Assortment planning and store space optimization: a practical application at a retailer

Alexandre Miguel Torres de Aragão

Master's Dissertation

Supervisor: Manuel Augusto de Pina Marques

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

Mestrado Integrado em Engenharia e Gestão Industrial

2017-07-17

“あきらめたらそこで試合終了だよ”

安西先生

Resumo

Gestão de gama e espaço são cada vez mais importantes para retalhistas à medida que as cadeias de abastecimento ficam mais *lean*, a competição mais feroz, e margens de lucro menores. Decidir que produtos expor em prateleira mostra a capacidade do retalhista de se adaptar a preferências individuais de cada consumidor e, além do impacto no lucro, uma gama adequada pode contribuir para a boa imagem da companhia e a sua cota de mercado no longo prazo.

O presente trabalho consiste numa otimização de gama e espaço de loja num retalhista eletrónico. Devido à sua natureza (baixa rotação de stock, ciclos de vida curtos e introdução dinâmica de produtos no mercado), uma gama desalinhada com as preferências do cliente pode significar custos excessivos e a geração de um enorme número de produtos descontinuados.

Há cinco fases principais na metodologia deste estudo: construção de *clusters* de lojas baseados em perfis de clientes, estimação da procura intrínseca de SKUs, determinação de taxas de substituição, a heurística de escolha de gama, e planeamento de espaço. No entanto, a metodologia de *clustering* não será coberta em detalhe ao longo deste trabalho, dado ser um procedimento já extensivamente estudado e não ser algo exclusivo a planeamento de gama ou de espaço.

A procura intrínseca é calculada usando um modelo MNL que utiliza preço, marca, e os dois atributos principais de cada categoria para estimar a cota de cada SKU candidato de cada categoria para a loja maior de cada *cluster*. De seguida, a aplicação de um modelo exógeno de substituição é feita possível através de uma estrutura desenvolvida que começa por definir a substituição entre níveis de atributos e constrói uma matriz de substituição entre produtos – fazendo assim possível a aplicação de um modelo exógeno de substituição a um problema de gestão de gama dinâmico. Depois, devido a restrições de gestão, uma heurística sub-ótima com base no próximo-melhor-SKU é construída para ordenar produtos de acordo com uma ponderação da sua contribuição marginal para vendas líquidas e diversidade extra trazida para a gama. Finalmente, com todos os produtos ordenados, um modelo matemático que maximiza a venda líquida total foi desenvolvido para possibilitar a otimização de alocação de espaço para categorias adjacentes, considerando restrições estratégicas e de equipamento.

Duas categorias foram escolhidas para apresentar resultados preliminares – TVs para resultados de gama, e tratamento de tecidos (ferros regulares, ferros com caldeira, tábuas/equipamentos de engomar, máquinas de coser) para resultados de espaço. Para as TVs, os resultados esperados são de um aumento de 4,2% na venda líquida total das 39 lojas de um cluster analisado. Em tratamento de tecidos, reorganizando o espaço alocado a cada tipo de produto na loja, um aumento de 8% em vendas totais (destes quatro tipos de produtos) é esperada. Resultados reais não estão disponíveis à data pois os testes piloto ainda não começaram. No entanto, os resultados estão alinhados com trabalhos prévios realizados na área e são plausíveis quando considerando os erros associados a cada parte da metodologia.

Abstract

Assortment and space planning are increasingly important for retailers as supply chains get leaner, competition fiercer and profit margins thinner. Deciding which products are on the shelf directly shows the capability of retailers to adapt to individual preferences of consumers and, besides the impact on the bottom line, an appropriate assortment can contribute to the image of the company and market share on the long term.

The current work consists on an assortment and space optimization at an electronics retailer. Due to the nature of electronics (slow stock rotation, short life cycles and dynamic introduction of products in the market), an assortment unaligned with customer preferences can cause excessive costs and generate an enormous number of discontinued products.

There are five main stages in the methodology of this work: customer profile-based store clustering, estimating intrinsic demand of SKUs, determining substitution rates, the assortment heuristic, and space planning. However, the clustering methodology is not thoroughly discussed in this study since it is a methodology already thoroughly studied and not exclusive to assortment and space planning.

Intrinsic demand is calculated using an MNL model that uses price, brand, and the two main attributes of each product category for share estimation of candidate SKUs of each category at the largest store of each cluster. Then, an exogenous substitution model is made possible by a newly developed framework that starts with attribute-based substitution matrixes and builds up to SKU-to-SKU substitution – making adaptation of an exogenous model to dynamic assortment planning possible. Afterwards, due to mostly manageability constraints, a suboptimal next-best-SKU-based heuristic is developed in order to rank SKUs according to their marginal contribution to sales and extra diversity brought into the assortment for each cluster. Finally, with all products ranked, a net sales maximizing mathematical model was developed to allocate store space to adjacent product categories, considering available store equipment and strategic constraints.

Two categories are chosen to present results – TVs for the assortment results and textile handling (regular irons, boiler irons, sewing machines and ironing boards) for space optimization results. For TVs, the expected results are an increase in 4,2% in total net sales across the 39 stores of one particular cluster. For textile handling, by reorganizing the space allocated to each category at one store, an increase of 8% in net sales across the four types of products is expected. Real results are not available to the date, as pilot testing has not yet started. However, results are in line with previous work in assortment planning and these results are plausible when considering the error associated with each part of the methodology.

Acknowledgements

First, I would like to thank LTPlabs for allowing me to take part in such a fascinating project. In particular, I am incredibly grateful to Professor Pedro Amorim for this irreplaceable opportunity which I will treasure forever.

I am also very thankful to Professors Teresa Bianchi de Aguiar and Manuel Pina Marques for the thorough review of this work. Your insightful input contributed tremendously to my growth and to the improvement of this study.

To Andreia Cardoso and Pedro Campelo: your help in this project was invaluable. Working with you made work an extraordinarily enjoyable experience. Thank you for all the time, knowledge, and dinners: it is going to be difficult to repay the huge debt I owe you two.

To all the friends I made during these six years, and especially to Paulo Pereira, Bruno Batista and João Martins: I hope you enjoyed spending your time with me as much as I enjoyed spending it with you. All the conversations and advice have allowed me to become one step closer to the person I want to be in the future.

To my family. Even if I do not express it often, I am very grateful that you are always present when I need you, even though I have been away myself for the past few years. No matter where I am, know that you are always in my heart.

Finally, Satona. I know these long university years were not a bed of roses - thank you for always trusting me and doing your best through all the hardships. I am truly glad that I am able to fulfill the promise I made to you 6 years ago, and I am looking forward to the future with you more than ever.

To everyone I met in Portugal, The Netherlands, Japan and Taiwan – know that you will always have a home wherever I am. Thank you.

Alexandre Aragão

Contents

Acronyms and symbols

B&M – Brick and mortar

A&S – Assortment and space

SKU – Stock keeping unit

OOS – Out of stock

OOA – Out of assortment

CA – Conjoint analysis

CM – Category manager

FA – Functional attributes

PA – Preference attributes

List of figures

List of tables

1. Introduction

1.1 Motivation

The rise in competition in the retail market witnessed in the recent years has made keeping profitability stable a challenge for retailers (Deloitte, 2017). Online retailing can no longer be seen as a separate line of business as customers now expect traditional brick and mortar (B&M) retailers to have an efficient omnichannel structure in place, thus turning what used to be local competition into a global market.

In order to offset competition-driven price reductions, retailers are tackling both the cost and revenue sides of the problem. Regarding costs, projects are constantly being carried out to make supply chains leaner and improve cost efficiency. Areas such as sourcing, purchasing, distribution and stock replenishment are improved on a regular basis and, for the most sophisticated players, additional enhancements only bring about marginal gains. As for revenues, several strategies are already put in place – the usage of loyalty cards is an example of how retailers try to retain regular shoppers, and dynamic pricing policies (regular and promotional) aligned with adequate marketing approaches are frequently used to capture new customers.

Even though these strategies succeed in getting the customer to come to the store, thus increasing the total number of visitants, they overlook a critical factor: the maximization of revenue per customer. Given the low margin retailers usually operate at, even a slight increase in revenue per customer has proven to be able to represent a rise in profitability of up to 50% (Kök and Fisher, 2007). Making sure that the right product is on the right shelf at the right time is increasingly important as physical and online competition are fierce and supply chains have reached a level of efficiency which makes customers less willing to wait for a product to be on the shelf of a particular store.

However, the operational constraints surrounding assortment management are very strict: even if a corporation is willing to incur in extra supplier costs by having a larger assortment and it is manageable from a supply chain perspective, assortment broadness will always be limited by available store/shelf space. Maximum space is a physically restrained variable and, as such, it is not subject to optimization; yet, the amount of space allocated to the various product categories can be chosen and it should be tightly linked to firms' growth strategies and the assortment width they intend to make available for the customer. Hence, space and assortment should not be dealt with separately and a symbiosis between both should be achieved in order to maximize profit for the company.

In nowadays fast paced markets, assortment and space planning are not a once-a-semester activity that category managers need to carry out – they need to be constantly analyzed and dynamic enough to allow for changes due to different seasons, customers and market fluctuations. An optimized assortment planning and space management lead to less depreciation costs, more revenue and greater customer satisfaction, consequently increasing net margin and boosting market share on the long term.

1.2 The project

This work was carried out as part of a consulting project at a national electronics retailer which goal was to create a tool to support assortment planning decisions and store space optimization across all stores. The urgency of this project was very high as the company was facing tremendous costs due to discontinued products at the stores (and some lost in the supply chain), and the complexity of managing each category was becoming too much for managers to handle manually and in an ad-hoc manner.

The commercial department of the firm was the main stakeholder of this project, but due to the degree of correlation between the commercial teams and all other activities, space, layout, merchandising and stock teams were also heavily involved. The project was divided into two main streamlines: assortment and space (A&S).

The development comprised three main stages: business requirement mapping, model development, and testing. Given the high level of involvement of the various parties, the first stage consisted of almost two months of bi-weekly workshops with three different teams: Assortment team #1, Assortment team #2, and a Space team. The main goals of these workshops were to first map the current A&S management processes, and then to devise a development strategy of the new tool – all relevant inputs, business rules and desired outputs/data visualization requirements were thoroughly discussed. The second stage consisted of the development of the proposed decision support system according to all needs of the company, which methodology will be presented further in this study. Finally, a pilot test on some product categories was done to measure the effectiveness and accuracy of the models.

Even though outside the scope of this work, it is worthy to mention that a full-scale implementation of the developed tool is expected at the over 150 stores of the firm in the near future for the most relevant product categories. The project's timeline is shown on Figure 1.

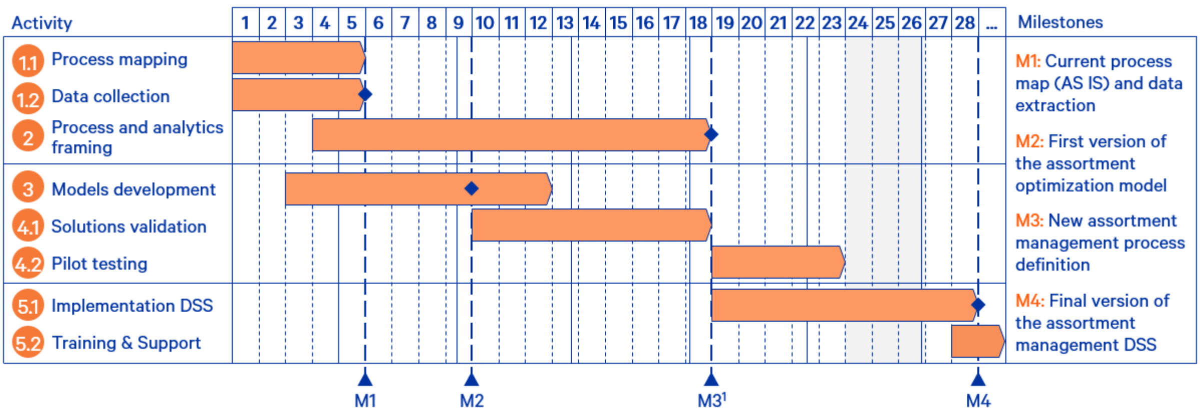

Figure 1 - The roadmap of the project (measured in weeks)

1.3 Dissertation structure

This work is organized in the following way. In Chapter 2, existing literature that is relevant for the problem will be carefully studied, and existing gaps which the current study intends to fill will be made clear. Chapter 3 intends to define the problems faced by the company and the opportunities of improvement identified in the project. Chapter 4 states the main assumptions made in this project and the methodology used to approach it. In Chapter 5, available results to the date are presented and discussed in light of all business constraints applied. In Chapter 6, conclusions of this work are presented and possible refinements and proposals of future work are also debated.

2. Literature review

This chapter intends to introduce the state of the art regarding assortment planning and store space optimization, as well as its advantages and challenges from both managerial and operational standpoints. In section 2.1, academic approaches and common practices concerning this problem will be broadly reviewed to familiarize the reader with currently used methodologies. A deeper dive into consumer choice modelling and each model's benefits and shortcomings will be taken in section 2.2. Finally, in section 2.3, practical applications of all discussed techniques will be presented along with its conclusions and repercussions in practice.

2.1 Assortment and store space management

A product assortment is defined by the set of products a retailer decides to have on the shelf of each store at different points in time, and assortment planning is done to maximize sales or gross profit (Kök et al., 2009). The choice of which products to display is not trivial and it cannot be simply based on the popularity or sales shares of products: the decision-maker should base shelf space decisions on substitution effects between stock keeping units (SKUs) and consumer choice behavior (Hübner et al., 2011). In a very significant paper, Gorsten and Gruen (2003) have estimated that when consumers are faced with an out-of-stock (OOS) or out-of-assortment (OOA) situation, only 45% will decide to substitute their original choice by another SKU on the shelf – meaning that an unfitting assortment might cause up to 55% of lost potential sales. Moreover, as omnichannel retail popularity keeps on growing, customers' willingness to wait is expected to get shorter.

