Quantitative Methods for Omnichannel Decision-Making

Andrey VASILYEV

To my parents.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Prof. Ralf W. Seifert. He has been extremely supportive throughout my PhD studies, providing me with every opportunity to succeed in my academic pursuits. His mentorship went beyond research, as he shared his knowledge and experience that have had a profound impact on my personal and professional growth. I consider myself fortunate to have had the opportunity to work under his guidance.

I would also like to extend my gratitude to my co-author, Dr. Sebastian Maier. He not only supported me in the process of publishing academic articles, but also kindly agreed to host me at UCL when a change of scenery was most needed.

I am truly grateful to Prof. Gaétan de Rassenfosse, Prof. Arnd Huchzermeier, Prof. Ulrich W. Thoneman, and Prof. Philippe Wieser, for their willingness to be part of my thesis committee.

Additionally, I wish to express my gratitude to the CDM professors, particularly Prof. Daniel Kuhn and Prof. Thomas Weber, who served as members of my jury for my proposal defense. The Convex Optimization course taught by Prof. Kuhn was an amazing source of inspiration.

I am greatly appreciative of the exceptional administrative support provided by Ilona Ball. Her kindness and thoughtfulness ensured a smooth and hassle-free experience throughout my PhD studies.

I was fortunate to embark on my PhD journey at the TOM chair alongside Yara Kayyali El Alem. I am grateful to have had such a wonderful companion during this time. I would also like to acknowledge other incredible TOM members – Rachel Lacroix, René Glogg, Anna Timonina-Farkas, Onur Demiray, and Buse Şen – who created a great environment of support and friendship at our chair.

A special thanks goes to my office mates: Philipp Schneider, Atin Aboutorabi, Andrea Xu, and Yanan Niu. Our discussions on research and other topics have always been motivating and enjoyable. I would also like to extend my gratitude to the remarkable colleagues I met earlier in my PhD journey: Charles Ayoubi, Paul van Baal, and Ling Zhou.

I wish to express my appreciation to my friends from my BSc years – Yaroslav Sachek,

Acknowledgements

Abstract

Omnichannel retail has emerged as the new standard in today's commerce landscape, with retailers integrating their physical and online channels to enhance the customer shopping experience. However, such integration presents significant challenges for retailers, particularly in relation to optimizing their product assortments. This thesis comprises three chapters, each addressing different aspects of omnichannel assortment optimization under various assumptions. In the first chapter, we introduce the multichannel attraction model (MAM), a discrete choice model designed specifically for omnichannel environments. Focusing on a dual-channel setting, we formulate the assortment optimization problem under the MAM as a mixed-integer linear program, and provide a computationally efficient heuristic method for solving large-scale instances of this problem. We also describe general effects of the implementation of widely-used omnichannel initiatives on the MAM parameters and explore the properties of optimal assortments through numerical experiments.

In the second chapter, we generalize our modeling framework to the case of a retailer managing both an online store and a network of physical stores. Additionally, we incorporate demand stochasticity and inventory management considerations into the assortment optimization problem under the MAM. We show that overlooking the demand variability can result in suboptimal assortment decisions due to the demand pooling effect. We derive complexity results for the assortment optimization problem, which we then formulate as a mixed-integer second-order cone program. We also develop two heuristic algorithms based on different relaxations of the formulated optimization problem. Our findings indicate that an increasing coefficient of variation of demand has a dual effect on optimal assortment sizes, initially causing a decrease in online assortment size due to rising costs, followed by an increase in online assortment size because of the demand pooling effect.

Finally, in the third chapter, we address a key limitation of the MAM by developing a modeling framework for omnichannel assortment optimization that accounts for basket shopping behavior of customers. We model customer choices using a Markov random field – in particular, the Ising model – which captures pairwise demand dependencies as well as the individual attractiveness of each product. We provide theoretical insights into the

Abstract

structure of optimal assortments based on the graphical representation of the Ising model, and develop a customized metaheuristic algorithm for obtaining high-quality solutions to the assortment optimization problem. Lastly, we perform an extensive numerical analysis to gather insights into the properties of optimal assortments and evaluate the benefits of omnichannel assortment optimization as opposed to optimizing assortments in siloed channels.

Keywords: omnichannel retailing, assortment optimization, discrete choice modeling, integer programming, newsvendor model, demand pooling, Ising model, basket shopping behavior

iv

Zusammenfassung

Der Omnichannel-Retailing hat sich als neuer Standard in der heutigen Handelsland-schaft etabliert. Einzelhändler integrieren ihre physischen und Online-Kanäle, um den Service für die Kunden zu verbessern. Eine solche Integration stellt jedoch erhebliche Herausforderungen für Einzelhändler dar, insbesondere in Bezug auf die Optimierung ihrer Produktsortimente. Diese Dissertation ist in drei Kapitel unterteilt, die jeweils unterschied-liche Aspekte der Omnichannel-Sortimentsoptimierung mit verschiedenen Anforderungen behandeln. Im ersten Kapitel stellen wir das Multichannel-Attraktionsmodell (MAM) vor, ein diskretes Entscheidungsmodell, das speziell für den Omnichannel-Handel entwickelt wurde. Für den Dual-Channel-Vertrieb formulieren wir das Problem der Sortimentsopti-mierung unter dem MAM als mixed-integer lineares Programm und entwickeln hierfür eine effiziente heuristische Lösungsmethode, die auch besonders für Problemstellungen mit großer Produktvielfalt geeignet ist. Wir beschreiben auch die Auswirkungen der Implementierung von weit verbreiteten Omnichannel-Initiativen auf die MAM-Parameter und untersuchen die Eigenschaften optimaler Sortimente anhand numerische Experimente.

