UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

MACHINE LEARNING METHODS TO SUPPORT CATEGORY MANAGEMENT DECISIONS IN THE RETAIL INDUSTRY

TESIS PARA OPTAR AL GRADO DE
DOCTOR EN SISTEMAS DE INGENIERÍA

LUIS ALBERTO ABURTO LAFOURCADE

PROFESOR GUÍA:
MARCEL GOIC FIGUEROA
PROFESOR GUÍA 2:
ANDRÉS MUSALEM SAID

MIEMBROS DE LA COMISIÓN:
RICHARD WEBER HAAS
JOCHEN DE WEERDT

SANTIAGO DE CHILE
2019

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE DOCTOR EN SISTEMAS DE INGENIERÍA
POR: LUIS ALBERTO ABURTO LAFOURCADE
FECHA: 2019
PROF. GUÍA: MARCEL GOIC FIGUEROA ANDRÉS MUSALEM SAID

MACHINE LEARNING METHODS TO SUPPORT CATEGORY MANAGEMENT DECISIONS IN THE RETAIL INDUSTRY

La introducción de herramientas de Machine Learning para resolver problemas de Big Data, han creado diversas oportunidades, en particular en la industrial del retail. Esta tesis está dividida en dos ensayos donde se usan herramientas de Maching Learning para resolver problemas de Gestión de Retail de manera novedosa.

En el primer ensayo se analizan datos transaccionales para mejorar la formulación de restricciones para el problema de optimización de precios en categorías. Usando un método de Reglas de Asociación se identifica qué reglas han sido consistentemente aplicadas, y se evalúan cuales están asociadas a mejores desempeños en la categoría. Basado en estas reglas de precios, se construye un conjunto de precios factibles basados en datos, y se combinan con rutinas de optimización para entender como esta información puede complementar el análisis econométrico tradicional. Cuando combinamos nuestra metodología con distintos métodos para estimar respuesta de clientes a cambios de precio, encontramos que modelos de demanda simples como doble log, son muy sensibles a esta definición de factibilidad de precios. Modelos más sofisticados como Lasso o Modelo Jerárquico Bayesiano son menos sensibles a los cambios de restricciones, pero de igual forma son afectados por la definición de precios factibles. Nuestros resultados numéricos muestran que la metodología propuesta no solo lleva a soluciones de precio más realistas, sino que también son más robustas a variabilidad en los datos, llevando mejores resultados para este problema de negocio.

El segundo ensayo presenta un enfoque de Machine Learning para estudiar interrelaciones entre categorías de productos y detectar motivaciones latentes de compra. Este se basa en un modelo llamado Latent Dirichlet Allocation (LDA), el cual ha sido ampliamente usado en análisis de texto, para extraer tópicos de documentos, midiendo la probabilidad de coocurrencia entre palabras. En nuestro contexto de retail, se extraen motivaciones de compra en vez de tópicos latentes de textos, analizando relaciones entre categorías de productos en vez de palabras, usando una base de datos de transacciones en vez de documentos. La contribución de esta investigación es la aplicación de LDA en datos de supermercados, modificando el modelo básico para lograr tres objetivos diferentes. Primero, se usa un modelo LDA estándar para extraer y describir las motivaciones de compra. Segundo, el modelo de LDA se amplía con un enfoque supervisado, para estimar conjuntamente las motivaciones de compra y la relación de estas con el tamaño monetario de la canasta. Por último, el modelo LDA se generaliza para permitir que cada motivación dependa de características demográficas y de compra de los clientes. Estas motivaciones de compra identificadas son fundamentales para mejorar decisiones de promociones y recomendación de productos.

En resumen, los resultados propuestos en esta tesis, entregan soluciones para entender el comportamiento de compra de clientes y así apoyar decisiones de precio y promociones, y mejorar la competitividad en la industria del retail.

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE DOCTOR EN SISTEMAS DE INGENIERÍA
POR: LUIS ALBERTO ABURTO LAFOURCADE
FECHA: 2019
PROF. GUÍA: MARCEL GOIC FIGUEROA ANDRÉS MUSALEM SAID

MACHINE LEARNING METHODS TO SUPPORT CATEGORY MANAGEMENT DECISIONS IN THE RETAIL INDUSTRY

The introduction of machine learning to deal with big data problems have created several business opportunities, in particular in the retail industry. This thesis is divided in two essays where machine learning are used to deal with retail management problems in a novelty way.

In the first essay we analyze transactional data to improve the formulation of constraints for multiproduct pricing optimization. Using an association rules approach we identify what business rules have been consistently applied before and evaluate which ones are associated to better business performance. Based on these pricing rules, we build a data-driven set of feasible prices and combine it with standard price optimization routines to understand how this information can complement the traditional econometric analysis of the demand. When combining our methodology with different approaches to estimate customer responses to price changes, we found that simple demand models such as the double log model are very sensitive to the definition of the feasible set. More sophisticated approaches such as Lasso or the Hierarchical Bayesian Model are less sensible to the addition of more linear constraints, but still are affected by the definition of the feasible set. Adding more constraints to the feasible set can only lead to smaller values of the profit function. Interestingly, our numerical results indicate that the methodology does not only leads to more realistic price solutions, but they are also more robust to data variations leading to better business performance.

The second essay presents a machine learning approach to study interrelationships among product categories and to detect latent shopping motivations. We rely on Latent Dirichlet Allocation (LDA), which has been widely used in text mining to extract topics from documents, measuring the probability of co-occurrence of words. In our retail context, we will extract latent shopping motivations instead of latent text topics, analyzing relationships among product categories, instead of words, in a transaction instead of a document database. The contribution of this research is then to apply LDA in a retailing setting, modifying the basic model to achieve three different goals as follows. First, a standard LDA model will be used to detect and describe motivations. Second, the basic LDA model will be extended to jointly estimate the latent shopping motivations and the relationship between these motivations and basket size using a supervised approach. Finally, the LDA model will be generalized to allow purchase motivations to depend on customer and shopping trip characteristics.

In sum, the results proposed in this thesis provide novelty solutions based on customer preferences to support price and promotions decisions for practitioners and managers in the retail industry.

Gracias a la familia que me crió, y a la nueva que viene...

Acknowledgements

La última vez que entré a la Universidad a estudiar fue hacía 20 años. Volver a hacerlo después de ese tiempo era un desafío importante que implicó audacia, romper temores y mucho esfuerzo. Quiero agradecer a mucha gente que me acompañó en este lindo y duro viaje del doctorado. Primero a mi familia, papás, mis sobrinos, mi futura señora Paloma. Su amor y apoyo fueron fundamentales en todo momento.

A los miembros de esta comisión que sin su constante soporte no habría encontrado soluciones a los problemas que nos planteamos en un comienzo. Muchas gracias Marcel por tu paciencia y apoyo. Andrés gracias por tu exigencia y guía para llevar a cabo nuestro paper. Richard y Jochen, thank you so much for your valuable feedback and support during the last year. Muchas gracias a todos por la confianza en mi y en los proyectos que llevamos a cabo. Al resto de los profesores del programa y de la facultad, de los cuales aprendí mucho también. Con cariño recuerdo la dedicación y capacidad de los profesores Olivares, Noton, Correa, Ordoñez, Schwartz y Montoya.

Gracias a mis compañeros del doctorado. Muchos de ellos de otra generación, a alguno le hice clases años atrás, fueron apoyo de los frustrados códigos y problemas que teníamos en nuestros trabajos. Son gente linda, sencilla, sana, que estoy seguro serán grandes académicos en el futuro. En particular gracias a Dana, Vero, Edu, Edu2, Cristian, Renny, Alvaro, prof Bucarey, prof Canales, Javier, Valeria, Ricardo, Coni, Pablo y Seba. Particular agradecimiento a mi compañero de oficina Daniel Yung, por su apoyo y amistad todo este tiempo. Gracias también al Instituto Sistemas Complejos de Ingeniería, a Carla y Alejandra. Su apoyo para presentar nuestros trabajos en el extranjero y hacer proyectos aplicados fue fundamental en el desarrollo de esta tesis. Gracias al Estado de Chile que a través de Conicyt y su programa de Doctorado Nacional, me financió la beca 21150091 y el costo del programa.

Gracias al Departamento de Ingeniería Industrial y a la administración del programa. Gracias a Linda, Pablina y Fernanda que me ayudaron y salvaron en miles de ocasiones durante estos años. Gracias enormes por su cariño y dedicación.

A wonderful experience inside the PhD program was living abroad. Thanks to the Needs Project I could live in Belgium for five months working at KULeuven doing a research internship. This was one of the most valuable experience in the PhD. I knew wonderful people, had interesting talks, and learn a changing experience. Particular thanks to the Professor Jochen De Weerdt and his team at Liris Center (Monique Snoeck and Bart Baesens).

Table of Contents

List of Tables

List of Figures

Introduction

The recent introduction of new technologies to deal with big data problems have created several business opportunities across different industries (Agarwal and Dhar, 2014). In particular, retail companies register a huge amount of everyday decisions made by their customers that face every product categories to make their shopping (Erevelles et al., 2016). This information have been used to build statistical models to uncover consumer preferences and support several marketing decisions for product category management, such as price, promotions, assortment, layout, among others (Bradlow et al., 2017). Nevertheless, the application of these statistical models are difficult to scale to industry problems and hard to understand or apply by practitioners(Lilien et al., 2013; Lilien, 2011). On the other hand, machine learning methods have been used in many applications to predict consumer behavior like CRM, customer lifetime value prediction, churn detection, cross selling, customer acquisition, among others (Ngai et al., 2009). Machine learning tools are becoming popular because of their powerful capacity for handling great amount of cases (transactions, customers) and considering many variables (attributes, products for example) to solve many predictive problems. Recent literature in Marketing (Chintagunta et al., 2016; Desai et al., 2012; Hauser et al., 2006) has pointed out there is a great opportunity to use these novelty developments in machine learning to solve traditional marketing problems, improving the scale, dealing with millions of customer transactions, and the capacity to be applied and used for operational and tactical decisions by practitioners and business managers. This thesis consider two chapters about retail management problems covered with machine learning approaches. Both methodologies use transactional data to discover consumer behavior and complementary and substitution patters between products, brands and categories.

The first chapter is about pricing decisions. Multiproduct pricing is one of the key decisions in retail management to improve category performance. Traditional approaches for optimizing prices involve the use of transactional data to estimate a demand system as a function of historical prices and then, conditional on the demand system, store managers can decide all the product price levels in the assortment to maximize profits over the complete product category.

Category pricing presents several challenges. For example, to capture cross-elasticities, demand systems tend to be highly parameterized and potential endogeneity of prices might lead to parameters with signs different than expected. Additionally, when solving for optimal prices, first order conditions frequently suggest extreme solutions which are beyond of what managers consider reasonable. Oftentimes, to find reasonable solutions, managers rely on business rule to guide the solution and incorporate some tactical and strategic factors that cannot be explicitly included in the demand model. In this research we analyze transactional data to identify what business rules have been consistently applied before and evaluate which ones are associated to better business performance. Based on these pricing rules, we build a data-driven set of desirable prices and combine it with standard price optimization routines to understand how this information can complement traditional econometric analysis. To deal with the computational burden of identifying a potentially large set of rules, we adopt the Apriori algorithm developed in the context of market basket analysis.

When combining our methodology with different approaches to estimate customer responses to price changes, we found that simple demand models such as the double log model are very sensitive to the definition of the feasible set. More sophisticated approaches such as the L1 Regularization or the Hierarchical Bayesian Model are less sensible to the addition of more linear constraints, but still are affected by the definition of the feasible set. While more constraints can only lead to smaller values of the profit function. Our numerical results indicate that the methodology does not only leads to more realistic price solutions, but they are also more robust to data variations leading to better business performance. In summary, in this research we present a novel approach to pricing optimization. This approach is easy to implement and not only provide mangers with a reliable automatic mechanism to decide about prices of multiple products, but also more consistent decisions.

The second chapter use machine learning to analyze consumer behavior and detect shopping trip motivations. In this context,

Both essays give a novel use of popular machine learning methods to solve price and promotions decisions in the retail industry. The proposed methodologies have two important advantages: they are simple to understand for managers using rules and maps as tools to show consumer behavior patterns to profit in the supermarket. Also, these methodologies are based in popular machine learning methods implemented in most of the statistical software in the industry, so their implementation should be easy for practitioners in the industry. Both methods proposed give new approaches to apply machine learning algorithms in marketing decisions, using transactional information to extract consumer preferences to support and improve business results for category management.

Chapter 1

Category Pricing Optimization Using Data Driven Constraints

1.1 Introduction to Multiproduct Pricing

Pricing is one of the most relevant decisions in marketing and category management in the retail industry (Roberts et al., 2014). Specifically, supermarket chains have to set prices for many products over all the stores every week. Also, these prices change over time due to promotions originated from suppliers as well as the retail chain itself to shift traffic and sales. In this context, pricing decisions made by category managers are complex not only because of the large number of decisions to make, but also because retailers want to use explicit relationships between the products inside a category.

Price affects not only the demand of the product itself, but also the demand of complementary and substitute products. These “side effects” in product promotions, also described in literature as cannibalization, can reduce profit performance in the total category (Walters, 1991). Cannibalization happens when a promotion is defined for a specific product with average margin, increasing its sales, but changing and replacing consumer preferences from high value products to this promoted average profit product. These promotions can reduce overall performance of the category due to substitute patterns between competitive brands. Similarly, price reduction at product level must consider substitution patterns with other product prices with different format of the same brand. For example, products with larger formats should have lower price per unit of weight of volume than smaller products (e.g. orange juice of 500 ml should cost more per ml than the same orange juice brand of 1 Lt.). This case shows how individual promotions break consumer incentives for buying larger formats, producing other sources of cannibalization in the category performance. This is a good example of the operational complexity present when dealing with pricing changes in every day price definition. Managers can easily fall in these pricing mistakes because retailers deal with many products, stores and weeks to set prices. A typical supermarket chain must deal with 60,000 different products (SKU’s) X 50 stores X 52 weeks = 156 millions of weekly pricing decisions every year. This is a measure of the high complexity that category mangers face in operational pricing decisions.

To solve this pricing problem, research uses the vast amount of data available in retail transactions. Scanner data has valuable transactional patterns that can lead to support many category management decisions, specifically pricing (Bucklin and Gupta, 1999). In particular, retail companies record a huge amount of shopping decisions made by their customers. (Erevelles et al., 2016). This information has been used to build statistical models to uncover consumer preferences and support several marketing decisions for product category management, such as price, promotions, assortment, layout, among others (Bradlow et al., 2017).

Most traditional researches about pricing use this data available to calibrate sophisticated econometric models and measure how prices affect demand and therefore category profit. The hard task is to construct an efficient and general model to include most relevant factors that affects demand, especially own and cross price sensitivity effects. Using the expected profit as an objective function into an optimization framework, we can find the optimal prices that maximize the category performance.

1.1.1 Practical problems in pricing optimization

The application of these statistical models are difficult to scale to industry problems and hard to understand or apply by practitioners (Lilien et al., 2013; Lilien, 2011). Multiproduct pricing presents several challenges. To capture cross-elasticities, demand systems tend to be highly parameterized and potential endogeneity of prices might lead to parameters with signs different than expected. Additionally, when solving for optimal prices, first order conditions frequently suggest extreme solutions which are beyond of what managers consider reasonable.

Oftentimes, to find reasonable solutions, managers rely on business rules to guide the solution and incorporate some tactical and strategic factors that cannot be explicitly included in the demand model. For example, decision makers can require that every price must remain in a certain distance of its historical level or impose a maximum deviation from the weighted price average. they add business rules and ad-hoc constraints to limit the movement of pricing to stay in a vicinity the original price vector (10% for example). Nevertheless the optimal solutions are very sensible to these restrictions because constraints are frequently binding. An important and central problem of this research is how to define these constraints to use in the optimization framework. Essentially, these rules are subjectively defined and as they limit the price space they also restrict the maximal profit to achieve. It would be desirable a data driven approach to define these constraints or rules that prices should respect and improving the optimal solution.

A second problem is the robustness and stability of the solution proposed. Depending on the values estimated in the cross elasticity matrix, the optimal price vector solution for the category can be very different with limited perturbation in the elasticity matrix. An example can be found in Figure 1.1 with simulations obtained from demand parameters estimated in Montgomery (1997) for a subset of orange juice category products. Each point in the scatter plot shows the optimal price solution using a draw of the probability distribution estimated for the cross elasticity matrix in the Hierarchical Bayesian demand model. The optimization was constrained using as bounds a 1% and 400% of the original prices. The diagonal of the cross scatter plot shows the frequency distribution of the solution. As the reader can notice, an important fraction of the optimal solutions are bounded, so the optimal value depends on how constraints are defined. This problem defines another performance criterion related to how many of the optimal prices get a corner solution, especially conditioned to the constraints already defined.

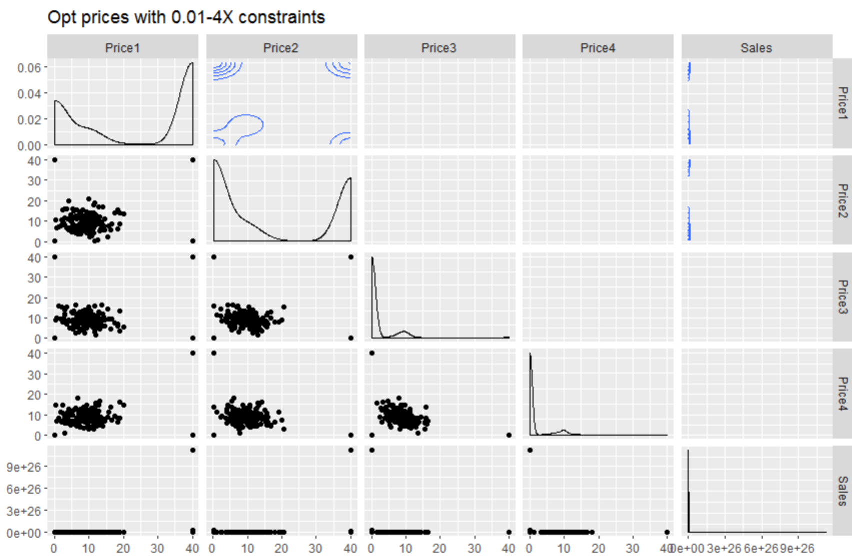

Figure 1.1: Optimal prices scatter plot using constraints between 40% from original prices.

To address these problems of uncertainty in the solution obtained, most relevant research has focused on improving the estimation of the elasticity matrix or improving the formulation of the demand models. For example, Montgomery (1997) uses hierarchical bayesian modeling pooling information from all stores to strengthen the estimation of elasticity parameters. Manchanda et al. (1999) uses a Multivariate Probit method to decompose the relationship between two product of a category, identifying price effects, natural cross selling and others consumer specific patterns. As we already mentioned, most of the commercial software of pricing introduces the use of constraints to pricing decisions into the optimization framework (Kopalle et al., 2009). These rules are known as pricing rules and work as a method from practitioners to introduce business knowledge to avoid non sensical optimal solutions, very far from actual values (Kunz and Crone, 2015).

Looking for a good set of constraints for this pricing optimization problem could be a hard problem due to the large space of solutions. Also, there is an important problem when evaluating millions of supermarket transactions to compare pricing between products to evaluate rule performance. Machine learning appears as a good solution for this problem because of their powerful capacity for handling great amount of cases (transactions, customers) and considering many variables (attributes, products for example) to solve many predictive and search problems.

While almost all marketing and economics literature on retail pricing has focused on deriving precise estimates on the underlying demand system that determine the profit function π, very little work has been conducted to discuss how constraints in the optimization setting should be defined and what is its role in pricing decisions. The contribution of this article goes exactly in this direction. We describe a data-driven methodology to help managers to identify business rules that are not only consistent with their strategic view of the market, but that can also lead to better profits.

1.1.2 Machine Learning and data driven pricing rules

The recent introduction of new technologies to deal with big data problems have created several business opportunities across different industries (Agarwal and Dhar, 2014).