As such, there is a need for retailers to understand consumers' preferences and adapt their product offerings. Up until recently, conjoint analysis (CA) was a widely used technique for multi-attribute utility modeling. It consists on carrying out a survey in which several product choice sets are shown to consumers, who choose their favorite option from each set. Their preference levels can then be estimated for the different attribute levels (Green et al., 2001). As CA is of easy interpretation and based on logical concepts, it was widely used by marketers in the past to fundament management decisions regarding assortment planning. However, retailers can no longer consider assortment decisions to be static – the proliferation of various products in the market in the past few years has forced retailers to become able to dynamically change assortments to adapt to consumers' needs (Rooderkerk et al., 2013). Furthermore, consumer preferences are constantly changing and continuously performing CA is neither sustainable nor an efficient strategy.

In order to cope with these market adversities, numerous new methodologies have been developed, leveraging on the existing sales/stock data to support assortment management decisions. The two main streams of research are on Multinomial Logit (MNL) models and exogenous demand (ED) models (Yücel, 2008). MNL consists of a discrete choice model with multi-attribute utility estimation that has as base the premise that consumers always choose a product that maximizes the total utility they get from their purchase. Exogenous models are probabilistic substitution models that assume an initial intrinsic demand for each SKU and then estimate substitution probabilities based on sales and stock-out data (Hübner et al., 2011). The specificities of each model will be discussed thoroughly in the following section.

Concerning space management, it is necessary to make the distinction between shelf space planning and store layout planning. Shelf space planning problems are related to the amount of facings each SKU has in a given shelf (Kök et al., 2007). Store layout planning refers to the allocation of floor space to each category of products such that total net profit/sales is maximized (Fisher, 2010). Considering the application of this work at an electronics retailer, with almost no leeway for SKU facing optimization (as each SKU has only one due to low stock rotation and broad assortments), store layout planning will be the main focus of this study.

Product category space planning is also heavily dependent on substitution patterns within each category. When space managers are making space allocation decisions, they often favor productive categories – however, not considering product substitution rates, frequently leads to not optimal decisions (Fisher, 2010). Observing Figures 1 and 2, even though category 2 is more productive than category 1 with less SKUs, due to differences in substitution patterns, for total net sales optimization category 1 should have more floor space, as the marginal improvement in net sales is larger after 10 SKUs are allocated. Usually, in the cases seen in literature, as shelf space is a constraint (i.e. there is no flexibility for layout optimization), this is still not a major research area.

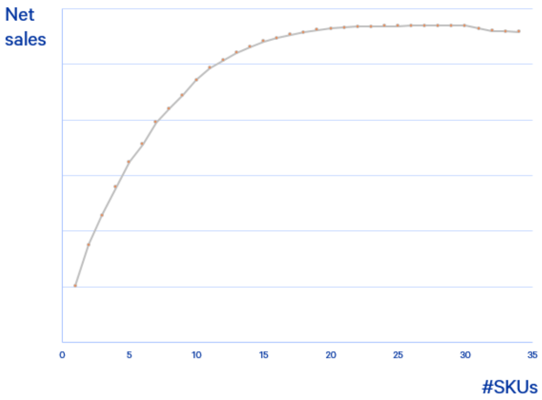

Figure 2 - Category space elasticity for Category 1

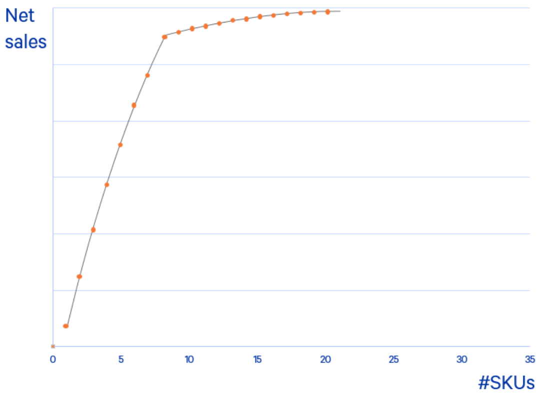

Figure 3 - Category space elasticity for Category 2

2.2 Consumer choice modelling

As forecasting intrinsic demand of SKUs (i.e. demand that is free of any substitution effects) and substitution rates between various products is a requirement for the project, it becomes necessary to analyze the different options for choice modelling reviewed in literature. In recent publications, the most popular approaches are the Multinominal Logit Models and the Exogenous Substitution Models (Hübner et al., 2011).

Multinomial logit models are based on the premise that the utility that consumers give a particular product k (U_k) can be separated in a deterministic component (v_k) and a random component ($\mathcal{E}_k$), in which the random component follows a Gumbel distribution. The deterministic component v_k can also be expressed as the sum of the individual utilities of each attribute level (Equation 2.1).

$$U_k = v_k + \mathcal{E}_k \tag{2.1}$$

These models focus mainly on the attributes of each product and how these influence consumers' choice at the moment of purchase (Guadagni et al., 1983).

The probability P of an individual choosing each one of the alternatives k can be written as equation (2.2).

$$P_k = \frac{e^{v_k}}{\sum_j e^{v_j}}\ , where\ j\ are\ all\ products\ available \tag{2.2}$$

To be able to compute the utilities of each attribute level, a Maximum Likelihood Estimation (MLE) approach is applied over historical sales data or consumer surveys done by the company.

One of the biggest advantages of these models is their conceptual simplicity: consumers are rational and will always try to maximize the utility they get from any given purchase. However, these models have two major drawbacks that need to be considered.

The first one is the underlying assumption that the utilities of attributes considered for the analysis are linear and independent from each other. While this might hold true for some types of products (e.g. the flavor of a yogurt is not dependent on its price), it is a very bold assumption in others (e.g. the size of a yogurt is highly correlated with price). In order to overcome this issue, several alternatives have been tested over time. Bentz and Merunka (2000) integrated artificial neural networks with this kind of model to try to understand the primary relations between attributes. Even though results were not disappointing, the complexity and instability introduced in the model were proven to be extremely relevant, making it a not robust approach for the problem. On the other hand, a widely used solution for this problem is the application of nested MNL models. A "nest" is considered to be a group of alternatives that can be substitutes of each other, and substitution outside the nest does not happen (Resmevichientong et al., 2014). Though this sounds like an appealing alternative, the fact that it requires previous knowledge about the market and the alternatives being considered makes it difficult to apply to very dynamic product categories. Moreover, substitution inside each nest is still affected by attribute correlations, which means that nested models can indeed mitigate the aforementioned problem, but not completely eliminate it.

The second issue with MNL models is even more relevant. Taking a look at equation (2.2), one can understand that regardless the number of new alternatives introduced in the market (or old alternatives removed), the ratio of choice probability between two alternatives *m* and *n* will remain constant. This is called the independence of irrelevant alternatives (IIA) property (Hübner et al., 2011). One clear example that illustrates this property is the Red Bus/Blue Bus paradox, introduced by McFadden (1974). Suppose that the probability of a given consumer of going to work by car is ½ and the probability of taking a red bus is ½ as well (given that these are the only alternatives). Now assume that a blue bus is introduced in the market similar to the red bus). An MNL model will predict the probabilities of taking each means of transportation to be 1/3. It makes sense that people taking the red bus would divide between both buses and people taking cars would be unaffected, but according to an MNL model, initial probability ratios remain constant. This means that, if a bus company introduced buses of numerous colors, it could theoretically eliminate the use of cars (Cheng et al., 2007).

IIA is a property that cannot be ignored when applying an MNL model, especially when the difference in SKUs in one category is so relevant such that substitution patterns are not the same across the category. However, due to the capability of estimating market shares of combinations of attributes before they are released in the market, MNL is still a broadly used methodology and it is integrated in a plethora of marketing software currently available in the market.

Exogenous substitution models directly estimate demand for each different SKU, thus not using an attribute-based approach. The formulation is simple: consumers choose product *i* from product set $N = \{1,2,3,\ldots,P\}$ and if *i* is not available, they choose product *j* with a probability of μ_{ij} (Hübner et al., 2011). Therefore, the core of exogenous substitution models will be the substitution matrix formed by all μ_{ij} ($i,j \in N$). These models allow for both OOA and OOS substitution and are usually only applied with one round of substitution – it is assumed that the customer leaves after not finding his/her second option. Kök (2009) has proven that this assumption is not too restraining.

The capability of distinguishing substitution patterns between SKUs in one product category is a definite advantage of exogenous models over MNL models. However, there are three core issues regarding exogenous models that are yet to be addressed in literature.

The first one is related to the estimation of all μ parameters in the substitution matrix. In order to be able to accurately estimate the existing N^2 substitution parameters, raw sales data are very often not enough – for instance, if two products *i* and *j* are present at every store, it is not possible to estimate the respective substitution rates. Being this the core of exogenous models, it is a central issue that has to be addressed before anything else. Kök (2007) deals with this matter by assuming substitution to be proportional to the market share of each product on the shelf (i.e. the most popular products get more substitution demand) at a dutch food retailer, and compares the results with a uniform substitution model (i.e. all products absorb the same substitution demand): results were slightly better for most categories. While this solution might be robust for food retailing, it might not be true for electronics as different consumers usually have very different preferences and the average popularity cannot be used as a proxy for the particularities of taste of individual customers.

The second downside is the inability of dealing with new products in assortments. In nowadays dynamic markets, SKUs are introduced and removed from shelves on a daily basis, and a static substitution matrix cannot deal with new SKUs if there is not either a) previous information on its substitution patterns, or b) sales data from other stores. This is one of the reasons why most papers take a static approach to assortment planning when using exogenous models (Kök et al., 2009).

The third existing problem regards the lack of sales forecasting capabilities of exogenous models. Whereas MNL models can predict market shares for each product depending on what other products are on the shelves, exogenous models only predict substitution patterns between SKUs, and do not address the issue of calculating their intrinsic demand. Hence, to successfully use a customized exogenous substitution model, there is a prior need of estimating intrinsic demand for each SKU in the assortment. Smith and Agrawal (2000) address this issue by assuming a distribution for the demand of each SKU and estimate its parameters based on historical stock-out and sales data, and then defining the optimal assortment and stock levels in the same model.

Exogenous substitution models deal with the main flaw of MNL models (IIA property) but bring about other core issues which are very problematic to deal with if the assortment problem is dynamic. These shortcomings are harder to solve than those of the MNL models, which is why exogenous models are comparatively less used than MNL models in recent literature.

2.3 Practical applications of assortment planning methodologies

Kök and Fisher (2007) have produced one of the most influential works in assortment planning. By using historical sales and stock data from a dutch food retailer, they have used and exogenous substitution model as a base to optimize the assortment per store, as well as the number of facings for each SKU. First, it is assumed that the substitution effects in the store with the largest assortment are not relevant. Second, it is supposed that substitution rates are proportional to the store popularity of each product; then, using OOS and OOA data, they compute δ, the probability of substitution happening at all (1- δ being the probability of the customer leaving without attempting substitution). OOS, OOA and footfall data are also used to estimate a demand function per customer for each SKU. Finally, they use an iterative greedy heuristic to estimate the optimal assortment and SKU facings per store, and their results suggest it is possible to have uplift of 50% in net profit.

Fisher and Vaidyanathan (2011) showed an innovative approach using an exogenous model. By working together with retailers in the tire industry, they developed a model that took into account factors such as the impossibility of substitution between tires of different sizes. This introduced a new dimension to assortment planning problems as it not only considered choice sets of SKUs, but it also specified substitution rates at an attribute-level. The result was a growth in revenue between 3.6% to 5.8%. This paper was of very easy interpretation from both an academic and managerial perspective, and the reason why there were no researchers building on top of it is thought to be the lack of capability to scale this work to numerous categories – a thorough analysis of each category is necessary and it was considered to be too time consuming.

Smith and Agrawal (2000) developed an exogenous model for optimizing assortment and inventory levels with shelf space constraints. Again, the demand function is estimated for each consumer, and the probability of a product being on shelf is predicted using a Markov chain model. By assuming that customers arrive at the store according to a negative binomial distribution, they formulate the profit function as a Newsboy model and determine the optimal stock quantities and list of products. They arrive at very interesting results: for instance, for some stores, listing the most popular product was sometimes not part of the optimal solution, which can be motivated by it being a low-priced SKU that was absorbing demand of more expensive alternatives.

Rusmevichientong et al. (2014) have extended the traditional MNL model to include the diversity between consumers – instead of assuming a mean utility vector (thus assuming every customer is probabilistically identical), they modeled mean utilities as random variables. This particular variant of the MNL, the mixture logit, allows them to go around some of the issues with the original MNL models. However, due to the assumption that mean utilities are stochastic, they conclude the problem no longer has an optimal assortment solution and it is computationally difficult to obtain a robust result.

Sinha et al. (2013) have used an MNL model using aggregate data on 3000 SKUs (wine) across several customers for two years at a weekly level to summarize the importance of each level of each attribute and then built an attribute map to assess the uniqueness of each SKU, thus estimating the nontransferable demand of each. This allowed them to understand the impact of product positioning in an assortment and to build sets of similar products (much like the nests in nested MNL models). They used the results to decide which kind of wine to produce and obtained fascinating results: even though market share decreased in 14%, revenue increased by 0.7%, thus showing that quantity sales are not always tightly tied to revenue growth.

Miller et al. (2010) also base their work on an MNL model but introduce some changes. First, they construct on fuzzy set theory and implement fuzzy utility thresholds to build consideration sets. Second, they use a factor to account for each retailer's market share in the market in choice modelling (thus making consumers more willing to substitute when the retailer has more market share) as it is argued that consumers are not knowledgeable to the point of knowing the assortment of every other competitor and will be tempted to assume that a big player will have a more complete product offering. However, their model did not take into account promotional activity nor stock-out data, which caused some of the final results to be not so satisfactory.

