 weeks. As the data do not follow immediately from the provided data format, we spent ample time to carefully compile the data for our subsequent analysis. To evaluate the private-label positioning of various chains, we choose the product categories in which all four retailers carry private labels and discard those categories in which the private label sales are zero for many weeks.

For our exploratory analysis, we introduce a two-level hierarchical market share model to analyse the data. The first level is a market share attraction model, which is usually used to study brand shares, but here we use it to investigate chain shares within each product category. We incorporate a second level to measure the impact of the private-label positioning on baseline market shares, national-brand price sensitivity, and private-label price sensitivity. In this second level we also control for various chain and category characteristics. A Bayesian method is used to jointly estimate the parameters in the two levels of the model.

The outline of this chapter is as follows. In Section 3.2 we review the relevant literature, where some prior studies allow us to form expectations on the size and sign of certain effects. These expectations are presented in Section 3.3 together with our conceptual framework, which is useful for the subsequent data analysis. In Section 3.4 we discuss our data, propose our empirical methodology, and we give further information on the estimation procedure. We relegate technical details to Appendix B. In Section 3.5 we present our empirical results. Section 3.6 concludes this chapter with various recommendations for retailers. In this section we also provide limitations of our study and suggest a few further research topics.

3.2 Literature

The literature contains studies that correlate private-label use to store loyalty at the individual consumer level. Some studies find a positive correlation between private-label use and store loyalty. As to the causal direction, there are major differences in the focus of different studies. For example, Ailawadi et al. (2001) focus on the causal effect of store loyalty on private-label usage. They use perceptual data to measure store loyalty and private-label usage and find the two constructs to be positively related. Their explanation is that store-loyal consumers trust their chosen store, become familiar with its private labels, and therefore become more likely to use these private labels (see also Richardson et al. 1994). Bonfrer and Chintagunta (2004) also empirically find that store loyalty increases private-label choice probabilities at the individual consumer level. On the other hand, Corstjens and Lal (2000) focus on the reverse, that is, the private label's ability to build store loyalty. They put forward an analytical model to show that, under certain conditions, a high-quality private label can generate store loyalty and enhance profitability. They empirically show that store loyalty (measured by the share of a household's expenditures) positively depends on private-label penetration at the household level. Seenivasan et al. (2016) also discover a monotonic positive relationship between private label loyalty and store loyalty.

Nevertheless, there is other empirical evidence that private-label use is not necessarily positively correlated with store loyalty. Richardson (1997) finds that private-label users may actually be loyal to private-label products in general and not to the specific brand of a particular retailer. This implies that consumers may not differentiate between different retailers' private labels. Thus there will be no differential effect of private labels that would lead to increased retailer market shares. Moreover, Ailawadi et al. (2008) find that there is a nonlinear relationship between private-label use (measured by a household's private-label share) and store loyalty (measured by share-of-wallet). Private-label use and store loyalty both affect each other and both effects have an inverted-U shape. Koschate-Fischer et al. (2014) obtain similar results.

The above mixed findings can perhaps be explained by differences in the private-label positioning. When a retailer uses low-price low-quality private labels to capture the price-sensitive segment in the market, the private-label users may be especially attracted by the low prices and may not be loyal to any retailer. In contrast, a quality or premium private label strategy can build up store loyalty because consumers may link the chain's image with the quality private labels they offer.

Our study differs from the prior literature in three important aspects. First, we explicitly focus on retail competition and study the market share at the chain level using a hierarchical market share model. Second, we develop a systematic measurement of the retailer's private-label positioning across categories. We will elaborate on this measurement in Section 3.3.1. Third, we investigate the contribution of the private label to chain equity by studying the differential effect of the private-label's positioning strategy.

3.3 Conceptual Framework and Expectations

The expectations pertaining to the impact of price on the retailer's market share are quite straightforward. The retailer's market share will be negatively influenced by own price and positively influenced by competitor's price and lagged own price (due to stockpiling). This will hold for the price of the national brands as well as for the price of the private label. We distinguish between national-brand and private-label price changes to see how the market shares are affected by each. For example, a negative private-label price effect implies that a decrease in private-label price not just attracts consumers from national brands within the chain, but actually contributes to the chain's market share. The magnitude of the price effects may depend on the private-label positioning and retailer and market characteristics. In the following subsections we will mainly discuss the impact of private-label positioning and other control variables on baseline chain shares and on own national-brand and private-label price sensitivities.

3.3.1 Private-label positioning variables

The prior literature (Sayman et al. 2002; Meza and Sudhir 2010) defines the private-label positioning as how different or similar a private label is compared to a specific national brand within the same market. We aim to compare a retailer's private-label positioning across categories and retailers. To this end we need a more detailed description of the positioning. Below we develop a three-dimensional measure.

Breadth of private-label program When retailers extend their private-label programs across a larger number of categories, they make their private-label concept more salient to consumers. On the one hand, this signals effort and commitment to the program (Dhar and Hoch, 1997). Thus, a broad private-label program is expected to increase a chain's baseline share and make the chain share less sensitive to private-label price changes. On the other hand, a very salient private-label concept emphasises price and in turn may make the retailer's share actually more sensitive to private-label and national-brand price. The breadth of the private-label program is a retailer-specific factor and therefore does not distinguish between categories.

Private-label price positioning A low-price private-label strategy emphasises the price competition between retailers, and thus makes a retailer's share more sensitive to private-label price changes. At the same time, a low-price private label may attract more price-sensitive consumers who actually visit the stores of a chain for the cheap private-label products. Therefore a low-priced private label may increase a retailer's baseline share. The segment of consumers who only buy the low-price products may not be affected by national-brand price promotions as even the discounted national-brand products may still be considered as too expensive. The private-label price positioning is specific to a certain product category. The price gap between national brands and private labels may vary substantially across categories because of the different market characteristics of each category (Meza and Sudhir, 2010). To disentangle the differences across categories from the overall retailer's positioning, we measure the national brand-private label price differential relative to average category price, or in other words the relative price gap of the private label.

Private-label assortment size Having a large number of products in the private-label's assortment is an indication of the retailer's effort to create differentiation in its private label and to satisfy various consumer needs. Consumers are more likely to find what they need in a larger assortment (Kahn, 1998). Thus we expect that a large-private label assortment contributes to a chain's baseline share. The differentiation created by a large number of private-label products may also make consumers less sensitive to private-label prices. With a large assortment size, the private label can attract more consumers from the national brand and thus make the market share less sensitive to national-brand price. The assortment size varies across categories. To disentangle category differences from chain positioning, we measure the assortment size relative to the average category level.

3.3.2 Chain/category-specific control variables

We also consider variables to control for the impact of other chain characteristics. These variables are chain- and category-specific, including overall assortment size, overall price level, and promotion intensity. Similar as for private-label price and assortment size, we measure all these variables relative to the average category level.

Chain/category assortment size Retailers offer different assortment sizes across categories. On the one hand, consumers have heterogeneous preferences and variety-seeking tendencies, a wider assortment size would meet the diverse needs of consumers better and thus increase market share (Kahn, 1998). While on the other hand too many choices would increase a consumer's cognitive effort. If consumers feel overwhelmed and choose to buy nothing, a too large assortment could be detrimental to market share (Gourville and Soman, 2005). We will allow for such a nonlinear effect by including a quadratic form of assortment size in our model to capture the possible decrease in the marginal effect.

Chain/category price level The average price level of all brands within a category reflects the overall price positioning of a retailer. A relatively high price positioning may be due to the high service the retailer provides or to an assortment with more quality national brands. Better service may contribute to a retailer's baseline share, while more expensive products may do the contrary. In either case, the retailer aims to meet the special service or quality needs of part of the consumer population. These consumers are usually less price sensitive. Thus the total market share is expected to be less sensitive to both national-brand and private-label prices when the category price level is high.

Chain/category promotion intensity Frequent price promotional activities can build store traffic (Dhar and Hoch, 1997), and thus may contribute to a chain's baseline market share. On the other hand, intensive promotions lead to price-conscious consumers (Fok et al., 2006), therefore intensive promotions are expected to make market share more price sensitive.

3.3.3 Category-specific control variables

Finally we consider category-specific variables to control for market-level differences across categories. These variables are by definition the same across retailers, so they cannot have an impact on a chain's baseline market share. They may affect chain share price sensitivities.

Category sales High category dollar sales reflects either a high general price level or that large quantities are purchased (Fok et al., 2006). In both cases the category takes up a relatively large part of the consumer's budget. When such a product is on promotion, consumers are likely to accelerate their purchase or to stockpile the product in order to take advantage of the price deal (Macé and Neslin, 2004). Thus the chain share of such categories is expected to be more sensitive to price changes.

Category private-label development In some categories private labels have larger overall shares than in others. The differences in private-label penetration are mainly determined by differences in category characteristics (Dhar and Hoch, 1997). For example, in categories like paper towels and milk, private labels usually have high penetration; while in categories like shampoo and frozen pizza, private label share is relatively low. The private-label development is a good indicator for category characteristics. In categories with a higher penetration, chain share is expected to be more sensitive to private-label price. In categories with a lower penetration, chain share is expected to be more sensitive to national-brand price.

Category number of brands Categories with many national brands may either be very competitive or very differentiated. If the category is highly competitive, consumers are expected to be price sensitive. However, if the market is very differentiated consumers may be price insensitive.

To summarise the above we list all variables and their expected signs in Table 3.1. As the signs of price effects are usually negative, a negative moderating effect on price effects means that the price sensitivity is strengthened and a positive moderating effect decreases price sensitivity.

3.4 Methodology

In this section we first discuss our data set and next we present our hierarchical market share model.

3.4.1 Data

The data are part of the IRI Academic Data Set (Bronnenberg et al., 2008). We spent quite some time to compile the original scanner data, chain and store information, and product information files into a format that fits our research needs. We use

Table 3.1: Predicted Impact on Chain Share and Price Coefficients

Variables	Expected Impact on: Baseline Market Share	National Brand Price Coefficient	Private label Price Coefficient
- *Private label [PL] Positioning Factors*			
Breadth of PL Program	+	–	+/–
Relative Price Gap of PL	+	+	–
PL Assortment Size	+	+	+
- *Chain- and Category-specific Control Factors*			
Chain Category Assortment Size	+	+	+
Chain Category Assortment Size Squared	–	–	–
Chain Category Price level	+/–	+	+
Chain Category Promotion Intensity	+	–	-
- *Category-specific Control Factors**			
Category Sales		–	–
Category PL Development		+	–
Category Number of Brands		+/–	+/–

*: Category-specific factors are constant across chains and therefore cannot influence baseline share.

weekly transaction data from four main grocery chains in a US city area (one of the 50 IRI markets) spanned over 261 weeks from October 2001 to October 2006.[2] The sales of the four chains cover around 90% of the total sales of all grocery chains in the data set in this area. Among the 30 product categories on which the data set provides information, there are 21 categories for which the four retailers have continuous non-zero private label sales. For the remaining 9 categories[3], retailers either do not sell private labels or they witness many continuous zero sales for their private labels. We label these 9 categories as "PL-weak" categories. The data of the 9 "PL-weak" categories will only be used for calculating the breadth of a chain's private-label program. The other 21 categories, used to calibrate our hierarchical market share model, are: blades, coffee, cold cereal, diapers, facial tissue, frozen dinners/entrees, frozen pizza, household cleaners, hot dog, margarine/butter, mayonnaise, milk, mustard & ketchup, paper towels, peanut butter, salty snacks, shampoo, spaghetti/Italian sauce, sugar substitutes, toilet tissue, and toothbrushes.[4] The market shares are aggregated to the chain level. For each category, we sum the store weekly volume sales to get the chain weekly volume sales, which are then used to obtain chain weekly market shares. We use volume sales instead of dollar sales to calculate shares as we want to focus on explaining the market shares by the chains (private-label) positioning. Part of the positioning is the choice of the overall price level. As the price positioning is used as an explanatory variable in the second-level model, we choose to use a share measure that is independent of prices.

3.4.2 Hierarchical market share model

Our hierarchical model consists of two layers. The first layer is a market share attraction model describing the retailers' market shares in a specific category (see Fok

[2]We have chains 1, 2, 3, 4. But due to some unique characteristics of chain 4, part of the data of chain 4 are showing up under another chain name, for example chain 5 for some weeks. We sum the sales of chain 4 and chain 5 to get the "real" sales of chain 4 and take the larger assortment size among the two chains as the "real" assortment of chain 4. We get reasonable data series from this aggregation for most categories except for blades, diapers, shampoo, and toothbrush. These four categories witness a sudden surge in sales and assortment size during the weeks when data show up under chain 5. To control the effect of the data aliasing we include a dummy variable for all the categories. The dummy equals one for those weeks where part of the data of chain 4 shows up under chain 5, and zero elsewhere.

[3]These categories are beer, carbonated beverages, cigarettes, deodorant, laundry detergent, razors, soup, toothpaste, and yogurt.

[4]Among the 21 categories, sugar substitutes, household cleaners, and shampoo have zero private label sales entries during the initial period of the data set. This may be caused by private label entry before the start of the data set. To have a "clean" data set with mature private labels, we removed the first few observations with zero private label sales. For sugar substitutes, we removed the first 40 observations; for household cleaners and shampoo we removed the first 10.

et al., 2002 for a review and full discussion of the market share attraction model). The second layer associates chain baseline market shares and price-sensitivities across all categories and chains with private-label positioning variables and some other chain- and/or category-specific characteristics.

First layer: market share attraction model For each product category we specify a market share attraction model. Let $MS_{i,c,t}$ denote the market share of chain i ($i = 1, ..., I,$) in category c ($c = 1, ..., C,$) in week t, where I is the total number of chains in the market and C gives the number of categories. This market share is assumed to equal the attraction of the chain in the category in week t ($A_{i,c,t}$) divided by the sum of the attraction of all chains in the same category in week t, that is,

$$MS_{i,c,t} = \frac{A_{i,c,t}}{\sum_{j=1}^{I} A_{j,c,t}} \qquad \text{for } i = 1, ..., \ I. \tag{3.1}$$

The attraction in turn is associated with marketing-mix variables, lagged marketing-mix variables, and lagged market shares. We use the following variables: F denotes feature, measured by the percentage of volume sales associated with a product on feature; D denotes display, measured by the percentage of volume sales that are on display; NBP and PLP denote national-brand price and private-label price, respectively, measured by dollar sales divided by volume sales; and ST denotes the number of stores open for a particular chain in a particular week. The ST variable is included to control for the store availability of a chain. The attraction specification reads as

$$\begin{aligned} A_{i,c,t} = \exp &\left(\mu_{i,c} + \sum_{j=1}^{I} (\beta_{3,j,i,c} F_{j,c,t} + \beta_{4,j,i,c} D_{j,c,t}) + \varepsilon_{i,c,t} \right) \\ &\cdot \prod_{j=1}^{I} \left(NBP_{j,c,t}^{\beta_{1,j,i,c}} PLP_{j,c,t}^{\beta_{2,j,i,c}} ST_{j,t}^{\gamma_{j,i,c}} NBP_{j,c,t-1}^{\alpha_{1,j,i,c}} PLP_{j,c,t-1}^{\alpha_{2,j,i,c}} MS_{j,c,t-1}^{\alpha_{3,j,i,c}} \right), \end{aligned} \tag{3.2}$$

where the error term $(\varepsilon_{1,c,t}, ..., \varepsilon_{I,c,t})'$ follows a normal distribution with mean zero and $I \times I$ covariance matrix Λ_c. The constant $\mu_{i,c}$ represents the baseline attraction of chain i in category c, which is the attraction level of a chain corrected for all the explanatory variables in (3.2). The attractions are monotonically related to baseline market shares. A large $\mu_{i,c}$ means a large baseline attraction and consequently a high baseline market share. Hereafter we will use the term "baseline market share" to refer to $\mu_{i,c}$. The $\beta_{k,j,i,c}$ ($k = 1, 2, 3, 4$) parameters represent the effect of marketing instrument k of chain j on the attraction of chain i, $\gamma_{j,i,c}$ represents the effect of number of stores of chain j, and the $\alpha_{p,j,i,c}$ ($p = 1, 2, 3$) parameters represent the effect of lagged variables.

To estimate the parameters, we take one chain as the benchmark and linearize the model. By taking chain I as the base chain and calculating the difference between the natural logarithm of $MS_{i,c,t}$ and $MS_{I,c,t}$, we get

$$\begin{aligned}\log MS_{i,c,t} - \log MS_{I,c,t} = \mu_{i,c} - \mu_{I,c} \\ + \sum_{j=1}^{I}(\tilde{\beta}_{3,j,i,c}F_{j,c,t} + \tilde{\beta}_{4,j,i,c}D_{j,c,t} + \tilde{\beta}_{1,j,i,c}\log NBP_{j,c,t} \\ + \tilde{\beta}_{2,j,i,c}\log PLP_{j,c,t} + \tilde{\gamma}_{j,i,c}\log ST_{j,c,t} + \tilde{\alpha}_{1,j,i,c}\log NBP_{j,c,t-1} \\ + \tilde{\alpha}_{2,j,i,c}\log PLP_{j,c,t-1} + \tilde{\alpha}_{3,j,i,c}\log MS_{j,c,t-1}) + \eta_{i,c,t}\ ,\end{aligned} \tag{3.3}$$

for $i = 1, ..., I-1$. The linearization shows that only the parameter differences $\tilde{\mu}_{i,c} = \mu_{i,c} - \mu_{I,c}$, $\tilde{\beta}_{k,j,i,c} = \beta_{k,j,i,c} - \beta_{k,j,I,c}$, $\tilde{\gamma}_{j,i,c} = \gamma_{j,i,c} - \gamma_{j,I,c}$, $\tilde{\alpha}_{p,j,i,c} = \alpha_{p,j,i,c} - \alpha_{p,j,I,c}$ are identified. The error term in (3.3) becomes $\eta_{i,c,t} = \varepsilon_{i,c,t} - \varepsilon_{I,c,t}$. Given the assumption on the errors in the attraction specification, $(\eta_{i,c,t}, ..., \eta_{I,c,t})'$ follows a normal distribution with mean zero and $(I-1) \times (I-1)$ covariance matrix Σ_c.[5]

Model selection The market share attraction knows many restricted forms, and many more extensive specifications. For example, more lags of prices and sales may be used or lags of the other variables may also be considered. To arrive at the attraction specification in (3.2) we started from a "full model" with two period lags of feature, display, price, and market share. Next, Likelihood Ratio test are performed for each category on the number of lags for each model component. Combining the test results from all the categories, we obtain the specification in (3.2), that is no lag of feature or display, one lag of price, and one lag of market share.