Machine learning methods have been used in many applications to predict consumer behavior like CRM, customer lifetime value prediction, churn detection, cross selling, customer acquisition, among others (Ngai et al., 2009). Machine learning tools are becoming popular because of their powerful capacity to handle great amount of cases (transactions, customers) and considering many variables (attributes, products for example) for solving many predictive problems. Recent literature in marketing (Chintagunta et al., 2016; Desai et al., 2012; Hauser et al., 2006) has pointed out that there is a great opportunity to use these novel developments in machine learning to solve traditional marketing problems, improving the scale, dealing with millions of customer transactions, and the capacity to be applied and used for operational and tactical decisions by practitioners and business managers.

Using this motivation, this article proposes a methodology to analyze large scanner data and extract pricing patterns in an inexpensive and efficient way adapting a machine learning method called association rules. This approach extracts and represents demand patterns that describe the effect of pricing in a profit category. This methodology, in a simple and reliable way, discovers the interactions between products in order to detect high profit price space, and also gives insights to the managers in the form of what-if scenarios with the overall category assortment when they change a specific product price. The proposed methodology will look for two types of pricing rules. First, a set of rules that are consistent with previous decisions made by managers. In other words, we want rules with enough significance to support valid relations and especially, to generalize demand patterns out of the training dataset. Second, it identifies pricing rules that find a strong correlation between a price structure and profit scenarios greater than the average. When we mention price structure we are talking about rules for the category mean, or a specific product, or a constraint related to the price difference between a couple of products of the category assortment.

The proposed methodology in this work extracts decision rules and constraints for prices based on previous transactions. This method will define rules between product prices in order to obtain profitable scenarios. The approach proposes a mechanism to systematically look for a good set of constraints between prices that produced good performance in the transaction history of the category. In other words, the methodology presented will get pricing rules to avoid scenarios of bad performance in a category, like cannibalization or breaking incentives for larger formats.

Some of the potential advantages of using pricing rules as a methodology to describe and optimize pricing decisions are:

- Intuitive and easy to understand: Compared with own and cross elasticity matrix, pricing rules provides a more simple, intuitive and comprehensive way to understand a relation between prices and the category performance inside the category. Pricing rules works as price scenarios comparisons for individual products or pairs of products.

- Easy to implement and actionable: Pricing rules can be considered as constraints of feasible prices in the category. They can be easily applied to a transactional database using structured languages such as SQL, so they are simple to implement and to check or validate with new data.

- Easy to validate and compare with business intuition: Pricing rules are easy to check and confront with previous business rules inside the category in terms of what product, brand, format are more valuable than other.

- Easy to extract from scanner data: Using the association rules framework, it is easy and fast to extract thousands of pricing rules, sort them to select most relevant and valuable rules to apply in a certain category.

- Robustness: Pricing rules are build analyzing millions of transactions comparing performance category for different pricing scenarios in the historical database. Therefore, pricing rules extracted are robust based on historical performance of different pricing policies and their impact in the category demand.

When combining our methodology with different approaches to estimate customer responses to price changes, we found that simple demand models such as the double log model are very sensitive to the definition of the feasible set. More sophisticated models such as the LASSO or the Hierarchical Bayesian model, are less sensible to the addition of more linear constraints, but still are affected by the definition of the feasible set. Adding more constraints to the feasible set can only lead to smaller values of the profit function. Therefore if our construction of the feasible set is an accurate representation of what managers are really willing to consider, then the reduction in profit derived from adding data-driven constraints gives an estimate of the overestimation of optimal profits when using unconstrained profit optimization. Interestingly, our numerical results indicate that the methodology does not only lead to more realistic price solutions, but they are also more robust to data variations.

In essence, in this research we analyze transactional data to identify what business rules have been consistently applied before and evaluate which ones are associated to better business performance. Based on these evaluations, we build a data-driven set of feasible prices and combine it with standard price optimization routines to understand how this information can complement traditional econometric analysis of the demand.

The rest of the article is structured as follows. First, the paper describes in detail a literature review focusing on the pricing approaches available and its limitations. We also review bibliography related to pricing rules and the machine learning methods we will use to extract data driven pricing rules. In the next chapter we describe the methodology proposed adapting the market basket analysis framework to correlations between price structure and profit performance. We apply the methodology proposed to a supermarket transactional database analyzing a specific product category. The results maximizing profit performance using different constraints and scenarios are evaluated. Finally we draw some conclusions from the results and limitations about the proposed methodology.

1.2 Literature Review

To organize the literature review, we revise previous research on three main topics. First, we explore how other investigations have proposed different optimization methodologies to solve the multi-product problem. Second, we explore how the concept of pricing rules has been used in previous work. Finally, we explore the machine learning methodology and the data driven approaches we will use to extract pricing rules.

1.2.1 Category Pricing Optimization

Pricing problem has been addressed from many perspectives and researches including dynamic pricing, psychological and consumer behavior motivations, long term and brand effects, among others Grewal et al. (2011). In this research we focus on multiproduct pricing where there is an important interrelation effect between products inside a category. According to Manchanda et al. (1999), these effects have many different sources: the first one is related to product prices affecting demand of other products inside the category (cross elasticity effects). Second, there is some natural complementary effect between products inside a category (e.g. mayonnaise and ketchup in the dressing category), and finally interrelation effects related to consumer characteristics. This research also estimated that the relevant cross price effects happen inside the category. In other words, other category promotions or price reduction doesn't have a significant impact into the category demand in study. The typical setting where all these assumptions about demand behavior are found is in supermarket product categories. We can assume there is an automatic replenishment and stock outs don't affect consumer choice. As these products are bought regularly, strategic consumer behavior doesn't play an important role affecting demand. All these elements are important to consider when the demand model would be defined.

Several approaches have been proposed in the literature to solve pricing problems. Little and Shapiro (1980) use a theoretical approach to suggest that optimal prices comes from an equilibrium where consumers solve a short term utility maximization to decide which product and how much to buy, and where stores maximize profits subject to utility constraints. Reibstein and Gatignon (1984) solves the pricing problem for multiple brands showing the importance of dealing with cross elasticities and substitute patterns between products. This work has a two-step methodology: first, a product demand model based on the different prices of the category in order to recover own and cross product elasticities. And second, the definition of prices in a product category through a profit maximization setting as defined by equation 1. This two step procedure is a standard in the pricing literature and applications related and we will also adopt it to solve the problem. Montgomery (1997) face the pricing problem using a hierarchical bayesian approach for demand estimation, getting robust estimations of the sparse cross elasticity matrix. Finally, Montgomery and Goic (2010) strengthen the formulation of the bayesian demand model, incorporating information into the prior definitions.

1.2.2 Business Rules for Pricing

There are many examples in literature of the use of business rules as pricing constraints in the optimization problem. Using the two step approach mentioned before, Vilcassim and

Chintagunta (1995) works with scanner data to calibrate a Multinomial Logit model for the yogurt category. This is one of the first papers using the concept of "pricing rules" defined as rules of thumb for constant markup for each brand defined by the supermarket.

Hawtin (2003) discuss why different business rules are needed when solving practical pricing problems and evaluate how the optimal solutions are affected by those rules. Kunz and Crone (2015) extend the work of Hawtin highlighting two elements. First, a side benefit of using business rules is the adherence of the optimal solution to business. And second, they evaluate the economic impact of including business rules in a multi-objetive settings. Ferreira et al. (2015) set pricing rules to further ensure the customer is getting a great deal, restricting the upper bound to be no more than the maximum of $15 or 15% greater than the lower bound. Natter et al. (2007) discuss the automatic use of pricing models requires additional business rules. Their price recommendations, for instance, are restricted to a maximum of 15% price increases in each pricing round. Levy et al. (2004) discusses the use of pricing rules in retailing and describe the limitations of its use, including the slow update to more recent competitive scenarios and the substantial deviation from the optimal solution.

A common practice in the retail industry is to use heuristic rules to directly determine prices. For example markup where prices directly track marginal costs, or competitive prices where prices closely follow prices from competitors (Levy et al., 2004). It also describes that optimization routines for pricing includes business rules to restrict the search space to price solutions. Ma and Fildes (2017) describes another application using business rules. It is conceptually similar to Cohen et al. (2017), but they claim that in the demand function they use cross-product influences. From a methodological perspective, it is interesting that they use a regularized estimators to ensure realistic promotional parameters. While not discussed explicitly, they realize that price solutions are sensible to parameter estimates, which is one of the motivations for using business rules. From a marketing perspective it is interesting to note that they consider not only prices as decision variables, but also feature and display. Beyond pricing, it is well established that managers learn from past experience and they translate such knowledge into simple rules (Bingham and Eisenhardt, 2011).

Another good example of pricing rules is found in Cohen and Perakis (2018). They use use (i) discrete price ladders, (ii) limited time of promotional discounts for a given time and (iii) minimum separating periods between consecutive promotions. These rules are now implemented in Oracle Retail Pricing Tools (Oracle, 2018). Cohen and Perakis (2018) extends the set of business rules already considered by Cohen et al. (2017). For example this consider a set of products that should be discounted together, spacing out promotions of sets of similar items by a minimal number of separating periods.

The use of business rules is not confined to academic research and it is indeed a widespread practice in the retail industry. In fact Rosenberg and Sills (2019) propose a method to identify business rules that might be implicit in historical pricing (this is fairly similar in spirit to this research but only looking price information, not its performance in profit). Here rules are classified in bounded rules providing ranges for individual prices and comparative rules imposing restrictions between families of products. Among the rules considered they include how closely the retailer should match prices from the competitor and dynamic rules referred to the maximum and minimum deviation from a current price to be considered. Also Rosenberg

(2018) propose a revenue optimization approach to decide prices while complying with a set of business rules. Moreover, they assign a monetary value to each business rule to decide whether they should be considered in the decision process. Handorf et al. (2016) does not deal with pricing rules, but with product design rules. It can be used to say that "while we focus on pricing rules, the conceptual framework can be expanded to other types of business rules (e.g. product and channel coordination)".

1.2.3 Association Rules

The methodology proposed in this paper for extracting profitable pricing rules in a product category is based on a machine learning approach called Association Rules. This framework proposed by Agrawal and Srikant (1994) is one of the most popular and most cited data mining algorithms and it can be used for describing relationships between attributes or characteristics in a database. The most common application of this model is the use of association rules in a retail transactional database or scanner data, to extract purchasing rules between products which might be, useful for creating bundles of products or cross product promotions (Han et al., 2011) . In the market basket analysis framework, the main idea is to extract "interesting" rules between complementary products (e.g. if I by pasta, it is likely that I will also buy tomato sauce). Association rules uses different statistical measures in order to capture how interesting a product combination is. There are other applications of association rules using spatial (Han et al., 1997), temporal (Han et al., 1999), intrusion detection (Treinen and Thurimella, 2006) and web log databases (Mobasher et al., 2001). For a recent survey of association rules and its applications see Kotsiantis and Kanellopoulos (2006).

The problem to evaluate complementarity between products in a supermarket is combinatorial and computationally expensive as the number of products increase in the supermarket. To identify importance and level of complementarity of each solution in the rule space, Agrawal and Srikant (1994) and Brin et al. (1997) propose a set of performance metrics to evaluate the predictive value of a rule as described in table 1.1. A, B, and Z are products. N_z is the number of transactions with Product Z, and N is the total number of transactions.

Measure	Expression
Support	$Supp(Z) = \frac{Nz}{N}$
Confidence	$Conf(A \to B) = \frac{Supp(A \cap B)}{Supp(A)}$
Lift	$Lift(A \to B) = \frac{Conf(A \to B)}{Supp(B)} = \frac{Supp(A \cap B)}{Supp(A)Supp(B)}$

Table 1.1: Interesting measures for association rules

Support measures the purchase probability of a product given itemset. Confidence is the conditional probability of buying product B given a purchase in product A. Lift measures the increase in probability of buying one of the products, when the other is present in the basket.

As previously mentioned, exploring and extract association rules between products is a combinatorial and time consuming problem. Agrawal and Srikant (1994) proposed a fast method to extract interesting rules condition to a minimal support. Years later, Han et al. (2000) proposed FP-Growth, a tree exploration method to improve efficiency and time for extracting rules.

We will translate the traditional market basket analysis setting into the pricing optimization, to extract instead of rules between products, rules between prices and profit performance of the category.

Using association rules we can extract rules between specific price structure and profit scenarios better than the average, and with enough data support to have significant and robust results. These pricing rules can be used as constraints in the price optimization setting in order to guide the search of optimal prices to good price spaces with good probable solutions, and also avoiding or removing from feasible space, scenarios with lower profit than the average.

1.3 Methodology

For simplicity, in our methodology we consider that the set of business rules Ω is a convex polytope and therefore we restrict our attention to business rules that can be expressed as linear inequalities. Moreover, we decompose the set of feasible prices in two subsets Ω_o and Ω_p such that $\Omega = \Omega_o \cap \Omega_p$. Here, Ω_o corresponds to the set or order rules that consider relationships of prices that are consistently present in the transactional data, but that not necessarily lead to better performance. For example, if a product category contains regular and premium brands, we would observe that on most days the price of a regular brand p_r is lower than the price of a premium brand p_p. In that case, we will consider that $p_r \leq p_p$ for all p in Ω_o. While most of these constraints might be already known by managers, some of them might not be obvious and our methodology facilitates to organize a potentially large number of such rules. The set Ω_p contains what we label as performance rules. These rules correspond to relations of prices that are only present in a fraction of cases, but when active lead to better profits. For example, if we observe that when the average prices of brands A and B exceed a certain threshold τ sales and profits of the category drops, then we include the corresponding constraint and say that $p_A + p_B \leq 2\tau$ for all p in Ω_p. We will measure separatelly the performance of this two kind of rules and also we will analyze how they are selected to define feasible prices.

Given the large number of potential inequalities that can be added to the feasible set, the problem of identifying and evaluating business rules can be computationally demanding. However, a number of efficient methods such as the Apriori (Agrawal and Srikant, 1994) and the Dynamic Itemset Counting algorithms have been developed to find such association rules in the context of market basket analysis. In this article we apply such developments to populate the feasible set and evaluate its impact in retail pricing. In this context, we search for price rules of the form $p_{\mathrm{i}} \leq p_j$ and its relationship with events of good profit performance. In a general form, this pricing rule has form $(a'P \leq b) \longrightarrow (\pi(p_{\mathrm{i}}, p_j) > \overline{\pi})$ and evaluate them using metrics of support, confidence and lift. While we require that rules in the performance set Ω_p have relative high values for all three metrics, order rules in Ω_o only requires support being close to hundred percent of the cases. To evaluate the impact of using data-driven constraints in category pricing optimization, we apply the aforementioned methodology on two years of transactional data for several categories in a supermarket chain. In all categories we identified several order rules showing that the proposed methodology is effective in automatically identifying relevant facets of market structures. Similarly, the methodology also provide numerous performance rules. Importantly, those rules have explanatory power in identifying price solutions associated to larger profits. In fact, in a validation test, price solutions in Ω_p exhibit 30% more profits than solutions that violate those prescriptions. Moreover, in some cases, the set of business rules derived from our methodology is dense enough to guarantee that any feasible price solution is on average only 28% from the optimal solution.

1.3.1 Association Rules for Pricing

The idea is to explore interesting pricing rules between sets of products that compete with each other in a certain category, which can produce a good or bad performance in terms of revenue. The basic structure to explore price relations and get interesting measures is given by:

$$(a'P \leq b) \longrightarrow (\pi(p_i, p_j) > \overline{\pi})$$

The methodology will search across weekly observations of prices that produce a combined profit greater than the historical average of profit of these two products. The interesting association rule to obtain is a specific comparison between a couple of products inside a category (price of product i greater than price of product j for example), with a profit performance greater than the average for this couple of products. The intuition behind the methodology proposed is shown in Figure 1.2. This illustrates a region of profit scenarios across different prices for two products (P1 and P2). Each data-point represents scenarios of low and high profit for different instances of P1 and P2. Blue lines represents the pricing rules, reducing the feasible price space where high profit performance is more likely to happen.

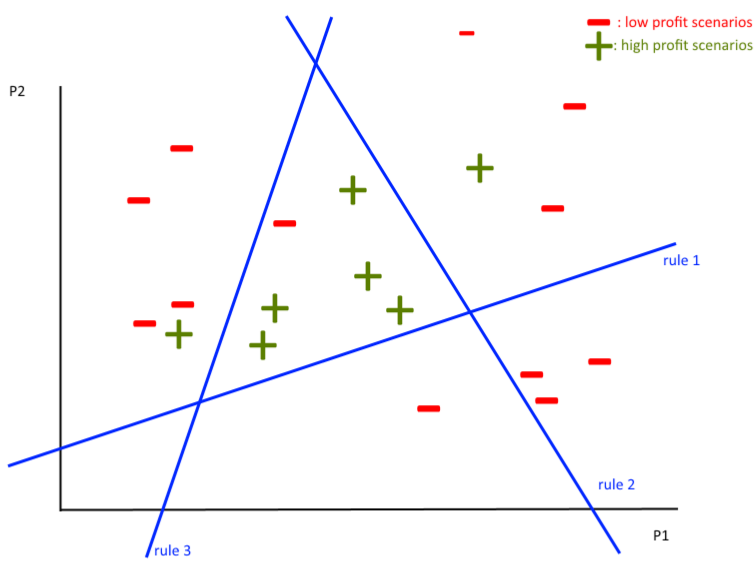

Figure 1.2: Profit scenarios over price space and pricing rules as constraint to allocate better prices to product category.

The approach needs a couple of considerations and constraints for comparing product prices. First, to compare prices of different product formats, prices will be normalized by the size of each product. Second, we consider as a product category, the most desegregated group of products available in the product category aggregation levels.

The evaluation of pricing rules requires and adaptation of the typical measures used in market basket analysis (Tan et al., 2002). In this case we will measure support, confidence and lift as follow:

Pricing Rule Measure	Expression
$Supp(\text{Price}P_\text{i} \leq \text{Price}P_j)$	$\frac{Card(\text{Price}P_\text{i} \leq \text{Price}P_j)}{Card(Totalweeks)}$
$Conf(\text{Price}P_\text{i} \leq \text{Price}P_j \longrightarrow \pi_{\text{i}j} > \overline{\pi})$	$\frac{Card((\text{Price}P_\text{i} \leq \text{Price}P_j) \cap (\pi_{\text{i}j} > \overline{\pi}))}{Card(\text{Price}P_\text{i} \leq \text{Price}P_j)}$
$Lift(\text{Price}P_\text{i} \leq \text{Price}P_j \longrightarrow \pi_{\text{i}j} > \overline{\pi})$	$\frac{Conf(\text{Price}P_\text{i} \leq \text{Price}P_j)}{Supp(\pi_{\text{i}j} > \overline{\pi})}$

Table 1.2: Interesting Measures for Pricing Rules

Please observe that $\pi_{\text{i}j}$ is the profit obtained when Price i is lower or equal than Price j. Support measures the percentage of weeks where the price rule between products is met in the total number of weeks where both products were traded together. This measure will only consider weeks where both products were sold in the transactional historical database. Confidence will measure the probability of having a weekly profit of both products greater than historical average, conditioning that the price rule between products is met, this is $P_\text{i} > P_j$. Lift measures the times that the probability of having $pi_{\text{i}j} > \overline{\pi}$ increases, when $P_\text{i} \leq P_j$ happens.

1.3.2 Pricing Rule Taxonomy

Any price constraint could be evaluated, but we require an explicit definition of the pricing rule space. We focus on creating rules for single prices, comparing differences between products, not only if $P_\text{i} > P_j$, and rules for the average of all prices of the category assortment. Rules for average price are interesting to keep under control category prices as a group. Single product price rules help to measure performance of ranges of different prices. At last, ranges for differences of prices between two products give a sense of prices of some brand should be larger than others.

We consider the following taxonomy of linear pricing constraints described in table 1.3.