Finally, Mahajan and Van Ryzin (2000) have used MNL to estimate individual utilities and then build choice sets for consumers and built a heuristic to optimize on-hand stock based on exogenous substitution existing for each choice set. Their results show that retailers should stock more of the most popular SKUs as they absorb more substitution demand and it is cheaper to have more on-hand stock of an already existing product than acquiring a new substitute variant as backup.

3. The problem

This chapter intends to describe the problems the retailer was facing that made them undertake an assortment and space planning project. It is divided into three sections: a description of the main stakeholders at the company, the main processes related to assortment and space planning, and problems/opportunities for improvement.

3.1 Stakeholders

There are two departments involved in the assortment and space planning processes: the commercial department and the space department. The main stakeholders of the commercial department are the category managers (CM) and as for the space department, there are three teams involved: the micro-space team, the assortment team, and the merchandising team, as is shown in Figure 4.

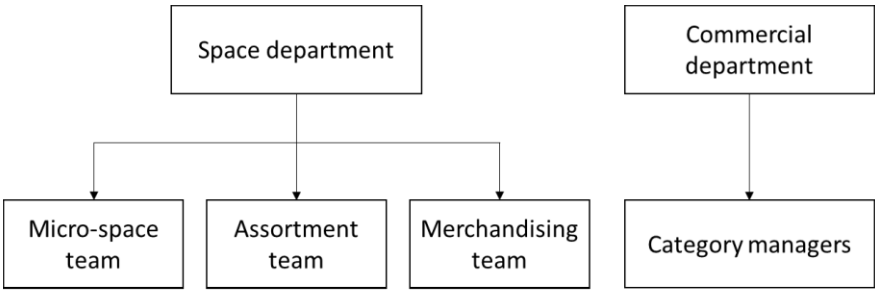

Figure 4 - Main stakeholders of the project

The micro-space team is responsible for space allocation decisions of every product category at the stores. The only constraint they have is a maximum space for each business unit; however, they also follow the guidelines of the CMs and the assortment team regarding the ideal space for each category at each store.

The assortment team is in charge of evaluating store KPIs for every category and making proposals regarding the ideal number of SKUs from each category that should be in each store. They also decide the percentage of discontinued SKUs to be displayed in each store and discuss with the CMs how stores should be clustered to maximize their goals. Their role is to make a link between CMs and the space team to make sure processes advance in a smooth fashion.

The merchandising teams are in control of writing the merchandising manuals that include information about how each SKU should be displayed at the stores – i.e., the number of SKUs per piece of equipment, category adjacencies, etc.. They are also responsible for building the planograms for the product categories that require them to do so.

Commercial managers are the only stakeholders regarding product selection. They first decide which SKUs they want to have stock of, and then which products to display at each store. They are also responsible for deciding when and how promotions should happen and, in general, have the final say in every decision regarding their category. For instance, even though the assortment team has the duty of analyzing stores and proposing ideal numbers of SKUs, CMs can challenge this decision and eventually changing it if the assortment team agrees with his/her reasoning.

3.2 Main processes

There were four main processes analyzed during the project: full assortment revisions, fast line up revisions, store layout revisions and dynamic layout adjustments. The first two are strictly related to assortment per se; the latter two regard mostly store space.

3.2.1 Full assortment revisions

A full assortment revision is a process triggered once a year for most categories, and it consists on a full review of every aspect related to assortments: the redefinition of ideal store spaces, the percentage of discontinued products, facings replications, suppliers and, of course, SKUs themselves. This process is usually aligned with suppliers' line up updates so that the retailer can change its line up accordingly to meet the market's latest trends. A line up is defined as the active SKUs each stakeholder has at each moment in time.

The process starts with a complete store/category review by the assortment teams in order to evaluate the performance of the category in the previous years. KPIs such as gross margin returns on footage/inventory. Another very important KPI to assess whether the number of SKUs is suited for a particular store is assortment productivity – the ratio between the number of SKUs with at least one sale and the number of SKUs displayed per month.

Next, the assortment team develops a proposal for the percentage of discontinued products on shelf for the next year, that depends on each category's capability of selling older products and the category's role in the company – the rule of thumb is that important category's that contribute to the company's image and to a significant portion of net sales will have a smaller percentage of discontinued products.

Third, in light of all previous analysis, the assortment team defines which should be the space for new SKUs that, ideally, should be available at each store (hereafter referred to as *objective space*). At this stage, there might be inputs from the CMs regarding category-specific market growth.

After all these considerations, the assortment team will group stores with similar objective spaces (without taking into consideration any other difference between stores) and form an *assortment cluster* – stores with similar objective spaces that are supposed to have the same active physical assortment. At the same time, the CM and the merchandising teams will review merchandising criteria (and, if existent, planograms) and change it according to the category's needs and annual goals.

Finally, there is a pre-validation meeting involving CMs and the assortment and merchandising teams to present individual results and validate them with all participants. The CM has the final say – if he/she does not agree with, for instance, an objective space reduction, it is possible to challenge this decision and change it. The CM also has the possibility of subdividing assortment clusters – in case there is a clear difference regarding the type of customer of stores within the same assortment cluster, the CM can choose to divide it into two to be able to make better product allocation decisions – however, this is motivated on just an ad-hoc analysis based on the CM's empirical knowledge.

When assortment clusters are decided, the CM proceeds to choosing the SKUs for each cluster. This is a manual process based on the CM's knowledge of the category and the company's strategic guidelines – several business rules concerning percentage of price quartiles of SKUs displayed, the participation of each brand in the assortment and attribute diversity are applied. For the SKU selection itself, usually CMs start by choosing the physical assortment of the largest assortment cluster and gradually remove SKUs to form the assortments of the smaller clusters, always having the imposed business rules in mind. This means that assortments of larger and smaller clusters are typically concentric.

3.2.2 Fast line up revisions

A fast line up revision is a process that can be triggered by three main reasons: a product launched by a supplier outside of an assortment revision period, an SKU performing extremely bad in a physical assortment, or an SKU performing exceptionally well in the online channel. Unlike a full assortment review, store space changes are not allowed – for an SKU to enter a physical assortment, the CM must swap it for another one. Moreover, because stores are grouped in assortment clusters, it is not possible to allocate a product to just one store (granted the store is not alone in its assortment cluster) – the CM has to allocate the new SKU to every store in the cluster.

Furthermore, it is not possible to change the cluster of a store in this process (i.e., the CM cannot expand or reduce its category in a store, as category space remains constant).

3.2.3 Store layout revisions

As with every retailer, store layout needs to be reviewed periodically. In this case in particular, stores are reviewed at least once every 18 months, and these layout revisions are distributed throughout the year due to manageability constraints.

The process is started by the micro-space team, which decides the minimum and maximum percentages of space each business unit should have, according to previous performance and strategic guidelines. These indications are passed on to the macro-space team (no contact during this project), which creates a blueprint of the store, already including the proposed/available equipment for each business unit.

The blueprint is returned to the micro space team, which has to allocate product categories to shelves. The first step is to divide the available business unit space into space for each *ecosystem* – a group of product categories which use cases are related (e.g. toasters and milkshake machines are in the “breakfast” ecosystem). After the space for each ecosystem is set, the micro space manager must decide the space for each category. If the space is enough such that all ideal spaces (set during the full assortment review) can be set, no extra work is needed; however, if the available store space is not enough to allocate the ideal space for every category, there is a need to analyze the involved categories in order to decide which will get less space than requested. The micro space manager will gather the involved CMs and the assortment team to go over the most relevant KPIs and try to find an agreement between CMs – if such is not possible, he/she will make the decision based on previous data and market growth perspectives. No decision support tool is currently in place to help micro-space managers take this decision.

After all these space trade-offs are dealt with and there is a pre-final version of the store layout, it will be sent for validations of business unit managers and, if approved, will be forwarded to the budgeting team to estimate costs regarding new equipment and other relevant expenses.

3.2.4 Dynamic layout adjustments

As sometimes the store will stock out on existing discontinued products and gain extra space for the affected categories, there is a need to slightly adjust space so that empty shelves are not shown to the final customer. There are two main solutions for this issue. The first one is buying new equipment/switching equipment with other product category, knowing that the budget for this kind of operation is very low and usually the cost burden is the store's responsibility. The second option is to choose to display more discontinued products or to do more facing replications for the same product category; yet, stores need to be careful not to lose product variety nor to display too many discontinued products, as it might hurt the company's image in the market.

3.3 Problems and opportunities for improvement

During the time spent mapping current processes with the teams, eight main opportunities were identified as potential candidates to be dealt with during the project.

The current store clustering methodology (1) is the first issue to address. The current processes only consider space constraints when deciding on assortment clusters, which are not representative of the type of customer that purchases at each store; in other words, store space is, at most, representative of the amount of traffic the store can generate, and not the kind of traffic. For instance, consumers in some regions of the country do not use gas heating, whereas people in other regions do. With the current methodology, if stores in these two very different areas have the same space for heating equipment, they will end up with the same assortment which will not be appropriate for a large part of the customers. Average purchasing power of the region the store is located in is also expected to be a relevant factor when deciding if a store should have a high or low-end assortment type. It is believed that including these variables in the clustering process will allow the retailer to improve its consumer targeting strategies, thus increasing productivity and net sales (and reduce the amount of obsolete stock, as SKUs are adapted to the kind of demand of each store).

The lack of a robust methodology for estimating SKU demand (2) is also considered to be an issue of utmost importance. Even though a forecasting software is implemented, it only takes into consideration each SKUs sales history, not relating a single product's performance with the rest of the assortment that is on the shelf – this is a major shortcoming given that assortment changes are rather common and the demand of individual products is affected by the assortment' configuration of the store. On top of this, the current software is not able to predict demand for completely new products in the market. For some categories, new products comprise 90% of the assortment after full assortment reviews, which means that CM are choosing SKUs for their assortment without any analytical support regarding sales forecasting or how adding each SKU to the assortment will impact the demand of the other products.

This leads to the third topic identified as critical: the absence of a suggestion tool for SKU selection (3). CMs currently choose SKUs based on their tacit knowledge of the market, purchased market data, and unstructured market information received from suppliers. Furthermore, even though the CM tries to ensure product variety in the assortment by considering products' dissimilarity, no methodology for evaluating cannibalization and substitution effects between products is in place. Hence, the current process does not allow for an estimation of percentage of satisfied customers, or for an assortment planning that maximizes net profit/revenue or other goals considered by the manager.

On the space optimization side, severe limitations were found concerning space trade-offs between product categories (4) during the store layout revision process. The micro-space managers take into account past performance of the categories in the stores and albeit it serves as a proxy to how the category will perform in the future, it is not enough to assess how the category will behave with more/less SKUs on the shelf: in order to do this, the manager needs to know the space elasticity curve of the category. Some categories might be high performers with a small number of SKUs on the shelf (i.e. with extremely high substitution rates), but adding extra products to the shelf will not improve total revenue significantly; on the other hand, others might be performing worse with less SKUs, but a space increase might contribute significantly for the category's total revenue. Hence, a performance review of categories with their present store space is not enough to decide which should get extra store space. Also, costs of adding an extra SKU to the shelf are not currently being considered: this is especially relevant if categories with high display, stocking and logistic costs (e.g. refrigerators) are being compared with lower cost categories (e.g. vacuum cleaners).

Additionally, as processes were mapped, the need of using data from several sources became apparent: external market studies, internal sales data, supplier information are all heavily relied on for any decision regarding assortment or space. As such, treatment and integration of data from several data sources (5) is also considered to be a critical issue in order to make A&S processes leaner and obtain reliable results, thus being able to have a consistent decision-making process. This integration has to contemplate not only several sources, but also different levels: product category data, competition data, consumer data, store-level data, suppliers' data and strategic internal data.

Furthermore, besides every CM using different data sources for category evaluation, they use different KPIs for assessing performance. The lack of standardization in this area is also a critical matter, as space managers need to standardize the KPIs of every category in order to compare them. After discussing it with the teams, all members agreed that a standard visualization tool of every KPI related to A&S was a necessity (6) and some of the essential functionalities were discussed: easy visualization of the impact when adding constraints to the assortment (e.g. suppliers' requests), end of lifecycle information, comparisons of different time periods and SKU underperformance warnings. As CMs manage more than one hundred stores at the same time, it is difficult to track every SKU and a tool like this could represent a significant uplift in efficiency.

There was also a major problem related to products' end of lifecycle which the project's team was unaware at first: since this is an electronics retailer, there is a non-negligible amount of discontinued products being generated at a very fast rate (mainly during full assortment revisions), and there is no regulated phase in/out of SKUs integrated with the stocks and promotional teams (7). The CMs do not communicate their intentions of removing products of the assortment to the other teams beforehand, which means that preventive measures cannot be taken to avoid having excessive stock of the respective SKUs. This is a problem that exists across categories and is seen as the issue with the highest cost reduction potential in the whole company.

Finally, space managers consider that the current top-down approach for space allocation (starting by defining the areas of each business unit) does not cope with the space needs of each product category (8). In some business units, product categories have more space than needed; in others, all categories have substantial growth potential but are limited by the business unit maximum size. The application of a bottom-up approach to space allocation was discussed in order to try to guarantee that the maximum number of categories have the ideal store space defined by the respective CM.

During the course of the project, all these problems were carefully analyzed and issues (1) to (6) were chosen to be addressed. Problems (7) and (8) were also thought to be extremely relevant but out of the scope of an assortment planning project – point (7) is strictly related to reverse logistics and (8) involves many other stakeholders which were not involved. Nevertheless, these difficulties were always taken into consideration when thinking of solutions for the other issues. In this work in particular, the methodology implemented to solve points 1 to 4 will be thoroughly studied, and its impact in the bottom line of the retailer measured.

4. Methodology

The following chapter describes the methodology used to optimize the product assortment of the aforementioned retailer. First, an overview of the methodology will be addressed in section 4.1. Subsequently, main assumptions, intrinsic demand forecasting, substitution rates, the assortment heuristic, and space planning will be explained in sections 4.2, 4.3, 4.4, 4.5, and 4.6, respectively.