Im zweiten Kapitel erweitern wir unser Modell für den Fall eines Einzelhändlers, der sowohl einen Online-Shop als auch ein Netzwerk von physischen Geschäften betreibt. Darüber hinaus integrieren wir eine stochastische Nachfrage und Bestandsmanagement-Überlegungen in das Sortimentsoptimierungsproblem im MAM-Modell. Wir zeigen, dass die Vernachlässigung der Nachfrageschwankungen aufgrund des Poolingeffekts zu subop-timalen Sortimentsentscheidungen führen kann. Wir leiten Komplexitätsergebnisse für das Sortimentsoptimierungsproblem ab, das wir dann als mixed-integer Programm zweiter Ordnung formulieren. Zusätzlich entwickeln wir auch zwei heuristische Algorithmen, die auf verschiedenen Annährungen des formulierten Optimierungsproblems basieren. Un-sere Ergebnisse zeigen, dass ein zunehmender Variationskoeffizient der Nachfrage einen zweigeteilten Effekt auf die optimalen Sortimentsgrößen hat. Zunächst nimmt die Größe des Online-Sortiments aufgrund steigender Kosten ab, danach folgt eine Zunahme der Online-Sortimentsgröße aufgrund des Nachfragebündelungseffekts.

Im letzten Kapitel entwickeln wir ein Modellierungsframework für die Optimierung des Sortiments im Omnichannel-Einzelhandel, das das Einkaufsverhalten der Kunden berücksichtigt, die im Alltag typischerweise eine bestimmte Kombination von Produkten,

einen Warenkorb, kaufen. Diese Erweiterung ist in den Standard-MAM-Modellen nicht enthalten. Wir modellieren die Entscheidungen der Kunden unter Verwendung eines Markovschen Zufallsfelds, insbesondere des Ising-Modells, das die paarweise Nachfrageabhängigkeiten sowie die individuelle Attraktivität einzelne Produkte erfasst. Wir liefern theoretische Einblicke in die Struktur optimaler Sortimente, indem wir das Ising-Modell graphisch darstellen und einen angepassten metaheuristischen Algorithmus zur Erlangung hochwertiger Lösungen für das Sortimentsoptimierungsproblem entwickeln. Abschließend führen wir eine umfangreiche numerische Analyse durch, um Einblicke in die Eigenschaften optimaler Sortimente zu gewinnen und die Vorteile der Omnichannel-Sortimentsoptimierung im Vergleich zur Optimierung von Sortimenten in einzelnen Kanälen zu bewerten.

Stichwörter: Omnichannel-Retailing, Sortimentsoptimierung, diskrete Entscheidungsmodelle, ganzzahlige Optimierung, Newsvendor-Modell, Nachfragebündelung, Ising-Modell, Warenkorbanalyse

Contents

Contents

List of Figures

List of Tables

Introduction

The rise of omnichannel retail has had a profound impact on both companies and customers, revolutionizing the way they interact with each other. Through a seamless integration of multiple sales channels, omnichannel retailers offer customers the convenience of easily accessing products and services both online and offline. Consequently, customers have the freedom to choose between visiting brick-and-mortar (B&M) stores, browsing online platforms, or utilizing mobile apps to make their purchases. They also enjoy the flexibility of seamlessly transitioning between sales channels, such as checking in-store product assortments on a retailer's website. A common illustration of omnichannel functionality is the buy-online-and-pick-up-in-store (BOPS) service, also known as click-and-collect, which allows customers to reserve products online for convenient collection at a retail store. Another example of an omnichannel initiative is installing digital help desks in B&M stores to provide customers with immediate access to information about the retailer's online store, including its assortment, prices, and delivery options.

As the boundaries between digital and physical retail environments become blurred, companies have to implement omnichannel initiatives not to lose market share. According to the Global Shopper Trends Report by iVend Retail (2019), 81.4% of consumers reported using BOPS, representing a growth of nearly 30% from the previous year's survey. The study of Sopadjieva et al. (2017), based on a survey of 46,000 customers who made a purchase between June 2015 and August 2016, found that 73% of participants used multiple channels during their shopping journey compared to 20% of store-only shoppers and 7% of online-only shoppers. It also revealed that omnichannel retailers are more likely to retain customers. In fact, customers who had an omnichannel shopping experience took 23% more repeat shopping trips to the retailer's stores within 6 months after the purchase than those who shopped through a single channel. It is therefore not surprising that the State of Omnichannel Retail report (Brightpearl, 2017) found that 87% of retailers agree that omnichannel is crucial to their business success.