Next, we further test the models for other restrictions. Well-known restrictions are: restricted competition, where the marketing instrument of chain i is assumed to not influence chain j, and restricted effects, where on top of the restricted competition assumption the impact of a marketing instrument on the attraction is assumed to be constant across chains (see Fok et al. 2002). We tested these assumptions for the

[5] Given no restrictions on Λ_c, the covariance of $(\varepsilon_{1,c,t}, ..., \varepsilon_{I,c,t})'$, there are no restrictions on Σ_c either. The relationship between Σ_c and Λ_c is given by $\Sigma_c = L\Lambda_c L'$, where $L = (\mathbf{I}_{I-1} \vdots -\mathrm{i}_{I-1})$ with $\mathbf{I}_{I-1}$ an identity matrix and i_{I-1} a unity vector.

price components using F-tests. It turns out that for most categories we cannot reject the null hypotheses that a chain's attraction is only influenced by its own national-brand and private-label price[6]. Note that although the price competition effect on attraction is restricted, a chain's market share is still influenced by all other chain's prices because the denominator in (3.1) is a sum of all chain's attractions. Given this assumption, we obtain as the final first-level model[7]

$$\begin{aligned}\log MS_{i,c,t} - \log MS_{I,c,t} = \mu_{i,c} - \mu_{I,c} & \\ & + \sum_{j=1}^{I}(\tilde{\beta}_{3,j,i,c} F_{j,c,t} + \tilde{\beta}_{4,j,i,c} D_{j,c,t} \\ & + \tilde{\gamma}_{j,i,c} \log ST_{j,c,t} + \tilde{\alpha}_{3,j,i,c} \log MS_{j,c,t-1}) \\ & + \beta_{1,i,c} \log NBP_{i,c,t} - \beta_{1,I,c} \log NBP_{I,c,t} \\ & + \beta_{2,i,c} \log PLP_{i,c,t} - \beta_{2,I,c} \log PLP_{I,c,t} \\ & + \alpha_{1,i,c} \log NBP_{i,c,t-1} - \alpha_{1,I,c} \log NBP_{I,c,t-1} \\ & + \alpha_{2,i,c} \log PLP_{i,c,t-1} - \alpha_{2,I,c} \log PLP_{I,c,t-1} + \eta_{i,c,t}\ .\end{aligned} \tag{3.4}$$

Note that we can now identify all the parameters that relate to the price effects. The baseline market share parameters $\mu_{i,c}$ are however still not identified separately.

Second layer: explaining baseline market share and market share price sensitivity In the second layer of the model we relate the baseline market shares and the price effects to a set of moderating variables. The second layer consists of

[6]Among all the 21 categories, only 1 category, namely peanut butter, rejects the restricted competition assumption on national-brand price; 7, namely frozen pizza, household cleaners, milk, peanut butter, spaghetti/Italian sauce, sugar substitutes, and toothbrushes reject the restricted competition assumption on private-label price ; 7, namely blades, diapers, hot dog, mayonnaise, mustard & ketchup, peanut butter, and toothbrushes, reject the competition restriction on lag national-brand price; 2, namely peanut butter and toilet tissue, reject the competition restriction on lag private-label price. There is only one category, peanut butter, that rejects all the restrictions on national-brand price, private-label price, lag national-brand price, and lag private-label price.

[7]Due to the earlier mentioned data problem with chain 4, we include the dummy variable Dum_t mentioned in footnote 2 in equation (3.4) when we estimate the model. Dum_t equals to one for those weeks when part of the data of chain 4 shows up under chain 5. Dum_t is not shown in the equation.

three linear equations:

$$\mu_{i,c} = X'_{\mu,i,c}\theta_\mu + \xi_{\mu,i,c} \quad (3.5)$$

$$\beta_{1,i,c} = X'_{\beta,i,c}\theta_1 + \xi_{1,i,c} \quad (3.6)$$

$$\beta_{2,i,c} = X'_{\beta,i,c}\theta_2 + \xi_{2,i,c} \quad (3.7)$$

for $i = 1, 2, \ldots, I$ and where $(\xi_{\mu,i,c},\ \xi_{1,i,c},\ \xi_{2,i,c})'$ is normally distributed with mean zero and a 3×3 covariance matrix Ω. As listed in Table 3.1, we explain the baseline market share of chain i in category c ($\mu_{i,c}$) by three private-label positioning factors (Breadth of program, Relative price gap of private label, and Private label assortment size) and three chain and category control factors (Chain assortment size, General price level, and Promotion intensity).

In the market share attraction model only the differences $\tilde{\mu}_{i,c} = \mu_{i,c} - \mu_{I,c}$ are identified. However, (3.5) specifies a model for $\mu_{i,c}$. Implicitly, (3.5) implies that $\tilde{\mu}_{i,c} = (X_{\mu,i,c} - X_{\mu,I,c})\theta_\mu + \xi_{\mu,i,c} - \xi_{\mu,I,c}$. This shows that θ_μ is identified as long as the matrix formed by the differences of $X_{\mu,i,c} - X_{\mu,I,c}$ has full rank. This implies that for the baseline market shares we should not include a constant in (3.5). The advantage of the specification in (3.5) is that the parameters have a straightforward interpretation and that we obtain a natural correlation structure for the random shocks in $\mu_{i,c}$.

summarising, $X_{\mu,i,c}$ consists of six explanatory variables with no intercept. The explanatory variables for national-brand price sensitivities ($\beta_{1,i,c}$) and private-label price sensitivities ($\beta_{2,i,c}$) are the same, denoted by $X_{\beta,i,c}$, and include the six explanatory variables in $X_{\mu,i,c}$, a constant and three category control variables (Category size, Category private label development, and Number of brands).

In a related paper, Datta et al. (2017) use market share model to study sales-based brand equities and raise endogeneity issue as they allow the intercepts to change over time. They control for endogeneity by including quarterly dummies and Gaussian copulas. In our market share model, we include lagged marketing mix variables and lagged market shares of the focal retailer and all the competitors to account for the effects of possible strategic marketing mix adjustments over time. Our baseline market shares and effects of marketing-mix variables are all time invariant, therefore the endogeneity issue of Datta et al. (2017) is not present here.

3.4.3 Variables in the second level of the model

In this subsection we give the details of the definitions of all explanatory variables in the second layer of the model. All these variables are standardized, that is, they have mean zero and unit variance, such that we can easily compare the effect sizes.

The **Breadth of a private-label program** is measured by the number of categories in which a retailer carries a private label. All four chains carry private labels in the 21 categories we use to estimate our model parameters. In the 9 "PL-weak" categories there are cases where a retailer introduced a private label in a category within the time span of the data set or even cancelled the private label some weeks after the introduction. For each chain each category, we calculate the ratio of number of weeks with positive private label sales and total number of weeks. This ratio equals 1 for the 21 categories we use to estimate our model and is smaller than 1 for the 9 "PL-weak" categories. Then we sum the ratios across all the 30 categories for each chain.

The **relative price gap of a private label** is obtained by comparing the prices. For the private label of chain i in category c we use

$$PLPriceGap_{ic} = \frac{NBP_{ic} - PLP_{ic}}{NBP_{ic}}, \tag{3.8}$$

where the private-label price of chain i in category c (PLP_{ic}) is the average price over time defined by

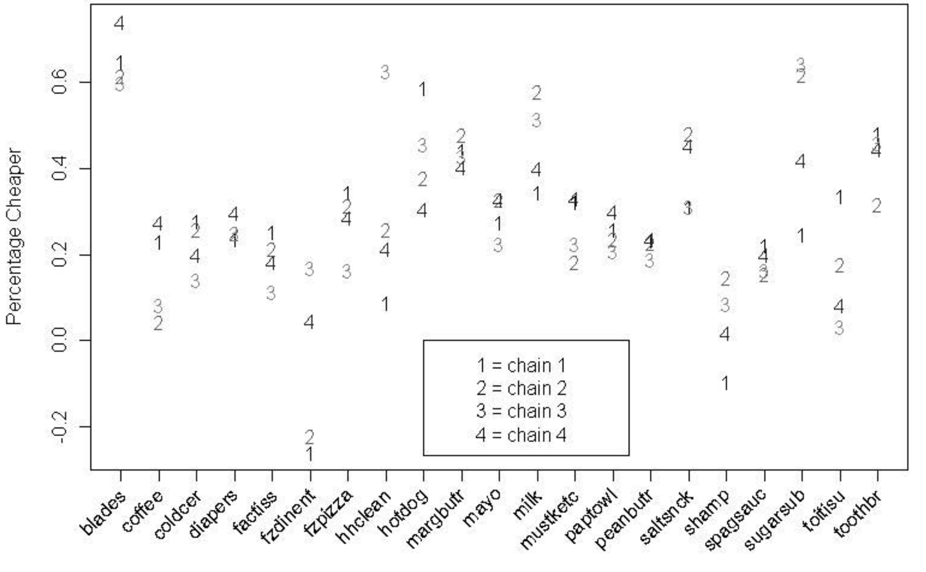
$$PLP_{ic} = \frac{1}{T}\sum_{t=1}^{T} \frac{\$salesPL_{ict}}{volsalesPL_{ict}}. \tag{3.9}$$

A similar calculation is used for the average national-brand price (NBP_{ic}). In Figure 3.1 we present a graphical summary of this variable. There is a large variance in the level of private label price gap across the 21 categories. This variance largely comes from category differences. We therefore subtract category average from $PLPriceGap$ to get a chain's private-label positioning relative to the average category level, that is,

$$RPLPriceGap_{ic} = PLPriceGap_{ic} - \frac{1}{I}\sum_{i=1}^{I} PLPriceGap_{ic}. \tag{3.10}$$

The **private-label assortment positioning** is quantified by the average number of UPCs (Universal Product Code) a private label has in a category over time (referred to as $PLassor_{i,c}$). Figure 3.2 shows the assortment size across categories and chains. To disentangle the category difference from chain positioning, we com-

Figure 3.1: Relative price gap of private label – How much cheaper is the private label relative to the national brand

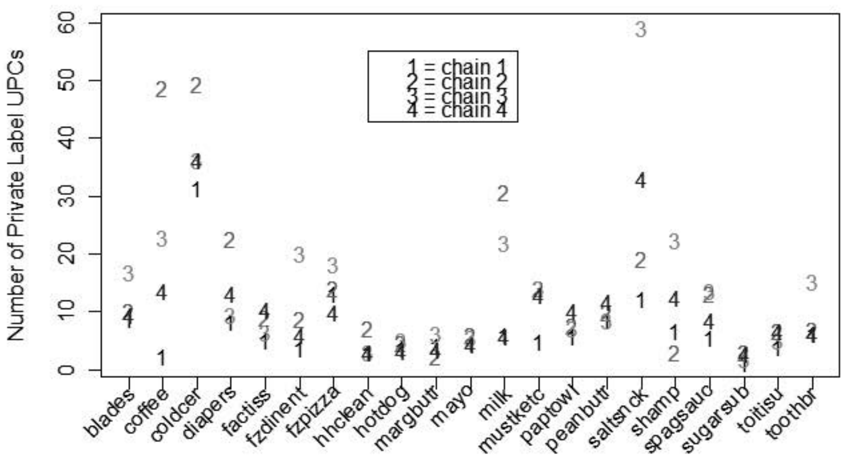


pute

$$RPLassor_{ic} = \frac{PLassor_{ic}}{\frac{1}{I}\Sigma_{i=1}^{I} PLassor_{ic}} - 1\,. \tag{3.11}$$

We measure the private label assortment positioning by the percentage that a chain's private label's assortment is larger or smaller than the average category level.

Figure 3.2: private label Assortment Size



To obtain the **chain/category assortment size** we first calculate the average number of UPCs that chain i has in category c over time: $assor_{ic}$. Next we compare $assor_{ic}$ to the category average value over all retailers and measure the relative percentages to get the chain's assortment positioning: $Rassor_{ic}$.

The **chain/category price positioning** is operationalized as follows. The price of chain i in category c, P_{ic}, is measured by the average dollar sales divided by average volume sales. Next we calculate the percentage a chain's price in a category is higher or lower than the average category level to get the chain's price positioning: rP_{ic}.

The original data contains an indicator for price promotions. The indicator equals one when the price cut is larger than 5%. We use this indicator to calculate the percentage of volume sales that are sold on promotion and use it as a proxy for **chain/category promotion intensity**. Again we subtract the average category level to get a relative measure.

Finally, the **category size** is measured by the sum of the average weekly dollar sales over all chains in a category. The **category private label development** variable is measured by the average of the weekly private label volume shares over all chains in a category. To count the **number of brands** we use the UPC descriptors at L5 level in the original data. This variable equals the average number of brands sold in a category.

3.4.4 Estimation

A Bayesian approach is applied to simultaneously estimate the parameters in the two layers of the model, that is, all the parameters in (3.4), (3.5), (3.6), and (3.7). We use the Gibbs sampler (Geman and Geman, 1984) with data augmentation (Tanner and Wong, 1987) to perform the Markov Chain Monte Carlo (MCMC) simulation and to obtain posterior results. We use uninformative priors for all parameters. Starting values for the chain are based on the results of a frequentist two-step estimation (that is first estimate the market share models for all categories, then stack all $\mu_{i,c}$, $\beta_{1,i,c}$, $\beta_{2,i,c}$ for $i = 1, ..., I$ and $c = 1, ..., C$, respectively and estimate three linear models). Details of the Bayesian estimation procedure are presented in Appendix B.

3.5 Empirical results

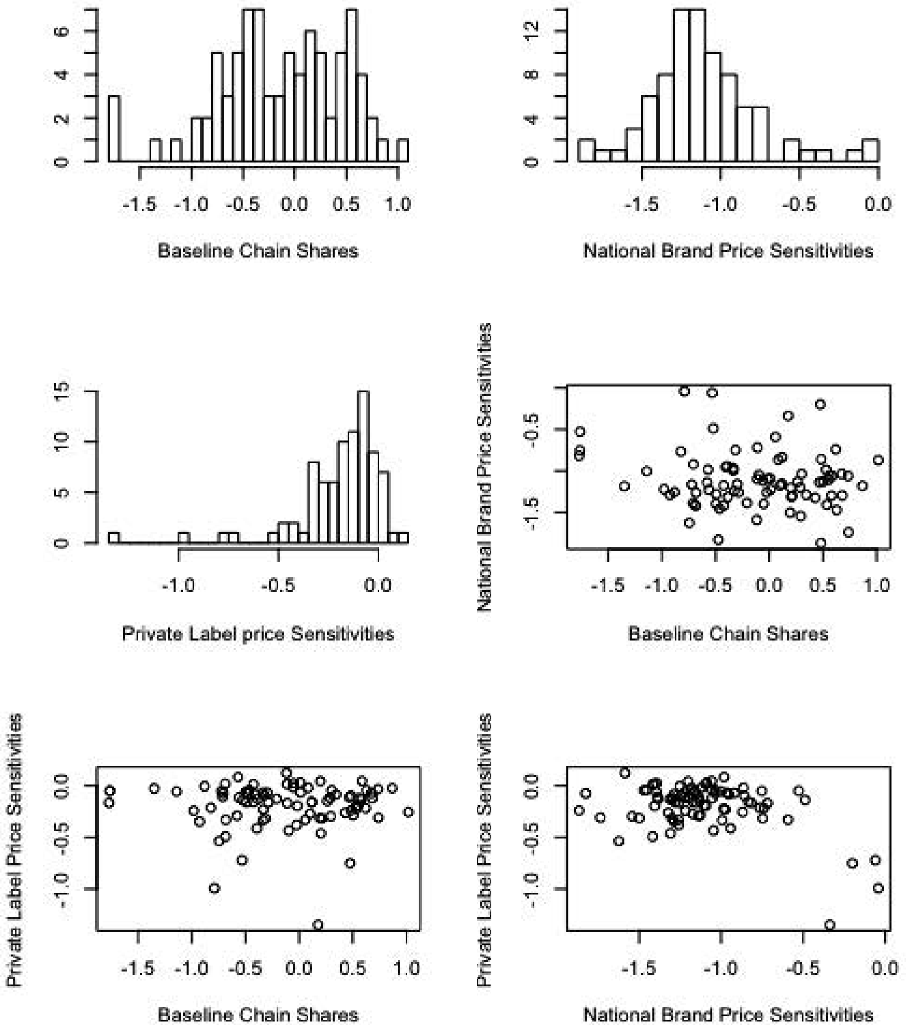
We obtained 150,000 draws from the Markov chain and discarded the first 50,000 draws as a burn-in period. Of the remaining draws we save each 20th draw to remove auto-correlation in the chain. Our posterior results are thus based on 5,000 draws. The simulation procedure is implemented in R. The Geweke test (Geweke, 1992) and the Heidelberger and Welch stationarity tests (Heidelberger and Welch, 1983) in the Bayesian output analysis (boa) package (Smith, 2007b) show that the Markov chain has converged. We have compared the results of this simultaneous estimation procedure with the two-step frequentist estimation. Our estimates are mostly in line with this conventional two-step approach, but are in general more efficient.The standard errors are smaller due to the simultaneous treatment of both levels of the model. As a result we have more significant national-brand price coefficients and private-label price coefficients. From the two-step approach, we get 78 significant national-brand price coefficients and 46 significant private-label price coefficients. From the Bayesian approach, we get 80 and 51 significant coefficients, respectively.

3.5.1 Results for first-layer: baseline market shares and price sensitivities

We summarise the estimates for the first-layer parameters in Figure 3.3. The distribution of the posterior means of the baseline market share is presented in the upper left plot in Figure 3.3. The plot shows the posterior means of the constants in the first-layer across all chains and categories ($\mu_{i,c}$). As expected the constants are all spread around zero. This matches equation (3.5) where the expected value of $\mu_{i,c}$ is restricted to be zero because a constant is not included in the matrix of explanatory variables. Note again that the mean of $\mu_{i,c}$ is not identified as only the differences of $\mu_{i,c}$ matter for market shares. Chains with a positive baseline chain share parameter will be more attractive than the other chains after controlling for price, marketing-mix variables, and number of stores. Our posterior results show that there are only a few significant non-zero constants, 8% of all the constants, 7 out of 84, are significant, that is, zero is not contained in the 95% HPD interval. More specifically, chain 4 has larger baseline shares in five categories: blades, diapers, mayonnaise, mustard & ketchup, and shampoo. The other three chains have constants close to zero in almost all categories, except for chain 1 in toilet tissue and chain 2 in salt snacks and mustard & ketchup. In these cases the baseline shares are significantly smaller. This implies that most market share differences across chains are explained by price and other marketing-mix instruments.

Figure 3.3 also shows the histograms of the posterior means for the national-brand and private-label price sensitivities, see the upper middle and upper right plot, respectively. As can be seen in the figure, almost all price effects are negative, which means market shares are negatively associated with both national-brand and private-label prices, just as expected. The dispersion in the national-brand price sensitivities is larger than that in the private-label price sensitivities. There are more significant national-brand price effects (80 out of 84 or 95%) than significant private-label price effects (51 out of 84 or 61%). The mean national-brand price effect (-1.11) is stronger than the mean private-label price effect (-0.19). Overall, the posterior probability that the national-brand price effect is stronger than that of the private label is 0.80. This implies that the market shares are more sensitive to national-brand prices than to private-label prices. In other words, price promotions in national brands are most effective in increasing the chain's market share.

Figure 3.3: Histograms and Scatter Plots



The bottom panel of Figure 3.3 shows scatter plots which present the possible correlations between the posterior means for the baseline shares and the two price sensitivities. As shown in the graphs, there is no obvious correlation between baseline market share and the two price effects (lower left and middle plots). There is a significant negative correlation (-0.39) between the two price effects (lower right plot), and this means that when a private-label price has a large impact on share, the national-brand prices tend to have a small influence on share. The source of this correlation will become clearer when we observe the results from the second-layer of the model.

3.5.2 Results from second-layer: private-label positioning effects and other moderating effects

The second layer of our model captures the effect of the positioning of the private label and other chain- and/or category-characteristics on baseline market shares and share price sensitivities. The posterior results are presented in Table 3.2, where the three columns give the moderating effects on the three dependent variables. As the moderators are standardized, we can see that the most important factor that affects baseline market share is chain/category assortment size, while the factor that influences price sensitivity the most is category private label share.