4.1 Methodology overview

The first issue to address was the current store clustering methodology. Assortment decisions based on space-based clustering are not capable to deal with store-by-store specificities (e.g. stores in distinct parts of the country will have different needs regarding heating equipment). Hence, the first step in the adopted methodology was to cluster stores by customer type for each product category. The variables used were sales percentage per price quartiles, brands and the levels of the main attribute of the respective product category. Moreover, as new stores are opening at a very high pace, exogenous variables such as average income, purchasing power, and average age of the region were associated with each cluster and used to predict clusters of new stores. This clustering methodology is not part of this work, thus not being described in detail, but it will serve as a basis for the remaining procedures.

Secondly, there is a need to estimate intrinsic demand of all candidate SKUs for an assortment. Due to the difference between SKUs in one category, an exogenous demand model that could differentiate substitution rates between products was initially thought to be more appropriate; however, due to the extremely dynamic nature of the market, with products going on and off shelf constantly, it was necessary to use an MNL model that could predict market shares based on the attributes considered. The MNL model was used to predict shares at the largest store of each cluster in case every candidate SKU was displayed on the shelf (i.e. intrinsic demand), taking as an input historical sales data.

However, even though an MNL model was used to predict intrinsic demand, an exogenous substitution model was applied to decide which products should be on the shelf for each cluster. First, the methodology in Kök et al. (2007) to estimate the percentage of customers who consider substitution was employed (with some adaptations). Afterwards, an attribute-based substitution matrix was developed taking into account similarity between the levels of every attribute and, consequently, inter SKU similarity.

With both intrinsic demand and substitution rates estimated, the assortment decisions can be made accordingly. In this particular case, a greedy heuristic based on the next-best-SKU is applied for each cluster of stores – SKUs are therefore ranked according to which is the best SKU to enter the assortment in each position. The next best SKU is decided by considering both contribution to net sales and the extra diversity it brings to the already existing assortment.

Finally, with the ranking of SKUs already decided, an optimization model was built to define which should be the store space for each category. The model’s decision is based on space/equipment available and the marginal contribution to net sales of each product category disputing the same space.

4.2 Main assumptions

For the application of the developed methodology, there are some assumptions made related to customer behavior.

First, in order to be able to model product substitution, the influence of the physical assortment at the store on customer choice is assumed to be negligible: it is assumed that the customer already has a set idea of his/her preferred attributes (e.g. price quartile, brand) and that this is defined before visiting the store. Moreover, it is supposed that consumers always choose an SKU that exists – i.e., no one goes to the store expecting to buy a flagship product for half the price, or a brand that does not exist.

ASSUMPTION (A1). *The consumers' choice of their favorite product is not influenced by what is on the shelf: they have a preferred set of attributes, and that set is necessarily present either in the physical or online assortments.*

Additionally, at the time of visiting a store, if the customers do not see what they prefer on the shelf, they might choose to substitute to another product with a certain probability. However, if the second preference is not available either, it is implicit that the customer does not proceed with a third substitution. This assumption is in line with the work of Kök and Fisher (2007), and it was proven not to be too restraining.

ASSUMPTION (A2). *The consumers will try to substitute once with probability δ, but if the second alternative is not available, the sale is lost.*

It is also imperative to define what is understood by "lost sales". Since the goal of this project is to maximize in-store purchases by optimizing assortment, it is assumed that, for practical purposes, a customer switching to the competition is no different than purchasing their preferred product at another store of the chain or through the online channel.

ASSUMPTION (A3). *All sales that are not made through the physical channel of the store the consumer is at are considered to be lost sales – including the online channel of the retailer or other stores in the chain.*

Furthermore, for the successful implementation of the methodology, it is essential to understand intrinsic demand of consumers (i.e. demand free of any substitution effects). As such, due to the necessity of using a proxy to estimate this demand, it will be considered that the store with the largest assortment for each cluster of stores of each product category has no relevant substitution demand – this will allow for more accurate substitution modelling for the smaller stores. With a broader assortment, it is not odd that consumers find their first choices more often - this supposition is also used by Kök and Fisher (2007).

ASSUMPTION (A4). *At the stores with the largest assortment of each cluster, substitution demand is negligible.*

Finally, to be able to estimate substitution effects between different SKUs/attributes, and since intrinsic demand is presumed to be known at the largest store of each cluster, it is also assumed that the intrinsic demand of small stores follows the same pattern (in percentage) as the largest store of its cluster. As store clusters will be based on customer types with the new methodology, this is not considered to be a daring assumption.

ASSUMPTION (A5). *The percentages of intrinsic demand of all SKUs at smaller stores is the same as the ones at the largest store of its cluster.*

Assumption A5 allows for a percental comparison between the intrinsic demand of each store; however, it is necessary to estimate absolute values of demand for every store to be able to calculate potential sales in each case. As such, it is necessary to consider a factor to scale absolute demand of the store with the largest assortment and make it comparable with smaller stores. The information that makes it possible to make the best approximation is considered to be footfall data from every store. It is debatable whether this data really represents the difference between stores for each product category – footfall data is general and it does not discriminate between categories. Also, stores located at places with more consumer traffic (e.g. shopping centers) are expected to have more visitors with no intention of purchasing. Regardless of these shortcomings, footfall data will be used as a scale factor between stores.

ASSUMPTION (A6). *The difference in demand between different stores for all categories can be scaled using general footfall data from comparable periods in time.*

4.3 Determining intrinsic demand

One of the main key issues faced in this project is the fact that the retailer is dealing with slow-moving but highly dynamic product categories, meaning that there is a need to be able to forecast demand of non-existing products with very scarce sales/stock-out data. Another relevant problem is the lack of individual-level data: even though store sales are known, there is no data regarding which customer bought each SKU. Hence, it was decided to apply an MNL choice model to calculate part worth utilities of levels of attributes.

However, MNL models are originally used for SKUs choice sets presented to customers, which is data not available to the team. Therefore, it is necessary to devise another way to successfully implement an MNL choice model. Since the number of stores per cluster is rather large, the implemented methodology requires that the model considers each store as an individual, and the shares of sales of each SKU to be its choice distribution – meaning that the utilities obtained will correspond to the utilities of the average customer of the store.

For this model to work, it is necessary to determine the length of time periods to use for the model. As assortment changes regularly, it was advised by the retailer's project team to not use extremely long periods; on the other hand, given that the SKUs are slow movers, a too short period of time might not be representative of the average demand patterns at each store. Hence, it was decided to use months as the base time period, and the model to be used for 12 months (to account for seasonality differences).

Another problem is the lack of information of on-shelf SKUs. Even though sales data is available, it is not representative of the SKUs that are in fact displayed at the store (since assortment productivity is usually low). The team had to cross sales data with stock data to estimate which SKUs were on shelf for each month: the assumption is that an SKU with at least 1 unit of stock during at least one week of the month (stock information is only available on a per week basis), is considered to have been on shelf for the whole month. Moreover, sales of SKUs that had no stock at the store are considered to be online sales and are not included in the choice model.

ASSUMPTION (A7). *SKUs with at least one unit of stock for at least one week in the month are considered to have been displayed on the shelf of that store during the whole month.*

The number of attributes, as well as the attribute levels considered in the analysis, is left to the CM to decide: since the industry is very dynamic, the most relevant attributes of a category of SKUs can change on a regular basis and it is considered necessary to give this kind of flexibility to the CMs in order for the decision support system to be used. However, having an excessive number of attributes in the analysis can make the model too sensitive and produce unreliable results. An example of the input data for the MNL choice model is presented below in Table 1.

Table 1 - Example of MNL input data

Store	Month	SKU	Attribute M	Attribute N	Sales share
1	Jan	A	M1	N2	30%
1	Jan	B	M3	N1	20%
1	Jan	C	M1	N1	50%
1	Feb	A	M1	N2	35%
1	Feb	B	M3	N1	25%
1	Feb	D	M2	N1	20%
1	Feb	E	M2	N2	20%

After the utility levels of each level are obtained, it is possible to predict the market shares at the largest store of each cluster of each candidate SKU by referring to Equation 2.2.

In order to obtain the absolute value of demand, it is needed to refer to assumption A4: since substitution demand is assumed to be negligible at the largest store, one can say that the number of units sold in the previous period matches the total number of customers that entered the store with intentions or purchasing (as every consumer finds the SKU they were looking for). Hence, the number of units sold at the store in the previous time periods can be considered to be the total market size for that store in the period analyzed. By getting an input from the CM about prospects of market growth, one can assess the number of customers that will try to purchase from each category at the store, and multiply that value by the market shares estimated by the MNL model, thus obtaining the intrinsic demand for all candidate SKUs at the largest store of each cluster.

Finally, it is also relevant to discuss how price is treated in the model. Even though price can be treated as a continuous variable in MNL models, due to big fluctuations in price, a continuous model was thought to be not robust enough. As such, price is divided in deciles and for each month, according to the average sales price of each SKU, they are placed in a decile (i.e. the same SKU can be in different deciles in distinct months). This way, some effects of promotional activity can also be captured in the model.

4.4 Estimating substitution rates

As mentioned in previous chapters, an MNL model cannot cope with different substitution rates according to similarity between SKUs; as such, it is necessary to consider an exogenous substitution model. Moreover, in order to be able to deal with the introduction of new products, the substitution matrix should have as basis the similarity between attributes and then build up to a full SKU-to-SKU matrix.

Before proceeding with this approach, there is a need of understanding the nature of distinct types of attributes and how consumers consider substitution for each type. As no literature using this kind of approach was found, it is necessary to develop a new framework regarding attribute classification.

First, there are attributes that are directly related to the usage of the product: the power of a toaster or the size of a TV are good examples. When customers visit a shop with a particular level of this kind of attributes in mind, it is not a matter of taste, but necessity – different consumers have different needs and these are in no way influenced by market preferences or trends. During the course of this work, these attributes will be called **functional attributes** (FA). FAs are considered to be orderable according to their functionality; however, this does not mean that they necessarily need to be numeric. For instance, the resolution of TVs is a categorical FA (e.g. HD Ready, Full HD, 4K). Furthermore, it is assumed that the substitution behavior between FAs is far more rigid than that of other attributes and consumers are only willing to substitute to adjacent levels of their original choice (i.e., a customer that intends to buy a 50’ TV will only consider switching to 45’ or 55’).

ASSUMPTION (A8). *For functional attributes, substitution only occurs between the consumer’s original choice and its adjacent levels.*

Besides functional attributes, there are other characteristics that do not influence the usage of the product: color or brand are examples of these. In this study, these will be referred to as **preference attributes** (PA). Unlike FAs, PA choice is highly influenced by the consumers’ peers and market trends – in general a customer will prefer more popular brands or the trendy colors. As such, substitution is assumed to be proportional to the market penetration of each level of the attribute. In other words, consumers will be more willing to substitute to the more popular alternatives.

ASSUMPTION (A9). *For preference attributes, substitution is proportional to the popularity of each level of the considered attribute.*

Having both types of attributes defined, it is possible to proceed to the estimation of substitution parameters for each attribute. For PAs, these parameters do not require estimation: they are the percentage of net sales that each level of the attribute generated in the previous time period. The example below (Equation 4.1) represents a substitution matrix between three brands (A, B and C) had they generated 70%, 20%, and 10% of sales, respectively.

$$\begin{array}{c} \\ A \\ B \\ C \end{array} \begin{array}{c} \begin{array}{ccc} A & B & C \end{array} \\ \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.7 & 0.2 & 0.1 \\ 0.7 & 0.2 & 0.1 \end{pmatrix} \end{array} \tag{4.1}$$

On the other hand, it is not trivial to estimate substitution effects between FAs. Based on the aforementioned definition of functional attributes, three substitution cases can be considered. First, when substitution occurs to both adjacent levels; second, when substitution only occurs to the level below the original choice; third, when substitution only occurs to the level above the original choice. Different FAs will have different substitution patterns according to its nature: for instance, consumers might only be willing to substitute to prices below their original choice, but they might only consider TVs with a resolution equal or above the one they initially searched for.

Let’s start by defining the attribute substitution matrix for FAs. Let ρ be the percentage of customers who choose to substitute to another SKU with the same level of attribute A which was their original choice. For a matter of practicality, let’s suppose that attribute A only has three levels, a_1, a_2 and a_3 ($a_1 < a_2 < a_3$). The three cases mentioned above can be represented by the substitution matrixes present in Equation 4.2 (both level substitution, forward substitution and backward substitution, respectively).

$$\begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \end{array} \begin{array}{c} \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{pmatrix} \frac{\rho+1}{2} & \frac{1-\rho}{2} & 0 \\ \frac{1-\rho}{2} & \rho & \frac{1-\rho}{2} \\ 0 & \frac{1-\rho}{2} & \frac{\rho+1}{2} \end{pmatrix} \end{array} \quad \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \end{array} \begin{array}{c} \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{pmatrix} \frac{\rho+1}{2} & \frac{1-\rho}{2} & 0 \\ 0 & \frac{\rho+1}{2} & \frac{1-\rho}{2} \\ 0 & 0 & 1 \end{pmatrix} \end{array} \quad \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \end{array} \begin{array}{c} \begin{array}{ccc} a_1 & a_2 & a_3 \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ \frac{1-\rho}{2} & \frac{\rho+1}{2} & 0 \\ 0 & \frac{1-\rho}{2} & \frac{\rho+1}{2} \end{pmatrix} \end{array} \tag{4.2}$$

Note that for the lowest and highest levels, due to having only one adjacency, the percentage of substitution related to the missing adjacency is absorbed by the level corresponding to the customer's original choice.

The next step consists in building the SKU-to-SKU substitution matrix. Though many approaches can be taken, in this study it is defined that the SKU-to-SKU matrix can be built by multiplying the values of the substitution probabilities of every attribute between the SKUs in each cell/column, and then normalizing the lines of the matrix to sum up to 1. As such, the SKU-to-SKU matrix can be written as in equation 4.3.