Ranges for average prices	$\gamma^L \leq \sum_\text{i} p_\text{i} \leq \gamma^U$
Ranges for individual prices	$\alpha^L \leq p_\text{i} \leq \alpha^U$
Ranges for differences of prices	$\beta^L \leq p_\text{i} - p_j \leq \beta^U$

Table 1.3: Taxonomy of pricing rules to extract using association rules

We perform a grid search for tuning the parameters $\gamma^L, \gamma^U, \alpha^L, \alpha^U, \beta^L$ and β^U, evaluating support and lift measures in thousands of rule combinations. We discard Order rules with pricing scenarios that have not been used before, and also discard Performance rules related to low profit combination of prices.

The pseudo code of the procedure methodology is in detailed in appendix.

Figure 1.3 shows some example of the pricing rules extracted in the three forms we explore. Two kind of charts represent first, a scatter plot of product category profit versus prices, with each data point corresponds a weekly profit and the observed price. Second row is a density plot of the observed prices. The three columns graph average category price, single product prices and difference between two product prices. For example, let's focus in the two graphs of the middle of the figure. Lower chart represents the density plot of the price of SKU 7027. This give us information about what is the dominion of the observed prices in the transactional history. Depending on that, two pricing rules are defined to define as feasible set, where the 95% density plot is located. The upper graph shows a scatter plot with each point representing a week with the observed price of the product 7027, and the profit obtained that week. The constraints defined the price space where better profits are obtained. These are the performance rules defined for this product.

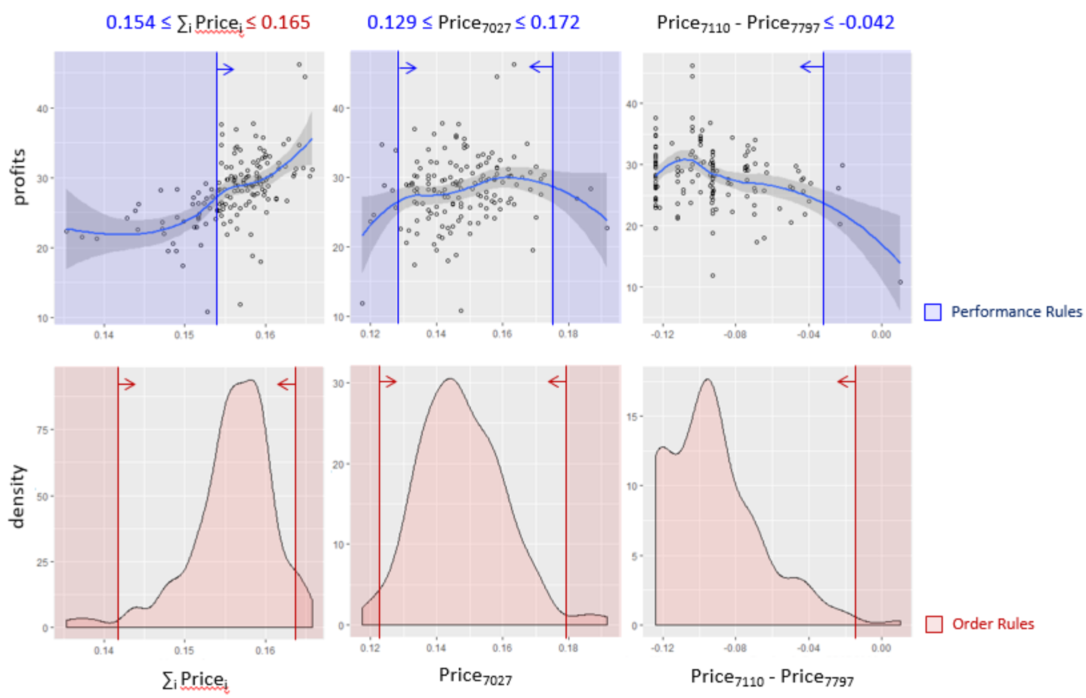

Figure 1.3: Examples of Order and Performance rules for Pricing Optimization

1.3.3 Other alternative methodologies to extract rules

We choose a non supervised approach to extract association rules between price structure and its relation with category profit. If the objective of the project should be to identify and extract rules to detect and predict good profit scenarios, a supervised approach should satisfy and probably get better solutions. There are many decision trees approaches (CART, C5.0, Chaid) (Han et al., 2011) that also can extract robust rules between predictors (product prices in our case) and a supervised label variable (category profit performance). An example of a decision tree is in figure 1.4.

The problem with this solution is that each leaf node in the tree that predicts expected profit, is a result of different sets of simultaneous pricing rules defining different disjoint feasible scenarios. It is clear that the implementation of this kind of pricing feasible solution

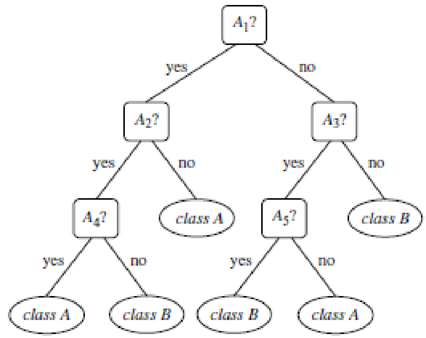

Figure 1.4: Supervised decision tree example. Source: Han et al. 2011.

into the price optimization setting is more complex and needs auxiliary variables. Also, our method delivers not only pricing rules correlated with category profit, but also order rules extracting previous decisions and implicit knowledge of pricing management for brands and formats. This is an important advantage of our non supervised approach compared with decision trees.

The association rules methodology is an efficient method to search for correlation in a high dimensional space. We use this method to find what price ranges and combinations of prices are correlated with profit scenarios greater than the average. The problem is this correlation is not enough. Maybe the high correlation scenario is produced due to many other factors beside prices. For example: apparel promotions, stock outs, store competitor promotions, holidays or seasonality, among others. The main goal is to identify the true causality of how prices affects the demand and therefore, the overall profit of the product category. Also, prices are decisions that are inducted by other information, so can this decision is an endogenous variable that makes it hard to find a causal relationship between prices and profit. This is certainly a limitation of our model and also for any structural model that tries to find the true relationship between prices and profit. One of the most popular approaches to find the true causality (Arora et al., 2008) is to perform a experimental design procedure to explore randomly different price structures and measure its effect into demand and total category performance. In this sense, these pricing rules are a good starting point to perform experimental design. There were pricing rules discarded with good lift indicating good profit scenarios, but didn't have enough support or information to be a reliable pricing rule. This set of pricing rules are a good starting point to explore or conduct experimental design in order to confirm whether the rules have enough support to validate a good profit scenario.

1.4 Empirical Application

Here we describe the results obtained applying the proposed methodology to extract pricing rules and optimal prices for a specific category of a supermarket chain. First, we describe the data available and some descriptive statistics about prices and demand of the category analyzed. Then we apply the methodology and we extract different types of order and performance rules. We evaluate characteristics of them and their generalization capability into a test dataset. Finally, we apply sensitivity analysis of the optimization results conditioned to threshold parameters and also evaluating robustness for the optimal solution obtained.

1.4.1 Data description summary

Our approach uses aggregated sale data which is easy available for most of retailers. To illustrate that we apply the methodology to a supermarket transactional database provided by Dunnuhumby[1]. The total database has information about 79 different stores during three years (156 weeks). We focus our analysis on the Pretzels category and a single store, but can be easily extended to more stores. The database contains also information about unit sales, household visits, aggregated at weekly and store level. It also includes information about temporary price discounts and promotions.

SKU ID	Product Description	Share	Normalized price per unit		
			Mean	Min	Max
1111009477	PL MINI TWIST PRETZELS	16.5%	0.10	0.08	0.11
1111009497	PL PRETZEL STICKS	13.9%	0.10	0.08	0.11
1111009507	PL TWIST PRETZELS	5.3%	0.10	0.08	0.11
2840002333	RLDGLD BRAIDED HONEY WHT	6.7%	0.29	0.20	0.33
2840004768	RLDGLD TINY TWISTS PRTZL	5.4%	0.18	0.12	0.21
2840004770	RLDGLD PRETZEL STICKS	3.5%	0.18	0.12	0.21
7027312504	SHURGD PRETZEL RODS	6.4%	0.15	0.12	0.19
7027316204	SHURGD MINI PRETZELS	5.8%	0.11	0.08	0.13
7027316404	SHURGD PRETZEL STICKS	5.9%	0.11	0.09	0.14
7110410455	MKSL MINI TWIST PRETZELS	2.2%	0.14	0.13	0.17
7110410470	MKSL DUTCH PRETZELS	1.7%	0.14	0.12	0.17
7110410471	MKSL PRETZEL STICKS	1.7%	0.14	0.13	0.17
7797502248	SNYDR PRETZEL RODS	9.7%	0.24	0.15	0.25
7797508004	SNYDR SOURDOUGH NIBBLERS	8.5%	0.18	0.15	0.21
7797508006	SNYDR FF MINI PRETZELS	6.9%	0.18	0.15	0.21

Table 1.4: Descriptive statistics of Pretzels product category.

Using this information we calibrate different demand models in order to estimate price elasticity for each product of the category.

1.4.2 Extracted pricing rules

The apriori algorithm searches for relevant rules in terms of support, confidence and lift. To do that, we apply some thresholds to select interesting rules. First, we want some minimum support to obtain reliable and significant relationships between prices and category profit. To do so, we will discard rules that happen less than 5% of the total time. The second filter we apply is for the profit lift. We want pricing scenarios that improve the category profit by at least a certain threshold. An example of selected and non-selected rules are described in table 1.5.

[1] `https://www.dunnhumby.com/careers/engineering/sourcefiles`

Each row represents a rule related to the difference in price comparing a specific pair of products. β^L is the price cut value in the domain of price difference where the rule will be evaluated. With this value, the continuous variable (price difference) is transformed into a price event (whether the price difference is greater than the cutting point β^L has happened or not). Support measures the percentage of weeks where the pricing rule happens, in this case 95% of the weeks in the calibration set. The fourth column reports the average profit of the category when the pricing rule happens. Lift represents the times the profit increase compared with the overall profit. The last column shows which rules were selected or not depending on whether the value of lift is greater than the threshold set. For this base case scenario the lift threshold was 1.02.

Table 1.6 shows the number of performance rules evaluated and selected for each type (sum, individual and pair comparison between prices). We also report the average profit of the rule.

For example, the fourth row in table 1.5 describes the difference in price between products '2840004770 - (RLDGLD PRETZEL STICKS)' and '7027316404 - (SHURGD PRETZEL STICKS)', comparing two brands of pretzel sticks. The rule evaluated is the difference of this two products is greater than 0.1 (β^L). This rule happens in 95% of the weeks analyzed. The overall category profit when the rule happened was 35.09, and it is 25% greater than the average weekly profit observed in the transaction database.

Pricing rules not only help the pricing optimization to get a more robust solution, but also provide meaningful information about product features and comparison between them. Using the same example, it is providing information that there is a good profit scenario when the subtraction between RLDGLD PRETZEL STICKS and SHURGD PRETZEL STICKS prices, is greater than 0.1. This rule gives intuition about brand preferences and up selling patterns that can lead to improve profit scenario.

One important question to answer related to the pricing rules extracted is about the generalization capacity to find the same structure between prices and profit performance, in other datasets, stores or time frames. To evaluate this we calculate the average profit for the performance rules selected in a train dataset based of 80% of the weeks and compare it with the profit in a test set with the 20% remaining weeks. Results are in table 1.7. Performance

Prod_i - Prod_j	β^L	$\text{Sup}(P_i - P_j > \beta^L)$	$\Pi(P_i - P_j > \beta^L)$	$\text{CONF}((P_i - P_j > \beta^L \to \Pi \geq \overline{\Pi})$	$\text{Lift}(\Pi(P_i - P_j > \beta^L))$	Selected
1111009497 - 7110410455	-0.02	0.95	46.24	1.00	1.64	1.00
1111009497 - 7110410471	-0.02	0.95	46.24	1.00	1.64	1.00
1111009497 - 7110410470	-0.02	0.95	37.41	0.33	1.33	1.00
2840004770 - 7027316404	0.10	0.95	35.09	0.11	1.25	1.00
2840004770 - 7110410470	0.08	0.95	34.72	1.00	1.23	1.00
7797502248 - 7110410470	0.12	0.95	34.72	1.00	1.23	1.00
7797508004 - 7110410470	0.08	0.95	34.72	1.00	1.23	1.00
1111009477 - 7110410455	-0.02	0.95	34.30	0.11	1.22	1.00
2840004768 - 7797502248	0.00	0.95	22.98	0.02	0.82	0.00
2840004770 - 7797502248	0.00	0.95	22.98	0.02	0.82	0.00
7797508004 - 7797502248	0.01	0.95	22.98	0.02	0.82	0.00
7797508006 - 7797502248	0.01	0.95	22.98	0.02	0.82	0.00
7110410455 - 7797502248	-0.04	0.95	22.87	0.02	0.81	0.00
7110410470 - 7797502248	-0.04	0.95	22.87	0.02	0.81	0.00
7110410471 - 7797502248	-0.04	0.95	22.87	0.02	0.81	0.00

Table 1.5: Selected rules for differences between products

Performance rules	# rules extracted	# rules selected	% selected	Average Profit extracted	Average Profit selected
Price Mean	8	2	25%	28.34	31.407
Individual Prices	150	14	9.3%	29.16	30.23
Pairwise difference	2250	78	3.5%	28.05	29.46
Total	2408	94	3.9%		

Table 1.6: Descriptive statistics of performance rules selected

rules selected seems to have consistent correlations between prices and category profit, even in a dataset out of the training set.

Profit mean (est. dev)	Train	Test	% Difference
Selected Performance rules	29.22 (0.79)	30.44 (1.26)	4.16%
All dataset	28.12 (4.95)	28.39 (4.66)	0.96%

Table 1.7: Optimal profit for scenarios considering different sets of Order and Performance Rules.

1.4.3 Pricing Optimization

Now, we use the order and performance rules as constraints for different optimization scenarios. First, we describe different demand models used as inputs in the objective function. Second, we describe and analyze the optimization results obtained using the pricing rules. Finally, we perform some sensitivity analysis changing threshold parameters.

Demand models to evaluate

To evaluate the impact of using business rules, we need to calibrate a demand model to include in the objective function. We calibrate a double log weekly aggregated demand to capture price elasticities. The model and their parameters to estimate will be as follows:

$$ln(Q_{\text{i}}(P_j)) = \alpha_{\text{i}} + \sum_j \beta_{\text{i}j} ln(P_j) + Feature_{\text{i}} + Display_{\text{i}} + TPR_{\text{i}} \tag{1.1}$$

Where Feature, Display and TPR are dummy variables to indicate if a certain week, the product i appeared in store circular, a store promotion display, and a temporary price reduction respectively. This demand function $Q(P_{\text{i}})$ estimated feed the objective function of the following optimization problem:

$$Max_{P\in\Omega}\Pi(P) = \sum_{\text{i}} (P_{\text{i}} - c_{\text{i}})Q_{\text{i}}(P) \tag{1.2}$$

Where the feasible space Ω is defined by order and performance pricing rules extracted.

We use three different approaches to estimate the $\beta_{\text{i}j}$ cross elasticity matrix:

- Ordinary Least Square - OLS: We use OLS to estimate all the cross elasticity matrix parameters.

- Hierchical Bayesian Model - HBM: This model use a pool of information from all the stores to improve the elasticity estimation of the store analyzed.
- Regularization methods with LASSO regression: We add L1 regularization to OLS to include only the most relevant elements in the elasticity matrix. We select the λ parameter using λ_{min} minimizing the cross validation mean square error.

Profit optimization results

Calibrated demand models are the input to estimate the total profit of the product category, conditioned to a specific set of prices. To find the optimal price vector for this multi-product setting, maximize the expected total category profit. Table 1.8 shows the optimal profit results, considering different demand models and different sets of pricing rules used as constraints for the optimization problem. We add as a baseline value the average expected profit obtained for each demand model in the last four weeks of the dataset. We also report the comparison percentage between this profit baseline and each expected profit obtained for each optimization scenario.

To select a specific number of selected rules we need to define some thresholds. For order rules, we choose as initial parameter threshold the [0%-100%] of the domain of observed prices. This means that the feasible price space will be only in a combination of previous observed prices. As we already mention, for performance rules we choose a minimum support as 80% and minimum lift as 1.2.

Thus, we define four scenarios of constraint sets compared in the table. The first two sets represents order rules, these are constraints to reduce the space only where previous prices were observed. The second couple of scenarios add performance rules, these are sets to indicate that the price vector is correlated with a better profit performance than the average. The first scenario considers 30 pricing rules associated to, lower and upper bound for individual order rules. The second scenario adds upper and lower bounds for pairwise differences and also constraints for the observed mean of the category price. This results in a total of 452 rules. In the third scenario 14 performance rules are added to the global set of constraints, restricting individual prices depending on the better scenarios of profit observed in the past weekly category performance. The fourth and last scenario adds performance rules constraining the difference between products and also the total price average of the product category. The total number of constraints or pricing rules for each is also reported.

There are important findings comparing the optimal profit results for each scenario. Let's focus first on the optimal results using the OLS method. As it can noticed the optimized expected profit is far better than actual values represented in the baseline. It seems very unlikely to have this kind of improvement in profit as a result of the price optimization policy. This result is even constrained using only prices that have been previously observed. This is a good example of the hard work that pricing managers have in practice obtaining results weekly, mainly because of the uncertainty related to the estimation of elasticity matrix, and thus, the expected profit. The results have exactly the same problems for different estimation procedures (Hierarchical Bayesian Modelling and L1 regularization). Even though these alternative methods reduce the uncertainty compared with the OLS and get expected profit in the right direction, the incremental percentage of profit is still far from real values, getting unreliable profit and optimal price results.

Looking back the OLS column, the second scenario including all the order rules (mean and pairwise product difference rules) shows an important reduction in the optimal profit obtained. The optimal solution for this scenario is 51.5% better than the previous profit obtained. This result means that the addition of 422 rules are constraining the optimal price solution and getting a more robust solution. Nevertheless, the distance with real profit values of the baseline gives us some intuition that the solution is not sufficiently robust and is using unreliable elasticity parameters in order to get an unrealistic profit solution, even in this constrained price space.

The third scenario evaluated includes 14 performance rules for individual prices added to the previous set. These few rules get a tighter solution reducing the gap with respect the baseline to only 33.2%. The last scenario includes the previous constraints plus 79 new performance rules for category price mean and difference between products. This reduces the gap obtaining a more realistic and reliable solution for the product category profit. More robust solutions are obtained using Hierarchical Bayes Modeling and L1 regression. The gap is only 5% with respect the baseline of last profit values.

Robustness analysis

One of the capabilities of using pricing rules as constraints in the price optimization problem is to obtain a more reliable and robust solution, even when the elasticity matrix parameters are not well estimated. To evaluate that we will conduct a simulation optimizing different pricing problems considering different elasticity matrices. We will sample the elasticity matrices using a multivariate normal distribution considering the mean and the variance-covariance matrix estimated with the OLS model (both OLS matrices are included in the appendix). For each draw of the elasticity matrix, we will conduct a price optimization considering four scenarios of pricing constraints:

- Prod[0-100]: Individual order rules using from 0% to 100% the domain of previous prices
- Initial Price box: rules for individual prices considering a lower and upper bound at

Scenarios	#Rules	OLS	HBM	LASSO
Average last 4 weeks (baseline)		31.26	27.41	31.47
Order Rules				
(+) Individual Prices	30	199222.4	160.782	204.564
		(>1000%)	(486,6%)	(550.0%)
(+) Individual Prices, Pairwise Differences and Mean	452	64.48	47.75	46.5
		(51.5%)	(42.6%)	(32.3%)
Performance Rules				
(+) Individual Prices	466	46.78	35.47	37.64
		(33.2%)	(22.7%)	(16.4%)
(+) Individual Prices, Pairwise Differences and Mean	545	40.15	32.89	34.58
		(22.1%)	(16.7%)	(9.0%)

Table 1.8: Optimal profit for scenarios considering different sets of Order and Performance Rules.

30% of increase or decrease from the last value registered for the category.

- All-order: Consider the set of 452 order rules for individual, pairwise difference and price mean.
- All-rules: Consider the set of 545 order and performance rules for individual, pairwise difference and price mean.