First of all, we did not find significant impact of private-label positioning factors on chain baseline market shares. Although for some categories, for example blades and diapers, the chain (chain 4) that offers the cheapest private-labels has larger baseline market shares, a statistically significant correlation between positioning and market share cannot be established. The private-label price position of chain 4 in blades and diapers may well contribute to their larger baseline market shares, but the relationship cannot be statistically generalised to other chains and other categories. The only factor that has significant impact on baseline market share is chain/category assortment size. Retailers who have a larger assortment than the average category level have a larger baseline share in the category. This is consistent with the finding of many studies, see for example Bayus and Putsis (1999) and Kahn (1998). Although the coefficient of the square of assortment size is negative, it is not significant.

The private-label positioning variables do have an impact on price sensitivities. First, the breadth of a private-label program has a negative influence on the private-label price sensitivity (see coefficient -0.040 in the last column of Table 3.2). This implies that the market share of a chain with a broad private-label program is actually more sensitive to its own private-label price than a chain with a narrow program.

Table 3.2: Posterior Means of the Moderating Effects on Chain Share and Price Coefficients

Moderators	Effects on Baseline Market Share	 National Brand Price Coefficient	 Private Label Price Coefficient
Intercept		-1.119 (0.044) ***	-0.190 (0.018) ***
- Private Label [PL] Positioning Factors			
Breadth of PL Program	-0.119 (0.113)	-0.057 (0.041)	-0.040 (0.018) **
Relative Price Gap of PL	-0.096 (0.097)	0.063 (0.034) *	0.016 (0.014)
PL Assortment Size	-0.020 (0.135)	0.098 (0.046) **	0.010 (0.019)
- Chain- and Category-specific Control Factors			
Chain/Category Assortment Size	0.370 (0.125) ***	-0.038 (0.046)	-0.002 (0.018)
Chain/Category Assortment Size squared	-0.143 (0.096)	0.008 (0.029)	-0.003 (.0.12)
Chain/Category Price level	-0.073 (0.098)	0.083 (0.036) **	0.052 (0.014) ***
Chain/Category Promotion Intensity	0.128 (0.105)	-0.031 (0.039)	-0.037 (0.018) **
- Category-specific Control Factors			
Category Sales		-0.069 (0.050)	0.043 (0.020) **
Category PL Development		0.266 (0.045) ***	-0.223 (0.018) ***
Category Number of Brands		0.100 (0.047) **	-0.033 (0.020) *
Posterior Mean of % explained	28.1%	58.0%	86.7%

Notes: Posterior standard deviations are in parentheses.
***, **, and * = zero not contained in 99%, 95%, and 90% HPD intervals, respectively.

By carrying private labels in categories where other retailers do not sell a private label, categories like for example beer, carbonated beverages, and deodorant, a retailer apparently makes its private label concept more salient to consumers. Given that private-label prices are usually lower than national-brand prices, a more salient private label concept emphasises the importance of price, and therefore makes a chain's share more sensitive to private-label price. On the other hand, a retailer facing more price-sensitive consumers may choose to sell private-labels in broader categories in the first place. So this negative correlation arises from a mutual causation.

Secondly, for the price positioning of the private label, we find that it has a positive effect (0.063, moderately significant, in the 3rd column of Table 3.2) on national-brand price sensitivity. This means that when the price gap between a private label and national brands is large, the impact of national-brand price changes on market shares tends to be smaller than when the price gap is small. A reason for this is that the low-price private label strategy increases the proportion of consumers in store who are price-sensitive. For example, suppose that one retailer's private label is 20% cheaper than its national brands, and that another retailer's private label is 40% cheaper. Even when there is a 20% discount, the latter national brands are still more expensive than the private label. So price-sensitive consumers may still stick to the private label. They are therefore less likely to be affected by national-brand price changes.

Third, we find that the larger the private label assortment size, the less sensitive the chain's share is to its own national-brand price (0.098). When there are more choices for the private label, it is more likely that consumers will find a certain private label item that fits their preferences. These consumers would be less sensitive to national-brand prices.

Besides the private-label positioning effects, we do have some other interesting findings on the chain- and category-specific control variables. First, a chain with high average category price level is less sensitive to both national-brand and private-label prices (see coefficients 0.083 and 0.052 in the 3rd and last column of Table 3.2). This is consistent with our expectation that a chain with high general price has less price sensitive consumers. Second, more promotions make the market share more sensitive to private-label price (-0.037 in the last column of Table 3.2), while they have no significant impact on sensitivity to national-brand prices.

In addition, we find some significant category-specific control variables. First, just as we expected, the market share in categories with high private label share is more sensitive to private-label price and less sensitive to national-brand price compared to the market share in categories with low private label share (see the coefficients 0.266 and -0.223 at the bottom panel of Table 3.2). This is the main source of the negative correlation between national-brand and private-label price effects reported in Section 3.5.1. In categories with high sales, market share is less sensitive to private-label price (0.043). Finally, in categories with more national brands, market share is less sensitive to national-brand price and more sensitive to private-label price, although the latter effect is just moderately significant (0.100 and -0.033). This implies that more national brands create differentiation among the national brands, while they bring more competition for private labels.

3.6 Conclusion and Discussion

Private labels play an important role in a retailer's strategy. The interaction between private labels and national brands has been investigated in many studies, while the impact of private labels on the competition between retailers at the aggregate level has not been studied before. We studied whether private labels can differentiate a retailer from its competitors and thus can contribute to chain equity. To this end we investigated the impact of retailer's positioning of private labels across categories on their category-specific market shares and market share price sensitivities via a hierarchical market share model.

The results of our analysis show that the first level of our market share model fits quite well. Most variations in the chain and category specific market shares can be explained by the explanatory variables. Besides, we find that the market shares are more likely to be affected by national-brand prices than by private-label prices. Moreover, the effect size of national-brand price is stronger than that of private-label price. So generally speaking price promotions in national brands are most effective in increasing the chain's market share.

In the hierarchical part of our model we link the baseline market shares and the price sensitivities to the retailer's private-label positioning strategy. Retailers can position their private label in terms of price, assortment size, and the number of categories that carry the private label. Across retailers and categories there tends to be a large variation in this positioning. We did not find significant impact of private-label positioning factors on chain baseline market shares. Although for some categories, for example blades and diapers, the chain (chain 4) that offers the cheapest (relative to national-brands in the chain) private-labels also has the largest baseline market shares in the two categories, a statistically significant correlation between positioning and market share cannot be established. The private-label price position of chain 4 in blades and diapers may well contribute to their larger baseline market shares in the markets, but the relationship cannot be statistically generalised to other chains and other categories in our dataset. Instead, the total category assortment size contributes to a retailer's baseline market share positively. So retailers could increase their own attractiveness by enlarging the overall category assortment size. For the marginal effect of assortment size, we find a negative but not significant effect of squared assortment size. Given our data limitations, whether there is a decreasing marginal effect of assortment therefore remains ambiguous. We have to leave the determination of the optimal assortment size to further research.

How a retailer positions its private label does have an impact on the price sensitivities. First, by carrying a private label in a broad number of categories, retailers actually emphasise the price concept. This makes their market shares more price sensitive. On the other hand, a retailer facing more price-sensitive consumers may choose to sell private-labels in broader categories in the first place. So this negative correlation arises from a mutual causation. Second, the cheaper the private labels compared to the national brands, the less sensitive the chain's market share is to national-brand prices. If a retailer sells a very cheap private label, we expect that its national-brand promotions become less effective in gaining market share. Third, the larger the private label assortment size, the less sensitive is the chain's share to its own national-brand prices. If a retailer has large private label assortment compared to its competitors, we expect that increasing national-brand price has less impact on market share, and that national-brand promotions become less effective in stealing away market share from competitors.

In conclusion of the private-label positioning effect on chain and category specific equity, a relatively lower priced private-label programme with larger assortment would be the most effective in weakening the market share price sensitivity to national brands, thus contribute to the chain differentiation.

Besides the private-label positioning effects, we also have some other interesting findings. First, a chain with high average category price level is less sensitive to both national-brand and private-label prices. Second, more promotions make the market share more sensitive to private-label price, while such promotions have no significant impact on the sensitivity to national-brand prices. Third, the market share in categories with high private label share is more sensitive to private-label price and less sensitive to national-brand price compared to the market share in categories with a low private label share. Fourth, in categories with high sales levels, market share is less sensitive to private-label price. Finally, in categories with more national brands, market share is less sensitive to national-brand price and more sensitive to private-label price, which implies that more national brands create differentiation across the national brands, while they bring more competition for private labels.

Our research has some limitations, which suggest several avenues for future research. First, some chain and category factors are not constant in the medium or long term. For example, the assortment size could vary over time or from season to season. Our modeling setup however requires us to treat them as constant during the five year period of our data set. The impact of changes in the assortment size on the competition between retailers is an interesting issue for the future. Second, our results are bounded to one U.S. grocery market. How the private-label positioning influences the competition among retailers can also be studied in European markets, where private labels seem to play a more important role. Third, one retailer could have multiple private-label brands across categories, usually a cheap label and a quality label. Due to our data limitations, we have no information about multiple private labels. If a retailer has multiple private labels, then what we have studied is the impact of the average positioning of the private labels. The roles that multiple private labels have in the competition between retailers remains another future research direction.

Chapter 4

Forecasting own brand sales: Does incorporating competition help?

4.1 Introduction

Forecasts of brand sales are relevant to both retailers and manufacturers. Forecasts give an impression of what future sales patterns can look like, and it helps to understand the competition between brands. This can facilitate the brand level organisation for the retailers and help the manufacturers to gear changes in the future marketing mix.

Brand sales forecasts are often generated from econometric time series models (Hanssens et al., 2003), where the well-known SCANPRO model (Wittink et al., 1988) is an illustrative example. Such models usually include past sales and own marketing activities (current and past), but frequently also variables concerning past competitor behaviour are included, at least if one knows this competition. Such variables can substantially improve the predictive performance. As retailers have the most complete information regarding to sales and promotions, in this chapter we take a retailer's point of view and address various ways to include information on competitors for the prediction of within-store brand sales.

Our key conjecture is that in practice it is often not known which brands are effectively the main competitive brands. One may then resort to a couple of strategies. One option is to simply ignore competition. This makes the model simple to analyse, as there is no need for the sometimes cumbersome collection and preparation of data from competitors. A second strategy is to spend effort in studying which are the most relevant competitive brands. Data can be obtained, for example, by interviewing consumers or by analysing cross-promotion information. The latter approach can be rather successful, see Moon et al. (2007), Blattberg and Wisniewski (1989b), Sethuraman et al. (1999), and Sethuraman and Srinivasan (2002), among others.

The third strategy, which we will address in the present chapter, is to consider all other possible brands as potential competitors that might be relevant for the forecasts of the own brand. This approach is relevant if we do not know beforehand which brands have predictive content, and in this case we can let the data help to decide on this each time we make a forecast.

Naturally, this third strategy challenges the usual regression based forecast methodology. Common categories in FMCG markets can easily involve more than ten brands. When a typical SCANPRO-based regression model includes current and past sales as well as current and past marketing mix, then that amount of brands leads to the inclusion of more than one hundred variables. In this chapter, we wish to address the question whether agnostically including all other brands and using modern data science technologies would lead to better forecasts of one's own brand sales.

A simple way to summarise competitor variables is to take weighted averages across all competitors, where the weights can be obtained from the brands' market shares. This method dramatically reduces the number of additional variables to be included in the forecast model. Other methods to exploit the rich information on competitors in a more refined way include dimension reduction methods like Principal Components Analysis (PCA), shrinkage methods like the Least Absolute Shrinkage and Selection Operator (Lasso) and elastic-net, and tree-based methods like random forest and boosting. We will use a range of these methods and compare them against a simple benchmark model that does not include any competitive information.

The dimension reduction idea to extract a small number of factors for use in prediction has been widely used in forecasting macro-economic time series like production and inflation, starting with the seminal work of Stock and Watson (1999; 2002). Our modeling strategy includes various specification options, including variable selection, variable grouping, the choice of the estimation window, the choice of the number of factors and of the lag structure in the sales model. The shrinkage methods shrink the estimated coefficients towards zero compared to the least squares estimates (James et al., 2013). Among different types of shrinkage methods the Lasso shrinks the coefficients of unimportant predictors exactly to zero and therefore performs variable selection. These shrinkage methods have recently gained popularity in forecasting sales (Ma et al., 2016; Sagaert et al., 2018) and macro-economic time series (Li and Chen, 2014; Medeiros and Vasconcelos, 2016; Smeekes and Wijler, 2018) due to their superior forecasting performances in a high-dimensional data environment.

Different from all linear (and log-linear) models, tree-based models segment data into groups using a decision tree format. Both random forest and the boosting tree method combine a large number of trees to generate usually more accurate forecasts than a single tree does. These tree-based methods are suitable to analysing complex non-linear relationships as they do not impose a particular structure on the data.

Our empirical test of the different methods concerns weekly data for brands in 31 categories, where we have a total of 169 brands. Our main conclusion, where we summarise across all cases and settings, is that although the own-brand-only benchmark model performs reasonably well, forecast accuracy can be improved for most brands using a certain way of including competitive information. Among all the methods we tested, the random forest method and the two shrinkage methods, that is, the Lasso and the elastic net, show the best forecast performance in terms of accuracy.

The outline of the chapter is as follows. In Section 4.2, we describe methods to include large amounts of competitor variables, where we only focus on those methods that have shown to be most reliable in the available literature. In Section 4.3, we discuss the data that we use for our illustrations and we provide the details of our empirical methodology. Section 4.4 contains the forecast results comparison and section 4.5 draws the conclusion and discusses some potential future research areas.

4.2 Methods and models

In this section we discuss various ways of forecasting sales. We start with a straightforward linear time series model for log sales that will serve as the benchmark. Next we introduce models that account for competitive effects in different ways. Each of these models includes competition by summarising or selecting from all the competing brands in some way.

To forecast the sales at time t one period ahead, that is, for time $t+1$, we use the available past information up to time t as well as the marketing efforts of all brands at time $t+1$. This forecasting situation is relevant for retailers, as in general they have full information on all (in-store) marketing efforts in the forecasted period.

In forecasting sales, both recursive expanding window and rolling window approach are popular. The former approach expands the estimation window period by period, when one more period of data is included, we re-run all the necessary procedures that result in the final out-of-sample forecast. The procedures vary across methods, may include variable selection, decomposition, determining the number of factors and the number of lags, choosing the tuning parameters, re-estimating the coefficients, and finally predicting with the updated coefficients. The rolling window approach, on the other hand, fixes the length of the estimation window, shifts the window forward by period, and re-run the forecasting steps each shift. The rolling window approach is more suitable for varying coefficient situations, while the expanding window is better for constant coefficient situations. We will implement both in our forecasting procedure and see which one is better for our extensive dataset.

4.2.1 Benchmark model

In the benchmark model we only use information of the focal brand.This model is an autoregressive model of order L with explanatory variables written as ARX(L) for the sales of the brand (in logarithms). As explanatory variables we take seasonal dummies and the own marketing efforts. There are 13 seasonal dummies, each covers four consecutive weeks. Denoting the sales of the focal brand at time t by s_t, the one-period-ahead forecasting model for sales after the natural log transformation is

$$\ln(s_{t+1}) = \alpha + M'_{t+1}\beta + D'_{t+1}\mu + \sum_{l=1}^{L} \gamma_l \ln s_{t+1-l} + \epsilon_{t+1}, \tag{4.1}$$

where M_{t+1} is a vector of marketing instruments and D_{t+1} is a vector of seasonal dummies, both for time $t+1$. In general M_{t+1} will contain the brand's price (in natural logarithms) and display and feature variables. By including lagged sales, the model captures dynamic effects like stockpiling and purchase inertia. The number of lags L is chosen by minimising the Bayesian information criterion (BIC).

4.2.2 Average competitor model (ACM)

One very parsimonious way to include information of all competing brands is to summarise these competing brands in a few variables. In the average competitor model, we summarise this information by taking a weighted average of all competitive marketing variables (denoted by $\bar{M}_t^c$). The weights are given by the current market shares in week t, which vary per week. The competitive sales are also summarised by taking their weighted average (denoted by $\bar{s}_t^c$). These average competitive variables capture possible cross effects on sales. The average competitor model with L lags for own sales and Q lags for averaged competitor sales reads as

$$\ln(s_{t+1}) = \alpha + M'_{t+1}\beta + D'_{t+1}\mu + \sum_{l=1}^{L} \gamma_l \ln(s_{t+1-l}) + \bar{M}^{c\prime}_{t+1}\beta^c + \sum_{q=1}^{Q} \gamma_q^c \ln(\bar{s}^c_{t+1-q}) + \epsilon_{t+1}. \tag{4.2}$$

4.2.3 Principal component regression

The average competitor model summarises the competition by taking a market-share weighted average. Another weighting scheme is obtained by principle component analysis (PCA), where the weights are chosen to maximise the retained variance and can be used to find the "optimal" linear combination(s) of competitive variables. The "optimal" here means that the constructed linear combinations explain as much variance as possible of the competitive variables using a less number of components - usually much less than the number of original predictors. The obtained principal components can be added to the benchmark model to obtain a model that accounts for competition. We label this model as the Principal Components Regression (PCR) model. The number of components can be set to a fixed number or chosen data-drivenly.

Forecasting by means of principal components has proven to be very effective in macroeconomics, see for example the review chapters in Stock and Watson (2006) and Stock and Watson (2012). In our setting, the competitor information is first summarised by a number of principal components, and then these components and their lags together with the variables in the benchmark model are used to forecast sales of the focal brand.

Let N_t be a K-dimensional vector of competitive variables consisting of the information that is available at time t on all competitors for the variables price (in natural logarithms), display, feature, and sales (in natural logarithms). In our retailing setting, when at time t we forecast $\ln(s_{t+1})$, the prices and promotion variables at $t+1$ are already known, whereas sales are of course not yet known for time $t+1$. Therefore, N_t consists of mixed information for time t (sales) and $t+1$ (price and promotion). If the start of the observation period is denoted by $t = 1$, the competitor information that is available at time t is collected in the $t \times K$ matrix $N = (\; N_1, \quad \ldots, \quad N_t)' \;$, where each column (variable) of N is standardized to have mean zero and variance one. The leading r principal components (with $r < k$) of this matrix are collected in the $t \times r$ matrix $F = (F_1, \ldots, F_t)'$, where F_t is the vector of r principal components at time t. The principle component regression (PCR) is now given by

$$\ln(s_{t+1}) = \alpha + M'_{t+1}\beta + D'_{t+1}\mu + \sum_{l=1}^{L} \gamma_l \ln(s_{t+1-l}) + \sum_{g=0}^{G} F'_{t-g}\lambda_g + \epsilon_{t+1}. \qquad (4.3)$$

In our retailing application, the competitive variables can be classified into four groups: prices, feature variables, display variables, and sales. To exploit this grouped variable structure, we can also perform PCA separately per group. Grouped PCA has the advantage of yielding meaningful factors like a competitive price factor and a competitive sales factor. These factors and their lags can then be used in the forecast equation (4.3), and we label this method as Grouped PCR.

4.2.4 Forecast-oriented factor construction

The PCA factors capture maximal variance of the predictor variables irrespective of the target variable that is to be forecasted. It may help to construct the factors in a way that reflects the final forecasting objective. Possible options are to select predictors according to their relation with the target variable or to derive factor weights from the correlation of the predictors with the target variable.

We consider two variable selection methods, known as "hard" and "soft" thresholding as introduced by Bai and Ng (2008), to which we refer for technical details. These selection methods provide a ranking of the predictor variables according to their importance in predicting the target variable. The principal components are then constructed from the subset of variables that are found to be most important in this sense. The resulting (subset) principal components are then used in the forecast equation (4.3).