M_{sub}	substitution matrix
N	number of SKUs
ρ_k	ρ for each functional attribute k
$f_{i,j}$	probability of substitution of SKU i by SKU j
δ	probability of customer choosing to substitute at all

$$\mathbf{M_{sub}} = \delta \times \begin{array}{c} \\ 1 \\ \dots \\ N \end{array} \begin{array}{c} \begin{array}{ccc} 1 & \dots & N \end{array} \\ \begin{pmatrix} f_{1,1}(\rho_1, \dots, \rho_n) & \dots & f_{1,N}(\rho_1, \dots, \rho_n) \\ \dots & \dots & \dots \\ f_{N,1}(\rho_1, \dots, \rho_n) & \dots & f_{N,N}(\rho_1, \dots, \rho_n) \end{pmatrix} \end{array} \quad (4.3)$$

Having defined the SKU-to-SKU substitution matrix (for the three cases represented above), it is now possible to write the expression that defines the forecast $FD_{j,k}$ of SKU j for a particular store k, which is the sum of two components: its own intrinsic demand and the substitution demand gathered from missing SKUs (scaled by footfall data), and it is represented by Equation 4.4.

$FD_{i,k}$	forecasted demand of SKU i at store k
ID_i	intrinsic demand of SKU i at the largest store of the cluster
ff_k	ratio of the footfall of store k and the largest store in its cluster
OA_k	set of SKUs out of the assortment of store k

$$FD_{i,k} = \left(ID_i + \sum_{j \in OA_k} ID_j * M_{sub_{j,i}}\right) * ff_k \quad (4.4)$$

Note that the only variables in the forecast expression are δ and $\rho_1 \dots \rho_n$, as the coefficient of the substitution matrix can be written in function of these variables. Hence, it is also possible to write the expression for the mean absolute percentage error (MAPE) of the forecast for previous periods of time for which real sales data exist (equation 4.5).

A_k	assortment at store k
$RD_{i,k}$	real demand of SKU i at store k
K	total number of stores

$$MAPE = \sum_{k=1}^{K} \frac{\sum_{j \in A_k} \frac{|FD_{j,k} - RD_{j,k}|}{RD_{j,k}}}{|A_k|} \tag{4.5}$$

Finally, it is possible to optimize the MAPE over the substitution criteria in order to obtain the optimal δ^* and $\rho_1{}^* \ldots \rho_n{}^*$ (equation 4.6).

$$(\delta^*, \rho_1{}^* \ldots \rho_n{}^*) = \underset{\substack{0 \leq \delta \leq 1 \\ 0{,}33 \leq \rho_1 \ldots \rho_n \leq 1}}{\arg min} (MAPE) \tag{4.6}$$

It is relevant to note that the values of ρ have a lower bound of 0.33 because a value lower than this would mean consumers were more likely to change their original choice than to keep it which does not seem to make practical sense. Also, the optimization described above was run for all combinations of both level substitution, backward substitution and forward substitution in different FAs; this means that this method allows for different FAs having different substitution patterns.

On a final note, it has been argued whether price (exceptionally considered a FA in the model) substitution patterns behave like the ones described. Indeed, customers would like to pay as least as possible for their favorite products. However, it is assumed that a steep drop in price will also cause a drop in quality (i.e. the levels of the other attributes) of the SKU, and the customer will not be willing to change to more than one adjacent decile.

4.5 Assortment heuristic

With both intrinsic demand and substitution rates of products estimated, it is possible to start choosing the assortment for the stores. The optimal global solution naturally consists on estimating the ideal assortment per store; however, being the goal of this work a practical application, there are three main reasons why this was not implemented.

The first one concerns the manageability constraints. Each CM has to manage one category at over 150 stores, and having to validate 150 different assortments was stated to be too much of a burden. The second motive is related to the generation of discontinued products when doing fast line up revisions. If stores have similar assortments, it is easy to control which products go off shelf and plan ahead to mitigate its impact; however, if stores have personalized assortments, the number of distinct SKUs being discontinued at all stores might be too high to handle. Finally, the extra effort required from the stocks and logistics teams is not negligible if most categories decide to have tailored assortments.

The proposed heuristic is a next-best-SKU-based greedy heuristic that calculates, at each step, the SKU that will generate more marginal net sales if it is included in the assortment, taking into account substitution/cannibalization effects and the intrinsic demand calculated at the largest store of the cluster (according to assumption A5/A6, it is possible to scale this demand according to footfall data later for smaller stores). There are two main components to deciding what is the next SKU to enter an assortment: the scores representing contribution to net sales and contribution to attribute diversity.

The **contribution to net sales** is calculated taking into account three core factors: the SKUs intrinsic demand, the substitution demand it can absorb from OOA products, and the cannibalization demand it will take away from SKUs already in the assortment. Equation 4.7 represents the contribution to net sales of product *i* outside the assortment.

S_n	net sales score of SKU *n*
ID_n	intrinsic demand of SKU *n*
OA	set of candidate SKUs out of the assortment
IA	set of candidate SKUs in the assortment
$M_{sub_{n,m}}$	substitution factor from SKU *n* to SKU *m*
P_n	purchasing price of SKU *n*

$$S_i = ID_i * P_i + \sum_{j \in OA} ID_j * M_{sub_{j,i}} * P_i - \sum_{k \in IA} ID_i * M_{sub_{i,k}} * P_k \tag{4.7}$$

The **contribution to attribute diversity** is calculated by assessing the uniqueness of each attribute of SKU i when compared to the products in *IA*, and it is expressed in equation 4.8.

D_n	diversity score of SKU *n*
A	set of attributes being considered in the analysis
$K_{a,l}$	set of SKUs which have level *l* of attribute *a*

$$D_i = \sum_{a \in A} \frac{|IA| - |K_{a,l}|}{|IA|}, \qquad i \in K_{a,l} \tag{4.8}$$

The importance given to these two scores is not set: CMs can choose whether they want to build a rank of SKUs that focuses more on diversity, or more on maximizing net sales. Since every category has diverse needs (and these needs change over time), it was decided to leave the fine tuning of these parameters to the CM. Having this in mind, the best SKU i^* chosen to enter the assortment at each position is given by equation 4.9.

i^*	SKU with the highest score at each step of the heuristic
s_w	weight of net sales contribution
d_w	weight of attribute diversity contribution

$$i^* = \underset{i \in OA}{\arg max} \left(s_w * \frac{S_i}{\underset{i \in OA}{max}(S_i)} + d_w * \frac{D_i}{\underset{i \in OA}{max}(D_i)} \right) \tag{4.9}$$

This score varies between 0 and 1 (as the sum of s_w and d_w is equal to 1), and it will only take the value of 1 if there is one SKU that has maximum marginal contribution to net sales and brings maximum extra attribute diversity to the assortment (at each step). The algorithm is over when $\{OA\}=\emptyset$; when this condition is reached, all candidate SKUs will be in an order such that each SKU is the one that maximizes the net sales/diversity score at each position.

4.6 Space planning

If the current methodology is applied, a valuable piece of information is available to the space manager at the time of deciding store layouts: the space manager will know exactly which SKUs will be on the shelf for each category for any objective space. The goal of space planning is to maximize the total net sales of the categories involved, taking into account the different types (and quantities) of equipment available. It also needs to be taken into consideration whether equipment can be shared between categories. This property will be hereinafter referred to “equipment divisions”. For instance, a division of 3 means that the equipment can be divided into thirds of its maximum space.

The total space of the store, the types and capacity of display equipment are information the space manager possesses: as such, it is possible to build a linear programming model that optimizes total net sales (or profitability, if profitability data is available). Parameters and decision variables of the applied model are shown in Table 2.

Table 2 - Decision variables and parameters of the model

Decision variables:		
	$Q_{e,n}$	Number of equipment of type *n* for product category *e*
	N_e	Number of chosen unique SKUs for product category *e*
	M_e	Number of total SKUs for product category *e* (discontinued products and replications included)
	$P_{e,s}$	=1 if product category *e* has *s* SKUs
	$D_{e,n}$	Number of equipment divisions of type *n* for category *e* (integer)
	U_n	Total number of equipment of type *n*
	l_n	Total length occupied by equipment of type *n* (meters)
Parameters:		
	E	Number of categories
	N	Number of different types of equipment
	$S_{e,s}$	Number of SKUs per category *e*
	$NS_{e,s}$	Net sales for category *e* with allocated space *s*
	T	Total available linear space (meters)
	Max_e	Maximum number of SKUs of category *e*

Min_e	Minimum number of SKUs of category *e*
I_e	Percentage of discontinued products for category *e*
Du_e	Percentage of facing duplications for category *e*
$MaxEq_n$	Maximum number of equipment of type *n*
$MinEq_n$	Minimum number of equipment of type *n*
L_n	Length of equipment of type *n* (linear meters)
Div_n	Number of equipment divisions (between categories) for equipment of type *n*
$C_{e,n}$	Capacity (in number of SKUs) of equipment of type *n* for product category *e*

The model was subjected to several constraints regarding space, equipment, and number of SKUs of each product category. Regarding space, there are constraints related to total available space (4.10) and the space occupied by each type of equipment (4.11).

$$\sum_{n=1}^{N} l_n \leq T \tag{4.10}$$

$$U_n * L_n = l_n, \forall\, n = 1 \ldots N \tag{4.11}$$

Concerning equipment, there are constraints about its minimum and maximum number (4.12), to guarantee a non-negative number of equipment *n* for each product category *e* (4.13), to relate its possible number of divisions with its quantity (4.14), and to guarantee that if the capacity of a certain equipment for a specific product category is 0, the number of divisions of that equipment for that category is also 0 (4.15).

$$U_n \leq MaxEq_n\ , U_n \geq MinEq_n\ , \forall\, n = 1 \ldots N \tag{4.12}$$

$$Q_{e,n} \geq 0\ ,\ \ \forall\, e = 1 \ldots E, n = 1 \ldots N \tag{4.13}$$

$$D_{e,n} = Q_{e,n} * Div_n\ ,\ \ \forall\, e = 1 \ldots E, n = 1 \ldots N \tag{4.14}$$

$$\sum_{e=1}^{E} D_{e,n} \,/\, Div_n = U_n\ ,\ \ \forall\, n = 1 \ldots N$$

$$C_{n,e} \leq Q_{n,e} * MaxEq_n,\ \ \forall\, e = 1 \ldots E, n = 1 \ldots N \tag{4.15}$$

Regarding product categories, there are five kinds of constraints applied: guaranteeing that there is only one number of SKUs chosen (4.16), that the minimum and maximum number of available SKUs is respected (4.17), relating total SKUs and active ones for each category (4.18), relating types of equipment used with each product category (4.19), and relating the allocated space with the number of SKUs chosen (4.20).

$$\sum_{s=1}^{S} P_{e,s} \leq 1 \,, \forall\, e = 1 \ldots E \tag{4.16}$$

$$N_e \leq Max_e \,, \forall\, e = 1 \ldots E \tag{4.17}$$

$$N_e \geq Min_e \,, \forall\, e = 1 \ldots E$$

$$N_e \leq (1 - Du_e - I_e) * M_e \,, \quad \forall\, e = 1 \ldots E \tag{4.18}$$

$$\sum_{n=1}^{N} Q_{e,n} * C_{e,n} \geq M_e \,, \qquad \sum_{n=1}^{N} Q_{e,n} * C_{e,n} - 0{,}999 \leq M_e, \quad \forall\, e = 1 \ldots E \tag{4.19}$$

$$\sum_{s=1}^{S_e} P_{e,s} * S_{e,s} = N_e \,, \quad \forall\, e = 1 \ldots E \tag{4.20}$$

Finally, the objective function can be written as the maximization of the total net sales of the categories involved in the space optimization (4.21).

$$\sum_{e=1}^{E} \sum_{s=1}^{S} NS_{e,s} * P_{e,s} \tag{4.21}$$

5. Preliminary results

At the time this study was completed, store implementation was yet to be done. Hence, all results presented will concern the predicting accuracy of the proposed models and the forecasted improvements in net sales for the analyzed categories. Furthermore, considering the great number of product categories present at the stores, a full analysis was not possible given the time constraints of this work. As such, the results of two categories were chosen to be described in this chapter.

In order to pressure test the intrinsic demand forecasting and substitution methodologies to the greatest possible extent, a highly dynamic (both in SKU introduction and pricing) category was considered to be more appropriate – thus, the TV category was chosen.

However, the TV category does not usually share store space with other categories, making it unfit to test the efficiency of the space allocation model. After several iterations, it was decided to use the category of textile handling. This category includes four distinct types of products: simple irons, boiler irons, ironing boards and sewing machines, and it is a rather stable category concerning assortment changes, making it ideal for this assessment.

Identically to the methodology, the ensuing part of this chapter is divided into four sections – determining intrinsic demand, estimating substitution rates, assortment heuristic, and space planning.

5.1 Determining intrinsic demand

As described before, to estimate which products were on the shelves during each month, it is necessary to cross sales data with stock data. Two-year sales data was requested at the start of the project and was received briefly but, on the other hand, stock data only became a necessity later and the two years' worth of data were not provided in a timely fashion and, at the time of completion of this work, only stock data from January to April of 2017 was available. Hence, the first three months were used to calibrate the model and the month of April was defined to be the test set. Ideally, the test set would have to be one year long to account for seasonality; however, this also means that the results presented below can only improve when more data is available.

As previously mentioned, the first step in this methodology is to apply an MNL model to estimate the utilities of attributes' levels considered. The results of this analysis are shown below: the attributes considered were price decile (Figure 5), brand (Figure 6, with brand names not present due to privacy issues), TV size (Figure 7) and resolution (Figure 8).