Despite the prevalence of omnichannel retail in the modern commerce landscape, few retailers have adopted the omnichannel approach from their inception. Instead, many

businesses have undergone a transition from traditional single-channel retailing to modern omnichannel retailing. Prominent examples include Walmart, a traditional B&M retailer, and Amazon, an e-commerce giant. Initially, Walmart focused on B&M stores while Amazon sold exclusively online; both companies have since expanded their presence in the digital and physical retail realms (Rooderkerk and Kök, 2019). In particular, Walmart has heavily invested in IT-related technology and digital infrastructure to create a seamless customer shopping experience (Banker, 2021). Likewise, Amazon has opened physical bookstores and acquired Whole Foods, including its hundreds of physical stores, in a massive $13.7 billion deal (Debter, 2017).

The recent COVID-19 pandemic has also contributed to the shift to omnichannel retailing. This crisis forced traditional B&M retailers to develop their online capabilities to stay afloat. Nowadays, it has become clear that there is no reverting to previous modes of operation once customers have become used to a seamless shopping experience (Barr, 2021). Meanwhile, as a result of gradual implementation of omnichannel practices, few retailers have fully exploited the potential of the omnichannel approach. One major retailer we interviewed managed three different sales channels (hypermarkets, convenience stores, and online platforms), yet they were not considering cross-channel purchasing behavior of customers when planning their business operations. Consequently, they were unable to optimize their product assortment within an omnichannel framework. This highlights the need for further development of methodologies that can facilitate decision-making in omnichannel environments.

While the literature on decision-making in the retail sector is vast, there is a lack of research specifically focused on omnichannel commerce. From a methodological perspective, accounting for several integrated retail channels presents a challenging task. Often, methods designed for single-channel settings cannot be readily applied in contexts involving multiple channels. This difficulty is especially pronounced in the case of omnichannel assortment optimization, which is one of major challenges faced by omnichannel retailers. Determining which products should be offered in which channels is exceedingly difficult due to the inherent complexity of omnichannel systems. Any change in the product assortment within one channel can propagate throughout the entire system, triggering demand shifts in all other channels. Such complexity necessitates comprehensive analysis and advanced analytical techniques to effectively account for customer shopping behavior in an omnichannel environment when making assortment decisions.

This thesis comprises three chapters that focus on different aspects of omnichannel assortment optimization. Each chapter stands as an independent paper, with one paper already published (Vasilyev et al., 2023), and two papers being in preparation. Importantly,

notation in each paper has to be considered in isolation. In the first paper, we introduce the multichannel attraction model (MAM), upon which we build a modeling framework aimed at addressing the complexities of omnichannel assortment optimization. In the second paper, we extend the framework developed in the first chapter in multiple key directions. Notably, we incorporate demand stochasticity and inventory management considerations into the assortment optimization problem under the MAM. In the third paper, we utilize the Ising model to address omnichannel assortment optimization while taking into account the basket shopping behavior of customers. Collectively, these three papers offer a comprehensive study of assortment optimization in omnichannel environments. The abstracts for each of these papers are as follows:

1. **Assortment optimization using an attraction model in an omnichannel environment.**

   Making assortment decisions is becoming an increasingly difficult task for many retailers worldwide as they implement omnichannel initiatives. Discrete choice modeling lies at the core of this challenge, yet existing models do not sufficiently account for the complex shopping behavior of customers in an omnichannel environment. In this paper, we introduce a discrete choice model called the multichannel attraction model (MAM). A key feature of the MAM is that it specifically accounts for both the product substitution behavior of customers within each channel and the switching behavior between channels. We formulate the corresponding assortment optimization problem as a mixed-integer linear program and provide a computationally efficient heuristic method that can be readily used for obtaining high-quality solutions in large-scale omnichannel environments. We also present three different methods to estimate the MAM parameters based on aggregate sales transaction data. Finally, we describe general effects of the implementation of widely-used omnichannel initiatives on the MAM parameters, and carry out numerical experiments to explore the structure of optimal assortments, thereby gaining new insights into omnichannel assortment planning. Our work provides the analytical framework for future studies to assess the impact of different omnichannel initiatives.

2. **Optimizing omnichannel assortments and inventory provisions using an attraction model.**

   Assortment optimization presents a complex challenge for retailers, as it depends on numerous decision factors. Changes in assortment can result in demand redistri-

bution with multi-layered consequences. This complexity is even more pronounced for omnichannel retailers, which have to manage assortments across multiple sales channels. Choice modeling has emerged as an effective method in assortment optimization, capturing customer shopping behavior and shifts in demand as assortments change. In this paper, we utilize the multichannel attraction model – a discrete choice model specifically designed for omnichannel environments – and generalize it for the case of a retailer managing both an online store and a network of physical stores. We integrate assortment decisions with optimal inventory decisions, assuming stochastic demand. Our model shows that overlooking the demand variability can result in suboptimal assortment decisions due to the demand pooling effect. We derive complexity results for the assortment optimization problem, which we formulate as a mixed-integer second-order cone program. We then develop two heuristic algorithms based on different relaxations of the formulated optimization problem. We establish the conditions under which the two relaxations are equivalent to each other and the conditions under which they are also equivalent to the initial problem. Furthermore, we conduct an extensive numerical analysis to provide managerial insights. We find that an increasing coefficient of variation of demand has a dual effect on optimal assortment sizes, initially causing a decrease in online assortment size due to rising costs, followed by an increase in online assortment size because of the demand pooling effect. Finally, we evaluate the potential benefits of omnichannel assortment optimization compared to assortment optimization in siloed channels.