4.2.4.1 Hard thresholding

In hard thresholding, the importance of the k-th predictor variable $N_{k,t}$ is assessed by means of its t-value in the following regression equation, which controls for the predictors of the benchmark model

$$\ln(s_{t+1}) = \alpha + M'_{t+1}\beta + D'_{t+1}\mu + \sum_{l=1}^{L} \gamma_l \ln(s_{t+1-l}) + N_{k,t}\beta_k^c + \epsilon_{t+1}, \tag{4.4}$$

where $N_{k,t}$ denotes the information at time t of the k-th predictor variable, that is, the sales in period t or the price or promotion activity in period $t+1$ of a specific competitor. This regression is applied for each of the variables $k = 1, \ldots, K$ separately, and the variables are ranked according to their absolute t-value. A potential drawback is that this method may select similar variables, as each predictor is evaluated separately without considering the other predictors. While this disadvantage does not exist in the soft thresholding method, which we introduce next.

4.2.4.2 Soft thresholding

The soft thresholding method selects variables sequentially such that each next variable adds most information for the target variable after controlling for the previously selected variables. The soft thresholding method (Bai and Ng, 2008) selects variables using least-angle regression (LARS) (Efron et al., 2004) and determines the optimal number of variables using an information criterion, such as the BIC. After the subset $\tilde{M}^c$ is selected, the forecasting procedure is exactly the same as above, that is r factors are extracted using PCA and these factors are used in (4.3).

LARS is a forward stepwise regression proposed by Efron et al. (2004). At the first step it selects the variable that correlates most with the target variable. Next the coefficient of the first selected variable is set to zero. Starting from zero, the coefficient of the variable is moved towards its least squares value. This way the correlation between the variable and the residual moves towards zero. When the correlation between the variable and the residual equals the correlation between the residual and a second variable, this second variable is added to the "active set". In this second step, the coefficients of the two variables are moved together in a way that their correlations with the evolving residual are tied and moved towards zero. As soon as a third variable "catches up" in terms of the correlation with the residual, the third variable enters the set. The whole process stops when all the variables are included in the model and at this point we get the common least squares fit. We use LARS on the residuals of the benchmark equation (4.1) of regressing the focal brand sales on its own marketing instruments and own lagged sales. This leads to a ranking of all the competitor variables and then the actual number of selected variables is determined by the BIC.

4.2.5 Shrinkage methods

Different from information summarising methods, shrinkage methods fit a model that contains all the J predictor variables but impose a penalty term onto the least squares estimation to constrain the size of the coefficient estimates, therefore shrink the estimated values towards zero. The shrinkage penalty term can take different forms, the most commonly used are L1 norm (the Lasso), L2 norm (ridge regression), and a linear combination of both L1 and L2 norm (the elastic net). As the ridge estimation shrinks all the coefficients towards zero proportionally, it will not exclude any predictor. We would like to select important predicting variables especially from all the competitors and therefore we will use the Lasso and the elastic net methods in the chapter.

The Lasso regression proposed by Tibshirani (1996) has become very popular over the last decades for the purpose of dealing with high dimensional data, that is the number of variables is relatively large to the number of observations. By adding L1 norm penalty to the sum of squared errors, the Lasso shrinks the coefficients of unimportant variables exactly to zero, and therefore performs variable selection. Our regression model now can be written concisely as

$$\ln\left(s_{t+1}\right) = \alpha + \sum_{j=1}^{J} \beta_j x_{jt} + \epsilon_{t+1}, \tag{4.5}$$

where x_{jt}, $j = 1, ..., J$ represents all the J candidate predictors from the forecasted brand and all the competing brands, together with the seasonal dummies. The information available at time t includes own and competitors' sales at t and their lags, own and competitors' price and promotion variables at $t+1$, and seasonal dummies at $t+1$.

The penalised least squares estimate is obtained by minimising

$$\sum_{t=1}^{T}\left(\ln s_{t+1} - \alpha - \sum_{j=1}^{J} \beta_j x_{jt}\right)^2 + \lambda \sum_{j=1}^{J} \mid \beta_j \mid . \tag{4.6}$$

Here the tuning parameter λ is to adjust how strong the penalty is. A very large λ will generate all zero estimates. The parameter λ is usually chosen via cross-validation, which is a data driven method that tries to minimise out-of-sample squared prediction errors.

The Lasso is useful to identify which competitor(s) and what type of promotion(s) are important in forecasting the focal brand sales. However, when there are highly correlated predictors, the Lasso will randomly choose one from the correlated group. The elastic net solves this problem by adding L2 penalty to (4.6).

Zou and Hastie (2005) proposed the elastic net regression that minimises

$$\sum_{t=1}^{T}\left(\ln s_{t+1}-\alpha-\sum_{j=1}^{J}\beta_j x_{jt}\right)^2+\lambda\sum_{j=1}^{J}\left(\gamma\mid\beta_j\mid+(1-\gamma)\beta_j^2\right). \tag{4.7}$$

This is a compromise of the Lasso and the ridge regression and can encourage grouping of highly correlated predictors. Both tuning parameters λ and γ here can be done via cross-validation. The details of how to do cross-validation can be found in James et al. (2013).

4.2.6 Tree-based models

Unlike traditional regression analysis, tree-based methods impose no specific structure on the data. Instead, they use a series of splitting rules to form decision trees. In the end the data will be partitioned into different groups and for every observation that falls in the same group, the group mean or mode of the dependent variable would be its predicted value. Regression tree analysis uses a procedure called recursive binary splitting to construct the rules. Starting from the top of the tree, that is, all the observations are in one group, the method considers all the predictors and all the possible values of each predictor as the cutpoint to split the data into two groups. The selected split is the one that minimises the residual sum of squares (RSS). The same process repeats for each of the two partitioned groups to split the data further. The binary splitting continues until a pre-specified stopping criterion is satisfied. The final groups produced are called terminal nodes. We refer to Breiman et al. (1984) for more details.

Regression trees are easy to interpret as it mimics human decision-making process. In addition, it is good at picking up the nonlinearity in the data and potential complex interactions between independent variables. For instance, the effect of own price on sales may be not (log) linear but there is a cut point in price, below which it will lead to a boom in sales. According to James et al. (2013), one major problem with trees is their instability due to the high variance, that is, small changes in the data can produce very different splits, therefore the output tree can be highly divergent or unstable. So their prediction accuracy is typically not as good as traditional regression methods. However methods like bagging, random forest and boosting generate multiple trees and aggregate the outcomes to make predictions. These methods reduce the variance significantly and hence improve accuracy.

4.2.6.1 Random forests

The random forest method is proposed by Breiman (2001) on the basis of bagging (Breiman, 1996). Bagging first bootstraps (sample with replacement) from data and generates multiple bootstrapped samples, then fits a tree to each of the samples, and finally averages the predictions of all the decision trees. Random forest is different from bagging on the part of building trees. Instead of including all the predictors for growing each tree, the random forest method randomly chooses a subset from all the predictors for consideration at each split in a tree. This way it can avoid generating very similar trees as does the bagging method and consequently reduces the correlation between them. Averaging many uncorrelated trees reduces the variance further compared to averaging many correlated ones (Hastie et al., 2001). For the detailed procedure of random forest see Appendix C.1.

4.2.6.2 Boosting

Boosting is a method that can be applied to many statistical methods, here we use it on decision trees. Instead of fitting one large tree to the data set, boosting grows many small trees sequentially in a way that each tree is grown using the residuals from a previous tree. Then the new decision tree is multiplied with a shrinkage parameter and then added into the fitted function to update the residuals. This approach allows us to gradually fit the data and slowly improve the fitted function in areas where it does not fit well. The procedure is described more specifically in Appendix C.2.

Different from bagging and random forest, which construct relatively big trees for each bootstrap sample, boosting fits a small tree each time. This means that the number of splits of each tree takes a small value. According to James et al. (2013) one to four splits often work well. There are three meta-parameters to choose here, and these are the shrinkage parameter, the number of splits, and the total number of trees. We will run trial regressions to choose among the different combinations.

A summary of all the models is presented in Table 4.1. Since PCR, Hard thresholding, and Soft thresholding use two ways to determine the number of factors included, there are actually 13 methods in total that we consider for our data.

Table 4.1: Methods of incorporating competitor information

Name	Description	Equation	Meta-parameters
Benchmark model	An autoregressive model with own marketing mix variables	(1)	Number of own lagged sales L
Average competitor model (ACM)	Adding competitor variables to the benchmark model by taking a market-share weighted average of all competitors' sales and marketing variables	(2)	Number of own and competitors lagged sals L, Q
Principal component regression (PCR)	Summarising all competitive variables by a number of factors and adding the factors and their lags to the benchmark model	(3)	Number of own lagged sales L, number of factors, number of lagged factors G
Grouped PCR	Summarising the four competitive variable groups (price, feature, display, and sales) with one factor for each group and adding the factors and their lags to the benchmark model	(3)	Number of own lagged sales L, number of lagged factors G
Hard thresholding	Constructing factors on a subset of variables that are selected by the hard thresholding rule	(4) & (3)	Number of own lagged sales L, number of factors, number of lagged factors G
Soft thresholding	Constructing factors on a subset of variables that are selected by the soft thresholding rule	(4) & (3)	Number of own lagged sales L, number of factors, number of lagged factors G
Lasso	Least squares with a penalty of L1 norm	(5) & (6)	Tuning parameter λ
Elastic net	Least squares with a penalty of a linear combination of L1 and L2 norm	(5) & (7)	Tuning parameters λ and α
Random forest	Averaging tree models grown on multiple bootstrap samples, at each split of a tree, considering a random subset of all the predictors	(8)	Number of trees
Boosting tree	Growing trees sequentially in a way that each tree is grown using the residuals from a previous tree, then averaging over all the trees	(9)	Shrinkage parameter λ, number of splits in each tree, number of trees

4.3 Data

We apply the forecasting methods in Section 4.2 on data of one store gathered from the Information Resources Inc. (IRI) data set (Bronnenberg et al., 2008), which contains information of 31 consumer product categories[1] spanning over five years. The store we use is randomly chosen from medium-sized ones, having five years data, having at least two brands in each category, and having sufficient variation in price. Some of the IRI categories are defined in a broad way and thus contain different types of products. We split these broad categories into narrower ones and for each category we aggregate stock keeping units (SKUs) data to the brand level by weights. For a more detailed explanation of how the data is compiled we refer to Horváth and Fok (2013). Finally, we have 31 product categories and 169 brands and the number of brands in each category varies from 2 to 10.

To choose between the recursive expanding window and the rolling window approach in forecasting, we estimate the benchmark model, the ACM and all the factor models using both methods and compare the forecasting accuracies. For the recursive window estimation we started from using data over the first 108 weeks, among which the first 4 weeks are used to obtain the lags of sales up to 4 periods thus not used in the parameter estimation. The parameters are obtained from data over 104 weeks[2] and are then used to forecast the sales of one week ahead, that is week 109. Then the real data of week 109 are included in the sample and we re-estimate the model. The new parameters are in turn used to forecast the sales of week 110. This window expanding re-estimation process continues till the sales of the last week of the five year period is forecasted. For the rolling window forecast we estimate the models each time over a window of 104 weeks and forecast one week ahead. All the forecasted natural logarithms of sales are added by one-half of the mean squared prediction error and then exponentiated back into sales units (Ma et al., 2016; Cooper et al., 1999).

Due to the fact that there are some brands whose display and feature variables

[1] The categories are: beer, blades, carbonated beverages, cigarettes, coffee, cold cereal, deodorant, diapers, facial tissue, frozen dinners/entrees, frozen pizza, household cleaners, hot dog, laundry detergent, margarine/butter, mayonnaise, milk, mustard & ketchup, paper towels, peanut butter, razors, salty snacks, shampoo, soup, spaghetti/Italian sauce, sugar substitutes, toilet tissue, toothbrushes, toothpaste, and yogurt. We exclude category photo from the original data set.

[2] To choose the window width we tried the width of 1.5 years, 2 years, and 2.5 years in rolling window forecasts for all the 169 bands across 31 categories. The results show that there are much more brands performing better with 2 year window width than with 1.5 year window width (103 brands out of 169). While the number of brands performing better with 2.5 year window width compared with 2 year window width is not significant (87 out of 169). Therefore we choose 2 years as our rolling window width and as our starting window width of the recursive forecasting.

may not have enough variation in an estimation window, this may cause a problem for least squares based method including the shrinkage methods. We set a threshold of at least five distinct values for a display/feature variable to be included in an estimation. But this is not necessary for tree-based methods. So for the random forest and boosting methods, we include all the display/feature variables for all the estimation windows. Having few or even no variation will not influence the estimation process.

4.4 Forecasting procedure and accuracy evaluation

4.4.1 Forecasting accuracy evaluation

Forecasting accuracy is measured by Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and Mean Absolute Percentage Error (MAPE), defined as

$$RMSE = \sqrt{\frac{1}{T}\sum_t (y_t - \hat{y}_t)^2}, \tag{4.8}$$

$$MAE = \frac{1}{T}\sum_t |y_t - \hat{y_t}|, \tag{4.9}$$

and

$$MAPE = \frac{1}{T}\sum_t \frac{|y_t - \hat{y}_t|}{y_t}, \tag{4.10}$$

where T is the number of the out-of-sample forecasts. A smaller RMSE, MAE, or MAPE means a more accurate forecast. Here RMSE and MAE are used to select models. MAPE is useful to compare cross categories and understand how much the forecasts relatively deviate from the real values as it is not scale dependent.

To compare how much better or worse a competing model is relative to the benchmark , we use relative MAE (RelMAE)

$$RelMAE = \frac{MAE^{comp}}{MAE^{bench}}, \tag{4.11}$$

which measures the MAE of the competing model relative to the benchmark model, a value smaller than 1 means an improvement of the forecast accuracy.

To compare all the models on their overall performance, Davydenko and Fildes (2013) proposed a measure called Average Relative MAE (ARMAE), which can be obtained from

$$ARMAE^{comp} = \Big(\prod_{b=1}^{N} \frac{MAE_b^{comp}}{MAE_b^{bench}}\Big)^{\frac{1}{N}}. \tag{4.12}$$

This measures the MAE of the competing model relative to the benchmark model for each brand b and then takes a geometric mean over all the N brands. Similar to RelMAE, an ARMAE value smaller than 1 means an improvement over the benchmark model.

4.4.2 Forecasting procedure and selection of meta-parameters

All our models involve lagged predictors, for model (1) it is lagged own sales, for model (2) it is lagged own sales and lagged average competitor sales, and for model (3) it is lagged own sales and lagged factors. We include up to four lags for all the lagged variables and the actual number of lags is determined by the BIC for each model. Note that model (2) and (3) both have more than one lagged variable, and in this case the optimal number of lags of different variables within the same model can be different. Shrinkage methods and tree-based models will select useful predictors from all the candidates no matter whether they are lagged variables or not. So there is no need to use a information criterion for them.

For all the models that involve factors, except for Grouped PCR, we tried two ways of determining the number of factors, one is fixing the number at two and another is including up to four factors, and then choosing the number with best predictive performance. In the latter data-driven method, we forecast for models with either 1, 2, 3, or 4 factors respectively and then choose the number of factors resulting in the smallest RMSE over the most recent 26 weeks (half year) within each estimation window. Next, once the number of factors is decided, the model with the optimal number of factor lags (chosen by BIC) is used for one-week-ahead out-of-sample prediction.

The implementation of the shrinkage methods involves choosing optimal tuning parameters. For both Lasso regression and elastic net, the tuning parameter λ is determined by five-fold cross-validation. Firstly, the data in the estimation window is randomly divided into five parts. Then we leave one part out and obtain results for a grid of 100 λ-values using the remaining four parts. Next we forecast over the left out part. This process continues until we obtain the "out-of-sample" forecasts for all the five parts. Then the RMSE over the whole estimation window is calculated and the λ that results in the smallest RMSE is chosen. Finally the chosen λ is used in fitting on the whole estimation window and forecasting one week ahead out of the estimation window. Strictly speaking, cross-validation on time series data is not correct because the temporal dependencies are interrupted when the data is randomly divided into parts. But it is still widely used as a heuristic, for example see Li and Chen (2014) and Ma et al. (2016). The tuning parameters γ in elastic net regression can be chosen data-drivenly as well, but to limit the number of models needed to be considered, we set $\gamma = 0.5$.

We implemented random forest using the randomForests package in R. As the number of bootstrapped samples B will not cause overfitting, we choose a value of B that is large enough for the prediction error to settle down. Our trial estimation results from 31 brands, one from each category, show that the errors have settled down when B is larger than 100. In our final estimation we set B equal to 500.

We applied boosting tree model using the gbm package in R. As the number of trees Bneeds to be large enough for the errors to converge but not too large to overfit, we run a trial regression on one beer brand. We try four different numbers of split, and for each number of split, two different shrinkage parameters, so in total eight different combinations. An optimal combination in terms of convergence speed and out-of-sample RMSE is then chosen to for forecasting sales of all the brands. The details of this process can be found in Appendix C.3.

4.5 Results

For the benchmark model, the ACM, and all the factor models, we perform both expanding window and rolling window estimation and compare the forecast accuracies (measured with RMSE). The results from 169 brands show that the expanding window procedure is better for the majority of the brands. The number of brands for which expanding window estimation performs better varies from 107 (the benchmark model) to 132 (the ACM). This may due to the fact that in general there is no structural break in our data span. Hence increasing the number of observations improves the accuracy. All the results presented hereafter are therefore from expanding window estimations.

Since MAPE is not scale dependent, the MAPE values can be compared across categories and brands, which will give us an idea about how accurate the forecasts from different models are. Table 4.2 presents the comparison results with the left panel showing the mean, median, and the standard deviation of the MAPE values, the right panel showing the number of brands whose MAPEs are smaller than 0.1, 0.2, and 0.3 respectively. It can be seen that the random forest achieves the best result in terms of lowest mean, median, and standard deviation and the highest number of brands whose MAPE is smaller than 20% and 30%. Our benchmark forecasts made without competitor information have MAPE smaller than 30% for 135 (that is about 80%) of the brands. The median and the mean of the MAPEs are 18.9% and 24.1% respectively. The boxplot of the MAPEs from all the models are shown in Figure 4.1. It can been seen that the errors of random forest are the least dispersed. We also find that the outliers of all the models highly (but not completely) overlap, which means that some brand sales are just more difficult to forecast than others.

Next we compare the RMSEs and MAEs of all methods for each brand and present the results in Table 4.3. The first panel of the table shows the number of brands each method performs the best in terms of achieving the lowest RMSE or MAE among all the models. The second panel presents the number of brands for which each method performs better than the benchmark. Finally the third panel shows the time cost in terms of minutes used to estimate the model and to forecast all the 169 brands. As can be seen in the table, the benchmark model only excels in 7 or 6 cases, which means that for all the other 162/163 brands, forecast accuracy can be improved by incorporating competition in some way. From the first panel we can see that the random forest model takes the lead with the most number of brands showing the best performance, 33 measured with RMSE and 45 with MAE. The elastic net comes the second with 23 (RMSE) and 27 (MAE) best performances. The Hard thresholding with optimal number of factors follows behind with 20 (RMSE) and 17 (MAE) best performances. However if we compare all the models with the benchmark, then the elastic net takes the lead with improved forecast accuracy for 117 (RMSE) and 130 (MAE) brands. The Lasso is nearly as good as the elastic net. This implies that in general, the Lasso and elastic net can improve forecasts over the benchmark for majority of the brands. The random forest method does not show such general good performance, however for some of the brands where that the other models do not perform well, random forest can improve forecast accuracy substantially. It can also be seen in the table that for models involving factors, the method using an optimal number of factors in general performs better than the one using fixed number of factors. It is also worth mentioning that the ACM delivers quite good forecasts given how straightforward the idea (just averaging the competitors variables) behind it is.