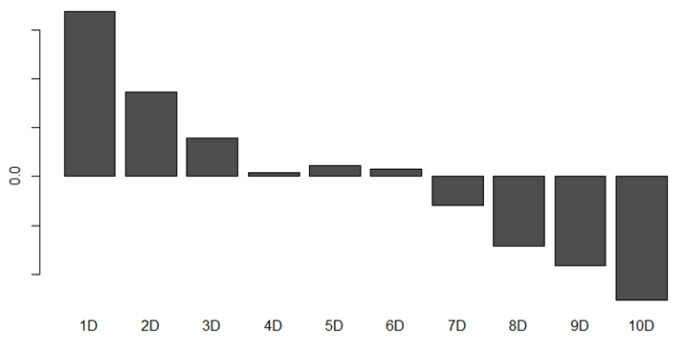

Figure 5 - Relative utilities of price deciles

Figure 6 - Relative utilities of brands

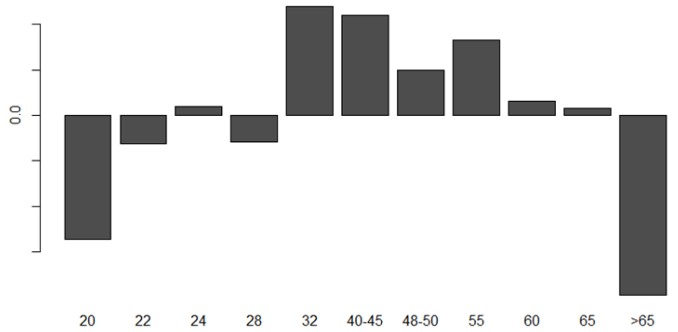

Figure 7 - Relative utilities of sizes

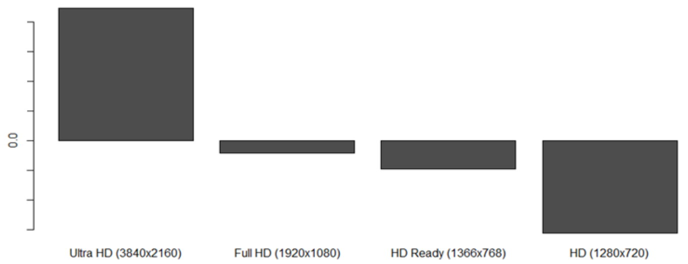

Figure 8 - Relative utilities of resolutions

The interpretation of the figures above is not trivial: for instance, taking a look at Figure 7, the fact that 60' has a higher relative utility than 65' does not necessarily mean that the first is better than the latter – it means that, keeping every other attribute constant, the average consumer prefers to purchase a 60' TV than a 65' one.

It is not easy to evaluate whether these results are accurate, as it would require a complete marketing study to do so. Yet, it is possible to analyze the two attributes which are known to have a clear preference order: price and resolution. While it is not correct to assume that the average consumer prefers certain brands or sizes, it is safe to say that the average customer prefers to pay cheaper prices and get the highest resolution possible. Observing Figures 5 and 8, it is clear that the results obtained follow this trend – there is a slight deviation in the 4th decile in figure 5, but this is thought to be caused by lacking data or promotional activity that could not be read by the model. Regarding resolutions, it is perceptible that the average consumer prefers higher resolutions (Ultra HD) over low resolutions (HD). Moreover, even though it is not an analysis as robust as for price and resolution, one can also draw conclusions from the size attribute: the average consumer prefers medium sized TVs because small ones (less than 32') are not big enough for a living room environment, and the extremely large ones are not adequate for the size of the average living room in Portugal.

Based on the examination of the breakdown of attributes above, it is assumed that the values estimated for the utilities of attributes are aligned with reality and it is reasonable to advance to intrinsic demand estimation.

As detailed before, the estimation of shares of each SKU can be done based on each product's attribute levels and the preliminary results are in Figure 9.

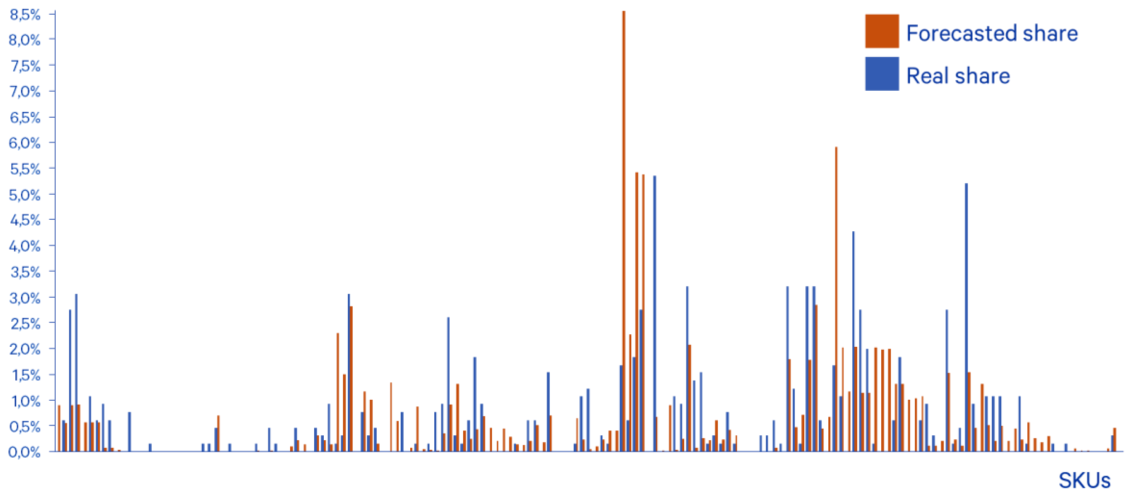

Figure 9 - Forecasting of store shares of SKUs at the largest store of a cluster

The mean absolute percentage error (MAPE) per store/month/SKU of this forecast is 82%, which was thought to be a rather high value. There are three main explications to why this value is high.

The first one is related to assumption A7 – SKUs with at least one unit of stock during the month are assumed to have been on shelf. However, it is known that some SKUs are only at stores for limited time periods (e.g. one week) during promotional activities, and others can simply run out of stock at any point during the month. It is extremely unlikely that all SKUs considered were displayed for the whole 30 days in the month, and some of the SKUs predicted to have a high share participation might simply not have been available for a significant amount of time.

The second reason concerns marketing activity. It is known that some brands have exclusive brand spaces in the store, and that the category managers choose distinct products every month to highlight. Again, this is something that the model cannot predict and this activity is too dynamic to be planned beforehand and provided as an input, as it depends on suppliers, market trends and competitors.

The third reason is linked to promotional activity. As pricing is very dynamic in this category, the sales price of each unit might be different on a daily basis, meaning the choice set each customer sees when he/she enters the shop is different – which directly impacts attribute level utility estimation.

While nothing can be done to effectively solve the first two points, promotional activity can be partially included in the model – as the sales price is known for every transaction, it is possible to calculate the average sales price of each SKU for each month in the test period. Hence, it was decided to include this extension in the model: instead of considering the regular price for each product in the choice set of each month, it was considered the decile in which the average sales price of each product for that particular month is located (for SKUs without sales, the regular price was considered). Even though this does not solve the fact that it is not known which products will have discounts in the following months, it is possible to assess price importance more accurately if this change is made. The results comparable to the ones on Figure 9 are presented on Figure 10.

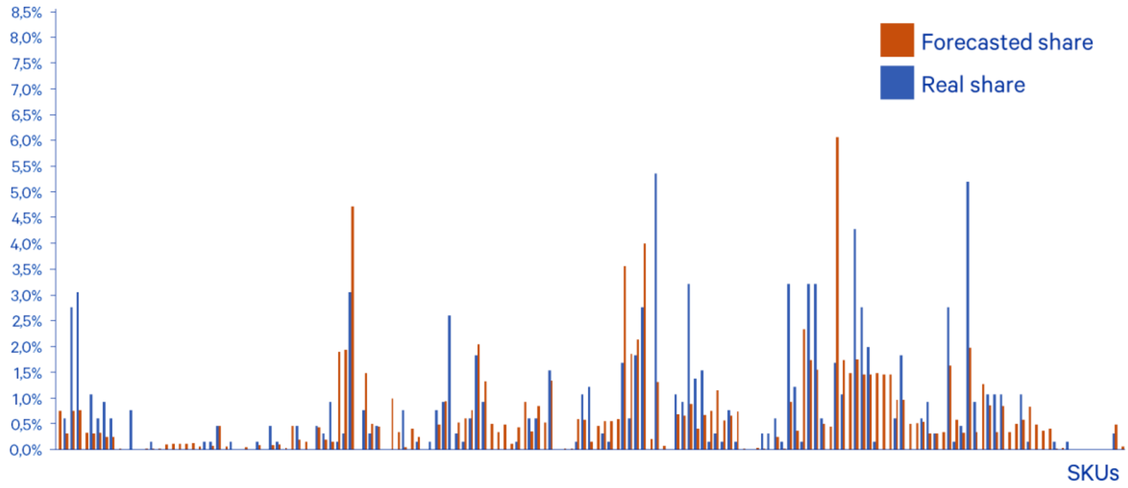

Figure 10 - Forecast of SKUs sales shares considering dynamic pricing

Not only the results are graphically more pleasing, the MAPE also drops 11 percental points to a value of 71%. This error exists due to the abovementioned reasons, which were not tackled during the course of the project. Nevertheless, even though the error might still seem rather high, it is considerably better than the one from the current forecasting system the retailer has in place, which has a store/month/SKU MAPE of over 100% for this category. Moreover, the current system uses 1 year data and considers seasonality effects for forecasting purposes.

On a final note, the attribute levels represented (Figures 5 to 8) are the ones calculated with this extension to the model.

5.2 Estimating substitution rates

The methodology described in the methodology (Section 4.4) was applied to predict the sales (total and per SKU) of stores with smaller assortments, based on the sales of the store with the larger assortment. In order to make the needed computations lighter, only the ρ for price and size were considered (resolution was left out for this part of the methodology assessment). The optimized values obtained were δ=0.91, ρ_{price}=0.79 and ρ_{size}=0,66.

The value for the store/month MAPE is 45% and the values for the real sales values and respective forecasts can be seen on Figure 11. This MAPE was calculated using optimized parameters and, again, a lower value was expected since for this case it is not being calculated on an SKU basis. However, even though not described in this study, it must be taken into account that stores are divided in clusters according to average customer profile. Therefore, it is more relevant to analyze only the results for the forecasts of the stores that are in the same cluster as the largest store being used. These results can be consulted in Figure 12 – in this case, the MAPE drops to 29%. This proves the efficiency of the store clustering methodology and the substitution model simultaneously. Regardless of results being rather satisfactory, it is still relevant to discuss why an error of this dimension exists. Besides all the motives explained in the previous section (difficulty of knowing what is on shelf, different promotional items) also applying to this case, there is one extra reason that might be increasing the error: the inherent flaw in footfall data. Even though footfall data is being used to scale stores' total demand, it is definitely not a good category proxy – for instance, if a store has twice as many visitors, it does not mean that twice as many people will try to buy items from each category. Yet, while no other market data is available, footfall is the only available information that allows to relate stores demands. It can be argued whether total store sales can also be used – and even though this is another valid solution, given that the most important assumption of this work is that SKU substitution is crucial for assortment planning, using sales data as proxy would go against what is considered to be the cornerstone of the project.