Figure 1.5 reports box plot charts of percentage deviation of each optimal price obtained from each elasticity matrix sampled, compared with the optimal price from the original solution with no variance. The idea is to measure how the uncertainty in the elasticity parameters change the optimal solution obtained. We perform this analysis for the OLS estimation procedure. The individual price report is included in the appendix, as well as the results using HBM.

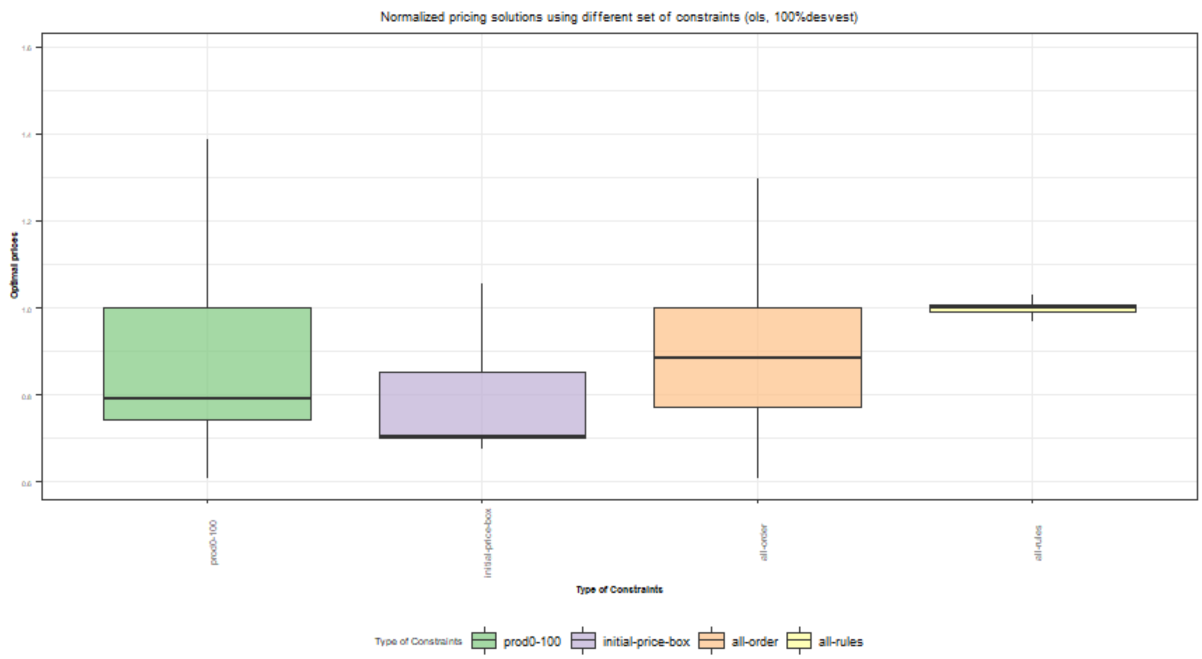

Figure 1.5: Normalized price solutions using different set of constraints. OLS Model

As the reader can notice the interquartile distance of the optimal prices distribution is considerable. Price variations are exceeding the 20% ranges, even for the first three constrained scenarios using some variation of order rules. When the constraint set includes performance rules, the deviation is reduced considerably giving a robust optimal price solution. This is an important contribution from pricing rules, especially the performance rules set. They constrain the problem giving a robust price solution, even for different values of the elasticity matrix.

Worst case scenario: Minimizing Performance

Another valuable results of including data driven pricing rules to profit category optimization is that they allow to avoid bad pricing decisions that practitioners can take in a myopic way. One of the problems to avoid in pricing decisions are the cannibalization of category profit, as we mention in 1.1. Pricing rules also can work as cutting planes to avoid this bad performance price space and cannibalization scenarios. To explore the worst result we won't evaluate the maximization of category profit, but rather the minimization of the profit. This result will be the worst possible solution of profit performance, for a constraint set and price space given. We can compare the minimum profit obtained with and without the pricing rules to understand the contribution of using this data driven set of constraints. Results are in table 1.9. We include the maximizing results already presented in table 1.8 as a contrast to compare.

Pricing Rules Scenarios	#Rules	OLS-Max (Profit Opt)	OLS-Min (Worst Case Scenario)
Last 4 weeks (baseline)		31.26 (Avg)	26.38 (Min)
Order Rules			
(+) Individual Prices	30	199222.4	3.36
		(>1000%)	(-87.2%)
(+) Individual Prices, Pairwise Differences and Mean	452	64.48	17.5
		(51.5%)	(-33.7%)
Performance Rules			
(+) Individual Prices	466	46.78	27.45
		(33.2%)	(4.0%)
(+) Individual Prices, Pairwise Differences and Mean	545	40.15	28.2
		(22.1%)	(6.8%)

Table 1.9: Evaluating worst case scenario minimizing profit, considering different set of pricing rules.

It is clear in the table how the use of order and performance rules reduces the feasible price space improving the minimum category profit obtained. It is interesting to see that the minimum profit solution using all order and performance rules has a better profit solution than the minimum profit obtained in the last four weeks (6.8% better). This result gives the intuition that in the last weeks of the pricing solution there were some specific prices with bad performance reducing profit of other substitute products. If pricing rules have been applied just as a control for actual pricing decisions, expected profit would be higher, without using any optimization tool. This is an interesting side benefit of pricing rules. They can work as control and check of feasibility of actual pricing decisions. It is not a must to use optimization tools to get profit from the application of these data driven pricing rules.

Threshold sensitivity analysis

It is interesting to analyze how the threshold parameters (minimum lift to select performance rules and price domain for order rules), can constrain or relax the final optimal solution obtained in the price optimization problem. To do that, we evaluated different scenarios of price optimization using different sets of pricing rules considering the OLS estimation procedure for elasticities.

Table 1.10 shows different optimization results considering two directions of changes in pricing rule thresholds. First, order rules change with the percentage of the original price domain observed constraining the original space ([0%-100%]) in two other more restricted spaces ([5%-95%] and [25%-75%]). Second, we also apply a different threshold to select less or more performance rules depending on their lift value. We add a more constrained scenario (Lift > 1.1) and a more relaxed scenario (Lift > 1.0), compared the original one (Lift > 1.02).

The results in table 1.10 show how the optimal profit is more constrained when more

	Price constraint set	# rules	Order scenario 1: Price domain [0%-100%]	Order scenario 2: Price domain [5%-95%]	Order scenario 3: Price domain [25%-75%]
Base performance scenario: Lift > 1.02	Individual Order rules	30	19222.4	17980.61	288.966
	All Order rules	452	64.48	44.69	33.32
	Ind Performance rules	466	46.78	40.58	28.79
	All perf & order rules	545	40.15 (*)	35.38	INFEASIBLE
Constr perfor. scenario 1: Lift > 1.1	Individual Order rules	30	19222.4	17980.61	288.966
	All Order rules	452	64.48	44.69	33.32
	Ind Performance rules	459	54.18	42.00	30.46
	All perf & order rules	477	35.89 (*)	35.79	INFEASIBLE
Unconstr Performance Scenario: Lift > 1.0	Individual Order rules	30	19222.4	17980.61	288.966
	All Order rules	452	64.48	44.69	33.32
	Ind Performance rules	469	45.82	42.05 (*)	32.3
	All perf & order rules	612	36.001	35.89 (*)	INFEASIBLE

Table 1.10: Optimal profit results considering different constraint sets using different thresholds for pricing rules.

pricing rules are applied. It seems that constraining the price domain space using a tight scenario of order rules, reduces more dramatically the optimal profit. This result is to some extend expected given that the performance rules we are adding to the constraint set have lower lift. This means that they are probably not active constraints in the final optimal solution because they don't lead to better profit scenarios than previous constraints with more lift.

1.5 Conclusions

This project entails the development of a novel data driven methodology to extract rules from transactional data, to get more robust solutions to the multi-product price optimization setting. The methodology proposed explores the transactional data of product category demand and prices to extract two different types of rules: order rules to formalize how prices have been decided in the past, and performance rules to indicate conditions that lead, on average, to better profit solutions.

We use the pricing rule set as constraints for the pricing optimization problem. To estimate the category profit conditioned to price decision, a double log demand model was calibrated using three different elasticity matrix estimation procedures (OLS, HBM and L1 regularization). When only order rules are applied as constraints, the profit solution obtained is still far from the actual values, giving intuition of an unrealistic solution and improving profit using uncertainty in the estimation of the elasticity matrix in the objective function. When both performance and order rules are applied, the solution is close to the previous values of profit. More robust estimations such as with the Hierarchical Bayesian Model and L1 regularization results in improvements of 16.7% and 9.0% respectively.

There is an additional value of pricing rules for practitioners and managers that is the interpretability. Order rules works as a structured method to recover previous managerial decision making about price categories. It is a way to extract from data what managers consistently decided in the past for the category price vector. We think that there is an important implicit knowledge hidden in previous decisions. For example, that a certain premium brand is consistently more expensive than the private label brand. Or bigger formats charge cheaper per size than smaller formats of the same brand. This kind of knowledge is extracted and converted into explicit knowledge in order pricing rules.

In the same way, performance rules give a structured way to understand where good profit scenarios are located in the price space solution. More than a cross elasticity term, it gives intuition about what happens with the expected profit if a certain distance in prices for a couple of products is decided upon.

The implementation of pricing rules also seems to be easier. For traditional SQL database servers, applying pricing rules as queries seems very practical to understand if a specific alternative price vector is feasible or not, and also evaluate what-if scenarios if some prices do not respect a specific rule. Using the methodology proposed, it is easy and fast to extract thousands of pricing rules, sort them to select the most relevant and valuable rules to apply in a certain category.

Pricing rules are built analyzing a database with millions of transactions comparing category performance for different historical pricing scenarios. In that sense, the most important contribution of this methodology is that pricing rules extracted provide a robust solution to price optimization based on historical performance of different pricing policies and their impact in the category demand.

1.5.1 Limitations and Future Work

This methodology provides a fast and robust approach to get reliable solutions for multi-product price optimization. There are some limitations and alternative approaches that are interesting to discuss.

There are some extensions we can do to the multi-product price optimization setting. The demand model and the definition of the price category can be modified adding other interesting controls and decision variables. For example, if we would have available information about past prices of competitors we can add this information not only into the cross elasticity matrix, but also into the pricing rules extracted. For example we can create rules about the correlation between our category performance and a specific popular product price of the competitor store. Using the same approach we can understand how a specific distance of a product of our portfolio and the competitor price, can affect our product performance.

Another way to extend the actual model is to include in the price optimization (and also into the constraints), all the information regarding promotions, extending the setting to a price and promotion optimization. The data analyzed includes information about in store promotional display, application of special shelf tag prices or if the product was included in the retail catalog. We can use our methodology to extend the pricing rules to promotional rules to correlate promotional actions and profit scenarios. Also, we can extract rules and understand how a mix of actions (price reduction and product catalog) can affect demand and therefore category profit. These rules can improve the robustness of the pricing and promotion optimization results.

Our pricing rule setting is applied to a specific problem in the supermarket industry to define prices for a multi-product category considering substitution and complementarity patterns between products. This setting can be extended to similar problems in the retail industry considering other kind of substitutions between products. For example we can extract pricing rules to constrain the dynamic pricing problem in the fashion goods categories where seasonality and timing is one of the most important variables to explain the demand and results. In the same way as we did with consumer goods, we can extract rules and understand how the price distance between weeks can affect the demand and the overall performance.

As a summary, in this research we analyzed transactional data to identify what business rules have been consistently applied before and evaluate which ones are associated to better business performance. Based on these pricing rules, we build a data-driven set of feasible prices and combine it with standard price optimization routines to understand how this information can complement the traditional econometric analysis of the demand. This novel approach for pricing optimization is easy to implement and not only provides managers with a reliable automatic mechanism to decide about prices of multiple products, but also enables more consistent decisions.

Chapter 2

Analyzing Shopping Behavior Using LDA in Transactional Data

2.1 Introduction

Retailers face the challenge of making frequent decisions for a large numbers of products. For example, retailers have to make frequent decisions about promotions, prices, replenishment, layout and assortment. Category management is a popular approach to deal with this complex problem (Nielsen, 1998). It separates the product portfolio and their decisions into independent and isolated business units, trying to replicate the concept of divide and conquer ("divide et impera"). This popular approach implicitly makes strong assumptions because it considers categories as fully independent from each other. This also implies that an important problem is how to partition the set of products offered by a retailer into product categories.

Going beyond the limitations of category management requires us to evaluate how product categories are related to each other. One approach for understanding these interrelationships is to use transactional data. It is clear that an important fraction of supermarket transactions include products from more than a single product category.

At the same time, shoppers can have different motivations to buy and make a trip to a retail store. For example, some customers make monthly grocery re-stocking purchases of groceries. Other motivations include buying fresh products for immediate consumption (Walters and Jamil, 2003). These different drivers for creating supermarket transactions are called "shopping trip missions" or "motivations" (ECR-Europe, 2011). Each of these motivations is associated with a different set for product categories, so studying the joint incidence of product purchases can give us an opportunity to uncover these latent drivers of shopping trips.

Knowledge and prediction of these shopping trip motivations should help retailers to make better decisions for each product category. For example, retailers could implement targeted marketing actions and promotions based on the predicted shopping trip motivations for each customer. Consequently, our research is focused on detecting shopping trip motivations and then using this information to predict consumer behavior.

The marketing literature offers some approaches to measure category interrelationships and detect shopping motivations behind transactions. One approach is by Manchanda et al. (1999) using Multivariate Probit models, who analyze a small number of product categories. Other methods rely on distance metrics such as K-means (Sarantopoulos et al., 2016) or multidimensional scaling (Musalem et al., 2018) in order to detect clusters of product categories.

Multidimensionality is a problem for most of these methods. A typical retailer has more than 400 product categories so it somewhat restrictive to reduce the analysis of interdependencies across categories to a Cartesian distance or make comparison of categories only pairwise, without explicitly modeling the overall distribution of joint purchase.

Another important element to consider is the probabilistic assignment of a specific transaction to a particular purchase motivation. With distance-based models it is not obvious how to estimate these membership probabilities. Instead, it would be useful to rely on statistical theory to estimate the assignment of transactions to purchase motivations. Similarly, we are also interested in using statistical theory based methods to related product categories and shopping motivations.

This essay presents a machine learning approach to study interrelationships among product categories and to detect latent shopping trip motivations. Machine learning has experienced a major growth in research and applications in the last years. Specifically in retailing, there are many opportunities to use the vast information generated in transactions and loyalty clubs to calibrate different supervised and unsupervised models (Bradlow et al., 2017). In particular, we propose both supervised and unsupervised methods to model the interrelationship among product categories. We rely on Latent Dirichlet Allocation (LDA) created by Blei et al. (2003), which has been widely used in text mining to extract topics from documents, measuring the probability of co-occurrence of words. In our retail context, we will extract latent shopping motivations instead of latent text topics, analyzing relationships among product categories, instead of words, in a transaction instead of a document database. Table 2.1 represents the transformation from one problem to another.

LDA for text analysis	Words	Documents	Text Topics
⇓	⇓	⇓	⇓
LDA for transactional analysis	Products	Transactions	Shopping motivations

Table 2.1: Relation between traditional LDA used for text analysis and LDA for transactional analysis

The contribution of this research is then to apply LDA in a retailing setting, modifying the basic model to achieve three different goals as follows. First, a standard LDA model will be used to detect and describe shopping motivations. Second, the basic LDA model will be extended to jointly estimate the latent shopping motivations and the relationship between these motivations and basket size using a supervised approach. Finally, the LDA model will be generalized to allow purchase motivations to depend on customer and shopping trip characteristics.

The rest of this essay is organized as follows. First, we review the relevant literature, particularly in terms of articles related to the detection of shopping trip motivations, machine learning and LDA, and latent models used in transactional settings. Second we describe how the basic LDA framework is extended to model shopping trip motivations, purchase incidence within each product category and purchase amount. We apply these latent models using a Bayesian framework calibrated with transactional data from a US supermarket chain. Finally, we compare the results, insights and limitations of the proposed methods.

2.2 Related Research

In this section we review the relevant literature for this paper. First, we discuss research about shopping trip motivations and analytical methods to estimate these motivations from transactional data. Second, the review focuses on the LDA and some applications in marketing. Finally, this review describes in detail some applications of LDA with transactional data.

2.2.1 Shopping trip motivations

Kollat and Willett (1967) is one of the first articles about shopping motivations. They used surveys to uncover shopping trip motivations. Buying motivations were primarily related to buying products that the customer forgot to purchase in previous period or products for which the customer ran out of inventory at home. Kahn and Schmittlein (1989) analyzed panel data and classified shopping motivations into quick and regular trips, describing them in terms of their basket size, frequency and purchase amount, among other features. Walters and Jamil (2003) use surveys to study how different metrics of consumer behavior (e.g., the number of promotions they are searching, the profitability of the shopping basket and store choice) depend on the purchase motivations. This study detects three types of motivations: weekly grocery shopping, replenishment, and special purchases motivated by specific categories or promotions. Bell et al. (2011) focus their research how unplanned category purchase depends on the shopping trip goals. They classified shopping motivations as concrete type (to take advantage of specific promotions) or as abstract type (to fill up on weekly needs).

Most of the traditional research about shopping trip motivations infers them applying questionnaires. However, to detect interrelationships between product categories and the latent drivers that produce these shopping trips it is also useful to consider data approaches based on transactional data.

In particular, the traditional approach to discover interactions between categories is based on econometric modeling and, more specifically, discrete choice models that rely on random utility theory. Manchanda et al. (1999) analyzed cross category effect using a multivariate probit model. This article focuses on detecting and measuring different source of cross-category effects: cross-price elasticities across categories leading to either complementarity or substitution between two categories; other non-price interrelationships among categories. Increasing the scope of categories analyzed, Russell and Petersen (2000) studies cross-category effects calibrating a multinomial logit model based on supermarket transactions. Both models are computationally expensive and would face an important challenge if they were escalated to the complete assortment of categories of a retailer.

An more scalable approach to estimate relationships between categories, is based on clustering techniques. Sarantopoulos et al. (2016) detects shopping missions using K-means. It uncovers seven different clusters with different proportions depending on store format. A similar work is done by Griva et al. (2018) generating a customer visit segmentation estimating ten different clusters for shopping missions. One of the challenges of using K-means is that its results come from a heuristic method providing only a local minimum solution depending on the starting point. There is no guarantee of a global minimum solution. Also, in high dimensional spaces, such as the multi-category problem we are facing, there is an important issue of using euclidean distance. As the number of dimensions (i.e., product categories) increase, the sparsity of the space also increases reducing the relative difference in the distances between pairs of observations (curse of dimensionality) Beyer et al. (1999). Other approaches to study interrelationships between product categories include models based on distance metrics. For example, Videla-Cavieres and Rios (2014) use market basket analysis and graph mining techniques to uncover groups of product categories. In a similar way, Musalem et al. (2018) uses multidimensional scaling to map, and clustering, to uncover and describe shopping motivations. This work detects four different motivations: groceries, hygiene products, fresh and immediate consumption products, and hedonic products.

2.2.2 Latent Dirichlet Allocation (LDA) and Marketing Applications

As mentioned in the introduction, this paper uses LDA to model latent shopping motivations. In this subsection, we briefly introduce the basic LDA model (Blei et al. (2003)) and review applications in marketing. This method, originally developed for text mining, detects latent topics in a document by focusing on the co-occurrence of words in a set of documents.

Latent Dirichlet Allocation model considers every document as a collection of words. Each word belongs to a certain topic with a given probability. Conditioning on a specific topic, some values for a word are more probable than others. Hence, each document contains a mixture of topics and the presence of each word in the document depends on these topics.

Previous research has used LDA for a variety of applications such as linguistics, political science, biomedicine, geography, software engineering, social networks, crime prediction and marketing (Jelodar et al., 2019). Focusing on this last field, Tirunillai and Tellis (2014) uses LDA to extract latent dimensions of product quality from different product reviews. Christidis and Mentzas (2013) builds a product recommendation system based on topics estimated by LDA. Büschken and Allenby (2016) uses an extension of the LDA model to extract topics from text restaurants and hotel reviews, in order to predict consumer ratings. Schröder et al. (2019) use LDA to create web usage profiles from Internet browsing histories and purchases, and then use these profiles to perform a customer segmentation via K-means. For a recent survey of LDA applications in Marketing see Reisenbichler and Reutterer (2019).