From the results we conclude that the Lasso, elastic net and the random forest are all good options to incorporate competition information to improve forecast accuracy. The former two shrinkage methods perform better than the benchmark for most of

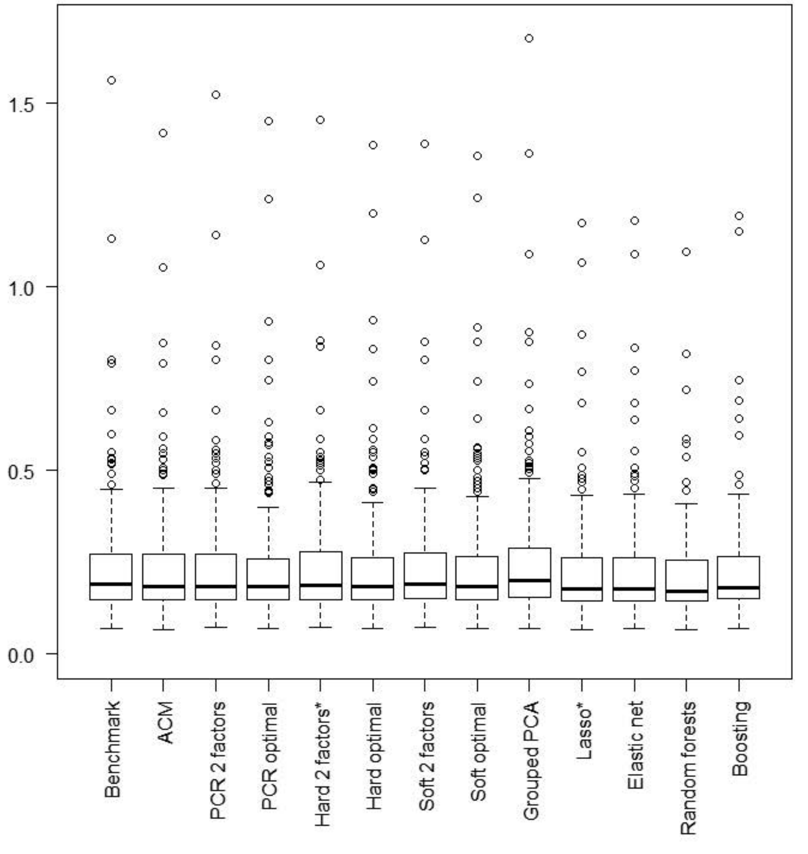
Table 4.2: Mean Absolute Prediction Error comparison of all the models

	MAPE			Number of brands whose MAPE		
Models	Mean	Median	SD	$< 10\%$	$< 20\%$	$< 30\%$
Benchmark	0.241	0.189	0.177	5	90	135
ACM	0.239	0.183	0.171	**7**	95	134
PCR 2 factors	0.241	0.185	0.176	6	94	134
PCR optimal	0.242	0.184	0.183	6	98	137
Hard 2 factors	0.674*	0.187	5.612*	5	94	132
Hard optimal	0.241	0.184	0.180	6	98	137
Soft 2 factors	0.253	0.189	0.226	4	90	132
Soft optimal	0.243	0.183	0.180	3	92	138
Grouped PCR	0.260	0.200	0.207	4	86	130
Lasso	0.259	0.178	0.447	5	101	140
Elastic net	0.229	0.176	0.160	5	100	140
Random Forest	**0.215**	**0.170**	**0.131**	4	**105**	**147**
Boosting	0.228	0.181	0.152	4	102	141

Note: The figures printed in bold show the best results in the column.

* An outlier, brand “mayo3”, has a MAPE value of 73.17, which leads to the high mean and standard deviation of the method.

Figure 4.1: Boxplot of Mean Absolute Percentage Errors (MAPEs)



* The Hard 2 factors model and the Lasso have one outlier each that is not shown in the plot, both are from brand "mayo3". The MAPE values are 73.17 and 5.67 respectively.

the brands. The computation burden, measured in minutes used, of the shrinkage methods is comparable to that of the factor models, however their performances are better than the latter. As to the random forest , when it captures the underlying non-linearity and interactions right, it can provide the most accurate forecasts. But the computation time is much longer than the shrinkage models. Nevertheless, nowadays parallel computing is easily available so the computation burden is not necessarily a problem.

To compare the three best performing models even further, that is, the Lasso, elastic net, and the random forest, we present the boxplot of their RelMAEs in Figure 4.2. Despite that the Lasso has one outlier "mayo3"[3], together with elastic net, it shows improvement (values smaller than 1) for the most brands, although the magnitude of the improvements are less compare to the random forest model, the deterioration is less as well. It means that the two shrinkage methods are the safest method to incorporate all the competition variables for forecasting sales here, although they are not always the best.

We have 31 product categories and each category has different characteristics and competition environments. To look into the model performances within each category, we calculate the ARMAE (average relative MAE, see (4.12)) of the best three competing models relative to the benchmark and present the results in Table 4.4. The lowest ARMAE of each row is printed in bold and a value smaller than 1 means that the method is better than the benchmark on average in the category. It can be seen that the random forest performs the best in most of the categories and particularly well in *cigarettes, mayonaise*, and *razors,* where the improvements over the benchmark are much larger than those of the Lasso and elastic net. While in categories like *beer, cold cereal, hotdog, paper towel, soup*, and *toilet tissue*, the random forest performs no better than the benchmark model.

From all the results shown in Tables 4.2, 4.3 and 4.4 we can say that the Lasso and elastic net outperform all the competing models for most brands. If one wants to choose a method that can utilise the competitors information to improve sales forecasting performance, then the two shrinkage methods are the best choice. These methods are particularly good at picking up useful predictors from a large amount of variables. For this reason, they can identify what types of promotion from which competing brand affect the focal brand's sales the most. However, for products that

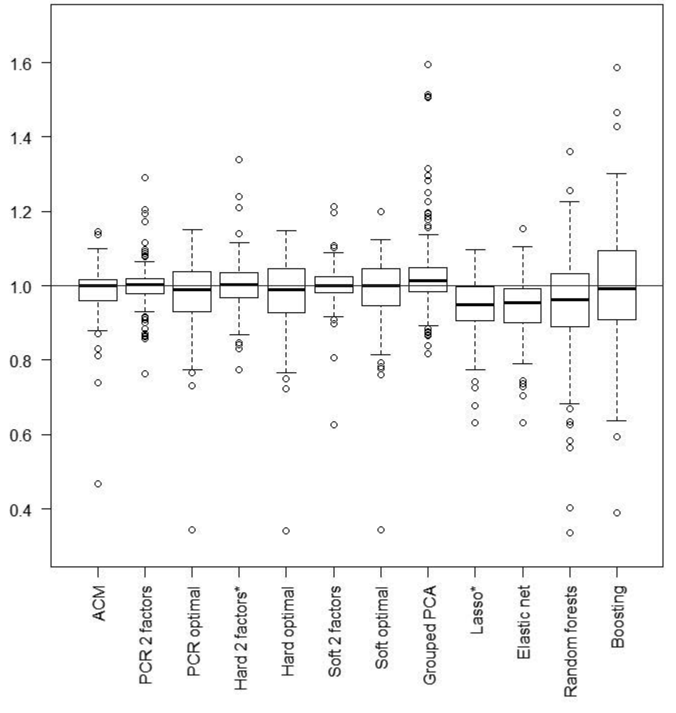
[3]The outlier value comes from an extremely high sales forecast at the early stage of the forecasting process. Even when we exclude the variables with less than five distinct values, there are still cases when a predictor's value is within a certain range in the estimation window, but the value of the predictor at the prediction week is far out of the this range. This situation can be ruled out by increasing the length of the estimation window.

Table 4.3: Number of brands each method achieving lowest or lower than the benchmark RMSE/MAE

Model	Number of brands of each model performs BEST among all the models		Number of brands of each model performs BETTER than benchmark		Total time cost in minutes
	RMSE	MAE	RMSE	MAE	
Benchmark	7	6	-	-	**5**
ACM	12	12	97	88	16
PCR 2 factors	8	4	79	80	20
PCR optimal	18	17	94	92	61
Hard 2 factors	7	4	82	81	28
Hard optimal	20	17	93	91	67
Soft 2 factors	6	3	88	83	52
Soft optimal	9	9	82	84	92
Grouped PCR	1	0	69	60	18
Lasso	17	13	115	**130**	30
Elastic net	23	27	**117**	**130**	30
Random Forest	**33**	**45**	89	108	276
Boosting	8	12	79	91	2651
Total	169	169	169	169	-

Note: The figures printed in bold show the best results in the column.

Figure 4.2: Boxplot of Relative Mean Absolute Errors (MAEs)



* The Hard 2 factors and the Lasso both have one outlier "mayo3" with value of 316.45 and 3.20 respectively, which are not shown in the plot.

do not promote much, for example razors, as we only include variables that have at least 5 distinct values[4], the valuable promotion information is excluded from the model at the early stage of forecasting. This information enters the model only when it has accumulated enough variation. However tree models do not require variables to have enough variation so they can include this type of information from the very beginning to contribute to forecasting. This is probably why the random forest model shows the least dispersed errors. On top of that, the random forest model mimics the process of aggregating much of people's decision making processes. For example, if a brand is on price promotion, some people would buy it straight away, some would like to see whether another brand they like is on promotion and some may also consider whether and how much they have already bought in the previous weeks. So the random forest can capture some underlying nonlinearities that linear models can hardly detect.

4.6 Conclusion and discussion

This study investigated whether incorporating competing brands information helps to improve sales forecasts. If yes, then how much value is added to the benchmark forecast model that excludes competitor variables? Furthermore, we compared different techniques of including competitors' sales and marketing activities.

The Average Competitor Model summarises all competitive information by weighted averages; Principal Component Regression (PCR) model summarises all competitive information by a number of factors and uses these factors and their lagged terms to forecast sales. The methods involve Hard and Soft thresholding constructing factors on a subset of variables that are selected by the hard and soft thresholding rule respectively. The method called Grouped PCR summarises the four variable groups, namely price, feature, display, and sales, with one factor for each group and uses these factors and their lags for forecasting. Among these methods, PCR model, Hard thresholding and Soft thresholding require choosing the number of factors. For these three models we experiment with a fixed number of factors (2 factors) and with choosing an optimal number based on past forecasting performance. We perform one week out-of-sample forecasts on 169 brands from 31 product categories. The results show that the data-driven method outperformed the fixed number of factors for all three models.

[4]As we aggregate the data to brand level, so the feature and display variables are measured in percentage of volume sold that is on promotion instead of dummies.

The Lasso and elastic net put a constraint to the coefficients such that the most influential predictors will be selected. Random forest and the boosting method are based on tree models, which are suitable to fit complex nonlinearity in the data and are like human decision making processes.

Our forecasting results show that the benchmark model can be improved in its forecast accuracy, measured with RMSE, MAE, MAPE, and RelMAE, by incorporating competitive information for 162 brands out of the total 169. Among different alternative models the Lasso, elastic net and random forest show the best forecasting performance. For most brands, the Lasso and the elastic net are better than the benchmark and they are a safe choice in terms of forecast accuracy. Random forest on the other hand can improve the forecast accuracy substantially for some of the brands, especially when the other methods do not perform well. One possible explanation is that for those brands which are not often promoted, the two shrinkage methods and all the other linear models do not use the very infrequent promotion information well. On the other hand, the random forest method performs really well for these type of brands.

One could imagine if the shrinkage methods can incorporate infrequent promotion information wisely, then their forecast accuracy will improve considerately. A possible way to solve this is to include these variables regardless of their limited variation, but set an upper bound for insanely high forecasted sales, for example ten times previous mean sales. However the choosing of the bound can be tricky. Another way is to form a committee of the three best performing models: the Lasso, elastic net, and random forest, and find a way to take use of the advantages of each of them. We will leave these issues for future research.

Table 4.4: The models forecasting accuracy ARMAE in categories

	Lasso	Elastic net	Random forest
beer	**0.994**	0.989	1.117
blades	1.025	1.021	**0.973**
carbbev	0.945	0.938	**0.932**
carbbevSelect2	1.008	0.977	**0.968**
cigets	0.914	0.921	**0.795**
coffee	**0.875**	0.880	0.947
coldcer	0.976	**0.974**	1.004
deod	**0.938**	0.942	0.973
diapers	0.911	0.916	**0.849**
factiss	0.924	0.931	**0.885**
fzdinent	0.842	0.837	**0.797**
fzdinentSelect2	0.918	0.908	**0.885**
fzpizza	0.945	0.947	**0.917**
hhclean	0.958	0.961	**0.899**
hotdog	1.006	**1.002**	1.111
laundet	0.915	**0.906**	0.949
margbutrSelect2	**0.931**	0.935	0.952
mayo	2.959	1.504	**0.674**
mustketc	**0.888**	0.895	0.926
mustketcSelect2	0.942	0.936	**0.879**
paptowl	0.994	**0.993**	1.076
peanbutr	0.957	0.954	**0.895**
razors	0.777	0.772	**0.618**
saltsnck	0.979	**0.970**	0.994
shamp	0.929	0.929	**0.899**
soup	0.957	**0.955**	1.035
spagsauc	**0.955**	0.965	0.971
sugarsub	**0.935**	0.937	0.968
toitisu	0.964	**0.961**	1.019
toothbr	0.938	0.942	**0.920**
toothpa	**0.929**	0.932	**0.929**
Overall	0.962	0.949	**0.941**
Number of categories a model has best performance on average	8	7	17

Appendix A

Bayes Estimation of the Hierarchical Bayes Ordered Probit model

In this appendix we use the following notation:

- J is the number of brands in each category,
- C is the total number of categories,
- T_{jc} is the number of observations (after discarding the first two observations as we use two lags in the OP model) for the individual OP model of brand j in category c,
- P_{jc} is the number of explanatory variables (including the intercept) given in (2.1) (first layer of the HB-OP model),
- L is the number of explanatory variables (including the intercept) given in (2.3) and (2.4) (second layer of the HB-OP model).

To analyse the HB-OP model, we need to rewrite it in a more compact form. We first consider brand j in category c and rewrite (2.1) in matrix notation as

$$\begin{aligned} y^*_{jc} &= X_{jc}\zeta_{jc} + \varepsilon_{jc} && \text{(A.1)} \\ &= \begin{pmatrix} X_{0jc} & X_{1jc} & X_{2jc} \end{pmatrix} \begin{pmatrix} \alpha_{jc} \\ \beta_{1jc} \\ \beta_{2jc} \end{pmatrix} + \varepsilon_{jc} \quad , && \text{(A.2)} \end{aligned}$$

where $\varepsilon_{jc} \sim N(0, I)$, I is a T_{jc} dimensional identity matrix, y^*_{jc} is a T_{jc} dimensional vector. $X_{jc} \equiv (X_{0jc}, X_{1jc}, X_{2jc})$ is a T_{jc} by P_{jc} matrix with one column of ones and $P_{jc} - 1$ columns of explanatory variables. X_{0jc}, X_{1jc}, and X_{2jc} are three matrices containing the explanatory variables corresponding to the three coefficient vectors α_{jc}, β_{1jc}, and β_{2jc}. $\zeta_{jc} \equiv (\alpha'_{jc},\ \beta'_{1jc},\ \beta'_{2jc})'$ is a P_{jc} dimensional coefficient vector, α_{jc} is a vector containing coefficients α_{0jc}, α_{1jc}, α_{2jc}, λ_{1jc}, λ_{2jc}, ν_{2jc}, ..., ν_{13jc}, μ_{1jc}, ..., μ_{10jc} in (2.1), β_{1jc} and β_{2jc} are two $J-1$ dimensional vectors containing the competitive response coefficients β_{1ijc} and β_{2ijc}, $i = 1, \ldots J$, $(i \neq j)$.

The model in (2.2) can be rewritten as

$$y_{jct} = \begin{cases} 0 & \text{if} -\infty < y^*_{jct} \leq 0 \\ 1 & \text{if}\, 0 < y^*_{jct} \leq \gamma_{2jc} \\ 2 & \text{if}\, \gamma_{2jc} < y^*_{jct} \leq +\infty \end{cases}, \tag{A.3}$$

for $t = 1, \ldots, T_{jc}$ and where y^*_{jct} is an element of vector y^*_{jc}.

Finally, (2.3) and (2.4) can be rewritten as

$$\text{vec}\Big(\ \beta_{1jc}, \quad \beta_{2jc}\ \Big) \sim N\left(\text{vec}\Big(\ Z_{jc}\theta, \quad Z_{jc}\delta\ \Big),\ \Sigma \otimes I_{J-1}\right), \tag{A.4}$$

where Z_{jc} is a $J-1$ by L matrix containing an intercept and the $L-1$ explanatory variables given in (2.3) and (2.4), θ and δ are two L-dimensional coefficient vectors, Σ is a 2-dimensional covariance matrix, and I_{J-1} is a $J-1$ dimensional identity matrix.

The complete data likelihood function $p(y_{jct},\ y^*_{jct},\ \beta_{1jc},\ \beta_{2jc}|\,\alpha_{jc},\ \gamma_{2jc},\ \theta,\ \delta,\ \Sigma)$ is proportional to

$$\begin{aligned}
&\prod_{c=1}^{C}\prod_{j=1}^{J}\Bigg\{\exp\Big(-\frac{1}{2}(y^*_{jc} - X_{jc}\zeta_{jc})'(y^*_{jc} - X_{jc}\zeta_{jc}) \times \\
&\prod_{t=1}^{T_{jc}}\Big(I(y^*_{jct} \leq 0)I(y_{jct} = 0) + I(0 < y^*_{jct} \leq \gamma_{2jc})I(y_{jct} = 1) + I(y^*_{jct} > \gamma_{2jc})I(y_{jct} = 2)\Big) \times \\
&\exp\left(-\frac{1}{2}\text{vec}\left((\beta_{1jc} - Z_{jc}\theta),\ (\beta_{2jc} - Z_{jc}\delta)\right)'\left(\Sigma \otimes I_{J-1}\right)^{-1}\text{vec}\left((\beta_{1jc} - Z_{jc}\theta),\ (\beta_{2jc} - Z_{jc}\delta)\right)\right) \\
&|\Sigma \otimes I_{J-1}|^{-1/2}\Bigg\},
\end{aligned} \tag{A.5}$$

where $I(\cdot)$ is an indicator function, which equals to one if the condition within the brackets is true, otherwise zero. We use uninformative priors for the model parameters, except for imposing an inverted Wishart prior on the Σ parameter.