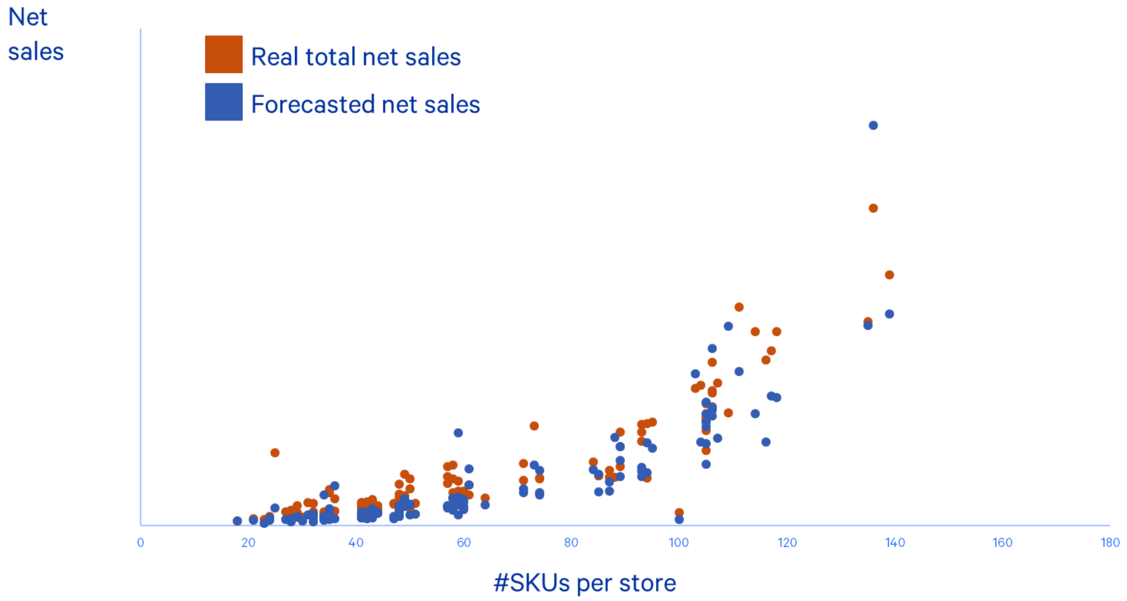

Figure 11 - Forecasts of total sales for all stores

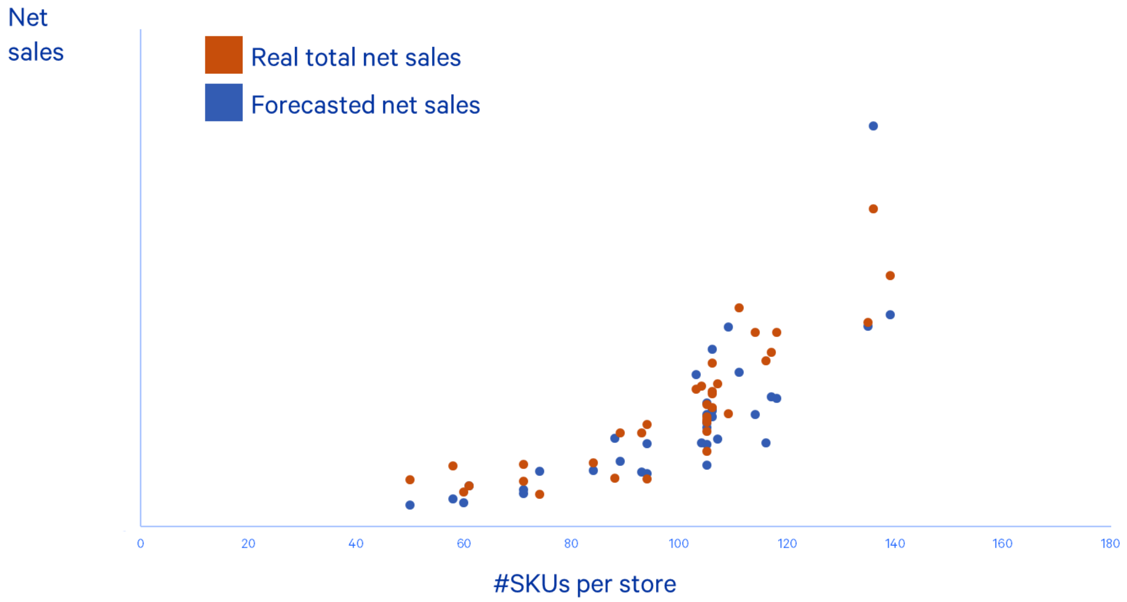

Figure 12 - Forecasts of total sales for stores in the same cluster

Even though MAPE has been the only metric referred to until now, when looking at Figure 12 there is something equally or more important that can be visually confirmed: the shape of the substitution curve. As can be seen, the predicted data follow the same tendency as that of real data, which means that the substitution patterns are being correctly understood by the model. This methodology's goal is to build a ranked list of SKUs as debated before, so more than predictive accuracy, it is important that the substitution is being correctly modelled.

Figure 13 is an example of a SKU-by-SKU forecast of quantity sales of a store in the same cluster as the largest one considered. The MAPE for this store is 53%, and the MAPE for all stores in the same cluster (product/month/store) is 63%.

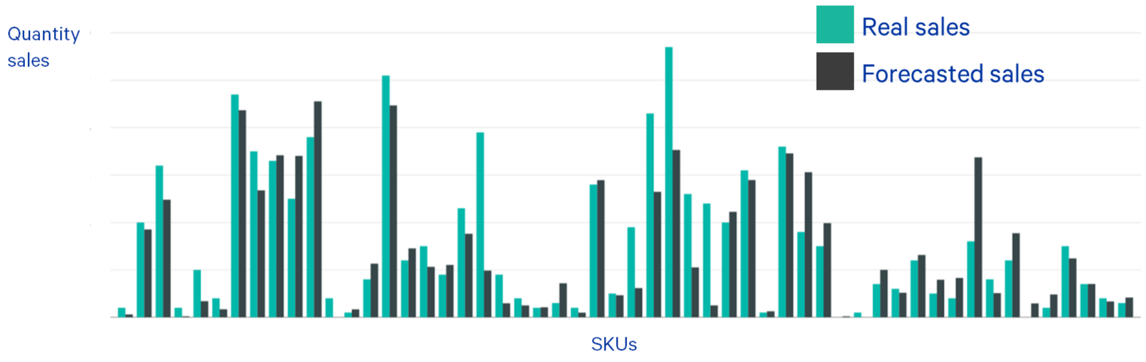

Figure 13 - SKU-based MAPE for a smaller store

As a final remark, note how these forecasts are based solely on four factors: the sales of a larger store, existing assortment at smaller stores, store profiles, and footfall data. No past data of previous sales of the smaller stores is being used for forecasting – which all other forecasting systems in the market use. Under these conditions, a combined MAPE of 63% is thought to be a very satisfying result.

5.3 Assortment heuristic

With validated results of intrinsic demand estimation and substitution rates, it is possible to proceed to calculating the optimal order of SKUs with the assortment heuristic. In this section, brands of SKUs will be omitted as before.

Table 3 shows a product ranking based purely on net sales maximization, whereas Table 4 contemplates a solution that gives 80% of weight to product diversity and 20% to net sales.

Table 3 - SKU rank focused on net sales

Rank	Price	Brand	Size
1	4D	A	32
2	8D	A	40-45
3	7D	A	40-45
4	10D	A	65
5	5D	A	32
6	8D	A	48-50
7	9D	B	55
8	7D	A	40-45
9	9D	A	55
10	7D	A	40-45
11	2D	A	24
12	10D	B	55
13	10D	B	65

Table 4 - SKU rank focused on SKU diversity (at 80%)

Rank	Price	Brand	Size
1	4D	A	32
2	9D	B	55
3	6D	C	40-45
4	2D	D	24
5	10D	A	65
6	1D	E	22
7	8D	A	48-50
8	7D	B	40-45
9	3D	E	32
10	8D	C	55
11	5D	A	40-45
12	8D	C	48-50
13	10D	B	65

Before moving forward to analyzing expected gains in net sales, it is pertinent to compare both solutions and understand how much in net sales is being lost by wanting to have a diverse assortment at the stores. Following this line of thought, Figure 14 shows the comparison between these two solutions and another one that equally weighs diversity and net sales.

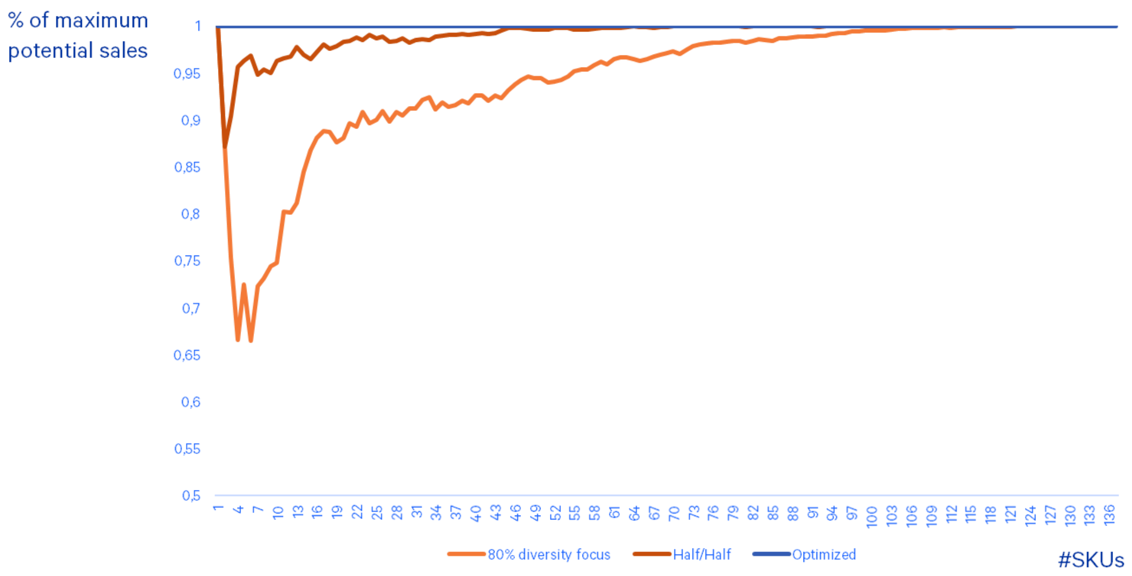

Figure 14 - Comparison of potential net sales of each solution

As the smallest store of the chain displays about 20 TVs, the 50%/50% is expected to increase SKU diversity without having a profound impact on the bottom line. On the other hand, a solution heavily focused on having a diverse assortment (which is very close to the current, manually implemented solution), will cause a 10% loss in total potential sales for all smaller stores (from 20 to around 40 SKUs), and will only achieve maximum sales potential at a space of 100 SKUs. Such severe results suggest that CMs need to analyze each category and deliberate whether extreme diversity is really needed and why – if it is a matter of corporate strategy, there is no optimization possible; if not, a careful analysis should be done in order to determine the minimum percentage allowable for the diversity weight of the solution.

Concerning the bottom line, the expected results if the optimized assortment is applied versus the sales with the current assortment per store are shown in Figure 15. The total sales values are omitted and results will be discussed on a percental basis.

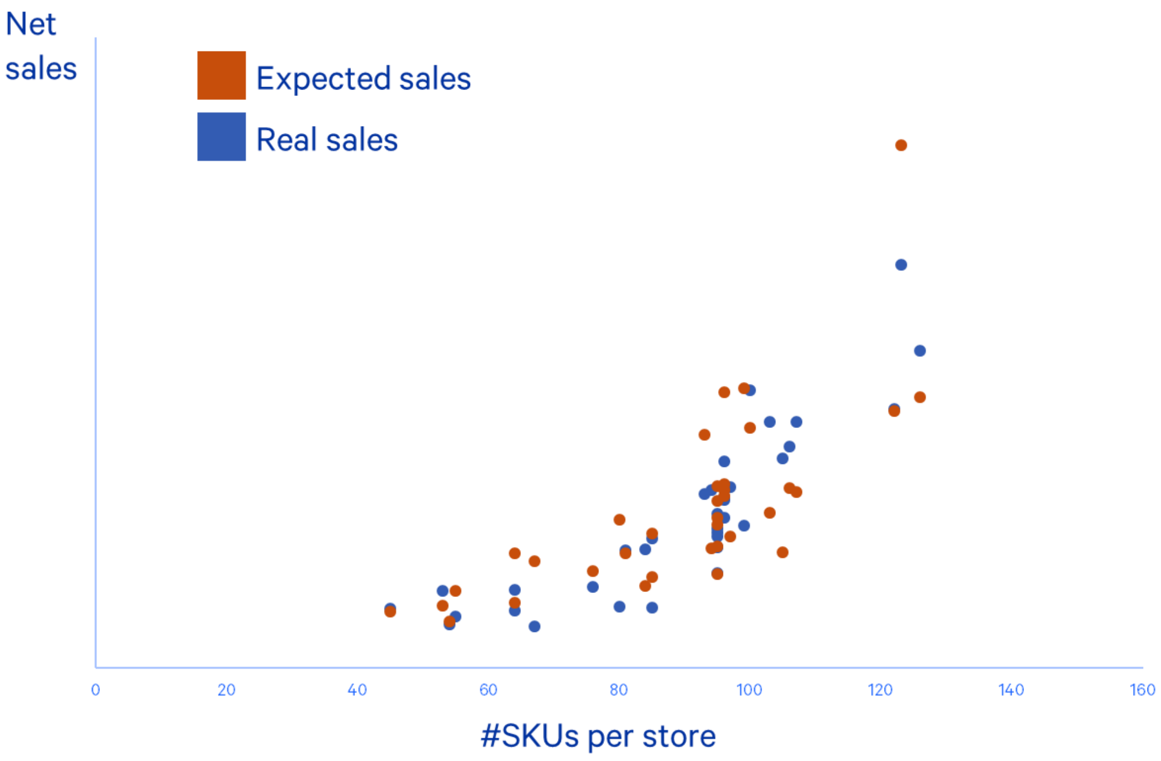

Figure 15 - Expected potential sales versus current sales

The average percentage increase per store in total net sales value calculated is 15%. However, this figure is not representative of the total extra value obtained by the company as the largest percental sales increases were obtained at the smaller stores which have lower absolute sales values – these results are in line with what was discussed regarding Figure 14, where it was stated that the smaller stores are the ones most affected by assortment changes, as it is where substitution effects have more impact.

Therefore, it is more relevant to calculate the percentage of total net gains for all these stores with the new assortment, which is equal to 4,2%. This result is also in line with what Kök and Fisher (2007) obtained for an assortment optimization project at a food retailer, and it is considered to be reasonable. However, in Figure 15, the results being compared concern the results of the planned active assortment versus all products sold at the stores during the same time period (including discontinued products). Though the necessity of evaluating the impact of including discontinued products in the optimized assortment is recognized, this is something that is only possible to measure if a pilot test is taken.

Regardless of the positive result, there are two important points to discuss concerning its assertiveness. The first one is related to the error of the approach – as discussed above, the MAPE of the substitution methodology is 29%, which is not encouraging when looking at a result as thin as a 4,2% improvement; the second issue is the effect of discontinued products.

The assortment planning methodology does not take into account substitution effects that happen between chosen active SKUs and discontinued products – ideally these should be included in the model, but due to the irregular nature of their behavior, it is not possible to build a robust model that takes their influence into account.

It can be reasoned that if the assortment heuristic is optimal, then it should optimize sales for every store. However, as described before, due to manageability constraints and other factors, the SKU rank is based on the sales potential at the largest store of each cluster – given that there is some intra-cluster dissimilarity, it is natural that some stores will have their sales decrease as the assortment is not optimized on a store basis.

On the other hand, as a final note, going back to Figure 12, one can estimate the bias of the forecast to be -12%: this is rather reassuring in the sense that if the model is underestimating sales, it might indicate that the improvement might indeed be quite close, or even higher than the proposed 4,2%. However, an exhaustive analysis is necessary to prove this concept.

5.4 Space planning

As mentioned in the beginning of this chapter, for this section the *Textile handling* category will be used as an example. To be able to apply the space allocation methodology, all previous steps were also undertaken: intrinsic demand estimation, substitution parameters calculation, and assortment optimization. The following steps presuppose that a ranking of SKUs is available for each one of the four types of products: simple irons, boiler irons, ironing boards and sewing machines.