3. **Omnichannel assortment optimization given basket shopping behavior.**

In markets where customers tend to purchase baskets of products rather than single products, assortment optimization is one of the major challenges for retailers. Removing a product from a retailer's assortment can result in a severe drop in aggregate demand if this product is a complement to other products. Accounting for the complementarity effect when making assortment decisions is especially challenging for omnichannel retailers, which have to manage product assortments across several channels. In this paper, we develop a modeling framework designed to tackle this problem. We model customer choices using a Markov random field – in particular, the Ising model – which captures pairwise demand dependencies as well as the individual attractiveness of each product. Using the Ising model allows us to leverage existing methodologies for various purposes including parameter estimation and efficient simulation of customer choices. We first consider a single-channel setting, in which we formulate the assortment optimization problem under this model and show that its decision version is NP-hard. We also provide several theoretical insights into the structure of the optimal assortments based on the graphical representation

of the Ising model, and develop a customized metaheuristic algorithm that can be used to obtain high-quality solutions to the assortment optimization problem. We then generalize our methodology and the theoretical results obtained for a single-channel setting to an omnichannel setting. Finally, we perform an extensive numerical analysis to gather insights into the properties of optimal assortments and evaluate the benefits of omnichannel assortment optimization.

1 Assortment Optimization Using an Attraction Model in an Omnichannel Environment

1.1   Introduction

Omnichannel retailing is a major trend in modern commerce. Its aim is to create a seamless customer shopping experience by integrating multiple retail channels with each other. One of the most common omnichannel initiatives is buy-online-and-pick-up-in-store (BOPS), also called click-and-collect. It allows customers to use online services to reserve a product for collection in a retail store. Other examples of omnichannel initiatives include providing online customers with in-store inventory availability information, and installing digital help desks in brick-and-mortar stores so that customers can readily access information about the retailer's online store, such as its assortment, prices and delivery options. With regard to supply chain management, one of the most prominent examples of channel integration is fulfilling a customer's online order from a local brick-and-mortar store to best leverage available inventories or respond in a timely fashion.

Modern consumers demand a variety of purchasing and delivery options, and retailers have to adapt their services to changes in customer expectations in order not to lose market share. For example, the rivalry between Amazon and Walmart induced the latter

to push the BOPS functionality in order to provide customers with an option comparable in convenience to Amazon's same-day delivery but without charging for shipping (Petro, 2020). According to the recent Global Shopper Trends Report by iVend Retail (2019), 81.4% of consumers reported using BOPS, which represents a growth of nearly 30% from the previous year's survey. The study of Sopadjieva et al. (2017), which is based on a survey of 46,000 customers who made a purchase between June 2015 and August 2016, found that 73% of participants used multiple channels during their shopping journey compared to 20% of store-only shoppers and 7% of online-only shoppers. It also revealed that omnichannel retailers are more likely to retain customers. In fact, customers who had an omnichannel shopping experience took 23% more repeat shopping trips to the retailer's stores within 6 months after the purchase than those who shopped through a single channel. It is therefore not surprising that the State of Omnichannel Retail report (Brightpearl, 2017) found that 87% of retailers agree that omnichannel is crucial to their business success.

However, implementing omnichannel strategies remains a difficult task for retailers, with estimating the effect of such strategies on demand being one of the key problems. Moreover, since an omnichannel environment is characterized by a high level of integration between retail channels, a change in assortment in one channel affects the demand across all channels, which makes assortment optimization extremely challenging. At the same time, omnichannel retailing is a rather recent research area with a relatively small number of analytical research papers published. In particular, there is a lack of works which consider the problem of demand and choice modeling in an omnichannel environment. Addressing this problem is an integral part of estimating the expected profit of a retailer, yet existing models do not account sufficiently and adequately for the complex nature of omnichannel shopping behavior. Our paper aims to fill this gap in the extant literature.

This paper makes the following contributions. We introduce a discrete choice model called the multichannel attraction model (MAM) that captures the complex shopping behavior of customers in an omnichannel environment. Importantly, the choice probabilities under the MAM are expressed through simple functional forms of the model parameters, making them easily interpretable. We prove that the assortment optimization problem under the MAM can be reformulated as a mixed-integer linear program. We propose a heuristic method to approximate its value in case of large-scale problems, and show numerically that our method is extremely efficient in terms of both computational performance and solution quality. We also present three different methods to estimate the MAM parameters based on sales history data, focusing on the case where only limited data are available. Next, we perform a sensitivity analysis of the model parameters in the two-channel case (with online and physical channels), which leads to insights into omnichannel assortment

optimization. For example, we show that a product with a relatively high unit profit in one of the channels may not be included in the corresponding optimal assortment, and vice versa – a product with a relatively low unit profit may be offered. We also analyze how the sizes of optimal assortments depend on the ratio of customers whose primary choice is to shop online to those whose primary choice is to go to a retail store, and on the proportion of customers willing to switch from one channel to another in case of absence of a certain product. We find that implementing the BOPS initiative can be unprofitable if the proportion of online customers using BOPS is too large compared to the additional traffic attracted to the offline channel. Finally, we demonstrate the benefits of omnichannel assortment optimization as opposed to optimizing siloed assortments in a multichannel environment.