One possible disadvantage of the LDA model is its inability to model correlations among topics. The proportion of one topic cannot be correlated with another topic because the use of the Dirichlet prior. To address this drawback, Blei and Lafferty (2006) proposed the use of CTM (Correlated Topic Modeling), which is similar to LDA but requires more parameters to explicitly consider correlations among topics.

2.2.3 Transactional applications using LDA

In terms of of LDA applied to retail transactions, Hruschka (2014) translates LDA for text mining setting to extract consumer profile from supermarket data. This paper extracts latent shopping motivations instead of text topics, analyzing relationships among product categories, instead of words, in a transaction database instead of a set of documents. Its particular application to a supermarket data finds ten different topics, and then uses these topics to generate product recommendations for some specific baskets.

Jacobs et al. (2016) use LDA to predict product demand for SKUs with low purchase frequency. To do that, they propose an extension of LDA using a hierarchical approach that accommodates demographic characteristics in the probabilities of topics. They calibrate this LDA model aggregating the information of all shopping trips for each customer.

An important difference of our work with respect to Jacobs et al. (2016) is that our approach individually models each shopping trip from each customer. This allows us to consider that the same customer can have different motivations across different shopping trips.

This is important because the same customer, or customer with same demographics, can have different shopping trip motivations depending on certain features of a shopping occasion (e.g., time of the day). For example, a customer making a shopping trip on a Friday night, might be more likely to have a hedonic shopping motivation compared with a transaction on a Sunday afternoon, where the customer might be more likely to visit the store to replenish her stock of groceries at home.

Our analysis can also accommodate dynamic behavior, including dependency among trips for the same customer. For example, recency information such as the time since the last shopping trip or the mix of latent motivations of the last shopping trip. If a customer made its last grocery shopping only a couple of days ago, the probability that a new transaction might once again be characterized by a grocery shopping motivations might be lower. With these examples we suggest it is valuable to consider an analysis at a transaction as opposed to a customer level.

One more difference between our approach and the existing literature is that we will proposed a supervised version of the LDA model to model the relationship between shopping trip motivations and variables of interest related to the shopping trip, such as basket size.

2.3 Methodology

This section details the LDA models formulated to analyze shopping behavior from transactional data. In particular, this essay will accomplish three key goals related to shopping trip motivations:

- Use traditional LDA models to estimate and describe shopping trip motivations from a supermarket transaction database.
- Model and explain shopping basket characteristics of interest (e.g., purchase amount) as a function of the latent purchase motivations.

- Relate purchase motivations to customer and shopping trip characteristics.

2.3.1 LDA for detecting shopping trip motivations

As we mentioned in the previous section, under an LDA model, every document is a collection of words, where each word belongs to a certain topic with a given probability. Conditioning on a specific topic, some values for a word are more probable than others. Figure 2.1 describes the model and its parameters. We use the following notation: M is the number of documents, N the total word instances in a document, K is the number of latent topics and V represents the set of words available or dictionary. In addition, the LDA model uses two prior hyper parameters, α and β, to specify Dirichlet distributions. The data generation process is as follows:

- Sample $\Phi_{kv} \sim \text{Dir}(\beta)$, where Φ_{kv} represents the probability of word v if the latent topic is k.
- For each document m, sample $\theta_m \sim \text{Dir}(\alpha)$, where θ_{mk} represents the probability of latent topic $k = 1, .., K$ within document m.
- For each word instance $n = 1, .., N$ in document m:
  - Sample Topic Z_{nm} $\sim$Multinom(θ_m)
  - Sample word from dictionary conditional on the topic for the word instance W_{nm} $\sim$Multinom($\Phi_k | Z_{nm}$)

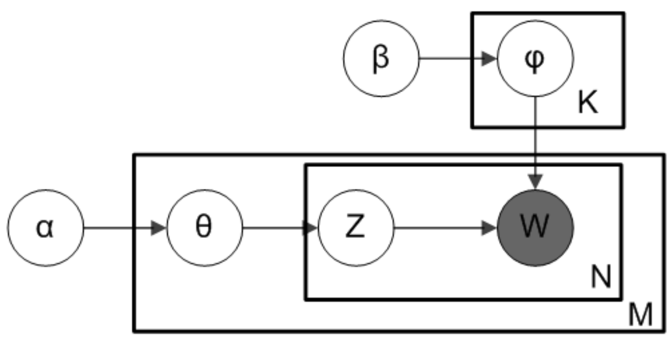

Figure 2.1: Model description and parameters used in Latent Dirichlet Allocation (Blei et al., 2003)

LDA estimates topic probabilistic distributions in a parsimonious manner, from word co-occurrence patterns observed across different documents. In our retail application, we will extract latent shopping motivations instead of latent text topics, analyzing relationships among product categories, instead of words. The LDA model will use a transactional database as an input specifying the the product categories purchased in each transaction. Note that in a retailing context, the interpretation of the output matrices of LDA has a differing meaning compared to text analysis. First, θ_{mk} represents the probability of latent shopping motivation k for transaction m. Φ_{kv} represents the probability of product category v under latent shopping motivation k.

Another model parameter that merits discussion is α, which is a vector of size K that controls the Dirichlet distribution for θ_{mk}, i.e. the probability of latent motivation k for transaction m. Typically, researchers choose symmetric values for this vector, such as $1/K$. Figure 2.2 shows how different alpha values produces different results for θ_{mk} sampling.

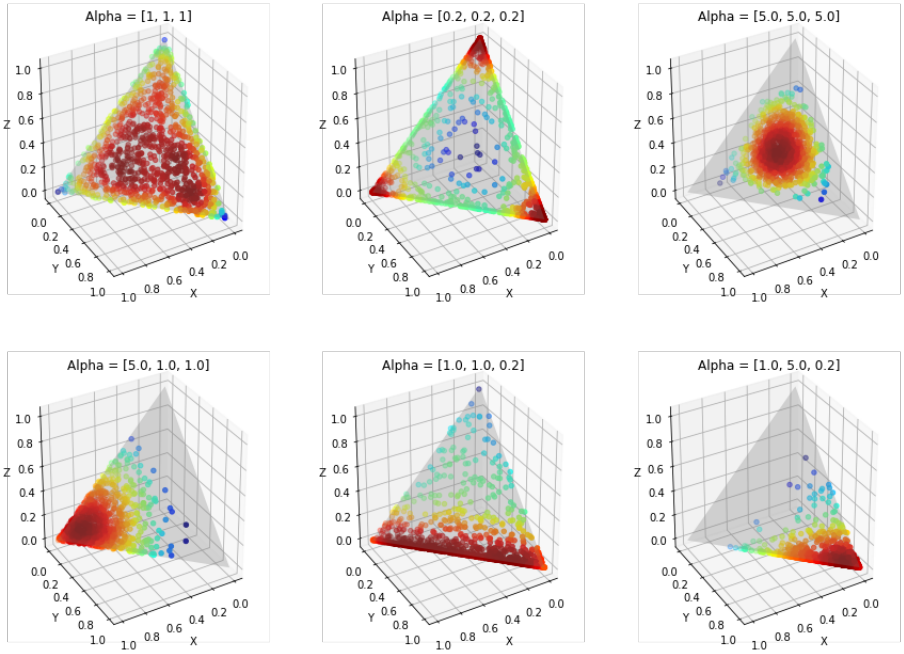

Figure 2.2: Variation of θ_{mk} sampling using different values of Alpha hyper-parameter of Dirichlet distribution. Source: (https://tinyurl.com/slhvr2o)

There are two main effects of the value of α on the resulting LDA model. First, the scale of the sum of the elements affects the fuzziness of the θ_{mk} output matrix. If all values of α are greater, the values of θ_{mk} for the same transaction m become more similar, which is consistent with each transaction containing a more even mix of purchase motivations. If, however, α is close to zero, the θ_{mk} probabilities become more extreme, which is consistent with each transaction being associated with a single purchase motivation. The second effect is related to the relative magnitude of the different components of α. The greater the value of a component, the greater the probability for that motivation, compared with the others.

In the estimation of our LDA model, α won't be fixed to a specific value, but instead it will be estimated from the data. Details about the estimation code used can be found in the Appendix.

Another relevant parameter is the number K of topics or shopping motivations to be extracted from the analysis. There is an important trade-off between complexity of the solution and capability to fit the data. A higher K will yield a more complex solution with more parameters to be estimated (K*(M+D+α)). Deveaud et al. (2014) propose a method to determine the optimal number of topics maximizing an information divergence criterion.

2.3.2 Supervised LDA for modeling shopping basket characteristics of interest

The second model implemented uses LDA motivations as a data embedding of the multi category information, in order to build a supervised LDA model for any feature of interest related to the shopping basket (e.g., purchase amount). Accordingly, we formulate a linear model of the variable of interest with the motivation probabilities θ_{mk} as independent variables. We define the following additional notation:

- Y_{ct} shopping basket characteristic of interest for transaction t for customer c. This variable is mean-centered to facilitate the estimation of the model parameters.
- ρ_k the importance of shopping motivation k as a predictor of Y_{ct}.
- σ the estimator of the standard deviation of the residuals of Y_{ct}.

With these definitions, we add the following modeling assumption to the basic LDA specification: $Y_{ct} \sim \mathcal{N}(\rho_k \theta_{mk}, \sigma^2)$.

In terms of the possible variables of interest, we can for example consider the number of product categories purchase or the contribution margin of the transaction. Our research will focus on using LDA to predict purchase amount. This KPI is relevant for the design of targeted promotional campaigns (1:1 marketing). If the retailer can anticipate and estimate the purchase amount of a particular transaction, they can focus different marketing actions depending on that prediction. For example, the retailer can recommend products to a customer with the goal of increasing basket size.

For identification purposes we will normalize ρ_1 to zero. Thus, when predicting the variable of interest, ρ_k needs to be interpreted in relation to the first motivation.

2.3.3 Hierarchical LDA for predicting shopping trip motivation

The objective of the third model is to model shopping trip motivations as a function of customer and shopping trip information. In the basic LDA model, the purchase motivation probabilities θ_m for all documents are drawn from the same Dirichlet distribution with parameter α. In the hierarchical model in this subsection, the purchase motivation probabilities for each document may change depending on the customer characteristics and other variables. This is similar to Jacobs et al. (2016), although in our application we model each trip separately so the purchase motivation probabilities for the same customer may differ across shopping trips.

Consider the following notation:

- D is the number of covariables used to predict shopping trip motivations.
- $X_{ct\mathrm{d}}$ is the value of covariate d for customer c in shopping trip t.
- $\delta_{\mathrm{d}k}$ is the coefficient of covariate d for estimating the probability of purchase motivation k.
- α_{ctk} is a Dirichlet parameter for predicting the probability of purchase motivation k for customer c during trip t. This parameter is in turn specified as follows: $\alpha_{ctk} = \lambda \frac{\exp(\sum_{\mathrm{d}=1}^{D} X_{ct\mathrm{d}} \delta_{\mathrm{d}k})}{\sum_{k=1}^{K} \exp(\sum_{\mathrm{d}=1}^{D} X_{ct\mathrm{d}} \delta_{\mathrm{d}k})}$, where λ is a scaling factor to modulate the magnitude of the sum of all the components of α.

In our setting, predictors of purchase motivations can be obtained from loyalty club information. For example, we can include customer demographics such as age, gender, family characteristics, income and occupation.

All these models described in this section were estimated using Bayesian methods and were coded in Stan. Details about the codes implemented can be found in the Appendix.

2.4 Empirical Application

We apply the proposed methodology to a supermarket transactional database provided by Dunnuhumby[1]. In the remainder of this section we present summary statistics of the data and then present the results of three alternative LDA models.

2.4.1 Data descriptive summary

LDA models will be calibrated using a sample observations in terms of customers, periods and products. Descriptive statistics are provided in the next table comparing the full data set and the sample we used. We note that the average basket size, both in dollar value and number of products is similar across the two columns.

	Full Sample	Subsample
Transactions	139,876	1,295
Households	801	50
Weeks	102	24
Stores	354	79
Categories	308	36
Products	67,904	949
Average amount spent ($) (st dev)	32.15 (38.32)	29.13 (36.31)
Average number of prod per ticket (st dev)	10.12 (13.03)	9.37 (12.65)

Table 2.2: Summary descriptive statistics for the transactional data set.

We note that the product hierarchy has four levels. At the more disaggregate level there are 67,904 stock keeping units (SKUs), which belong to 2,883 product subcategories. These subcategories in turn belong to 308 product categories which belong to 44 different departments. Our analysis will be done at the category level using the 36 most frequently purchased product categories and a sample of the total transactions considering a panel of 50 customers. Table 2.2 shows a statistics summary of the sample vs the complete dataset. Table 2.16 in the Appendix displays for each of the category the fraction of transactions in which they were purchased and the fraction of the total spending attributed to each category.

2.4.2 LDA to extract shopping trip motivations

We now present the results of applying the basic LDA model to the transactional data set. One difference with most applications is that rather fixing the values of the α parameter at a particular constant, we will estimate them from the data. Recall that these relative size

[1] https://www.dunnhumby.com/careers/engineering/sourcefiles

of the α components measure the frequency of these motivations in our data. In addition, the magnitude of these parameters characterizes the degree to which shopping trips are associated with a single (low values) or multiple motivations (high values). In our application we consider:

- $V = 36$ product categories in the product dictionary.
- $M = 1,295$ transactions.
- $N = 4,167$ product purchases.
- $K = 2$ purchase motivations

In addition, we set the $V = 36$ components of the β hyper-parameter to 0.01 and use a Uniform(0,3) prior distribution for each of the $K = 2$ components of α. The Bayesian estimation procedure was implemented running 4 chains for 5,000 iterations each, using the first half as the warm-up period. Convergence was assessed using the $\hat{R}$ statistic which measures the degree to which parallel chains with different starting values have converged to the same distribution. Values close to 1 of this statistic are consistent with the the Markov Chain achieving convergence. We note from our results that most of the model parameters have a value below 1.1 (see Table 2.3).

Min.	1st Qu.	Median	Mean	3rd	Max. Qu.
1.000	1.006	1.012	1.019	1.022	1.656

Table 2.3: $\hat{R}$ distribution for the basic LDA model parameters.

In terms of the results, we begin by discussing the Φ matrix (see Table 2.17 in the Appendix), which describes the probability of each product category under each motivation. A useful approach to interpret this parameter is to identify the most likely product categories for each shopping motivation. Table 2.4 shows the top ten product categories with higher probability of purchase under each motivation. From this table, we observe some differences between the two motivations. Product categories most likely to be purchased under motivation 1 seem related to hedonic goals or to (almost) ready to eat products. In addition, several product categories most likely to be purchased under motivation 2 are related to breakfast consumption.

The second matrix obtained from the basic LDA model is Θ, which indicates the probability of each motivation being present for each transaction. Table 2.5 shows summary statistics of the model results, including the posterior mean, standard deviation and $\hat{R}$ for the θ parameters for both motivations. The breakfast motivation has on average a 52.5% probability, while the hedonic motivation has a slightly lower prevalence of 47.4%. These results are consistent with the estimated values of α (see Table 2.6). Note that the estimated value for motivation 2 is slightly greater than that of motivation 1. Also, both components of α are estimated as being significantly greater than one, meaning that the probabilities for an specific transaction are very fuzzy between motivations. In other words, both motivations are likely to be present in each transaction.

Motivation 1:Hedonic	ϕ_1	Motivation 2:Breakfast	ϕ_2
soft_drinks	16,94%	fluid_milk_products	8,14%
bag_snacks	7,32%	tropical_fruit	7,80%
fluid_milk_products	7,03%	baked_breadbuns	7,38%
candy_-_checklane	6,74%	cold_cereal	5,32%
couponmisc_items	6,55%	eggs	4,97%
frozen_pizza	5,88%	cheese	4,86%
beers_ales	5,52%	deli_meats	4,68%
candy_-_packaged	5,47%	vegetables_-_shelf_stable	4,46%
lunchmeat	5,24%	refrgratd_juicesdrnks	4,25%
ice_creammilksherbts	5,22%	canned_juices	4,04%

Table 2.4: Most likely product categories for each purchase motivation using LDA.

	Mean_M1	SD_M1	$\hat{R}$_M1	Mean_M2	SD_M2	$\hat{R}$_M2
Min	0.1032	0.07658	1.000	0.2084	0.07658	1.000
1st Qu.	0.3941	0.17631	1.006	0.4282	0.17631	1.006
Median	0.4799	0.20361	1.012	0.5201	0.20361	1.012
Mean	0.4743	0.19566	1.017	0.5257	0.19566	1.017
3r Qu.	0.5718	0.22125	1.021	0.6059	0.22125	1.021
Ma	0.7916	0.25118	1.129	0.8968	0.25118	1.129

Table 2.5: Estimation results for θ using LDA.

2.4.3 LDA to predict shopping basket size

The second model we consider corresponds to a supervised LDA model where purchase motivations are used to predict basket size (dollar amount). This LDA model uses as independent variables, the motivation probabilities for each transaction estimated by the LDA. The model jointly estimates the purchase motivations and the degree to which they predict basket size.

We apply a log transformation to the basket size to reduce the skewness of the original distribution. Figure 2.3 shows the histogram and the box plot of the log transformation of the purchase amount. The estimation of the model for the log of purchase amount is given by $Log(Amount) = \rho_1 + \rho_2\theta_{m2} + \varepsilon$. Finally, the dependent variable was normalized subtracting its mean and dividing it by its standard deviation.

Results are based on three parallel chains with 10,000 iterations each. In terms of estimation convergence, Table 2.7 shows summary statistics of the $\hat{R}$ statistic.

To characterize the estimated motivations, we show in Table 2.8 the ten product categories most likely to be purchased under each motivation. Detailed results about the Φ parameters are available in Table 2.18 in the Appendix.

The results are substantially different from the ones obtained with the basic LDA model. Here the motivations and baskets are more fuzzy. Also motivation 1 includes food products

	Mean	SD	2.5%	97.5%	$\hat{R}$
α_1	1.71	0.16	1.39	1.96	1.32
α_2	1.87	0.10	1.61	2.00	1.03

Table 2.6: Posterior mean and standard deviation for Alpha

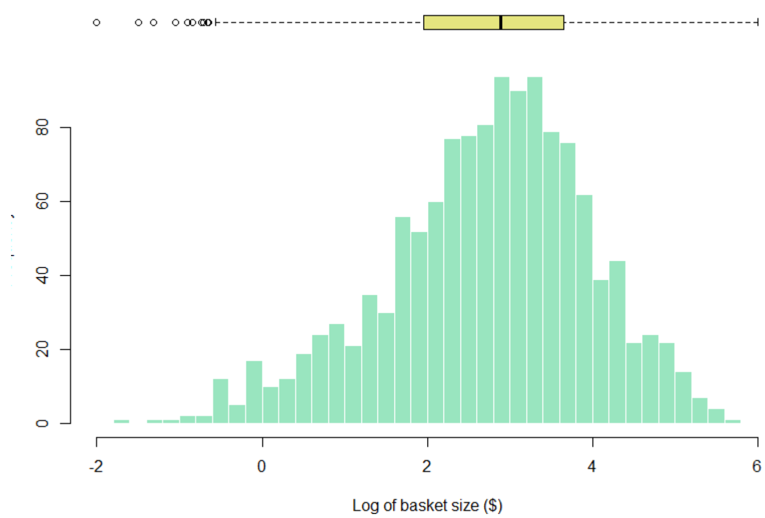

Figure 2.3: Histogram and Box Plot for the log of basket size amount

that can consumed with little effort. Motivation 2 includes ethnic and fresh products.

Table 2.9 shows the estimated distribution of the α parameters. The values of both components are significantly different, representing a bigger prevalence of motivation 1. These values are consistent with those for θ. The average θ value is estimated at 65.7% and 34.22% for motivations 1 and 2, respectively. Detailed results for the distribution of θ are shown in table 2.19 in the Appendix.