- *Sampling* y^*_{jct}

The full conditional posterior distribution of y^*_{jct} is a truncated normal distribution, where the truncation points are conditional on the value of y_{jct}. The distribution can be written as

$$p(y^*_{jct}|\, y_{jct},\, \zeta_{jc},\, \gamma_{2jc},\, \theta,\, \delta,\, \sigma_1^2,\, \sigma_2^2) \propto \begin{cases} N(x'_{jct}\zeta_{jc},\, 1)I(y^*_{jct} \leq 0) & \text{if } y_{jct} = 0 \\ N(x'_{jct}\zeta_{jc},\, 1)I(0 < y^*_{jct} \leq \gamma_{2jc}) & \text{if } y_{jct} = 1 \\ N(x'_{jct}\zeta_{jc},\, 1)I(y^*_{jct} > \gamma_{2jc}) & \text{if } y_{jct} = 2 \end{cases}, \tag{A.6}$$

where x'_{jct} is the t-th row of X_{jc}. To sample from the truncated normal distribution efficiently we use the inverse cumulative distribution function (CDF) technique. For example, to simulate y^*_{jct} from $N(x'_{jct}\zeta_{jc},\, 1)$ truncated by $(\gamma_{1jc},\, \gamma_{2jc}]$, we first generate u from $U[0,\, 1]$, then solve y^*_{jct} from

$$u = F(y^*_{jct}) = \frac{\Phi(y^*_{jct} - x'_{jct}\zeta_{jc}) - \Phi(\gamma_{1jc} - x'_{jct}\zeta_{jc})}{\Phi(\gamma_{2jc} - x'_{jct}\zeta_{jc}) - \Phi(\gamma_{1jc} - x'_{jct}\zeta_{jc})}\,, \tag{A.7}$$

where $F(y^*_{jct})$ is the CDF of y^*_{jct}, and get

$$y^*_{jct} = F^{-1}(u) = x'_{jct}\zeta_{jc} + \Phi^{-1}\left(u\left(\Phi(\gamma_{2jc} - x'_{jct}\zeta_{jc}) - \Phi(\gamma_{1jc} - x'_{jct}\zeta_{jc})\right) + \Phi(\gamma_{1jc} - x'_{jct}\zeta_{jc})\right). \tag{A.8}$$

- *Sampling* ζ_{jc}

We define $A = P^{-1}$, such that $PP' = \Sigma \otimes I$ and P is a lower diagonal matrix. We then left multiply $\text{vec}\begin{pmatrix} \beta_{1jc}, & \beta_{2jc} \end{pmatrix}$ with matrix A such that

$$A \cdot \text{vec}\begin{pmatrix} \beta_{1jc}, & \beta_{2jc} \end{pmatrix} \sim N\left(A \cdot \text{vec}\begin{pmatrix} Z_{jc}\theta, & Z_{jc}\delta \end{pmatrix},\, I\right). \tag{A.9}$$

So (A.2) together with (A.9) can be rewritten as

$$\begin{pmatrix} y^*_{jc} \\ A \cdot \text{vec}\begin{pmatrix} Z_{jc}\theta, & Z_{jc}\delta \end{pmatrix} \end{pmatrix} = \begin{pmatrix} X_{0jc} & X_{1jc}\ X_{2jc} \\ 0 & A \end{pmatrix} \begin{pmatrix} \alpha_{jc} \\ \beta_{1jc} \\ \beta_{2jc} \end{pmatrix} + \varepsilon^*_{jc}\,, \tag{A.10}$$

where $\varepsilon^*_{jc} \sim N(0,\, I)$. If we set

$$V_{jc} = \begin{pmatrix} y^*_{jc} \\ A \cdot \text{vec}\begin{pmatrix} Z_{jc}\theta, & Z_{jc}\delta \end{pmatrix} \end{pmatrix} \text{ and } W_{jc} = \begin{pmatrix} X_{0jc} & X_{1jc}\ X_{2jc} \\ 0 & A \end{pmatrix} \tag{A.11}$$

then we get

$$V_{jc} = W_{jc}\zeta_{jc} + \varepsilon^*_{jc}\,. \tag{A.12}$$

Thus it can be easily be seen that the full conditional posterior distribution of ζ_{jc}equals

$$p(\zeta_{jc}|y^*_{jct},\, y_{jct},\, \gamma_{2jc},\, \theta,\, \delta,\, \sigma_1^2,\, \sigma_2^2) \propto N\Big((W'_{jc}W_{jc})^{-1}W'_{jc}V_{jc},\, (W'_{jc}W_{jc})^{-1}\Big). \tag{A.13}$$

- *Sampling* γ_{2jc}

From (A.3) we can see that γ_{2jc} is greater than all y^*_{jct} whose corresponding y_{jct}value equals 1 and smaller than all y^*_{jct} whose corresponding y_{jct} equals 2 for all t. Thus the full conditional posterior distribution of γ_{2jc} follows a uniform distribution with lower bound $\max_t(y^*_{jct}|\, y_{jct} = 1)$, and upper bound $\min_t(y^*_{jct}|\, y_{jct} = 2)$.

- *Sampling* θ *and* δ

We stack all $J \cdot C$ elements of β_{1jc} and β_{2jc}, $(j = 1, \ldots, J,\ c = 1, \ldots, C)$ and rewrite (A.4) into matrix notation as

$$\begin{pmatrix} \beta_1 & \beta_2 \end{pmatrix} = Z \begin{pmatrix} \theta & \delta \end{pmatrix} + \begin{pmatrix} \eta_1 & \eta_2 \end{pmatrix}. \tag{A.14}$$

Thus the full conditional posterior distribution of $vec\,(\theta,\ \delta)$

$$p\left(\text{vec}\,(\theta,\ \delta)\,|y^*_{jct},\ y_{jct},\ \gamma_{2jc},\ \zeta_{jc},\ \sigma_1^2,\ \sigma_2^2\right) \propto N\Big(\text{vec}\left((Z'Z)^{-1}Z'\beta_1,\ (Z'Z)^{-1}Z'\beta_2\right), \Sigma\otimes(Z'Z)^{-1}\Big). \tag{A.15}$$

- *Sampling* Σ

Σ is the covariance matrix of the error term in (A.14). To improve convergence we impose an inverted Wishart prior on the Σ parameter with scale parameter $K_1 I$ and degrees of freedom K_2 . Here we set K_1 to 0.1 and K_2 to 5 such that the influence of the prior on the posterior distribution is marginal (for a discussion, see Hobert and Casella 1996). Σ can be sampled from an inverted Wishart distribution with scale parameter $((\beta_1\, \beta_2) - Z(\theta\, \delta))'\, ((\beta_1\, \beta_2) - Z(\theta\, \delta)) + K_1 I$ and $(J-1) \cdot J \cdot C + K_2$ degrees of freedom.

Appendix B

Bayes Estimation of the Hierarchical Bayes Market Share model

In this appendix we present the details of our MCMC sampler for the hierarchical market share model. We first rewrite the model and next present the sampling steps for all parameters.

Recall that we denote the number of chains by I, the number of categories by C, and the number of observations for category c by T_c. The market share attraction model for category c at time t contains $(I-1)$ linear equations for log relative market shares (see Equation 3.4). We write the system of equations for category c at time t as

$$\begin{pmatrix} \log MS_{1,c,t} - \log MS_{I,c,t} \\ \log MS_{2,c,t} - \log MS_{I,c,t} \\ \vdots \\ \log MS_{I-1,c,t} - \log MS_{I,c,t} \end{pmatrix} = J \begin{pmatrix} \mu_{1,c} \\ \mu_{2,c} \\ \vdots \\ \mu_{I,c} \end{pmatrix} + J\mathrm{diag}(\log NBP_{c,t}) \begin{pmatrix} \beta_{1,1,c} \\ \beta_{1,2,c} \\ \vdots \\ \beta_{1,I,c} \end{pmatrix} +$$

$$J\mathrm{diag}(\log PLP_{c,t}) \begin{pmatrix} \beta_{2,1,c} \\ \beta_{2,2,c} \\ \vdots \\ \beta_{2,I,c} \end{pmatrix} + (\mathbf{I}_{I-1} \otimes \mathrm{vec}(F_{c,t}, D_{c,t}, \log ST_{c,t}, \log MS_{c,t-1})')\tilde{\beta}_c +$$

$$J\mathrm{diag}(\log NBP_{c,t-1}) \begin{pmatrix} \alpha_{1,1,c} \\ \alpha_{1,2,c} \\ \vdots \\ \alpha_{1,I,c} \end{pmatrix} + J\mathrm{diag}(\log PLP_{c,t-1}) \begin{pmatrix} \alpha_{2,1,c} \\ \alpha_{2,2,c} \\ \vdots \\ \alpha_{2,I,c} \end{pmatrix} + \begin{pmatrix} \eta_{1,c,t} \\ \eta_{2,c,t} \\ \vdots \\ \eta_{I,c,t} \end{pmatrix}, \tag{B.1}$$

where, for example, $NBP_{c,t}$ denotes the $I \times 1$ vector of the national brand prices, diag(x) gives a diagonal matrix with the elements of x along the diagonal, and

$$J = \begin{pmatrix} 1 & & & -1 \\ & 1 & & -1 \\ & & \ddots & \vdots \\ & & 1 & -1 \end{pmatrix}, \tag{B.2}$$

such that, for example,

$$J\mathrm{diag}(\log NBP_{c,t}) = \begin{pmatrix} \log NBP_{1,c,t} & & & -\log NBP_{I,c,t} \\ & \log NBP_{2,c,t} & & -\log NBP_{I,c,t} \\ & & \ddots & \vdots \\ & & \log NBP_{I-1,c,t} & -\log NBP_{I,c,t} \end{pmatrix}. \tag{B.3}$$

By grouping all explanatory variables in $X_{c,t}$ and all parameters in b_c, we summarise this equation as

$$y_{c,t} = X_{c,t} b_c + \eta_{c,t}. \tag{B.4}$$

Finally, we stack all observations over time to obtain $y_c = X_c b_c + \eta_c$, where $y_c = (y'_{c,1}, \cdots, y'_{c,T})'$ and $\eta_c = (\eta'_{c,1}, \cdots, \eta'_{c,T})' \sim N(0, \mathbf{I}_{T_c} \otimes \Sigma_c)$.

For the same category, we stack the I elements $\mu_{i,c}$, $\beta_{1,i,c}$, and $\beta_{2,i,c}$, and rewrite the second layer of the model (3.5), (3.6), (3.7) as

$$\mathrm{vec}(\mu_c,\ \beta_{1,c},\ \beta_{2,c}) \sim N\left(\mathrm{vec}\left(X_{\mu,c}\theta_\mu,\ X_{\beta,c}\theta_1,\ X_{\beta,c}\theta_2\right),\ \Omega \otimes \mathbf{I}_I\right), \tag{B.5}$$

where $X_{\mu,c} = (X_{\mu,1,c}, \cdots, X_{\mu,I,c})'$ and $X_{\beta,c} = (X_{\beta,1,c}, \cdots, X_{\beta,I,c})'$.

Sampling b_c

Define $A_c = P_c^{-1}$ such that $P_c P'_c = \mathbf{I}_{T_c} \otimes \Sigma_c$ and left multiply (B.4) by A_c such that

$$A_c y_c \sim N\left(A_c X_c b_c,\ \mathbf{I}_{T_c(I-1)}\right). \tag{B.6}$$

Furthermore define $B = Q^{-1}$ such that $QQ' = \Omega \otimes I_I$ and left multiply $\mathrm{vec}(\mu_c,\ \beta_{1,c},\ \beta_{2,c})$ by B such that

$$B\mathrm{vec}(\mu_c,\ \beta_{1,c},\ \beta_{2,c}) \sim N\left(B\mathrm{vec}\left(X_{\mu,c}\theta_\mu,\ X_{\beta,c}\theta_1,\ X_{\beta,c}\theta_2\right),\ \mathbf{I}_{3I}\right). \tag{B.7}$$

Equation (B.6) together with (B.7) can now be rewritten as $V_c = W_c b_c + \eta^*_c$, where $\eta^*_c \sim N(0,\ \mathbf{I})$,

$$V_c = \begin{pmatrix} A_c y_c \\ B\mathrm{vec}\left(X_{\mu,c}\theta_\mu,\ X_{\beta,c}\theta_1,\ X_{\beta,c}\theta_2\right) \end{pmatrix}, \text{ and } W_c = \begin{pmatrix} A_c X_c \\ B \quad 0 \end{pmatrix}. \tag{B.8}$$

The full conditional posterior distribution of b_c therefore is $N\left((W'_c W_c)^{-1} W'_c V_c,\ (W'_c W_c)^{-1}\right)$.

Sampling Σ_c

Then the full conditional posterior distribution of Σ_c is Inverted Wishart with T_c degrees of freedom, and scale parameter

$$S_c = \sum_{t=1}^{T^c} (y_{c,t} - X_{c,t} b_c)(y_{c,t} - X_{c,t} b_c)'. \tag{B.9}$$

Sampling θ_μ, θ_1, and θ_2

We rewrite (B.5) for category c as $M_c = Z_c\theta + \xi_c$, where $\xi_c = \left(\xi'_{\mu,c},\, \xi'_{1,c},\, \xi'_{2,c} \right)' \sim N(0,\ \Omega \otimes \mathbf{I}_I)$, $M_c = \left(\mu'_c, \beta'_{1,c}, \beta'_{2,c} \right)'$, $Z_c = \begin{pmatrix} X_{\mu,c} & & \\ & X_{\beta,c} & \\ & & X_{\beta,c} \end{pmatrix}$, and $\theta = \left(\theta'_\mu,\, \theta'_1,\, \theta'_2 \right)'$. The full conditional posterior distribution of θ therefore is

$$N\left((\sum_{c=1}^{C} Z'_c(\Omega^{-1} \otimes \mathbf{I}_I)Z_c)^{-1}(\sum_{c=1}^{C} Z'_c(\Omega^{-1} \otimes \mathbf{I}_I)M_c),\ (\sum_{c=1}^{C} Z'_c(\Omega^{-1} \otimes \mathbf{I}_I)Z_c)^{-1} \right). \tag{B.10}$$

Sampling Ω

Given all b_c, Σ_c, for $c = 1, \ldots, C$ and θ, the posterior distribution of Ω follows an inverted Wishart distribution with IC degrees of freedom and scale parameter $\sum_{i=1}^{I}\sum_{c=1}^{C} e_{i,c}e'_{i.c}$, where $e_{i,c} = (\mu_{i,c} - X'_{\mu,i,c}\theta_\mu,\ \beta_{1,i,c} - X'_{\beta,i,c}\theta_1,\ \beta_{2,i,c} - X'_{\beta,i,c}\theta_2)'$.

Appendix C

Some Data-mining Methods

C.1 Random forest

The random forest procedure is described as following:

1. Bootstrap number B of subsamples from the data.

2. For each bootstrapped sample, construct a big tree using recursive binary split, at each split consider only a random sample of p predictors from the total J predictors, where $p \approx \sqrt{J}$.

3. Average over all B trees to obtain the out-of-sample prediction

$$\ln \hat{s}_{t+1} = \hat{f}(x_t) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^b(x_t), \tag{C.1}$$

where f stands for the tree function, f^b is the b-th tree generated from the b-th bootstrapped sample.

C.2 Boosting tree

The procedure of boosting tree is described as following, where the fitted function is denoted by $\hat{f}$, and B is the total number of trees generated sequentially:

1. Set starting fitted values to be zero, $\hat{f}(x) = 0$, so the starting residuals are the target variable $r_t = \ln s_t$ for all t in the estimation set.

2. For $b = 1, 2, \ldots, B$, fit a tree $\hat{f}^b$ with d splits to the estimation set that uses residuals r as target variable.

3. Update $\hat{f}$ with $\hat{f} + \lambda\hat{f}^b$, where λ, the shrinkage parameter, is a very small number and so $\lambda\hat{f}^b$ is a shrunken version of $\hat{f}^b$ obtained in step 2.

4. Update residuals r_t with $r_t - \lambda\hat{f}^b(x_t)$.

5. After number B of trees are generated, average over all the models

$$\hat{\ln s} = \hat{f}(x) = \sum_{b=1}^{B} \lambda\hat{f}^b(x). \tag{C.2}$$

For the tree based methods, it is not necessary to use the natural logarithm of sales. But to be consistent with all the other models, we still use the natural logarithm form.

C.3 Choosing meta-parameters of boosting tree

We try shrinkage parameter $\lambda = 0.01$ and 0.001, the number of splits of each tree or interaction depth $d = 1$ to 4. So there eight different models. The trial results from the eight models show that for $\lambda = 0.01$, the out-of-sample RMSE settled after 1000 to 2000 iterations depending on the d value, while for $\lambda = 0.001$, the error has not settled yet after 5000 trees. So we choose $\lambda = 0.01$ to save time of computation. The trial results also show that when interaction depth $d = 3$, which means the model allows up to 3-way interactions, the out-of-sample RMSE is the lowest compared to $d = 1$, 2, and 4. Therefore we use $d = 3$ and $\lambda = 0.01$ for all the brands sales forecasts. To avoid overfit due to too large number of iterations B, we set $B = 2500$ such that the errors have converged and then use five-fold cross-validation in each estimation window to choose optimal B.

Bibliography

Ailawadi, K. L., Beauchamp, J. P., Donthu, N., Gauri, D. K., and Shankar, V. (2009). Communication and promotion decisions in retailing: a review and directions for future research. *Journal of retailing*, 85(1):42–55.

Ailawadi, K. L. and Harlam, B. (2004). An empirical analysis of the determinants of retail margins: The role of store-brand share. *Journal of Marketing*, 68(1):147–165.

Ailawadi, K. L., Kopalle, P. K., and Neslin, S. A. (2005). Predicting competitive response to a major policy change: combining game-theoretic and empirical analyses. *Marketing Science*, 24(1):12–24.

Ailawadi, K. L., Neslin, S. A., and Gedenk, K. (2001). Pursuing the value-conscious consumer: Store brands versus national brand promotions. *Journal of Marketing*, 65(1):71–89.

Ailawadi, K. L., Pauwels, K., and Steenkamp, J.-B. E. (2008). Private-label use and store loyalty. *Journal of Marketing*, 72(6):19–30.

Bai, J. and Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146:304–317.

Bayus, B. and Putsis, W. (1999). Product proliferation: An empirical analysis of product line determinants and market outcomes. *Marketing Science*, 18(2):137–153.

Blattberg, R. C. and Wisniewski, K. (1989a). Price-induced patterns of competition. *Marketing Science*, 8(4):291–309.

Blattberg, R. C. and Wisniewski, K. J. (1989b). Price-induced patterns of competition. *Marketing science*, 8(4):291–309.

Bogomolova, S., Szabo, M., and Kennedy, R. (2017). Retailers' and manufacturers' price-promotion decisions: Intuitive or evidence-based? *Journal of Business Research*, 76:189–200.

Bonfrer, A. and Chintagunta, P. K. (2004). Store brands: Who buys them and what happens to retail prices when they are introduced? *Review of Industrial Organization*, 24(2):195–218.

Breiman, L. (1996). Bagging predictors. *Machine learning*, 24(2):123–140.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1):5–32.

Breiman, L., Friedman, J., Olshen, R., and Stone, C. (1984). Classification and regression trees. wadsworth int. *Group*, 37(15):237–251.

Bronnenberg, B. J., Kruger, M. W., and Mela, C. F. (2008). Database paper: The IRI marketing set. *Marketing Science*, 27(4):745–748.

Bucklin, R. and Gupta, S. (1999). Commercial use of UPC scanner data: Industry and academic perspectives. *Marketing Science*, 18(3):247–273.

Cooper, L. G., Baron, P., Levy, W., Swisher, M., and Gogos, P. (1999). "promocast" : A new forecasting method for promotion planning. *Marketing Science*, 18(3):301–316.

Corstjens, M. and Lal, R. (2000). Building store loyalty through store brands. *Journal of Marketing Research*, 37(3):281–291.

Datta, H., Ailawadi, K. L., and Van Heerde, H. J. (2017). How well does consumer-based brand equity align with sales-based brand equity and marketing-mix response? *Journal of Marketing*, 81(3):1–20.