The remainder of this paper is organized as follows: In Section 1.2, we review the two main streams of literature related to our research. In Section 1.3, we present a discrete choice model for omnichannel retailing, referred to as the MAM, and provide the intuition behind it. In Section 1.4, we formulate the corresponding assortment optimization problem as a mixed-integer linear program. An efficient heuristic method for solving the assortment optimization problem for very large numbers of products is provided in Section 1.5. The subsequent section is devoted to estimating the parameters of the MAM. In Section 1.7, we describe the impact of implementing widely-used omnichannel initiatives on the MAM parameters and present a numerical study which investigates the structure of optimal assortments. We summarize our contributions and discuss future research directions in Section 1.8.

1.2 Theoretical Background and Related Literature

1.2.1 Related Discrete Choice Models

Since discrete choice modeling is a vast and complex area of research, here we review only the works most relevant to our paper. A good introduction to discrete choice modeling can be found, for example, in Ben-Akiva and Lerman (1985) and Train (2002).

The multinomial logit model (MNL) formulated by McFadden (1973) is one of the most prominent discrete choice models. For clarity and introduction of notation, we provide a short formal description of the MNL. The choice probabilities under this model are derived as follows. Let $\mathcal{N} = \{1, 2, \ldots, n\}$ denote a set of products, $S \subseteq \mathcal{N}$ denote an offered set, and 0 denote the no-purchase alternative. A customer selects either one product from the offered set, or the no-purchase alternative. Each alternative $j \in \mathcal{N} \cup \{0\}$ has utility U_j, which is given by the sum of $\hat{U}_j$, a constant representing the known part

of the utility, and ξ_j, which is a Gumbel-distributed random variable representing the unobserved part of the utility. Random variables ξ_j are assumed to be independent and identically distributed (i.i.d.) for all the alternatives, and are commonly considered to be normalized so that their mean is zero and the variance is $\pi^2/6$. Furthermore, it is assumed that each customer selects the alternative with the highest utility among the available choices (the no-purchase alternative is available by default). Then, it can be shown that the probability of a customer selecting product j from the offered set S is

$$\pi_j(S) = \frac{e^{\hat{U}_j}}{e^{\hat{U}_0} + \sum_{k \in S} e^{\hat{U}_k}}. \tag{1.1}$$

The MNL can be viewed as a special case of the basic attraction model (BAM) developed by Luce (1959). Let v_j represent the "attractiveness" value of product j, and v_0 represent the attractiveness value of the no-purchase alternative. Under the BAM, the probability of a customer selecting product $j \in S$ is the ratio of the attractiveness value of product j to the sum of attractiveness values of all available alternatives, that is

$$\pi_j(S) = \frac{v_j}{v_0 + \sum_{k \in S} v_k}. \tag{1.2}$$

Clearly, the MNL choice probabilities (1.1) take the form (1.2) if we set the attractiveness value of each product j to $v_j = e^{\hat{U}_j}$, and the attractiveness value of the no-purchase alternative to $v_0 = e^{\hat{U}_0}$.

Gallego et al. (2014) proposed a generalization of the BAM called the general attraction model (GAM). As noted by the authors, the BAM may be too optimistic in estimating recapture probabilities as it ignores the possibility that a customer can choose to buy product $j \in \mathcal{N} \backslash S$ from another vendor or at a later time. For that reason, they modified formula (1.2) in the following way:

$$\pi_j(S) = \frac{v_j}{v_0 + \sum_{k \in S} v_k + \sum_{i \in \mathcal{N} \backslash S} w_i}, \tag{1.3}$$

where $w_i \in [0, v_i]$ represents the "shadow attractiveness" value of getting product $i \in \mathcal{N} \backslash S$ from another source. The meaning of the shadow attractiveness is that a customer does not consider the opportunity of buying a product somewhere else as long as it is present in the assortment. If the product is not available, however, a customer can decide to purchase it from another source. Note that under the GAM, the no-purchase probability is $\pi_0(S) = \frac{v_0 + \sum_{i \in \mathcal{N} \backslash S} w_i}{v_0 + \sum_{k \in S} v_k + \sum_{i \in \mathcal{N} \backslash S} w_i}$. Further note that the case $w_i = 0 \ \forall i \in \mathcal{N}$ results

in the BAM, and the case $w_i = v_i \; \forall i \in \mathcal{N}$ leads to the independent demand model as the choice probability of any product $j \in S$ does not depend on S.