We now consider the results for ρ, which measures the association between motivation 2 and the purchase amount of each transaction. Table 2.10 shows the estimation results for this parameter, where ρ_1 is an intercept to predict purchase amount and ρ_2 is the coefficient for the probability of motivation 2. Both parameters are significant at the 95% level. Negative values of ρ_2 indicate an inverse relationship between purchase amount and motivation 2. A possible explanation about this is that low effort food tends to be expensive leading to bigger purchase amounts compared with products prevalent for Motivation 2.

In terms of the ability of the model to predict purchase amount, Figure 2.4 shows a comparison between actual and estimated values of amount spent, which reveals a 97.5% correlation between both sets of values.

2.4.4 Hierarchical LDA to predict shopping trip motivation

The third model aims to explain and predict shopping motivations as a function of customer and shopping trip characteristics. In our dataset we will use five different consumer characteristics: age, Income, household size. We add two dynamic characteristics related to

Min.	1st Qu.	Median	Mean	3rd Qu.	Max. Qu.
0.9998	1.0014	1.0035	1.0058	1.0073	1.4599

Table 2.7: $\hat{R}$ distribution for LDA regression posterior parameters

Motivation 1: Low Effort Food	ϕ_1	Motivation 2: Ethnic & Fresh	ϕ_2
deli_specialties_(retail_pk)	12.86%	hispanic	26.43%
sald_drsng_sndwch_sprd	12.48%	baked_sweet_goods	12.06%
crackers_misc_bkd_fd	10.11%	deli_specialties_(retail_pk)	7.67%
candy_-_packaged	6.30%	baked_bread_buns_	6.48%
candy_-_checklane	6.21%	refrgratd_juices_drnks	4.75%
frozen_pizza	6.19%	sald_drsng_sndwch_sprd	3.90%
eggs	5.59%	baby_foods	3.69%
beef	5.42%	soft_drinks	3.32%
cold_cereal	5.18%	candy_-_checklane	3.13%
frzn_meat_meat_dinners	3.92%	tropical_fruit	3.12%

Table 2.8: Most likely product categories for each purchase motivation using LDA Regression.

shopping behavior for each customer. In particular we also include the last purchase amount and recency (i.e., the number of days since the last shopping trip). Descriptive statistics of these variables are shown in Table 2.11. All these variables were standarized substracting their means and dividing them by their standard deviations.

The estimation is based on 3 parallel chains with 15,000 iterations each. The distribution of the $\hat{R}$ statistic of all the estimated parameters is shown in Table 2.12. Most of the estimated parameters converged with an $\hat{R} < 1.1$.

In terms of the latent purchase motivations detected from the data, estimation results for the Φ matrix can be found in table 2.20 in the Appendix. The most likely categories for each motivation are shown in Table 2.13. If we compare this solution with the one obtained using the basic LDA model, the motivations present some changes in terms of the products. Although motivation 1 is also related to hedonic products, there are only 4 product categories among the 10 most likely under motivation 1 for both models. The second motivation is even more different from the one obtained under the basic LDA model. Only two categories belong to the ten most likely to be purchased for both models (baked bread and refrigerated Juices). The products related to this motivation are more diverse and are not as strongly related to breakfast consumption.

In terms of the prevalence of each motivation, Table 2.14 shows summary statistics of the posterior mean, standard deviation and $\hat{R}$ of the θ matrix. As before, both motivations are approximately equally prevalent. If we compare these values with those obtained under the basic LDA model, the distribution is more sparse than with the basic LDA solution. Intuitively, the covariates allow the motivation probabilities to become more extreme. Therefore, the values of θ are more extreme compared with the basic LDA approach.

We now discuss the δ parameters, which measure the importance of each transactional

	Mean	SD	2.5%	97.5%	$\hat{R}$
α_1	2.83	0.14	2.47	3.00	1.02
α_2	1.49	0.15	1.23	1.85	1.08

Table 2.9: Posterior mean and distribution for alpha estimation using LDA Regression.

	mean	sd	2.5%	97.5%	$\hat{R}$
ρ_1	1.36	0.10	1.19	1.58	1.05
ρ_2	-3.97	0.19	-4.33	-3.59	1.01

Table 2.10: Posterior mean and distribution for ρ estimation using LDA Regression.

characteristic to predict a shopping motivation k. For identification and without loss of generality we impose that the all the delta values related to the last motivation are equal to zero. This means that the estimated parameters will describe the relative importance of transactional characteristics as predictors of the first purchase motivation.

Table 2.15 describes the posterior mean, deviation and $\hat{R}$ of the values estimated. The mean estimators for all the five customer characteristics are significantly positive. This implies that greater levels of age, income, household size, recency and last amount are associated with a greater probability of the hedonic purchase motivation.

2.5 Conclusions

This work uses a Latent Dirichlet Allocation framework to analyze shopping behavior based on transactional data. In our retail context, we extract latent shopping motivations analyzing relationships among product categories. The contribution of this research is then to apply LDA in a retailing setting to achieve three different goals as follows. First, we apply a standard LDA model to detect and describe shopping motivations. Second, the basic LDA model is extended to jointly estimate the latent shopping motivations and the relationship between these motivations and basket size using a supervised approach. Finally, the LDA model is generalized to allow purchase motivations to depend on customer and shopping trip characteristics.

We applied this methodology to a supermarket transactional database focusing on 36 different product categories. We extract two shopping motivations. The first motivation is related to hedonic products that represents 48% of the total purchased products. The second motivation extracted is related to breakfast product categories accounting for 52% of the total purchases.

The second model proposed is a supervised LDA approach using the motivation probabilities as independent variables to predict a variable of interest, in our application this corresponds to the dollar value of the basket. The motivations extracted changed with respect to the motivations obtained in LDA detecting shopping patterns related to Low Effort Food and Ethnic and Fresh products. The regression of amount spent shows a good fit of the data using only the motivations extracted with LDA.

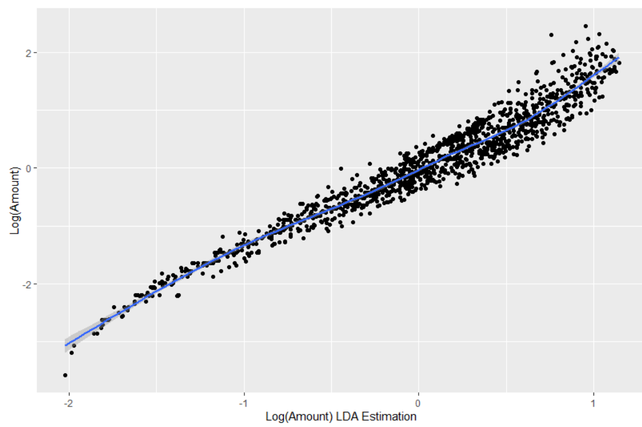

Figure 2.4: Scatter Plot of Real vs Predicted Normalized Log of Basket Amount using LDA Regression.

Age	Income	Household Size	Recency	Last Amount
Min. :22.00	Min. : 7.00	Min. :1.000	Min. : 0.000	Min. : 0.000
1st Qu.:30.00	1st Qu.: 30.00	1st Qu.:1.000	1st Qu.: 1.000	1st Qu.: 6.365
Median :40.00	Median : 62.00	Median :2.000	Median : 2.000	Median : 16.290
Mean :44.72	Mean : 68.73	Mean :2.284	Mean : 3.571	Mean : 27.891
3rd Qu.:50.00	3rd Qu.: 87.00	3rd Qu.:3.000	3rd Qu.: 5.000	3rd Qu.: 34.450
Max. :70.00	Max. :225.00	Max. :5.000	Max. :74.000	Max. :321.190

Table 2.11: Summary statistics for customer variable used to adapt LDA prior

The third model calibrated in this paper allows shopping motivations to depend on customer and shopping trip characteristics. In our setting we used age, income, family size, recency and the last purchase amount. These characteristics are positively associated with the first motivation (Hedonic products). Motivations extracted were different than results obtained with LDA. We also note that the distribution of motivation probabilities was more extreme than in the basic LDA, because of the use of customer demographics.

2.5.1 Limitation and Further Research

There are some limitations about these methods to analyze shopping behavior. We discuss them and propose further research to extend it.

The Hierarchical LDA used demographic and transactional characteristics of consumers to adapt the prior hyper parameter and obtain more flexibility to the model. We could improve this analysis using more sources of transactional and customer information. For example:

- Specific characteristics of the transaction: There is contextual information related to each transaction that can be useful to improve shopping behavior prediction. For example, temporal information about the transaction can be correlated with motivations. For instance, hedonic shopping motivations might be more likely during Friday nights. Particular example variables in this sense can be: time of the day, day of the week or day of the month.

Min.	1st Qu.	Median	Mean	3rd Qu.	Max. Qu.
0.9999	1.0018	1.0054	1.0131	1.0126	2.073

Table 2.12: Summary statistics for $\hat{R}$ of posterior mean estimated with hierarchical LDA

Motivation 1	ϕ_1	Motivation 2	ϕ_2
refrgratd_juices_drnks	7,17%	frzn_meat_meat_dinners	14,90%
deli_meats	6,42%	baked_bread_buns_	8,12%
candy_-_packaged	6,32%	baked_sweet_goods	7,44%
candy_-_checklane	6,16%	sald_drsng_sndwch_sprd	5,41%
coupon_misc_items	6,10%	bag_snacks	4,62%
beef	6,04%	deli_meats	3,84%
frozen_pizza	5,28%	candy_-_checklane	3,78%
hispanic	4,86%	soup	3,74%
fruit_-_shelf_stable	4,72%	hispanic	3,65%
cold_cereal	4,61%	_sauces	3,29%

Table 2.13: Top 10 categories for each motivation using Hierarchical LDA.

- Previous motivation preferences per customer: Another relevant source of information is related to the historical preferences of each customer. In particular, it would be relevant to store the average probability for each shopping motivation. For example, if a certain customer historically has more transactions related to healthy products, it would make sense to use this information to adapt the motivation probability for following transactions.
- Dynamic customer characteristics: A potentially useful source of information might be a transactional summary about previous shopping trips in terms of recency, frequency and monetary value (Bult and Wansbeek, 1995). This information would be useful to predict the motivations of future shopping baskets. For example, if a customer on average makes large purchases every 30 days, we may anticipate that in the next 30 days the customer will have a similar motivation. If instead a customer recently made a large purchase, this also gives information implying that a shopping trip a few days later are less likely to be again a large groceries replenishment basket.

It is also interesting to note that both extensions of the LDA method tried to predict shopping behavior (either basket size or shopping trip motivations). One important risk in the evaluation procedure of these methods is overfitting. This is the risk of over training machine learning models, fitting very well to the calibrated data, but loosing generalization capability and under performing in out of sample datasets. It would be useful to evaluate this prediction in a validation set and compare it with competitive prediction models built from other machine learning techniques, such as random forests or boosting trees. The benefit of LDA, is that it yields a parsimonious probability model that is easy to interpret. For example, we know that Motivation 1 is more important to predict basket size.

Another interesting analysis that could be performed in future research is related to the source of information used to extract shopping motivations. Some researches use polls or customer surveys to infer shopping motivations from customers. Our approach estimates

Mean_M1	SD_M1	$\hat{R}$_M1	Mean_M2	SD_M2	$\hat{R}$_M2
Min. :0.04386	Min. :0.06021	Min. :0.9999	Min. :0.1213	Min. :0.06021	Min. :0.9999
1st Qu.:0.34320	1st Qu.:0.24420	1st Qu.:1.0018	1st Qu.:0.3464	1st Qu.:0.24420	1st Qu.:1.0018
Median :0.50776	Median :0.29620	Median :1.0048	Median :0.4922	Median :0.29620	Median :1.0048
Mean :0.49183	Mean :0.28476	Mean :1.0135	Mean :0.5082	Mean :0.28476	Mean :1.0135
3rd Qu.:0.65361	3rd Qu.:0.34633	3rd Qu.:1.0112	3rd Qu.:0.6568	3rd Qu.:0.34633	3rd Qu.:1.0112
Max. :0.87875	Max. :0.36702	Max. :1.3463	Max. :0.9561	Max. :0.36702	Max. :1.3463

Table 2.14: Theta posterior distribution using Hierarchical LDA.

Variable	Parameter	Mean	sd	2.5%	97.5%	$\hat{R}$
Age	δ_{11}	0.0889	0.0562	0.0057	0.2204	1.0026
Income	δ_{21}	0.0694	0.0487	0.0030	0.1847	1.0036
Household Size	δ_{31}	0.0482	0.0358	0.0020	0.1322	1.0131
Recency	δ_{41}	0.0385	0.0310	0.0015	0.1149	1.0102
Last Amount	δ_{51}	0.0422	0.0348	0.0015	0.1270	1.0745
Intercept	δ_{61}	0.0257	0.0233	0.0007	0.0865	1.0002

Table 2.15: Posterior mean for δ parameter using HLDA.

shopping motivations based on transactional data. An opportunity to improve the predictions can appear if we mix both sources: customer survey and transactional data. If customers declare their shopping motivation, we can use this information into a supervised classification model. There is an extension of LDA that does this mix of sources, called Labeled-LDA (Ramage et al., 2009). If we could get that customer information about what they declare as shopping motivation, we can calibrate the Labeled-LDA in order to improve our predictions.

2.5.2 Managerial Implications

From a managerial perspective, the output of these models delivers important can be valuable to retailers. For example, it could be help in the process of creating targeted promotion campaigns in several ways:

- Recommendation Systems. In the absence of a recommendation system implementation, the LDA output can be used as a recommendation tool using the information estimated in the ϕ matrix. With a prediction of a particular motivation, this matrix provides the most likely products to be purchased. This information can be a good starting point as recommendation system. Second, if a recommendation system is already in use by a retailer, the LDA output may potentially improve the quality of the recommendations. These tools work based on collaborative filtering of previous transactions analyzing historical records. Using the LDA ouptut we can filter this historical clustering of records using the motivations detected. Hence, recommendations could be improved by conditioning on the purchase motivations predicted for each transaction.
- Promotions depending on Basket Size. Using the LDA regression model we can estimate more precisely the basket size conditional on a shopping motivation. Using this information retailers can design in-store marketing campaigns to increase the basket value via cross- and up-selling.
- Cross Category Promotions. An important value of detecting shopping motivations is

that we estimate category purchase probability, conditioned on a certain motivation. This information is important for category management. In this context, decisions such as price, promotion, assortment for a product category, are typically made independently from decisions for other product categories. LDA gives information about what are the most relevant categories for a particular shopping motivation. This information could be used to coordinate cross category initiatives and cross selling promotions.

In conclusion, LDA is a powerful and parsimonious tool that can be used not only for text analysis, but as shown in this paper for the analysis of shopping baskets from transactional data. In particular, the application of these models can help a researcher to detect shopping motivations and how they are related to customer or shopping trip characteristics. This information can be used to support cross-category management initiatives.

Conclusions

This thesis consider two essays about retail management problems covered with machine learning approaches. Both methodologies use transactional data to discover consumer behavior and complementary and substitution patters between products, brands and categories.

The first research entails the development of a novel data driven methodology to extract rules from transactional data, to get more robust solutions to the multi-product price optimization setting. The methodology proposed explores the transactional data of product category demand and prices to extract two different types of rules: order rules to formalize how prices have been decided in the past, and performance rules to indicate conditions that lead, on average, to better profit solutions.

We use the pricing rule set as constraints for the pricing optimization problem. To estimate the category profit conditioned to price decision, a double log demand model was calibrated using three different elasticity matrix estimation procedures (OLS, HBM and L1 regularization). When only order rules are applied as constraints, the profit solution obtained is still far from the actual values, giving intuition of an unrealistic solution and improving profit using uncertainty in the estimation of the elasticity matrix in the objective function. When both performance and order rules are applied, the solution is close to the previous values of profit. More robust estimations such as with the Hierarchical Bayesian Model and L1 regularization results in improvements of 16.7% and 9.0% respectively.

There is an additional value of pricing rules for practitioners and managers that is the interpretability. Order rules works as a structured method to recover previous managerial decision making about price categories. It is a way to extract from data what managers consistently decided in the past for the category price vector. We think that there is an important implicit knowledge hidden in previous decisions.

In the same way, performance rules give a structured way to understand where good profit scenarios are located in the price space solution. More than a cross elasticity term, it gives intuition about what happens with the expected profit if a certain distance in prices for a couple of products is decided upon.

The implementation of pricing rules also seems to be easier. For traditional SQL database servers, applying pricing rules as queries seems very practical to understand if a specific alternative price vector is feasible or not, and also evaluate what-if scenarios if some prices do not respect a specific rule. Using the methodology proposed, it is easy and fast to extract thousands of pricing rules, sort them to select the most relevant and valuable rules to apply in a certain category.

Pricing rules are built analyzing a database with millions of transactions comparing category performance for different historical pricing scenarios. In that sense, the most important contribution of this methodology is that pricing rules extracted provide a robust solution to price optimization based on historical performance of different pricing policies and their impact in the category demand.

As a summary, in the first research we analyzed transactional data to identify what business rules have been consistently applied before and evaluate which ones are associated to better business performance. Based on these pricing rules, we build a data-driven set of feasible prices and combine it with standard price optimization routines to understand how this information can complement the traditional econometric analysis of the demand. This novel approach for pricing optimization is easy to implement and not only provides managers with a reliable automatic mechanism to decide about prices of multiple products, but also enables more consistent decisions.

The second essay uses a Latent Dirichlet Allocation framework to analyze shopping behavior based on transactional data. In our retail context, we extract latent shopping motivations analyzing relationships among product categories. The contribution of this research is then to apply LDA in a retailing setting to achieve three different goals as follows. First, we apply a standard LDA model to detect and describe shopping motivations. Second, the basic LDA model is extended to jointly estimate the latent shopping motivations and the relationship between these motivations and basket size using a supervised approach. Finally, the LDA model is generalized to allow purchase motivations to depend on customer and shopping trip characteristics.

We applied this methodology to a supermarket transactional database focusing on 36 different product categories. We extract two shopping motivations. The first motivation is related to hedonic products that represents 48% of the total purchased products. The second motivation extracted is related to breakfast product categories accounting for 52% of the total purchases. The second model proposed is a supervised LDA approach using the motivation probabilities as independent variables to predict a variable of interest, in our application this corresponds to the dollar value of the basket. The motivations extracted changed with respect to the motivations obtained in LDA detecting shopping patterns related to Low Effort Food and Ethnic and Fresh products. The regression of amount spent shows a good fit of the data using only the motivations extracted with LDA.

The third model calibrated in this paper allows shopping motivations to depend on customer and shopping trip characteristics. In our setting we used age, income, family size, recency and the last purchase amount. These characteristics are positively associated with the first motivation (Hedonic products). Motivations extracted were different than results obtained with LDA. We also note that the distribution of motivation probabilities was more extreme than in the basic LDA, because of the use of customer demographics. In conclusion, LDA is a powerful and parsimonious tool that can be used not only for text analysis, but as shown in this paper for the analysis of shopping baskets from transactional data. In particular, the application of these models can help a researcher to detect shopping motivations and how they are related to customer or shopping trip characteristics. This information can be used to support cross-category management initiatives.

Both approaches contribute to solve retail problems with novelty approaches of Machine Learning. The use of these methods can lead to retailers to improve their pricing and promotion decisions, learning from consumer behavior and improving their performance in this competitive industry.

Bibliography

Agarwal, R. and Dhar, V. (2014). Editorial—big data, data science, and analytics: The opportunity and challenge for IS research. *Information Systems Research*, 25(3):443–448.

Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules. *Proceeding VLDB '94 Proceedings of the 20th International Conference on Very Large Data Bases*, 1215:487–499.

Arora, N., Dreze, X., Ghose, A., Hess, J. D., Iyengar, R., Jing, B., Joshi, Y., Kumar, V., Lurie, N., Neslin, S., et al. (2008). Putting one-to-one marketing to work: Personalization, customization, and choice. *Marketing Letters*, 19(3-4):305.