First, an example of the situation exemplified by Figures 2 and 3 will be shown – Figure 16 represents the space elasticity curves for regular irons and boiler irons, using the calculated optimized SKU ranking for both (sales values omitted).

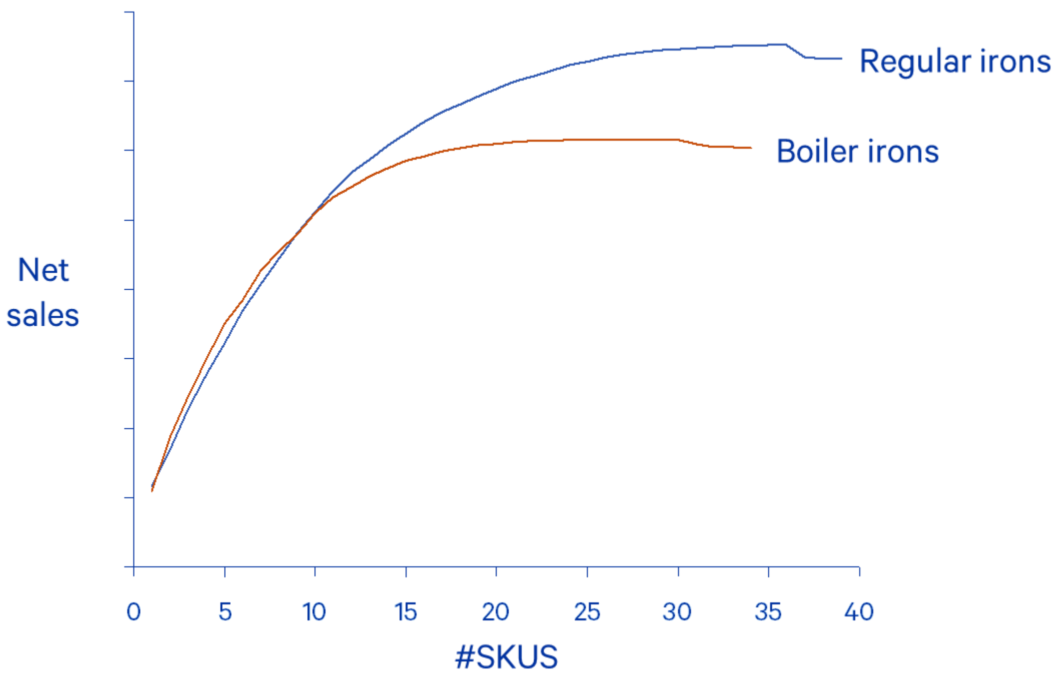

Figure 16 - Space elasticity curves of regular and boiler irons

There are two significant topics to point out regarding these curves. First, the fact that, indeed, different substitution behavior can generate distinct elasticity curves. In the case presented, boiler irons are more inter-substitutable than regular irons: as such, the curve is steeper in the beginning but stabilizes faster. Moreover, the reason why the curve of boiler irons stabilizes below the regular irons one is simply because the total demand for boiler irons is smaller.

The second topic is something that surprised the project team as well and it concerns the small drop in net sales at the end of each curve. At first, this was thought to be a problem in the substitution model – however, when an in-depth analysis was performed, the cause for this phenomenon was clear: as an assortment gets larger, the substitution it absorbs from OOA SKUs gets increasingly bigger (on a percental basis), which is why the space elasticity curve slowly stabilizes. It happens that some SKUs that enter the assortment in the final positions are cheaper alternatives to the ones already in – which means that all the substitution demand that was being transferred from the cheap SKU to the slightly more expensive alternatives will be cannibalized by the cheap one, reducing the total net sales value. These "toxic" SKUs exist in every category and, surprisingly, some of them are moderately popular products.

Regarding the space optimization itself, the model described before was applied to this product category and it returned the results Tables 5 and 6.

Table 5 - Types and quantities of each equipment for each type of product

	Equipment type						Total
	1	2	3	4	5	6	
Regular Irons	1,0	1,0	-	4,0	-	-	6,0
Boiler Irons	-	-	-	6,0	-	-	6,0
Ironing boards	-	-	-	1,0	-	-	1,0
Sewing machines	1,0	-	-	1,0	-	-	2,0
Total	2,0	1,0	-	12,0	-	-	15,0
Space (m)	2,0	1,0	-	16,0	-	-	19,0

Table 6 - Optimal number of SKUs and sales distribution per type of product

	# active SKUs	Sales potential €
Regular Irons	36	34,7%
Boiler Irons	30	28,4%
Ironing boards	4	8,1%
Sewing machines	6	28,8%
Total	76	100%

The store used for this example has a 19-meter-long corridor that needs to be divided by these four types of products. There are six different types of equipment, with different lengths and particularities (e.g. some can only be used for one type of SKUs, some can be shared and others do not). In order to estimate the increase in net sales by optimizing space, the total sales value was calculated with current space allocation, but with the SKUs of the optimized assortment (as different assortments cannot be compared solely on a space basis), and the total sales value considering the optimized assortment and space.

The predicted results show an increase of almost 8% in net sales of the whole product category (the aggregated sales of the four product types). By analyzing previous allocation and the optimized one, one can confirm that space was moved from the most populated product types (i.e., the ones that lose less demand if SKUs are removed), and space was allocated to categories that were underrepresented. These results are in line with what was discussed concerning Figure 16 – “toxic” SKUs are removed, thus improving slightly the net sales of the respective product type, and the extra space is allocated to a product type that will generate positive extra revenue by having space for additional merchandise.

Even though these results are very promising, it is not easy to measure them in practice: when space changes, the optimal assortment at that moment is allocated to the store, and, as argued above, both situations become incomparable. Yet, it might be possible to assess some of its effects with a tightly controlled pilot testing strategy.

6. Conclusions and future work

The work developed intends to coordinate assortment and space planning by first building a module capable of calculating intrinsic demand for every candidate SKU, estimating exogenous substitution demand, building a next-best-SKU-based suboptimal rank of products and, finally, using these to optimize space allocation of diverse product categories. The results were satisfactory – assortment changes for one of the tested categories are predicted to bring about a total increase in net sales of 4,2% across 39 stores, and space optimization for four product types at a store is expected to cause an 8% increase in net sales for the set of those same categories at that store.

The present state of the art in assortment planning focuses on MNL models due to their innate flexibility, which allows them to adapt to new SKUs in the assortment. Current exogenous models cannot handle this level of change; however, during the course of this work, a framework for classification of attributes was developed which allowed for an attribute-based substitution matrix, thus making the used exogenous model able to adapt to dynamic assortment planning realities. It is hoped that this work provides a significant contribution to future studies involving exogenous substitution models.

As discussed thoroughly before, there are several ways of enhancing the models proposed in this work, thus improving accuracy and gains in the bottom line of the retailer. In the MNL models, it is considered to be critical to understand which products were on the shelf during each month – with the current approach, products that were in the backroom (i.e. had stock) are considered to have been displayed as well. For modelling consumer choice, it is imperative to precisely know choice sets (if possible, on a weekly basis) to correctly model utility levels, thus improving the forecasting accuracy. Another extension proposed for the MNL model is the inclusion of past marketing activity. Some SKUs are forecasted to have a higher demand because they sold in the past due to promotional support that eventually stopped. Past marketing activity can be described as a discrete variable and can easily be included in the model as an SKU attribute – besides the improvement of the model, another very insightful result for the retailer would be the relative utility levels of distinct degrees of marketing activity.

For the exogenous substitution models, there are three main suggestions of development. First, the evaluation of other substitution patterns of different attributes. In this work, since the goal was a scalable, full-fledged application at a retailer of the system, trade-offs between assertiveness and run speed had to be made. However, from an academic perspective, it would be interesting to develop a model that optimizes substitution patterns without the constraints of substituting only to adjacent levels of functional attributes. Second, an in-depth analysis of the behavior of preference attributes is necessary to assess whether popularity based substitution is effective or not. For instance, it can be argued that customers looking for less popular levels are less willing to change (e.g. someone looking for a pink smartphone will be less willing to change to a black one than someone looking for a white phone). This kind of conjecture is plausible and can bring about meaningful results if tested. Finally, the way the SKU-to-SKU matrix is built can also be improved – in this work, it is considered that attributes are equally important and the SKU-to-SKU matrix is a simple multiplication-based transformation of the attribute substitution matrixes. However, it is a given that consumers are less willing to substitute some attributes than others. As such, it is thought to be relevant to test and optimize the weights of the different attributes for substitution: this can be straightforwardly implemented in the proposed methodology by adding weight parameters of attributes in the substitution matrix and optimizing them together with δ and ρ.

Regarding the assortment heuristic, there is one main improvement opportunity – the inclusion of discontinued products in the process. In order for this to be done, there would have to be a symbiosis with a reverse logistics methodology that allows to estimate the total costs of moving SKUs between stores. If it is possible to plan which discontinued products will be present at each store, it is believed that the full potential of the substitution benefits can be obtained.

Finally, for space planning, it is proposed that not only inter-shelf space is evaluated, but also intra-shelf space: the impact of having several facings for SKUs is ignored by the proposed method, but for categories with high stock rotation, this is expected to be relevant (e.g. keyboards and mice). There is a plethora of work on facings' space elasticity, and its coordination with this work, albeit not straightforward, is thought to be highly relevant.

Finally, it is important to mention that assortment planning is an unexplored, but extremely pertinent area of study as supply chains become leaner and profit margins lower. Following the results of this work, retailers should make some effort to understand which products should be on the shelves, as it has a significant and direct contribution to net sales and profit. The market trends in recent years show that, more than pleasing the masses, valuing the individuality of consumers is increasingly important, and assortment planning has undeniable impact in this area.

References

Deloitte UK (2017). A brave new world - The Retail Profitability Challenge.

Cheng, S., & Long, J. S. (2007). Testing for IIA in the Multinomial Logit Model. *Sociological Methods & Research*, *35*(4), 583–600. https://doi.org/10.1177/0049124106292361

Corsten, D., & Gruen, T. (2003). Desperately seeking shelf availability: an examination of the extent, the causes, and the efforts to address retail out-of-stocks. *International Journal of Retail & Distribution Management*, *31*(12), 605–617. https://doi.org/10.1108/09590550310507731

Fisher, M. - The New Science Of Retailing - Chapter Two.pdf. (n.d.).

Fisher, M. L., & Vaidyanathan, R. (2009). An Algorithm and Demand Estimation Procedure for Retail Assortment Optimization. *Management*, (2007), 1–36. https://doi.org/10.1287/mnsc.2014.1904

Fisher, M. L., Vaidyanathan, R., & Gu, A. (2009). *Retail Supply Chain Management* (Vol. 122). https://doi.org/10.1007/978-0-387-78902-6

Green, P. E., Krieger, A. M., & Wind, Y. (2001). Thirty Years of Conjoint Analysis: Reflections and Prospects. *Interfaces*, *31*(3–Supplement), 56–73. https://doi.org/10.1287/inte.31.4.56.9676

Guadagni, P. M., & Little, J. D. C. (1983). A Logit Model of Brand Choice Calibrated on Scanner Data CALIBRATED ON SCANNER DATA *, (September 2015).

Hübner, A. H., & Kuhn, H. (2012). Retail category management: State-of-the-art review of quantitative research and software applications in assortment and shelf space management. *Omega*, *40*(2), 199–209. https://doi.org/10.1016/j.omega.2011.05.008

Journal, D., & Complete, I. (2000). Neural networks and the multinomial logit for brand choice modellnig : A hybrid approach, *200*(February 1997).

Kök, A. G., & Fisher, M. L. (2007). Demand Estimation and Assortment Optimization Under Substitution: Methodology and Application. *Operations Research*, *55*(6), 1001–1021. https://doi.org/10.1287/opre.1070.0409

Mahajan, S., & van Ryzin, G. (2001). Stocking Retail Assortments Under Dynamic Consumer Substitution. *Operations Research*, *49*(3), 334–351. https://doi.org/10.1287/opre.49.3.334.11210

Mc Fadden, D. (1974). The Measurement of Urban Travel Demand. *Journal of Public Economics*, *3*, 303–328.

Miller, C. M., Smith, S. A., McIntyre, S. H., & Achabal, D. D. (2010). Optimizing and Evaluating Retail Assortments for Infrequently Purchased Products. *Journal of Retailing*, *86*(2), 159–171. https://doi.org/10.1016/j.jretai.2010.02.004

Rooderkerk, R. P., van Heerde, H. J., & Bijmolt, T. H. A. (2013). Optimizing Retail Assortments. *Marketing Science*, *32*(5), 699–715. https://doi.org/10.1287/mksc.2013.0800

Rusmevichientong, P., Shmoys, D., Tong, C., & Topaloglu, H. (2014). Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management*, *23*(11), 2023–2039. https://doi.org/10.1111/poms.12191

Sinha, A., Sahgal, A., & Mathur, S. K. (2013). Practice Prize Paper Category Optimizer: A Dynamic-Assortment, New-Product-Introduction, Mix-Optimization, and Demand-Planning System. *Marketing Science*, *32*(2), 221–228.

https://doi.org/10.1287/mksc.1120.0746

Smith, S. A., & Agrawal, N. (2000). Management of Multi-Item Retail Inventory Systems with Demand Substitution. *Operations Research*, *48*(1), 50–64. https://doi.org/10.1287/opre.48.1.50.12443

Society, T. E. (2016). Specification Tests for the Multinomial Logit Model Author (s): Jerry Hausman and Daniel McFadden Published by : The Econometric Society Stable URL : http://www.jstor.org/stable/1910997 Accessed : 04-08-2016 11 : 16 UTC Your use of the JSTOR archive ind, *52*(5), 1219–1240.

Yücel, E., Karaesmen, F., Salman, F. S., & Türkay, M. (2009). Optimizing product assortment under customer-driven demand substitution. *European Journal of Operational Research*, *199*(3), 759–768. https://doi.org/10.1016/j.ejor.2008.08.004