Gallego et al. (2014) also formulated the sales-based linear program (SBLP) for network revenue management under the GAM. The decision variables in the SBLP are sales quantities rather than the offered set. In the case of infinite capacity and a single market segment, the SBLP takes the form of the assortment optimization problem under the GAM. Below we provide the formulation of this problem since we refer to it further in the text. Let r_j be the gross profit per unit of product j and Λ be the total number of customers. Then, the following linear program can be used to find the optimal assortment if the choice probabilities are given by (1.3):

$$\max_x \sum_{j \in \mathcal{N}} r_j x_j \tag{1.4a}$$

$$\text{s.t.} \; \frac{\tilde{v}_0}{v_0} x_0 + \sum_{j \in \mathcal{N}} \frac{\tilde{v}_j}{v_j} x_j = \Lambda, \tag{1.4b}$$

$$\frac{x_j}{v_j} - \frac{x_0}{v_0} \leq 0 \qquad \forall j \in \mathcal{N}, \tag{1.4c}$$

$$x_0, x_j \in \mathbb{R}_{\geq 0} \qquad \forall j \in \mathcal{N}, \tag{1.4d}$$

where x_j is the sales quantity of product j, $\tilde{v}_j = v_j - w_j$ and $\tilde{v}_0 = v_0 + \sum_{j \in \mathcal{N}} w_j$. Constraint (1.4b) is the balance constraint, and constraints (1.4c) are the scale constraints.

Furthermore, the authors noted that the GAM is the limit of the nested logit model in which customers first select a nest constructed from offerings of the same product by different vendors and then select a vendor which offers this product, while assuming that the dissimilarity parameter of products in each nest tends to zero. The nested logit model, introduced by Domencich and McFadden (1975), is another well-known discrete choice model where each choice probability can be decomposed into the product of two standard logit probabilities: the probability that a certain nest is chosen and the probability that a certain alternative is chosen given the nest.

Finally, it is worth mentioning the Markov chain choice model in which the product substitution behavior of customers is represented by transitions in a Markov chain. Blanchet et al. (2016) showed that such a model provides a simultaneous approximation to all random utility choice models including the MNL, the nested logit model and mixed MNLs. They also proved that this approximation becomes exact in the case of GAM choice probabilities. However, the number of parameters of the general Markov chain choice model is $(n + 1)^2$, where n is the total number of products, and, moreover, there are no interpretable functional forms for the choice probabilities as computing them for a

certain assortment requires a matrix inversion where the matrix entries depend on the assortment.

1.2.2 Choice Modeling and Assortment Optimization in Omnichannel Retailing

As mentioned earlier, the problem of choice modeling in an omnichannel environment is underrepresented in the literature. Since most of the research conducted on omnichannel is either purely empirical or qualitative, there is a considerable lack of works presenting analytical models. Most of the analytical papers focus on supply chain and inventory management questions rather than on discrete choice modeling or assortment optimization in an omnichannel environment. For example, Schneider and Klabjan (2013) investigated conditions under which common inventory control policies are optimal in the presence of two sales channels. He, Xu, et al. (2020) developed a newsvendor model that considers cross-channel product returns for a dual-channel retailer. Several papers studied the effect of ship-from-store operations on the optimal inventory policy (Seifert et al., 2006), fulfillment policy (Bayram and Cesaret, 2021), or both policies combined (Govindarajan et al., 2021). However, to the best of our knowledge, very few papers address at least one of the following questions, which we consider jointly in this paper: discrete choice modeling in the presence of *multiple* retail channels; the impact of omnichannel initiatives on demand allocation; and, ultimately, omnichannel assortment optimization.

Cao et al. (2016) proposed a theoretical framework to analyze the effect of adding the BOPS channel (called "online-to-store" channel) to existing sales channels on demand allocation. A major limitation of their study is that it considers a single-product setup. The authors use utility functions associated with different channels to model customers' channel choices, where each utility function is a linear function of the following parameters: the product value, the price per unit of the product in the corresponding channel, the delivery cost (in the case of the online channel), as well as certain inconvenience costs and factors. The paper shows that it may not be profitable to implement the BOPS functionality for some products depending on their characteristics.

Gao and Su (2016) presented a somewhat more complex approach to analyze the impact of the BOPS channel on demand allocation. Similar to the work of Cao et al. (2016), the authors model customers' channel choices using utility functions associated with different channels, but they also account for the cross-selling effect and inventory management considerations. One of the key findings of their paper confirms that not all products are well-suited for the BOPS functionality. However, their model does not account for the product substitution effect in the case when a product is removed from the assortment.

More recently, Harsha, Subramanian, and Ettl (2019) studied the pricing problem in an omnichannel environment with a chain of brick-and-mortar stores in the presence of other (online) channels. In their paper, the authors first consider a single-product setup where a customer only selects a source to buy the product from. It is assumed that the brick-and-mortar stores are located in geographically distributed zones, meaning that customers' choices in a zone do not depend on the parameters of other zones. At the zone level, each customer obtains a utility for choosing a channel depending on the product's price in this channel. The choices of customers are defined by a BAM where all attractiveness values are expressed through positive and strictly increasing functions of price. The price optimization problem is then formulated as a mixed-integer linear program. The authors also extend their analysis to a multiproduct setup using a nested attraction model where nests correspond to channels and each nest is comprised of products included in the channel assortment. For their extension, they provide a mixed-integer linear program to solve the price optimization problem approximately. In a related paper, Harsha, Subramanian, and Uichanco (2019) studied the omnichannel price optimization problem in a single-product setup, whilst accounting for both exogenous cross-channel fulfillment flows and inventory constraints.