Bell, D. R., Corsten, D., and Knox, G. (2011). From point of purchase to path to purchase: How preshopping factors drive unplanned buying. *Journal of Marketing*, 75(1):31–45.

Beyer, K., Goldstein, J., Ramakrishnan, R., and Shaft, U. (1999). When is "nearest neighbor" meaningful? In *International conference on database theory*, pages 217–235. Springer.

Bingham, C. B. and Eisenhardt, K. M. (2011). Rational heuristics: the 'simple rules' that strategists learn from process experience. *Strategic management journal*, 32(13):1437–1464.

Blei, D. and Lafferty, J. (2006). Correlated topic models. *Advances in neural information processing systems*, 18:147.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, 3(Jan):993–1022.

Bradlow, E. T., Gangwar, M., Kopalle, P., and Voleti, S. (2017). The role of big data and predictive analytics in retailing. *Journal of Retailing*, 93(1):79–95.

Brin, S., Motwani, R., and Silverstein, C. (1997). Beyond market baskets: generalizing association rules to correlations. *ACM SIGMOD Record*, 26(2):265–276.

Bucklin, R. E. and Gupta, S. (1999). Commercial Use of UPC Scanner Data: Industry and Academic Perspectives. *Marketing Science*, 18(3):247–273.

Bult, J. R. and Wansbeek, T. (1995). Optimal selection for direct mail. *Marketing Science*, 14(4):378–394.

Büschken, J. and Allenby, G. M. (2016). Sentence-based text analysis for customer reviews. *Marketing Science*, 35(6):953–975.

Chintagunta, P., Hanssens, D. M., and Hauser, J. R. (2016). Editorial—marketing science and big data. *Marketing Science*, 35(3):341–342.

Christidis, K. and Mentzas, G. (2013). A topic-based recommender system for electronic marketplace platforms. *Expert Systems with Applications*, 40(11):4370–4379.

Cohen, M. and Perakis, G. (2018). Promotion optimization in retail. *Working Paper, Leonard N. Stern School of Business, New York University.*

Cohen, M. C., Leung, N.-H. Z., Panchamgam, K., Perakis, G., and Smith, A. (2017). The impact of linear optimization on promotion planning. *Operations Research*, 65(2):446–468.

Desai, P. S., Deighton, J., Rizley, R., and Keane, S. (2012). Introducing marketing science institute research priorities to" marketing science". *Marketing Science*, 31(6):873–877.

Deveaud, R., SanJuan, E., and Bellot, P. (2014). Accurate and effective latent concept modeling for ad hoc information retrieval. *Document numérique*, 17(1):61–84.

ECR-Europe (2011). The consumer and shopper journey framework. *ECR Europe.*

Erevelles, S., Fukawa, N., and Swayne, L. (2016). Big data consumer analytics and the transformation of marketing. *Journal of Business Research*, 69(2):897–904.

Ferreira, K. J., Lee, B. H. A., and Simchi-Levi, D. (2015). Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, 18(1):69–88.

Grewal, D., Ailawadi, K. L., Gauri, D., Hall, K., Kopalle, P., and Robertson, J. R. (2011). Innovations in retail pricing and promotions. *Journal of Retailing*, 87:S43–S52.

Griva, A., Bardaki, C., Pramatari, K., and Papakiriakopoulos, D. (2018). Retail business analytics: Customer visit segmentation using market basket data. *Expert Systems with Applications*, 100:1–16.

Han, J., Dong, G., and Yin, Y. (1999). Efficient mining of partial periodic patterns in time series database. In *Proceedings 15th International Conference on Data Engineering (Cat. No. 99CB36337)*, pages 106–115. IEEE.

Han, J., Koperski, K., and Stefanovic, N. (1997). Geominer: A system prototype for spatial data mining. *SIGMOD Rec.*, 26(2):553–556.

Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.

Han, J., Pei, J., and Yin, Y. (2000). Mining frequent patterns without candidate generation.

ACM SIGMOD Record, 29(2):1–12.

Handorf, T. W., Haussler, C., and Duffy, M. D. (2016). Methods and apparatus for maintaining business rules in a configuration system. US Patent 9,524,506.

Hauser, J., Tellis, G. J., and Griffin, A. (2006). Research on Innovation: A Review and Agenda for Marketing Science. *Marketing Science*, 25(6):687–717.

Hawtin, M. (2003). The practicalities and benefits of applying revenue management to grocery retailing, and the need for effective business rule management. *Journal of revenue and pricing management*, 2(1):61–68.

Hruschka, H. (2014). Linking multi-category purchases to latent activities of shoppers: analysing market baskets by topic models. *Marketing: ZFP–Journal of Research and Management*, 36(4):267–273.

Jacobs, B. J., Donkers, B., and Fok, D. (2016). Model-based purchase predictions for large assortments. *Marketing Science*, 35(3):389–404.

Jelodar, H., Wang, Y., Yuan, C., Feng, X., Jiang, X., Li, Y., and Zhao, L. (2019). Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia Tools and Applications*, 78(11):15169–15211.

Kahn, B. E. and Schmittlein, D. C. (1989). Shopping trip behavior: An empirical investigation. *Marketing Letters*, 1(1):55–69.

Kollat, D. T. and Willett, R. P. (1967). Customer impulse purchasing behavior. *Journal of marketing research*, 4(1):21–31.

Kopalle, P., Biswas, D., Chintagunta, P. K., Fan, J., Pauwels, K., Ratchford, B. T., and Sills, J. A. (2009). Retailer Pricing and Competitive Effects. *Journal of Retailing*, 85(1):56–70.

Kotsiantis, S. and Kanellopoulos, D. (2006). Association Rules Mining: A Recent Overview. *Science*, 32(1):71–82.

Kunz, T. P. and Crone, S. F. (2015). The impact of practitioner business rules on the optimality of a static retail revenue management system. *Journal of Revenue and Pricing Management*, 14(3):198–210.

Levy, M., Grewal, D., Kopalle, P. K., and Hess, J. D. (2004). Emerging trends in retail pricing practice: Implications for research. *Journal of Retailing*, 80(3).

Lilien, G. L. (2011). Bridging the academic–practitioner divide in marketing decision models. *Journal of Marketing*, 75(4):196–210.

Lilien, G. L., Roberts, J. H., and Shankar, V. (2013). Effective Marketing Science Applications: Insights from the ISMS-MSI Practice Prize Finalist Papers and Projects. *Marketing Science*, 32(March 2015):229–245,356–358.

Little, J. D. C. and Shapiro, J. F. (1980). A Theory for Pricing Nonfeatured Products in Supermarkets. *Journal of Business*, 53(3, Part 2: Interfaces Between Marketing and Economics):S199–S209.

Ma, S. and Fildes, R. (2017). A retail store sku promotions optimization model for category multi-period profit maximization. *European Journal of Operational Research*, 260(2):680–692.

Manchanda, P., Ansari, A., and Gupta, S. (1999). The "shopping basket": A model for multicategory purchase incidence decisions. *Marketing science*, 18(2):95–114.

Mobasher, B., Dai, H., Luo, T., and Nakagawa, M. (2001). Effective personalization based on association rule discovery from web usage data. In *Proceedings of the 3rd international workshop on Web information and data management*, pages 9–15. ACM.

Montgomery, A. L. (1997). Creating micro-marketing pricing strategies using supermarket scanner data. *Marketing science*, 16(4):315–337.

Montgomery, A. L. and Goic, M. (2010). Making better pricing decisions with implied priors. *Tepper School of Business Working Paper*.

Musalem, A., Aburto, L., and Bosch, M. (2018). Market basket analysis insights to support category management. *European Journal of Marketing*, 52(7/8):1550–1573.

Natter, M., Reutterer, T., Mild, A., and Taudes, A. (2007). Practice prize report—an assortmentwide decision-support system for dynamic pricing and promotion planning in diy retailing. *Marketing Science*, 26(4):576–583.

Ngai, E. W., Xiu, L., and Chau, D. C. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert systems with applications*, 36(2):2592–2602.

Nielsen, A. (1998). Eighth annual survey of trade promotion practices. *AC Nielsen, Chicago, Illinois*.

Oracle (2018). Oracle Retail Advanced Science Cloud Services Implementation Guide. `https://docs.oracle.com/cd/E76441_01/orase/pdf/orase/170/html/implementation_guide/offer-opt.htm`. [Online; accessed 17-March-2019].

Ramage, D., Hall, D., Nallapati, R., and Manning, C. D. (2009). Labeled lda: A supervised topic model for credit attribution in multi-labeled corpora. In *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*, pages 248–256. Association for Computational Linguistics.

Reibstein, D. J. and Gatignon, H. (1984). Optimal Product Line Pricing: The Influence of Elasticities and Cross-Elasticities. *Journal of Marketing Research*, 21(3):259.

Reisenbichler, M. and Reutterer, T. (2019). Topic modeling in marketing: recent advances and research opportunities. *Journal of Business Economics*, 89(3):327–356.

Roberts, J. H., Kayande, U., and Stremersch, S. (2014). From academic research to marketing practice: Exploring the marketing science value chain. *International Journal of Research in Marketing*, 31(2):127–140.

Rosenberg, A. N. (2018). System and method for integrating retail price optimization for revenue and profit with business rules. US Patent 9,940,649.

Rosenberg, A. N. and Sills, J. A. (2019). System and method for determining retail-business-rule coefficients from current prices. US Patent App. 10/181,138.

Russell, G. J. and Petersen, A. (2000). Analysis of cross category dependence in market basket selection. *Journal of Retailing*, 76(3):367–392.

Sarantopoulos, P., Theotokis, A., Pramatari, K., and Doukidis, G. (2016). Shopping missions: An analytical method for the identification of shopper need states. *Journal of Business Research*, 69(3):1043–1052.

Schröder, N., Falke, A., Hruschka, H., and Reutterer, T. (2019). Analyzing the browsing basket: A latent interests-based segmentation tool. *Journal of Interactive Marketing*, 47:181–197.

Tan, P.-N., Kumar, V., and Srivastava, J. (2002). Selecting the right interestingness measure for association patterns. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 32–41. ACM.

Tirunillai, S. and Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of Marketing Research*, 51(4):463–479.

Treinen, J. J. and Thurimella, R. (2006). A framework for the application of association rule mining in large intrusion detection infrastructures. In *International Workshop on Recent Advances in Intrusion Detection*, pages 1–18. Springer.

Videla-Cavieres, I. F. and Rios, S. A. (2014). Extending market basket analysis with graph mining techniques: A real case. *Expert Systems with Applications*, 41(4):1928–1936.

Vilcassim, N. J. and Chintagunta, P. K. (1995). Investigating retailer product category pricing from household scanner panel data. *Journal of Retailing*, 71(2):103–128.

Walters, R. G. (1991). Assessing the Impact of Retail Price Promotions on Product Substitution, Complementary Purchase, and Interstore Sales Displacement. *Source Journal of Marketing*, 55(2):17–28.

Walters, R. G. and Jamil, M. (2003). Exploring the relationships between shopping trip type, purchases of products on promotion, and shopping basket profit. *Journal of Business Research*, 56(1):17–29.

2.6 Appendix

2.6.1 Pricing rules extract procedure

Algorithm 1: Pricing Rules Extract Procedure

Input : Aggregate transactional data at product - weekly level
Input : Support and Lift threshold to select relevant rules
Output: Set of Pricing Rules for optimization: Order Rules Ω_o and Performance Rules Ω_p

foreach *Product* P_{i} *in the Product Category Set* **do**
    Normalize prices using product formats.
    Create Pricing features (absolute, differences and mean) for every week.
    Create price events per week discretizing continuous price features using parameters $\gamma^L, \gamma^U, \alpha^L, \alpha^U, \beta^L$ and β^U.
end foreach
Calculate total category profit for each week
Use Apriori algorithm over price events and profits per week, calculating support, confidence and lift measures
Use Lift threshold to select Performance Rules Ω_p
Use Support threshold to select Order Rules Ω_o

2.6.2 Robustness Analysis of optimal prices

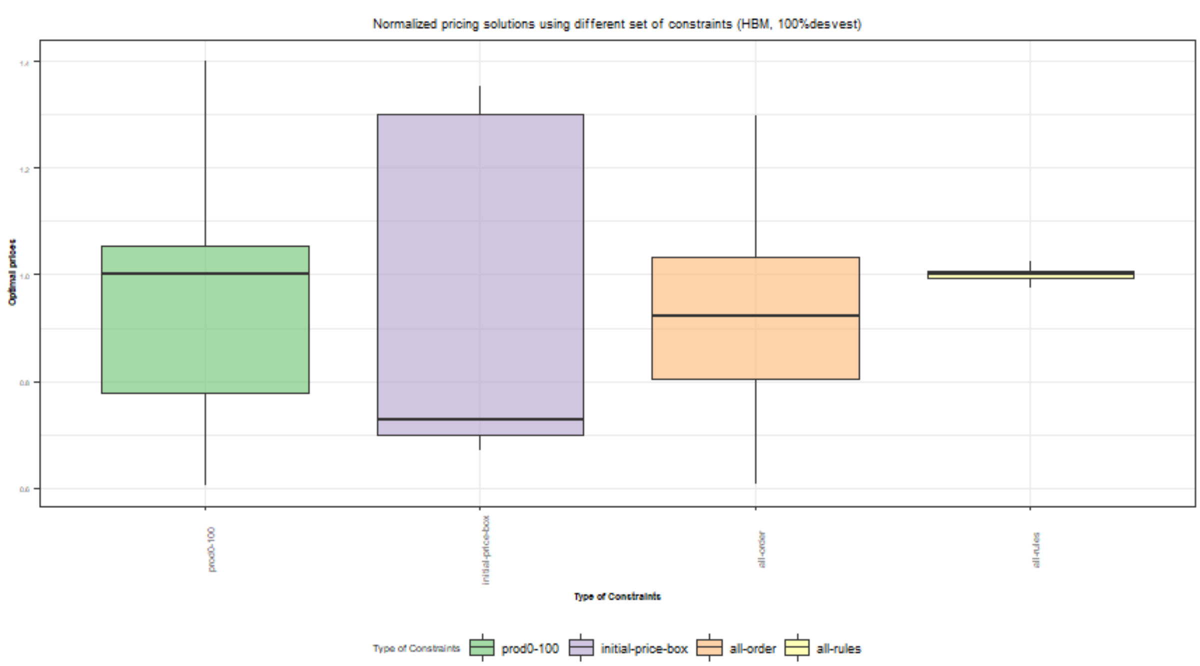

Figure 2.5: Normalized price solutions using different set of constraints. HBM Model

2.6.3 Robustness Analysis of optimal prices desegregated

2.6.4 Stan code for LDA

```
data {
  int<lower=2> K;                    // Num motivations
  int<lower=2> V;                    // Product category list
  int<lower=1> M;                    // num transactions
  int<lower=1> N;                    // Total products instances
  int<lower=1,upper=V> w[N];         // Product category n
  int<lower=1,upper=M> doc[N];       // Transaction ID for product n
  vector<lower=0>[V] beta;           // Product hyperparameter prior
}
parameters {
  vector<lower=0>[K] alpha;          // Motivation prior
  simplex[K] theta[M];               // Motiv dist for trx m
  simplex[V] phi[K];                 // Product dist for mot k
}
model {
  for (k in 1:K)
    alpha[k] ~ uniform(0,3);
  for (k in 1:K)
    phi[k] ~ dirichlet(beta);          // prior
  for (m in 1:M)
    theta[m] ~ dirichlet(alpha);       // prior
  for (n in 1:N) {
    real gamma[K];
    for (k in 1:K)
      gamma[k] = log(theta[doc[n], k]) + log(phi[k, w[n]]);
    target += log_sum_exp(gamma);      // likelihood;
  }
}
```

2.6.5 Stan code for LDA regression of basket amount spend

```
data {
  int<lower=2> K;                    // Num motivations
  int<lower=2> V;                    // Product category list
  int<lower=1> M;                    // Num transactions
  int<lower=1> N;                    // Total products instances
  int<lower=1,upper=V> w[N];         // Product category n
  int<lower=1,upper=M> doc[N];       // Transaction ID for product n
  vector<lower=0>[V] beta;           // Product hyperparameter prior
  vector[M] amount;
}
```

```
parameters {
  vector<lower=0>[K] alpha;        // Motivation prior
  simplex[K] theta[M];             // Motiv dist for trx m
  simplex[V] phi[K];               // Product dist for mot k
  vector[K] rho2;                  // Amount importance parameter
  real sigma;                      // Amount deviation parameter
}

transformed parameters {
vector[K] rho;                     // Amount parameters
rho[1]=0;
rho[2:K]=rho2[2:K];
}

model {
  sigma ~ uniform(0,2);
  rho2~normal(0,1);
  for (k in 1:K)
    alpha[k] ~ uniform(0,3);
  for (k in 1:K)
    phi[k] ~ dirichlet(beta);        // prior
  for (m in 1:M){
      amount[m] ~ normal(rho'*theta[m]+rho2[1],sigma);
      theta[m] ~ dirichlet(alpha);   // prior
  }
  for (n in 1:N) {
        real gamma[K];

    for (k in 1:K)
      gamma[k] = log(theta[doc[n], k]) + log(phi[k, w[n]]);
    target += log_sum_exp(gamma);   // likelihood;
  }
}
```

2.6.6 Stan code for Hierarchical LDA to predict shopping trip motivation

```
// version 2
// lambda=1 and fixing one of the intercepts to zero.
data {
  int<lower=2> K;                  // num motivations
  int<lower=2> V;                  // num product categories
  int<lower=1> M;                  // num transactions
  int<lower=1> N;                  // total product instances
  int<lower=1,upper=V> w[N];       // product n
```

```
  int<lower=1,upper=M> doc[N];   // trx ID for prod n
  vector<lower=0>[V] beta;        // Product prior
  int<lower = 1> D;               // num of covariables to predict Alpha
  matrix[M, D] x;                 // covariables data for each trx M

}
parameters {

  //vector<lower=0>[K] alpha;       // Alpha for each trx
  matrix[D, K-1] delta;             //importance of variable D in mot K
  simplex[K] theta[M];              // Motiv dist for trx m
  simplex[V] phi[K];                // Product dist for mot k
//   vector[K-1] inter;              // Intercept
//   real lambda;
}

transformed parameters {
//matrix[M, K] x_delta;
vector<lower=0>[K] alpha[M];      // Alpha for each trx
matrix[D, K] delta2;              // importance of variable D in mot K
matrix[M, K] x_delta;
//real x2[M, D+1];                // covariables data for each trx M
delta2[1:D,1:K-1]=delta;
for (d in 1:D)
delta2[d,K]=0;
//delta2[D,K]=0;

x_delta = x * delta2;

for (m in 1:M)
    //   for (k in 1:K)
    //      alpha[m,k] = lambda*inv_logit(x_delta[m,k]);;
    // alpha[m] = lambda*softmax(x_delta[m]');
    alpha[m] = softmax(x_delta[m]');
}

model {
// lambda  ~ uniform(0.1,15);
for (k in 1:K-1)
   delta[,k] ~ uniform(0,5);
for (k in 1:K)
   phi[k] ~ dirichlet(beta);
//for (k in 1:K)
  for (m in 1:M)
    theta[m] ~ dirichlet(alpha[m]);   // prior
  for (n in 1:N) {
    real gamma[K];
```

```
    for (k in 1:K)
      gamma[k] = log(theta[doc[n], k]) + log(phi[k, w[n]]);
    target += log_sum_exp(gamma);  // likelihood;
  }
}
```