Davydenko, A. and Fildes, R. (2013). Measuring forecasting accuracy: The case of judgmental adjustments to sku-level demand forecasts. *International Journal of Forecasting*, 29(3):510–522.

Dhar, S. K. and Hoch, S. J. (1997). Why store brand penetration varies by retailer. *Marketing Science*, 16(3):208–227.

Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, 32(2):407–499.

Fildes, R., Ma, S., and Kolassa, S. (2019). Retail forecasting: Research and practice. *International Journal of Forecasting.*

Fok, D., Franses, P. H., and Paap, R. (2002). Econometric analysis of the market share attraction model. In Franses, P. H. and Montgomery, A. L., editors, *Advances in Econometrics*, volume 16, chapter 10, pages 223–256. JAI Press.

Fok, D., Horváth, C., Paap, R., and Franses, P. H. (2006). A hierarchical bayes error correction model to explain dynamic effects of price changes. *Journal of Marketing Research*, 43(August):443–461.

Geman, S. and Geman, D. (1984). Stochastic relaxations, gibbs distributions, and the bayesian restoration of images. *IEEE Transaction on Patern Analysis and Machine Intelligence*, 6(6):721–41.

Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In *Baysian Statistics*, pages 169–193, New York. Oxford University Press.

Gourville, J. T. and Soman, D. (2005). Overchoice and assortment type: When and why variety backfires. *Marketing Science*, 24(3):382 – 395.

Guyt, J. Y. and Gijsbrechts, E. (2014). Take turns or march in sync? the impact of the national brand promotion calendar on manufacturer and retailer performance. *Journal of Marketing Research*, 51(6):753–772.

Hanssens, D. M., Parsons, L. J., and Schultz, R. L. (2003). *Market response models: Econometric and time series analysis*, volume 12. Springer Science & Business Media.

Hastie, T., Tibshirani, R., and Friedman, J. (2001). *The Elements of Statistical Learning.* Springer Series in Statistics. Springer New York Inc., New York, NY, USA.

Heidelberger, P. and Welch, P. D. (1983). Simulation run length control in the presence of an initial transient. *Operations Research*, 31(6):1109–1144.

Hobert, J. P. and Casella, G. (1996). The effect of improper priors on gibbs sampling in hierarchical linear mixed models. *Journal of the American Statistical Association*, 91(436):1461–1473.

Horváth, C. and Fok, D. (2013). Moderating factors of immediate, gross, and net cross-brand effects of price promotions. *Marketing Science*, 32(1):127–152.

Horváth, C., Leeflang, P. S., Wieringa, J. E., and Wittink, D. R. (2005). Competitive reaction- and feedback effects based on varx models of pooled store data. *International Journal of Research in Marketing*, 22(4):415–426.

James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An introduction to statistical learning*, volume 112. Springer.

Kahn, B. (1998). Dynamic relationships with customers: High-variety strategies. *Journal of the Academy of Marketing Science*, 26(1):45–53.

Keller, K. (1993). Conceptualizing, measuring, and managing customer-based brand equity. *The Journal of Marketing*, 57(1):1–22.

Kim, J., Allenby, G., and Rossi, P. E. (2002). Modeling consumer demand for variety. *Marketing Science*, 21(3):229–250.

Kopalle, P. K., Mela, C. F., and Marsh, L. (1999). The dynamic effect of discounting on sales: Empirical analysis and normative pricing implications. *Marketing Science*, 18(3):317–332.

Koschate-Fischer, N., Cramer, J., and Hoyer, W. D. (2014). Moderating effects of the relationship between private label share and store loyalty. *Journal of Marketing*, 78(2):69–82.

Leeflang, P. S. H. and Wittink, D. R. (1992). Diagnosing competitive reactions using (aggregated) scanner data. *International Journal of Research in Marketing*, 9(1):39–57.

Leeflang, P. S. H. and Wittink, D. R. (1996). Competitive reaction versus consumer response: Do managers overreact? *International Journal of Research in Marketing*, 13(2):103–119.

Leeflang, P. S. H. and Wittink, D. R. (2001). Explaining competitive reaction effects. *International Journal of Research in Marketing*, 18(1-2):119–137.

Li, J. and Chen, W. (2014). Forecasting macroeconomic time series: Lasso-based approaches and their forecast combinations with dynamic factor models. *International Journal of Forecasting*, 30(4):996–1015.

Ma, S., Fildes, R., and Huang, T. (2016). Demand forecasting with high dimensional data: The case of sku retail sales forecasting with intra-and inter-category promotional information. *European Journal of Operational Research*, 249(1):245–257.

Macé, S. and Neslin, S. A. (2004). The determinants of pre- and postpromotion dips in sales of frequently purchased goods. *Journal of Marketing Research*, 41:339–50.

Medeiros, M. C. and Vasconcelos, G. F. (2016). Forecasting macroeconomic variables in data-rich environments. *Economics Letters*, 138:50–52.

Meza, S. and Sudhir, K. (2010). Do private labels increase retailer bargaining power? *Quantitative Marketing and Economics*, 8(3):333–363.

Montgomery, D. B., Moore, M. C., and Urbany, J. E. (2005). Reasoning about competitive reactions: evidence from executives. *Marketing Science*, 24(1):138–149.

Moon, S., Kamakura, W. A., and Ledolter, J. (2007). Estimating promotion response when competitive promotions are unobservable. *Journal of Marketing Research*, 44(3):503–515.

Naik, P. A., Raman, K., and Winer, R. S. (2005). Planning marketing-mix strategies in the presence of interaction effects. *Marketing Science*, 24(1):25–34.

Narasimhan, C., Neslin, S. A., and Sen, S. K. (1996). Promotional elasticities and category characteristics. *Marketing Science*, 60:17–30.

Nielsen (2019). The rise and rise again of private label. resreport. accessed on 26 July 2019.

Pauwels, K. and Srinivasan, S. (2004). Who benefits from store brand entry? *Marketing Science*, 23(3):364–390.

Raju, J. S. (1992). The effect of price promotions on variability in product category sales. *Marketing Science*, 11(3):207–220.

Reibstein, D. J. and Wittink, D. R. (2005). Editorial: Competitive responsiveness. *Marketing Science*, 24(1):8–11.

Richardson, P. S. (1997). Are store brands perceived to be just another brand? *Jouranl of Product and Brand Management*, 6(6):388–404.

Richardson, P. S., Dick, A. S., and Jain, A. K. (1994). Extrinsic and intrinsic cue effects on perceptions of store brand quality. *The Journal of Marketing*, 58(4):28–36.

Russel, G. (1992). A model of latent symmetry in cross-price elasticities. *Marketing Letters*, 3:157–169.

Sagaert, Y. R., Aghezzaf, E.-H., Kourentzes, N., and Desmet, B. (2018). Tactical sales forecasting using a very large set of macroeconomic indicators. *European Journal of Operational Research*, 264(2):558–569.

Sayman, S., Hoch, S. J., and Raju, J. S. (2002). Positioning of store brands. *Marketing Science*, 21(4):378–397.

Seenivasan, S., Sudhir, K., and Talukdar, D. (2016). Do store brands aid store loyalty? *Management Science*, 62(3):802–816.

Sethuraman, R. and Srinivasan, V. (2002). The asymmetric share effect: An empirical generalization on cross-price effects. *Journal of Marketing Research*, 39(August):379–386.

Sethuraman, R., Srinivasan, V., and Kim, D. (1999). Asymmetric and neighborhood cross-price effects: Some empirical generalizations. *Marketing Science*, 18(1):23–41.

Smeekes, S. and Wijler, E. (2018). Macroeconomic forecasting using penalized regression methods. *International journal of forecasting*, 34(3):408–430.

Smith, B. J. (2007a). boa: An package for mcmc output convergence assessment and posterior inference. *Journal of Statistical Software*, 21(11):1–37.

Smith, B. J. (2007b). boa: An R package for MCMC output convergence assessment and posterior inference. *Journal of Statistical Software*, 21(11):1–37.

Srinivasan, S., Leszczyc, P. T. P., and Bass, F. M. (2000). Market share response and competitive interaction: The impact of temporary, evolving and structural changes in prices. *International Journal of Research in Marketing*, 17(4):281–305.

Srinivasan, S., Pauwels, K., Hanssens, D. M., and Dekimpe, M. G. (2004). Do promotions benefit manufacturers, retailers, or both? *Management Science*, 50(5):617–629.

Sriram, S., Balachander, S., and Kalwani, M. (2007). Monitoring the dynamics of brand equity using store-level data. *Journal of Marketing*, 71(2):61–78.

Steenkamp, J.-B. E. M., Nijs, V. R., Hanssens, D. M., and Dekimpe, M. G. (2005). Competitive reactions to advertising and promotion attacks. *Marketing Science*, 24(1):35–54.

Stock, J. H. and Watson, M. W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44:293–335.

Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes. *Journal of Business and Economic Statistics*, 20(2):147–162.

Stock, J. H. and Watson, M. W. (2006). Forecasting with many predictors. *Handbook of economic forecasting*, 1:515–554.

Stock, J. H. and Watson, M. W. (2012). Dynamic factor models. In *The Oxford Handbook of Economic Forecasting*. Oxford University Press.

Talukdar, D., Sudhir, K., and Ainslie, A. (2002). Investigating new product diffusion across products and countries. *Marketing Science*, 21(1):97–114.

Tanner, M. A. and Wong, W. H. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398):528–540.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1):267–288.

Van Heerde, H. J., Leeflang, P. S., and Wittink, D. R. (2000). The estimation of pre- and postpromotion dips with store-level scanner data. *Journal of Marketing Research*, 37(August):383–395.

Van Nierop, E., Fok, D., and Franses, P. H. (2008). Interaction between shelf layout and marketing effectiveness and its impact on optimizing shelf arrangements. *Marketing Science*, 27(6):1065–1082.

Wittink, D. R., Addona, M. J., Hawkes, W. J., and Porter, J. C. (1988). Scan* pro: The estimation, validation and use of promotional effects based on scanner data. *Internal Paper, Cornell University*.

Zou, H. and Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the royal statistical society: series B (statistical methodology)*, 67(2):301–320.

Concurrentie in het schap

Samenvatting

In dit proefschrift wordt de concurrentie op de retailmarkt voor verpakte consumptiegoederen vanuit verschillende invalshoeken onderzocht. In hoofdstuk 2 wordt bekeken hoe merken reageren op elkaars prijspromoties, waarbij de nadruk ligt op de asymmetrische reacties tussen merken met verschillende marktaandelen en prijsniveaus. Hiervoor wordt gebruikgemaakt van een Bayesiaans hiërarchisch geordend probitmodel (Hierarchical Bayes Ordered Probit model, HB-OP) om de modererende factoren op reacties te bestuderen. De resultaten tonen aan dat de intenties van de reactie worden beïnvloed door het relatieve marktaandeel van een merk, samen met enkele categoriespecifieke kenmerken. In hoofdstuk 3 wordt de competitie tussen retailketens en de rol van hun private label-merken onderzocht. We stellen een Bayesiaans hiërarchisch marktaandeelmodel (Hierarchical Bayes Market Share, HB-MS) voor om te onderzoeken hoe het marktaandeel van een detailhandelaar afhankelijk is van prijsveranderingen van landelijke merken en private labels. Ook wordt hiermee onderzocht hoe het basismarktaandeel en de prijsgevoeligheid worden beïnvloed door de positionering van private labels. Hoofdstuk 4 gaat over de vergelijking van traditionele omzetprognosemodellen met moderne technieken zoals factormodellen, Lasso, elastic net, random forests en boosting-methoden. We beschouwen alle mogelijke merken als potentiële concurrentie die mogelijk van nut is voor de verkoopprognoses van een bepaald merk. Deze aanpak is relevant wanneer we niet op voorhand weten welke merken een voorspellende waarde hebben, en in dit geval kunnen we de gegevens steeds laten meewegen bij het maken van een prognose. De nauwkeurigheid van de prognose van verschillende modellen wordt vergeleken voor een groot aantal merken.

About the Author

Wei Li was born and brought up in Hubei, China. She holds a Bachelor's and a Master's degree in International Economics from Wuhan University, China. In 2006, she joined the MPhil programme of Erasmus Research Institute of Management (ERIM) on marketing track. After she obtained a Master's degree in Business Research she continued her PhD study under the supervision of prof. dr. Dennis Fok and prof. dr. Philip Hans Franses. Her research is concerned with using quantitative methods to study competition in retail market. Her work has been presented at INFORMS Marketing Science Conferences. She moved to London in 2012 and worked for a marketing consultancy company based there. Later she moved again to Nottingham and now works for the University of Nottingham as a research and teaching fellow.

Portfolio

Wei Li

Sir Clive Granger Building
University Park
University of Nottingham
Nottingham NG7 2RD
United Kingdom
Email: wei.li@nottingham.ac.uk

Research Interests

My research interests lie in the application of econometric methods to study the firm and consumer behaviour in retail markets using large datasets, especially using machine learning methods to forecast firm performance and evaluating the forecast accuracy from different methods.

Teaching Expertise

I have experience teaching Econometrics, Statistics for Economics, Mathematics for Economics, and general economics at introductory level.

Work Experience

Nov 2020 - present: Research and Teaching Fellow, School of Economics, University of Nottingham

Sep 2018 - Sep 2020: Teaching Fellow, School of Economics, University of Nottingham
Sep 2016 - Sep 2017: Graduate Teaching Assistant, School of Economics, University of Nottingham
Sep 2012 - May 2016: Quantitative Forecast Manager, Future Foundation, London
Sep 2012 - Dec 2012: Econometrics Tutor, City University London

Education

2009 - 2012: PhD student in Marketing, ERIM, Erasmus University Rotterdam
2008: Master's Degree in M. Phil in Business Research, ERIM, Erasmus University Rotterdam
2006: Master's Degree in Economics, Wuhan University, China
2001: Bachelor's Degree in Economics, Wuhan University, China

Programming Languages and Softwares

Programming: R, SQL
Other Softwares: Stata, SPSS, Eviews

Other Honours

2006 - 2007: Scholarship of ERIM, Erasmus University Rotterdam
2005 - 2006: Huygens Scholarship, Nuffic, the Netherlands
1997 - 2004: People's Scholarship, Wuhan University, China

ERASMUS RESEARCH INSTITUTE OF MANAGEMENT (ERIM)

The ERIM PhD Series

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://repub.eur.nl/pub. ERIM is the joint research institute of the Rotterdam School of Management (RSM) and the Erasmus School of Economics (ESE) at the Erasmus University Rotterdam (EUR).

Dissertations in the last four years

Ahmadi, S., A motivational perspective to decision-making and behavior in organizations, Promotors: Prof. J.J.P. Jansen & Dr T.J.M. Mom, EPS-2019-477-S&E, `https://repub.eur.nl/pub/116727`

Akemu, O., Corporate Responses to Social Issues: Essays in Social Entrepreneurship and Corporate Social Responsibility, Promotors: Prof. G.M. Whiteman & Dr S.P. Kennedy, EPS-2017-392-ORG, `https://repub.eur.nl/pub/95768`

Albuquerque de Sousa, J.A., International stock markets: Essays on the determinants and consequences of financial market development, Promotors: Prof. M.A. van Dijk & Prof. P.A.G. van Bergeijk, EPS-2019-465-F&A, `https://repub.eur.nl/pub/115988`

Alserda, G.A.G., Choices in Pension Management, Promotors: Prof. S.G. van der Lecq & Dr O.W. Steenbeek, EPS-2017-432-F&A, `https://repub.eur.nl/pub/103496`

Arampatzi, E., Subjective Well-Being in Times of Crises: Evidence on the Wider Impact of Economic Crises and Turmoil on Subjective Well-Being, Promotors: Prof. H.R. Commandeur, Prof. F. van Oort & Dr. M.J. Burger, EPS-2018-459-S&E,

https://repub.eur.nl/pub/111830

Arslan, A.M., Operational Strategies for On-demand Delivery Services, Promotors: Prof. R.A. Zuidwijk & Dr N.A. H. Agatz, EPS-2019-481-LIS, https://repub.eur.nl/pub/126463

Avci, E., Surveillance of Complex Auction Markets: a Market Policy Analytics Approach, Promotors: Prof. W. Ketter, Prof. H.W.G.M. van Heck & Prof. D.W. Bunn, EPS-2018-426-LIS, https://repub.eur.nl/pub/106286

Balen, T.H. van, Challenges of Early Stage Entrepreneurs: the Roles of Vision Communication and Team Membership Change, Promotors: Prof. J.C.M. van den Ende & Dr M. Tarakci, EPS-2019-468-LIS, https://repub.eur.nl/pub/115654

Bernoster, I., Essays at the Intersection of Psychology, Biology, and Entrepreneurship, Promotors: Prof. A.R. Thurik, Prof. I.H.A. Franken & Prof. P.J.F Groenen, EPS-2018-463-S&E, https://repub.eur.nl/pub/113907

Blagoeva, R.R., The Hard Power Of Soft Power: A behavioral strategy perspective on how power, reputation, and status affect firms, Promotors: Prof. J.J.P. Jansen & Prof. T.J.M. Mom, EPS-2020-495-S&E, https://repub.eur.nl/pub/127681

Bouman, P., Passengers, Crowding and Complexity: Models for Passenger Oriented Public Transport, Prof. L.G. Kroon, Prof. A. Schbel & Prof. P.H.M. Vervest, EPS-2017-420-LIS, https://repub.eur.nl/pub/100767

Breugem, T., ÔCrew Planning at Netherlands Railways: Improving Fairness, Attractiveness, and EfficiencyÕ, Promotors: Prof. D. Huisman & Dr T.A.B. Dollevoet, EPS-2020494-LIS, https://repub.eur.nl/pub/124016

Bunderen, L. van, Tug-of-War: Why and when teams get embroiled in power struggles, Promotors: Prof. D.L. van Knippenberg & Dr. L. Greer, EPS-2018-446-ORG, https://repub.eur.nl/pub/105346

Burg, G.J.J. van den, Algorithms for Multiclass Classification and Regularized Regression, Promotors: Prof. P.J.F. Groenen & Dr. A. Alfons, EPS-2018-442-MKT,

https://repub.eur.nl/pub/103929

Chammas, G., Portfolio concentration, Promotor: Prof. J. Spronk, EPS-2017-410-F&E, https://repub.eur.nl/pub/94975

Chan, H.Y., ÔDecoding the consumerÕs brain: Neural representations of consumer experienceÕ, Promotors: Prof. A. Smidts & Dr M. A.S. Boksem, EPS-2019-493-MKT, https://repub.eur.nl/pub/124931

Dalmeijer, K., Time Window Assignment in Distribution Networks, Promotors: Prof A.P.M. Wagelmans & Dr R. Spliet, EPS-2019-486-LIS, https://repub.eur.nl/pub/120773

Dennerlein, T. Empowering Leadership and EmployeesÕ Achievement Motivations: the Role of Self-Efficacy and Goal Orientations in the Empowering Leadership Process, Promotors: Prof. D.L. van Knippenberg & Dr J. Dietz, EPS-2017-414-ORG, https://repub.eur.nl/pub/98438

Dolgova, E., On Getting Along and Getting Ahead: How Personality Affects Social Network Dynamics, Promotors: Prof. P.P.M.A.R Heugens & Prof. M.C. Schippers, EPS-2019-455-S&E, https://repub.eur.nl/pub/119150

Duijzer, L.E., Mathematical Optimization in Vaccine Allocation, Promotors: Prof. R. Dekker & Dr W.L. van Jaarsveld, EPS-2017-430-LIS, https://repub.eur.nl/pub/101487

Eijlers, E., Emotional Experience and Advertising Effectiveness: on the use of EEG in marketing, Prof. A. Smidts & Prof. M.A.S. Boksem, Eps-2019-487-MKT, https://repub.eur.nl/pub/124053

El Nayal, O.S.A.N., Firms and the State: An Examination of Corporate Political Activity and the Business-Government Interface, Promotor: Prof. J. van Oosterhout & Dr. M. van Essen, EPS-2018-469-S&E, https://repub.eur.nl/pub/114683

Feng, Y., The Effectiveness of Corporate Governance Mechanisms and Leadership Structure: Impacts on strategic change and firm performance, Promotors: Prof.