With regard to multichannel assortment optimization for multiple products, Bhatnagar and Syam (2014) presented an integer program to determine the optimal item allocation for a hybrid retailer that manages both a chain of physical stores and an online store. They found that the retailer's profitability can be increased by removing products with high carrying costs from the physical stores and making them available exclusively online, thereby reducing the inventory carrying costs. However, their model relies on a number of strong assumptions, including that the demand for each product is a fixed parameter, meaning that the product demands do not depend on the assortment.

A different angle on multichannel assortment optimization was provided by Dzyabura and Jagabathula (2018). They studied the problem of determining the subset of products from the retailer's online channel to offer in the offline channel in order to maximize the aggregate revenue. It is assumed that each product is defined by a set of attributes, and there is a utility associated with each attribute which depends on whether the product is offered in the offline channel. The intuition behind this approach is that the utilities of the attributes change when customers learn about products by inspecting them in a brick-and-mortar store. Their choice model is the MNL based on the utilities of the attributes. The paper shows that accounting for the impact of the retailer's offline assortment on the online sales can lead to substantial gains in expected revenue.

Lo and Topaloglu (2022) addressed the same problem as Dzyabura and Jagabathula (2018) but in a different setup. The key difference between these is that the work of Dzyabura and

Jagabathula (2018) assumes that there exists a product for every potential combination of feature values, while in the model of Lo and Topaloglu (2022) it is assumed that the product portfolio can be characterized by a features tree, where each leaf corresponds to a product, and its ancestors are the features. The authors consider a mixture of customers: offline customers who shop in the physical store, and online customers who first visit the physical store to inspect the products offered there and then choose a product from the full assortment offered online. They showed that the assortment optimization problem in this setup is NP-hard, and leveraged the features tree structure to provide a fully polynomial time approximation scheme (FPTAS) based on dynamic programming that allows to determine approximately optimal assortments.

Finally, Hense and Hübner (2022) studied omnichannel assortment optimization while taking into account both in-channel and cross-channel demand substitution. In contrast to our work, the authors consider the exogenous demand (ED) model instead of leveraging discrete choice modeling techniques. In their approach, the base demands are assumed to be pairwise independent, and if one product is not available, then the proportion of the demand that is substituted by another product is given by a parameter. Such an approach has certain advantages because, in addition to finding optimal assortments, it also allows the authors to determine optimal shelf space and inventory levels across channels. However, the ED model does not allow to capture some of the complexity of customers' product substitution behavior. For instance, according to this model, it is assumed that if a product is not available then its demand would be shifted to another product in the same channel, but the associated demand is lost entirely if this other product is also not available. Moreover, the relatively large number of parameters in their proposed approach makes the problem of estimating the parameters from sales history data particularly challenging.

1.3 Multichannel Attraction Model

1.3.1 Model Formulation

In this subsection we present a discrete choice model that captures the complex customer shopping behavior in an omnichannel environment. Our proposed model is a generalization of the GAM to a setup where a retailer can sell products across several channels. Importantly, by generalizing the GAM our model inherits a number of desirable features. First of all, the GAM itself generalizes the basic attraction model (BAM) and the multinomial logit model (MNL), which is arguably the most widely used discrete choice model. At the same time, choice probabilities under the GAM are formulated using simple closed-form expressions. We can also leverage the concept of shadow attractiveness

since it has an interpretation that is easily adaptable to the case of an omnichannel retailer. We refer to the proposed model as the multichannel attraction model (MAM).

The MAM is most suitable for the case of a multichannel retailer that offers a range of substitutable products of which customers select at most one product. For instance, a sneakers subdivision of a large apparel retailer is a useful illustration to keep in mind for further reading. For tractability reasons, we consider a retailer with two channels: an online store and a physical store (or a chain of physical stores). However, our model can be generalized to the case of a larger number of channels (see Appendix A.1). The main idea behind the MAM is to develop a framework which allows to manage assortments in both channels jointly, taking into account customers who switch from one channel to another if certain products are unavailable. We therefore separate customers into two groups: the first group comprises customers whose primary choice is to purchase a product in the first channel if all products are available in both channels, whereas customers from the second group shop through the second channel under the same condition. For both groups of customers, we model their choices using our proposed generalization of the GAM, where each shadow attractiveness value is divided into two parts which determine how likely the customers are to switch to another channel to buy the corresponding product.