Categories	% Ticket presence	% Dollar Value
SOFT_DRINKS	38,0%	3,5%
FLUID_MILK_PRODUCTS	33,4%	2,8%
BAKED_BREAD/BUNS/ROLLS	29,8%	1,8%
CHEESE	25,6%	2,2%
BAG_SNACKS	23,1%	1,7%
BABY_FOODS	18,2%	0,7%
BEEF	17,4%	3,7%
TROPICAL_FRUIT	16,3%	0,7%
CANDY_-_CHECKLANE	15,6%	0,4%
YOGURT	14,7%	0,8%
COLD_CEREAL	14,7%	1,5%
CANDY_-_PACKAGED	13,6%	1,0%
VEGETABLES_-_SHELF_STABLE	12,8%	0,5%
SOUP	12,7%	0,7%
DELI_MEATS	12,7%	1,8%
FROZEN_PIZZA	12,1%	1,3%
WATER_-_CARBONATED/FLVRD_DRINK	11,4%	1,3%
FRZN_MEAT/MEAT_DINNERS	11,4%	1,0%
CANNED_JUICES	10,4%	1,0%
FRUIT_-_SHELF_STABLE	10,1%	0,7%
COOKIES/CONES	9,8%	0,9%
CRACKERS/MISC_BKD_FD	9,7%	0,7%
LUNCHMEAT	9,7%	0,9%
EGGS	9,5%	0,4%
ICE_CREAM/MILK/SHERBTS	9,2%	0,9%
DRY_BN/VEG/POTATO/RICE	8,9%	0,5%
COUPON/MISC_ITEMS	8,8%	6,8%
SALD_DRSNG/SNDWCH_SPRD	8,6%	0,6%
REFRGRATD_JUICES/DRNKS	8,6%	0,7%
CONVENIENT_BRKFST/WHLSM_SNACKS	8,6%	0,7%
CONDIMENTS/SAUCES	8,3%	0,5%
BEERS/ALES	8,3%	2,6%
BAKED_SWEET_GOODS	8,0%	0,6%
HISPANIC	7,8%	0,5%
BAKING_MIXES	7,4%	0,4%
FRZN_NOVELTIES/WTR_ICE	7,3%	0,7%

Table 2.16: Summary descriptive statistics 2

	Mean_M1	SD_M1	2.5%M1	97.5%M1	Rhat_M1	Mean_M2	SD_M2	2.5%M2	97.5%M2	Rhat_M2
_cones	0.02	0.01	0.00	0.04	1.10	0.02	0.01	0.01	0.04	1.10
_sauces	0.00	0.00	0.00	0.00	1.00	0.04	0.00	0.03	0.05	1.06
baby_foods	0.02	0.01	0.01	0.03	1.59	0.00	0.00	0.00	0.01	1.66
bag_snacks	0.07	0.01	0.05	0.10	1.07	0.02	0.01	0.01	0.04	1.09
baked_bread_buns_	0.05	0.01	0.02	0.07	1.01	0.07	0.01	0.05	0.09	1.01
baked_sweet_goods	0.04	0.01	0.03	0.06	1.00	0.00	0.00	0.00	0.01	1.07
beef	0.05	0.01	0.03	0.07	1.07	0.03	0.01	0.02	0.05	1.08
beers_ales	0.06	0.01	0.04	0.07	1.10	0.00	0.00	0.00	0.00	1.00
candy_-_checklane	0.07	0.01	0.06	0.08	1.12	0.00	0.00	0.00	0.00	1.00
candy_-_packaged	0.05	0.01	0.03	0.08	1.21	0.01	0.01	0.00	0.02	1.39
canned_juices	0.00	0.00	0.00	0.00	1.01	0.04	0.00	0.03	0.05	1.04
cheese	0.05	0.01	0.03	0.07	1.05	0.05	0.01	0.03	0.07	1.05
cold_cereal	0.00	0.00	0.00	0.00	1.00	0.05	0.01	0.04	0.06	1.09
convenient_brkfst_whlsm_snacks	0.00	0.00	0.00	0.00	1.00	0.03	0.00	0.02	0.04	1.02
coupon_misc_items	0.07	0.01	0.05	0.08	1.09	0.00	0.00	0.00	0.00	1.01
crackers_misc_bkd_fd	0.00	0.00	0.00	0.01	1.01	0.04	0.00	0.03	0.05	1.02
deli_meats	0.00	0.00	0.00	0.01	1.01	0.05	0.00	0.04	0.06	1.08
deli_specialties_(retail_pk)	0.00	0.00	0.00	0.00	1.53	0.00	0.00	0.00	0.00	1.32
dry_bn_veg_potato_rice	0.00	0.00	0.00	0.01	1.03	0.03	0.00	0.03	0.04	1.04
eggs	0.00	0.00	0.00	0.00	1.00	0.05	0.00	0.04	0.06	1.10
fluid_milk_products	0.07	0.01	0.04	0.10	1.03	0.08	0.01	0.06	0.10	1.02
frozen_-_boxed(grocery)	0.00	0.00	0.00	0.00	1.56	0.00	0.00	0.00	0.00	1.36
frozen_pizza	0.06	0.01	0.05	0.07	1.07	0.00	0.00	0.00	0.00	1.00
fruit_-_shelf_stable	0.00	0.00	0.00	0.00	1.00	0.03	0.00	0.02	0.04	1.04
frzn_meat_meat_dinners	0.04	0.01	0.03	0.05	1.08	0.00	0.00	0.00	0.00	1.01
hispanic	0.00	0.00	0.00	0.00	1.00	0.03	0.00	0.02	0.03	1.04
ice_cream_milk_sherbts	0.05	0.01	0.04	0.07	1.05	0.00	0.00	0.00	0.01	1.03
lunchmeat	0.05	0.01	0.04	0.06	1.07	0.00	0.00	0.00	0.00	1.02
refrgratd_juices_drnks	0.00	0.00	0.00	0.00	1.00	0.04	0.00	0.03	0.05	1.04
sald_drsng_sndwch_sprd	0.00	0.00	0.00	0.00	1.00	0.04	0.00	0.03	0.05	1.10
soft_drinks	0.17	0.01	0.14	0.20	1.07	0.01	0.01	0.00	0.03	1.44
soup	0.00	0.00	0.00	0.00	1.00	0.03	0.00	0.03	0.04	1.06
tropical_fruit	0.00	0.00	0.00	0.01	1.02	0.08	0.01	0.06	0.09	1.12
vegetables_-_shelf_stable	0.00	0.00	0.00	0.00	1.00	0.04	0.00	0.04	0.05	1.05
water_-_carbonated_flvrd_drink	0.01	0.01	0.00	0.03	1.48	0.04	0.01	0.02	0.05	1.36
yogurt	0.00	0.00	0.00	0.00	1.00	0.03	0.00	0.02	0.04	1.06

Table 2.17: Phi posterior mean, standard deviation and Rhat for LDA.

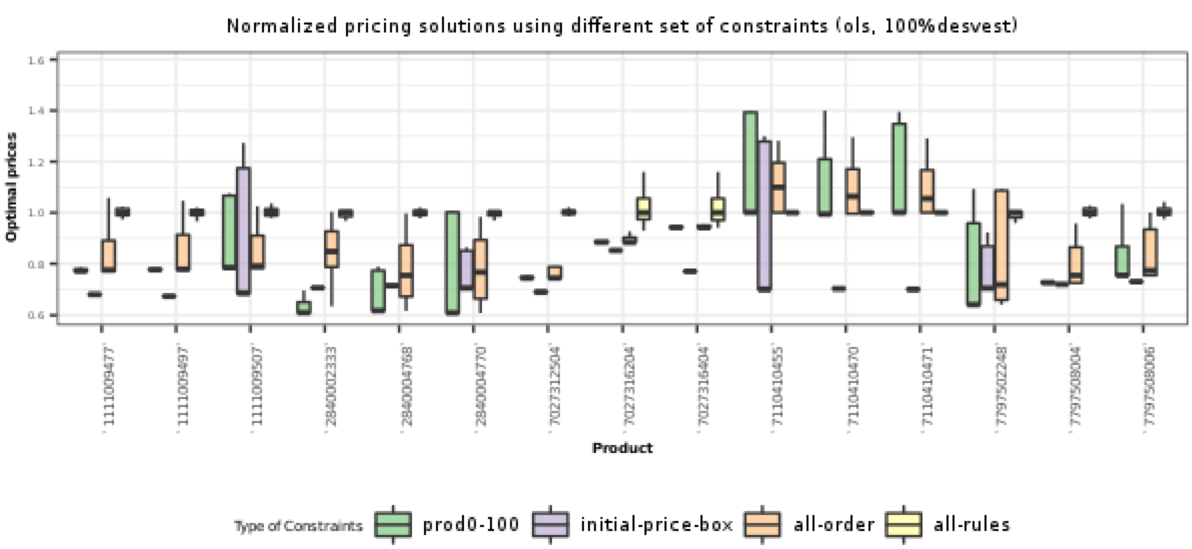

Figure 2.6: Normalized price solutions using different set of constraints. OLS Model

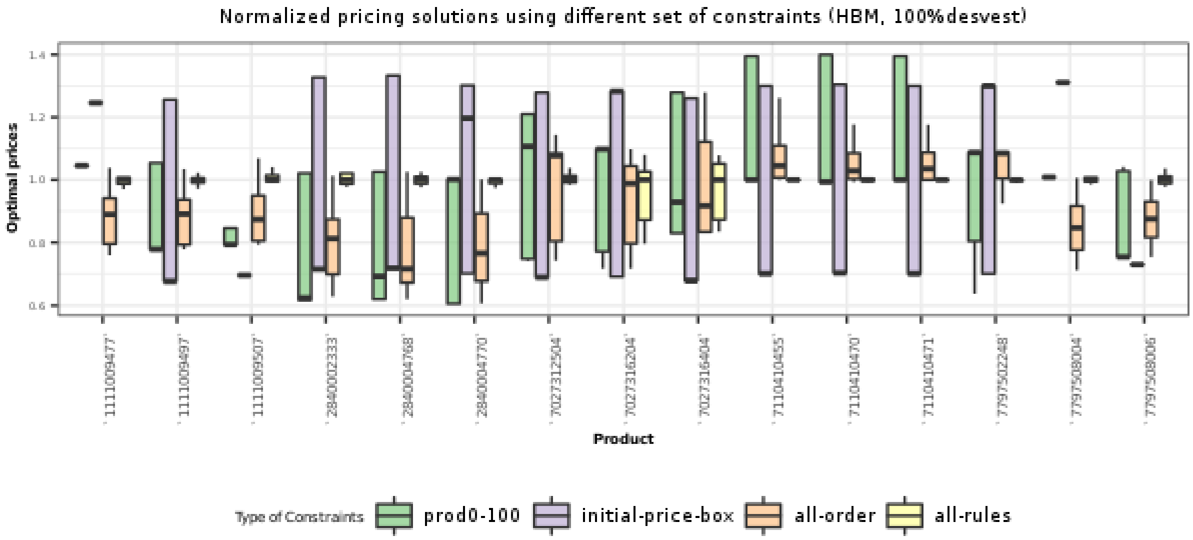

Figure 2.7: Normalized price solutions using different set of constraints. HBM Model

	Mean_M1	SD_M1	2.5%M1	97.5%M1	Rhat_M1	Mean_M2	SD_M2	2.5%M2	97.5%M2	Rhat_M2
_cones	0.03	0.00	0.02	0.04	1.00	0.03	0.00	0.02	0.03	1.00
_sauces	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
baby_foods	0.03	0.00	0.02	0.03	1.00	0.04	0.00	0.03	0.04	1.01
bag_snacks	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
baked_bread_buns_	0.01	0.00	0.01	0.02	1.00	0.06	0.01	0.05	0.08	1.00
baked_sweet_goods	0.00	0.00	0.00	0.01	1.00	0.12	0.03	0.07	0.17	1.01
beef	0.05	0.01	0.04	0.06	1.02	0.00	0.00	0.00	0.00	1.00
beers_ales	0.01	0.02	0.00	0.05	1.02	0.00	0.00	0.00	0.00	1.00
candy_-_checklane	0.06	0.01	0.05	0.08	1.00	0.03	0.00	0.03	0.04	1.00
candy_-_packaged	0.06	0.02	0.02	0.10	1.01	0.00	0.00	0.00	0.01	1.00
canned_juices	0.02	0.00	0.02	0.03	1.00	0.02	0.00	0.02	0.03	1.01
cheese	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
cold_cereal	0.05	0.00	0.04	0.06	1.02	0.02	0.00	0.01	0.03	1.01
convenient_brkfst_whlsm_snacks	0.00	0.00	0.00	0.00	1.00	0.00	0.01	0.00	0.03	1.01
coupon_misc_items	0.00	0.00	0.00	0.01	1.04	0.02	0.00	0.01	0.02	1.00
crackers_misc_bkd_fd	0.10	0.02	0.07	0.13	1.03	0.00	0.00	0.00	0.00	1.00
deli_meats	0.00	0.00	0.00	0.00	1.02	0.01	0.01	0.00	0.02	1.09
deli_specialties_(retail_pk)	0.13	0.01	0.10	0.16	1.03	0.08	0.02	0.04	0.11	1.06
dry_bn_veg_potato_rice	0.02	0.00	0.01	0.03	1.02	0.03	0.00	0.01	0.03	1.01
eggs	0.06	0.02	0.03	0.09	1.01	0.00	0.01	0.00	0.04	1.02
fluid_milk_products	0.03	0.00	0.02	0.04	1.01	0.03	0.00	0.02	0.04	1.02
frozen_-_boxed(grocery)	0.00	0.00	0.00	0.00	1.00	0.00	0.01	0.00	0.03	1.03
frozen_pizza	0.06	0.00	0.05	0.07	1.00	0.03	0.00	0.02	0.03	1.01
fruit_-_shelf_stable	0.00	0.01	0.00	0.04	1.01	0.00	0.00	0.00	0.00	1.00
frzn_meat_meat_dinners	0.04	0.00	0.03	0.05	1.00	0.03	0.01	0.01	0.04	1.02
hispanic	0.00	0.00	0.00	0.00	1.00	0.26	0.03	0.21	0.32	1.00
ice_cream_milk_sherbts	0.02	0.00	0.02	0.03	1.01	0.03	0.00	0.02	0.03	1.00
lunchmeat	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
refrgratd_juices_drnks	0.00	0.00	0.00	0.00	1.00	0.05	0.01	0.04	0.06	1.03
sald_drsng_sndwch_sprd	0.12	0.02	0.10	0.16	1.05	0.04	0.02	0.00	0.08	1.05
soft_drinks	0.03	0.00	0.02	0.03	1.00	0.03	0.00	0.03	0.04	1.00
soup	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.00
tropical_fruit	0.03	0.00	0.03	0.04	1.01	0.03	0.00	0.03	0.04	1.03
vegetables_-_shelf_stable	0.00	0.00	0.00	0.02	1.01	0.00	0.00	0.00	0.01	1.00
water_-_carbonated_flvrd_drink	0.00	0.00	0.00	0.00	1.46	0.02	0.00	0.02	0.03	1.00
yogurt	0.00	0.00	0.00	0.01	1.46	0.00	0.00	0.00	0.00	1.00

Table 2.18: Phi posterior mean, standard deviation and Rhat for LDA Regression.

Mean_M1	SD_M1	Rhat_M1	Mean_M2	SD_M2	Rhat_M2
Min. :0.1478	Min. :0.04302	Min. :0.9998	Min. :0.05435	Min. :0.04302	Min. :0.9998
1st Qu.:0.5549	1st Qu.:0.10194	1st Qu.:1.0014	1st Qu.:0.21319	1st Qu.:0.10194	1st Qu.:1.0014
Median :0.6870	Median :0.11689	Median :1.0035	Median :0.31300	Median :0.11689	Median :1.0035
Mean :0.6577	Mean :0.11086	Mean :1.0052	Mean :0.34228	Mean :0.11086	Mean :1.0052
3rd Qu.:0.7868	3rd Qu.:0.12572	3rd Qu.:1.0072	3rd Qu.:0.44512	3rd Qu.:0.12572	3rd Qu.:1.0072
Max. :0.9456	Max. :0.14472	Max. :1.0695	Max. :0.85223	Max. :0.14472	Max. :1.0695

Table 2.19: Theta posterior summary statistics obtained using LDA Regression.

	Mean_M1	SD_M1	2.5%M1	97.5%M1	Rhat_M1	Mean_M2	SD_M2	2.5%M2	97.5%M2	Rhat_M2
_cones	0.02	0.01	0.01	0.04	1.10	0.00	0.00	0.00	0.01	1.01
_sauces	0.02	0.01	0.01	0.03	1.09	0.03	0.00	0.03	0.04	1.02
baby_foods	0.02	0.01	0.01	0.03	1.07	0.00	0.00	0.00	0.01	1.02
bag_snacks	0.02	0.00	0.02	0.03	1.03	0.05	0.00	0.04	0.06	1.01
baked_bread_buns_	0.01	0.01	0.00	0.03	1.68	0.08	0.01	0.06	0.11	1.06
baked_sweet_goods	0.01	0.01	0.00	0.02	1.64	0.07	0.01	0.06	0.09	1.06
beef	0.06	0.01	0.04	0.08	1.02	0.00	0.00	0.00	0.00	2.07
beers_ales	0.03	0.01	0.02	0.04	1.01	0.00	0.00	0.00	0.00	1.59
candy_-_checklane	0.06	0.01	0.04	0.08	1.02	0.04	0.01	0.02	0.05	1.07
candy_-_packaged	0.06	0.01	0.05	0.08	1.01	0.02	0.00	0.01	0.03	1.07
canned_juices	0.02	0.01	0.01	0.04	1.02	0.00	0.00	0.00	0.00	1.01
cheese	0.02	0.00	0.01	0.03	1.01	0.03	0.00	0.02	0.03	1.01
cold_cereal	0.05	0.01	0.03	0.06	1.00	0.02	0.01	0.00	0.03	1.06
convenient_brkfst_whlsm_snacks	0.04	0.01	0.03	0.05	1.01	0.01	0.00	0.01	0.02	1.08
coupon_misc_items	0.06	0.01	0.05	0.07	1.03	0.00	0.00	0.00	0.00	1.02
crackers_misc_bkd_fd	0.00	0.00	0.00	0.00	1.01	0.02	0.00	0.02	0.03	1.00
deli_meats	0.06	0.01	0.04	0.08	1.31	0.04	0.01	0.02	0.05	1.03
deli_specialties_(retail_pk)	0.01	0.01	0.00	0.02	1.72	0.01	0.00	0.00	0.02	1.04
dry_bn_veg_potato_rice	0.04	0.01	0.03	0.06	1.04	0.03	0.01	0.01	0.04	1.01
eggs	0.02	0.00	0.01	0.03	1.04	0.02	0.00	0.01	0.03	1.01
fluid_milk_products	0.00	0.00	0.00	0.02	1.03	0.01	0.01	0.00	0.03	1.17
frozen_-_boxed(grocery)	0.04	0.00	0.03	0.05	1.02	0.03	0.01	0.02	0.04	1.15
frozen_pizza	0.05	0.01	0.04	0.07	1.01	0.00	0.01	0.00	0.02	1.17
fruit_-_shelf_stable	0.05	0.01	0.04	0.06	1.01	0.03	0.00	0.02	0.04	1.07
frzn_meat_meat_dinners	0.00	0.00	0.00	0.02	1.02	0.15	0.01	0.12	0.18	1.05
hispanic	0.05	0.01	0.04	0.06	1.04	0.04	0.01	0.02	0.05	1.05
ice_cream_milk_sherbts	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	1.04
lunchmeat	0.03	0.00	0.02	0.03	1.01	0.03	0.00	0.03	0.04	1.01
refrgratd_juices_drnks	0.07	0.01	0.06	0.09	1.01	0.03	0.01	0.01	0.05	1.21
sald_drsng_sndwch_sprd	0.00	0.00	0.00	0.00	1.00	0.05	0.01	0.04	0.07	1.16
soft_drinks	0.01	0.01	0.00	0.02	1.03	0.01	0.01	0.00	0.02	1.56
soup	0.03	0.00	0.02	0.04	1.02	0.04	0.01	0.03	0.05	1.29
tropical_fruit	0.02	0.01	0.00	0.04	1.10	0.02	0.01	0.00	0.04	1.28
vegetables_-_shelf_stable	0.03	0.01	0.02	0.04	1.08	0.03	0.01	0.02	0.04	1.24
water_-_carbonated_flvrd_drink	0.00	0.00	0.00	0.01	1.00	0.00	0.00	0.00	0.00	1.01
yogurt	0.00	0.00	0.00	0.00	1.01	0.03	0.00	0.02	0.03	1.02

Table 2.20: Phi posterior mean and standard deviation for Hierarchical LDA model