F.A.J. van den Bosch, Prof. H.W. Volberda & Dr J.S. Sidhu, EPS-2017-389-S&E, https://repub.eur.nl/pub/98470

Frick, T.W., The Implications of Advertising Personalization for Firms, Consumer, and Ad Platfroms, Promotors: Prof. T. Li & Prof. H.W.G.M. van Heck, EPS-2018-452-LIS, https://repub.eur.nl/pub/110314

Fytraki, A.T., Behavioral Effects in Consumer Evaluations of Recommendation Systems, Promotors: Prof. B.G.C. Dellaert & Prof. T. Li, EPS-2018-427-MKT, https://repub.eur.nl/pub/110457

Gai, J., Contextualized Consumers: Theories and Evidence on Consumer Ethics, Product Recommendations, and Self-Control, Promotors: Prof. S. Puntoni & Prof. S.T.L. Sweldens, EPS-2020-498-MKT, https://repub.eur.nl/pub/127680

Ghazizadeh, P. Empirical Studies on the Role of Financial Information in Asset and Capital Markets, Promotors: Prof. A. de Jong & Prof. E. Peek, EPS-2019-470-F&A, https://repub.eur.nl/pub/114023

Giurge, L., A Test of Time; A temporal and dynamic approach to power and ethics, Promotors: Prof. M.H. van Dijke & Prof. D. De Cremer, EPS-2017-412-ORG, https://repub.eur.nl/pub/98451

Gobena, L., Towards Integrating Antecedents of Voluntary Tax Compliance, Promotors: Prof. M.H. van Dijke & Dr P. Verboon, EPS-2017-436-ORG, https://repub.eur.nl/pub/103276

Groot, W.A., Assessing Asset Pricing Anomalies, Promotors: Prof. M.J.C.M. Verbeek & Prof. J.H. van Binsbergen, EPS-2017-437-F&A, https://repub.eur.nl/pub/103490

Hanselaar, R.M., Raising Capital: On pricing, liquidity and incentives, Promotors: Prof. M.A. van Dijk & Prof. P.G.J. Roosenboom, EPS-2018-429-F&A, https://repub.eur.nl/pub/113274

Harms, J. A., Essays on the Behavioral Economics of Social Preferences and Bounded

Rationality, Prof. H.R. Commandeur & Dr K.E.H. Maas, EPS-2018-457-S&E, https://repub.eur.nl/pub/108831

Hendriks, G., Multinational Enterprises and Limits to International Growth: Links between Domestic and Foreign Activities in a FirmÕs Portfolio, Promotors: Prof. P.P.M.A.R. Heugens & Dr. A.H.L Slangen, EPS-2019-464-S&E, https://repub.eur.nl/pub/114981

Hengelaar, G.A., The Proactive Incumbent: Holy grail or hidden gem? Investigating whether the Dutch electricity sector can overcome the incumbentÕs curse and lead the sustainability transition, Promotors: Prof. R.J. M. van Tulder & Dr K. Dittrich, EPS-2018-438-ORG, https://repub.eur.nl/pub/102953

Jacobs, B.J.D., Marketing Analytics for High-Dimensional Assortments, Promotors: Prof. A.C.D. Donkers & Prof. D. Fok, EPS-2017-445-MKT, https://repub.eur.nl/pub/103497

Jia, F., The Value of Happiness in Entrepreneurship, Promotors: Prof. D.L. van Knippenberg & Dr Y. Zhang, EPS-2019-479-ORG, https://repub.eur.nl/pub/115990

Kahlen, M. T., Virtual Power Plants of Electric Vehicles in Sustainable Smart Electricity Markets, Promotors: Prof. W. Ketter & Prof. A. Gupta, EPS-2017-431-LIS, https://repub.eur.nl/pub/100844

Kampen, S. van, The Cross-sectional and Time-series Dynamics of Corporate Finance: Empirical evidence from financially constrained firms, Promotors: Prof. L. Norden & Prof. P.G.J. Roosenboom, EPS-2018-440-F&A, https://repub.eur.nl/pub/105245

Karali, E., Investigating Routines and Dynamic Capabilities for Change and Innovation, Promotors: Prof. H.W. Volberda, Prof. H.R. Commandeur & Dr J.S. Sidhu, EPS-2018-454-S&E, https://repub.eur.nl/pub/106274

Keko. E, Essays on Innovation Generation in Incumbent Firms, Promotors: Prof. S. Stremersch & Dr N.M.A. Camacho, EPS-2017-419-MKT, https://repub.eur.nl/

pub/100841

Kerkkamp, R.B.O., Optimisation Models for Supply Chain Coordination under Information Asymmetry, Promotors: Prof. A.P.M. Wagelmans & Dr. W. van den Heuvel, EPS-2018-462-LIS, https://repub.eur.nl/pub/109770

Khattab, J., Make Minorities Great Again: a contribution to workplace equity by identifying and addressing constraints and privileges, Promotors: Prof. D.L. van Knippenberg & Dr A. Nederveen Pieterse, EPS-2017-421-ORG, https://repub.eur.nl/pub/99311

Kim, T. Y., Data-driven Warehouse Management in Global Supply Chains, Promotors: Prof. R. Dekker & Dr C. Heij, EPS-2018-449-LIS, https://repub.eur.nl/pub/109103

Klitsie, E.J., Strategic Renewal in Institutional Contexts: The paradox of embedded agency, Promotors: Prof. H.W. Volberda & Dr. S. Ansari, EPS-2018-444-S&E, https://repub.eur.nl/pub/106275

Koolen, D., Market Risks and Strategies in Power Systems Integrating Renewable Energy, Promotors: Prof. W. Ketter & Prof. R. Huisman, EPS-2019-467-LIS, https://repub.eur.nl/pub/115655

Kong, L. Essays on Financial Coordination, Promotors: Prof. M.J.C.M. Verbeek, Dr. D.G.J. Bongaerts & Dr. M.A. van Achter. EPS-2019-433-F&A, https://repub.eur.nl/pub/114516

Kyosev, G.S., Essays on Factor Investing, Promotors: Prof. M.J.C.M. Verbeek & Dr J.J. Huij, EPS-2019-474-F&A, https://repub.eur.nl/pub/116463

Lamballais Tessensohn, T., Optimizing the Performance of Robotic Mobile Fulfillment Systems, Promotors: Prof. M.B.M de Koster, Prof. R. Dekker & Dr D. Roy, EPS-2019-411-LIS, https://repub.eur.nl/pub/116477

Leung, W.L., How Technology Shapes Consumption: Implications for Identity and Judgement, Promotors: Prof. S. Puntoni & Dr G Paolacci, EPS-2019-485-MKT,

https://repub.eur.nl/pub/117432

Li, X. Dynamic Decision Making under Supply Chain Competition, Promotors: Prof. M.B.M de Koster, Prof. R. Dekker & Prof. R. Zuidwijk, EPS-2018-466-LIS, https://repub.eur.nl/pub/114028

Liu, N., Behavioral Biases in Interpersonal Contexts, Supervisors: Prof. A. Baillon & Prof. H. Bleichrodt, EPS-2017-408-MKT, https://repub.eur.nl/pub/95487

Maas, A.J.J., Organizations and their external context: Impressions across time and space, Promotors: Prof. P.P.M.A.R Heugens & Prof. T.H. Reus, EPS-2019-478-S&E, https://repub.eur.nl/pub/116480

Maira, E., Consumers and Producers, Promotors: Prof. S. Puntoni & Prof. C. Fuchs, EPS-2018-439-MKT, https://repub.eur.nl/pub/104387

Mirzaei, M., ÔAdvanced Storage and Retrieval Policies in Automated WarehousesÕ, Promotors: Prof. M.B.M. de Koster & Dr N. Zaerpour, EPS-2020-490-LIS, https://repub.eur.nl/pub/125975

Nair, K.P., Strengthening Corporate Leadership Research: The relevance of biological explanations, Promotors: Prof. J. van Oosterhout & Prof. P.P.M.A.R Heugens, EPS-2019-480-S&E, https://repub.eur.nl/pub/120023

Nullmeier, F.M.E., Effective contracting of uncertain performance outcomes: Allocating responsibility for performance outcomes to align goals across supply chain actors, Promotors: Prof. J.Y.F.Wynstra & Prof. E.M. van Raaij, EPS-2019-484-LIS, https://repub.eur.nl/pub/118723

Okbay, A., Essays on Genetics and the Social Sciences, Promotors: Prof. A.R. Thurik, Prof. Ph.D. Koellinger & Prof. P.J.F. Groenen, EPS-2017-413-S&E, https://repub.eur.nl/pub/95489

Peng, X., Innovation, Member Sorting, and Evaluation of Agricultural Cooperatives, Promotor: Prof. G.W.J. Hendriks, EPS-2017-409-ORG, https://repub.eur.nl/pub/94976

Petruchenya, A., Essays on Cooperatives: Emergence, Retained Earnings, and Market Shares, Promotors: Prof. G.W.J. Hendriks & Dr Y. Zhang, EPS-2018-447-ORG, https://repub.eur.nl/pub/105243

Plessis, C. du, Influencers: The Role of Social Influence in Marketing, Promotors: Prof. S. Puntoni & Prof. S.T.L.R. Sweldens, EPS-2017-425-MKT, https://repub.eur.nl/pub/103265

Pocock, M., Status Inequalities in Business Exchange Relations in Luxury Markets, Promotors: Prof. C.B.M. van Riel & Dr G.A.J.M. Berens, EPS-2017-346-ORG, https://repub.eur.nl/pub/98647

Pozharliev, R., Social Neuromarketing: The role of social context in measuring advertising effectiveness, Promotors: Prof. W.J.M.I. Verbeke & Prof. J.W. van Strien, EPS-2017-402-MKT, https://repub.eur.nl/pub/95528

Qian, Z., Time-Varying Integration and Portfolio Choices in the European Capital Markets, Promotors: Prof. W.F.C. Verschoor, Prof. R.C.J. Zwinkels & Prof. M.A. Pieterse-Bloem, EPS-2020-488-F&A, https://repub.eur.nl/pub/124984

Reh, S.G., A Temporal Perspective on Social Comparisons in Organizations, Promotors: Prof. S.R. Giessner, Prof. N. van Quaquebeke & Dr. C. Troster, EPS-2018-471-ORG, https://repub.eur.nl/pub/114522

Riessen, B. van, Optimal Transportation Plans and Portfolios for Synchromodal Container Networks, Promotors: Prof. R. Dekker & Prof. R.R. Negenborn, EPS-2018-448-LIS, https://repub.eur.nl/pub/105248

Romochkina, I.V., When Interests Collide: Understanding and modeling interests alignment using fair pricing in the context of interorganizational information systems, Promotors: Prof. R.A. Zuidwijk & Prof. P.J. van Baalen, EPS-2020-451-LIS, https://repub.eur.nl/pub/127244

Schie, R. J. G. van, Planning for Retirement: Save More or Retire Later? Promotors: Prof. B. G. C. Dellaert & Prof. A.C.D. Donkers, EOS-2017-415-MKT,

https://repub.eur.nl/pub/100846

Schouten, K.I.M. Semantics-driven Aspect-based Sentiment Analysis, Promotors: Prof. F.M.G. de Jong, Prof. R. Dekker & Dr. F. Frasincar, EPS-2018-453-LIS, https://repub.eur.nl/pub/112161

Sihag, V., The Effectiveness of Organizational Controls: A meta-analytic review and an investigation in NPD outsourcing, Promotors: Prof. J.C.M. van den Ende & Dr S.A. Rijsdijk, EPS-2019-476-LIS, https://repub.eur.nl/pub/115931

Smolka, K.M., Essays on Entrepreneurial Cognition, Institution Building and Industry Emergence, Promotors: P.P.M.A.R. Heugens, & Prof. J.P. Cornelissen, Eps-2019-483-S&E, https://repub.eur.nl/pub/118760

Straeter, L.M., Interpersonal Consumer Decision Making, Promotors: Prof. S.M.J. van Osselaer & Dr I.E. de Hooge, EPS-2017-423-MKT, https://repub.eur.nl/pub/100819

Stuppy, A., Essays on Product Quality, Promotors: Prof. S.M.J. van Osselaer & Dr N.L. Mead. EPS-2018-461-MKT, https://repub.eur.nl/pub/111375

Suba?i, B., Demographic Dissimilarity, Information Access and Individual Performance, Promotors: Prof. D.L. van Knippenberg & Dr W.P. van Ginkel, EPS-2017-422-ORG, https://repub.eur.nl/pub/103495

Suurmond, R., In Pursuit of Supplier Knowledge: Leveraging capabilities and dividing responsibilities in product and service contexts, Promotors: Prof. J.Y.F Wynstra & Prof. J. Dul. EPS-2018-475-LIS, https://repub.eur.nl/pub/115138

Toxopeus, H.S. Financing sustainable innovation: From a principal-agent to a collective action perspective, Promotors: Prof. H.R. Commandeur & Dr. K.E.H. Maas. EPS-2019-458-S&E, https://repub.eur.nl/pub/114018

Turturea, R., Overcoming Resource Constraints: The Role of Creative Resourcing and Equity Crowdfunding in Financing Entrepreneurial Ventures, Promotors: Prof. P.P.M.A.R Heugens, Prof. J.J.P. Jansen & Dr. I. Verheuil, EPS-2019-472-S&E,

https://repub.eur.nl/pub/112859

Valboni, R.,ÕBuilding Organizational (Dis-)Abilities: The impact of learning on the performance of mergers and acquisitionsÕ,Promotors: Prof. T.H. Reus & Dr A.H.L. Slangen, EPS-2020-407-S&E, https://repub.eur.nl/pub/125226

Vandic, D., Intelligent Information Systems for Web Product Search, Promotors: Prof. U. Kaymak & Dr Frasincar, EPS-2017-405-LIS, https://repub.eur.nl/pub/95490

Verbeek, R.W.M., Essays on Empirical Asset Pricing, Promotors: Prof. M.A. van Dijk & Dr M. Szymanowska, EPS-2017-441-F&A, https://repub.eur.nl/pub/102977

Visser, T.R. Vehicle Routing and Time Slot Management in Online Retailing, Promotors: Prof. A.P.M. Wagelmans & Dr R. Spliet, EPS-2019-482-LIS, https://repub.eur.nl/pub/120772

Vlaming, R. de.,Linear Mixed Models in Statistical Genetics, Prof. A.R. Thurik, Prof. P.J.F. Groenen & Prof. Ph.D. Koellinger, EPS-2017-416-S&E, https://repub.eur.nl/pub/100428

Vries, H. de, Evidence-Based Optimization in Humanitarian Logistics, Promotors: Prof. A.P.M. Wagelmans & Prof. J.J. van de Klundert, EPS-2017-435-LIS, https://repub.eur.nl/pub/102771

Wang, R., Corporate Environmentalism in China, Promotors: Prof. P.P.M.A.R Heugens & Dr F. Wijen, EPS-2017-417-S&E, https://repub.eur.nl/pub/99987

Wang, R., Those Who Move Stock Prices, Promotors: Prof. P. Verwijmeren & Prof. S. van Bekkum, EPS-2019-491-F&A, https://repub.eur.nl/pub/129057

Wasesa, M., Agent-based inter-organizational systems in advanced logistics operations, Promotors: Prof. H.W.G.M van Heck, Prof. R.A. Zuidwijk & Dr A. W. Stam, EPS-2017-LIS-424, https://repub.eur.nl/pub/100527

Wessels, C., Flexible Working Practices: How Employees Can Reap the Benefits for

Engagement and Performance, Promotors: Prof. H.W.G.M. van Heck, Prof. P.J. van Baalen & Prof. M.C. Schippers, EPS-2017-418-LIS, `https://repub.eur.nl/`

Wiegmann, P.M., Setting the Stage for Innovation: Balancing Diverse Interests through Standardisation, Promotors: Prof. H.J. de Vries & Prof. K. Blind, EPS-2019-473-LIS, `https://repub.eur.nl/pub/114519`

Wijaya, H.R., Praise the Lord!: Infusing Values and Emotions into Neo-Institutional Theory, Promotors: Prof. P.P.M.A.R. Heugens & Prof. J.P. Cornelissen, EPS-2019-450-S&E, `https://repub.eur.nl/pub/115973`

Williams, A.N., Make Our Planet Great Again: A Systems Perspective of Corporate Sustainability, Promotors: Prof. G.M. Whiteman & Dr. S. Kennedy, EPS-2018-456-ORG, `https://repub.eur.nl/pub/111032`

Witte, C.T., Bloody Business: Multinational investment in an increasingly conflict-afflicted world, Promotors: Prof. H.P.G. Pennings, Prof. H.R. Commandeur & Dr M.J. Burger, EPS-2018-443-S&E, `https://repub.eur.nl/pub/104027`

Ye, Q.C., Multi-objective Optimization Methods for Allocation and Prediction, Promotors: Prof. R. Dekker & Dr Y. Zhang, EPS-2019-460-LIS, `https://repub.eur.nl/pub/116462`

Yuan, Y., The Emergence of Team Creativity: a social network perspective, Promotors: Prof. D. L. van Knippenberg & Dr D. A. Stam, EPS-2017-434-ORG, `https://repub.eur.nl/pub/100847`

Zhang, Q., Financing and Regulatory Frictions in Mergers and Acquisitions, Promotors: Prof. P.G.J. Roosenboom & Prof. A. de Jong, EPS-2018-428-F&A, `https://repub.eur.nl/pub/103871`

This book investigates competition in the retail market of consumer packaged goods from different angles. Chapter 2 studies how brands react to each other's price promotions, the focus is put on the asymmetric reactions between brands with different market shares and price levels. To this end a Hierarchical Bayes Ordered Probit model (HB-OP) is employed to study the moderating factors on reactions. The results show that the reaction intentions are influenced by brands' relative market shares, together with some category specific characteristics. Chapter 3 investigates the competition between retail chains and the role of their private label brands. We propose a Hierarchical Bayes Market Share (HB-MS) model to investigate how a retailer's market share depends on price changes by national brands and private labels, and how the baseline market share and the price sensitivities are influenced by private-label positioning. Chapter 4 aims to compare traditional sales forecasting models with modern techniques like factor models, Lasso, elastic net, random forests and boosting methods. We consider all possible brands as potential competitors that might be useful for the sales forecasts of a focal brand. This approach is relevant if we do not know beforehand which brands have predictive content, and in this case, we can let the data help to decide on this each time we make a forecast. The forecasting accuracy of a variety of models are compared across a large number of brands.

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

ERIM PhD Series
Research in Management

Erasmus University Rotterdam (EUR)
Erasmus Research Institute of Management
Mandeville (T) Building
Burgemeester Oudlaan 50
3062 PA Rotterdam, The Netherlands

P.O. Box 1738
3000 DR Rotterdam, The Netherlands
T +31 10 408 1182
E info@erim.eur.nl
W www.erim.eur.nl