Let $\mathcal{N} = \{1, 2, \ldots, n\}$ denote a set of products which can be offered in both channels, and let 0 denote the no-purchase alternative. Furthermore, let $c \in \mathcal{C} = \{1, 2\}$ denote a channel index, $\bar{c} = \mathcal{C} \backslash \{c\}$ denote the other channel index, and $S_c \subseteq \mathcal{N}$ denote the set of products offered in channel c. By type-c customers we mean customers whose primary choice would be to shop in channel c if all products were available in both channels. For type-c customers, we use the following notation:

- $v_j^{(c)}$: attractiveness value of purchasing product $j \in S_c$ in channel c;
- $v_0^{(c)}$: attractiveness value of the no-purchase alternative;
- $u_i^{(c)} + w_i^{(c)}$: shadow attractiveness value of purchasing product $i \in \mathcal{N} \backslash S_c$ from another source (that is, either from channel $\bar{c}$ or from another retailer);
- $u_i^{(c)}/(u_i^{(c)} + w_i^{(c)})$: proportion of customers switching to channel $\bar{c}$ out of those willing to purchase product i outside of channel c (if $i \notin S_c$).

Superscript (c) indicates type-c customers' characteristics (who might shop in both channels), whereas subscript c refers to channel-specific features. Note that $u_i^{(c)}$ and $w_i^{(c)}$ are not in themselves attractiveness values. They provide an idea of how likely customers are to switch to another channel or to go to another retailer, but they do not represent utilities of different alternatives as such (see Subsection 1.3.2 for more details).

Also, similarly to the GAM, we assume that $u_j^{(c)} + w_j^{(c)} \in [0, v_j^{(c)}]$ for all products $j \in \mathcal{N}$, meaning that the shadow attractiveness value of purchasing product j from another source (including channel $\bar{c}$) does not exceed the attractiveness value of purchasing this product in channel c. This also implies that $\mathcal{N}$ is a set of substitutable products, and discarding a product from the assortment in channel c increases the demand for the remaining products generated by type-c customers. It is important to note that the demand substitution assumption is conventional for a setup where each customer purchases at most one product. The opposite effect – when discarding a product leads to a lower demand for some other products – is also possible in such a setup, e.g. if one product highlights the advantages of another product, thus creating a synergy. A recent example of a modeling framework devoted to assortment optimization in the presence of the product synergy effect can be found in Lo and Topaloglu (2019). However, accounting for this effect is outside the scope of our research.

Considering customers of two types associated with two channels is rational for two main reasons. Firstly, these two types of customers are likely to have noticeably distinct shopping preferences, which can be captured by different sets of parameters associated with different customer types. Secondly, the two flows of customers can differ considerably in volume (e.g., the number of customers associated with the online channel can be several times higher or lower than the one associated with the offline channel), and this has to be taken into account when making assortment decisions. Since the MAM is inherently a mixture of models where each model is associated with a customer type, it would be straightforward to extend the model by considering more types of customers. Nevertheless, in this work we focus on two types of customers in order to keep the model tractable.

We define the choice probabilities under the MAM given assortments in both channels in the following way. The probability that a type-c customer buys product j in channel c is

$$
\pi_{cj}^{(c)}(S_c) = \begin{cases} \dfrac{v_j^{(c)}}{v_0^{(c)} + \sum\limits_{k \in S_c} v_k^{(c)} + \sum\limits_{i \in \mathcal{N} \setminus S_c} (u_i^{(c)} + w_i^{(c)})} & \text{if } j \in S_c, \\ 0 & \text{otherwise;} \end{cases}
\tag{1.5}
$$

and the probability that a type-c customer buys product j in channel $\bar{c}$ is

$$
\pi_{\bar{c}j}^{(c)}(S_c, S_{\bar{c}}) = \begin{cases} \dfrac{u_j^{(c)}}{v_0^{(c)} + \sum\limits_{k \in S_c} v_k^{(c)} + \sum\limits_{i \in \mathcal{N} \setminus S_c} (u_i^{(c)} + w_i^{(c)})} & \text{if } j \in S_{\bar{c}} \setminus S_c, \\ 0 & \text{otherwise.} \end{cases}
\tag{1.6}
$$

It is straightforward to check that for each customer type, the sum of all choice probabilities (including the no-purchase probability) equals one.

Now, suppose that the total expected number of type-c customers is $\Lambda^{(c)} = \int_0^T \lambda^{(c)}(t)dt$, where $\lambda^{(c)}(t)$ is the arrival rate of type-c customers at time t, and T is the time horizon. Let $x_{cj}^{(c)}(S_c) = \pi_{cj}^{(c)}(S_c)\Lambda^{(c)}$ denote the expected number of type-c customers purchasing product j in channel c, and let $x_{\bar{c}j}^{(c)}(S_c, S_{\bar{c}}) = \pi_{\bar{c}j}^{(c)}(S_c, S_{\bar{c}})\Lambda^{(c)}$ denote the expected number of type-c customers purchasing product j in channel $\bar{c}$. Hereafter, we slightly abuse the notation by writing just $x_{cj}^{(c)}$ and $x_{\bar{c}j}^{(c)}$. Note that the overall probability that a customer (of any type) buys product j in channel c is as follows:

$$\pi_{cj}(S_c, S_{\bar{c}}) = \frac{x_{cj}^{(c)} + x_{cj}^{(\bar{c})}}{\Lambda^{(c)} + \Lambda^{(\bar{c})}}.$$

Importantly, we focus on the modeling setup with unlimited inventories, i.e., we do not consider the possibility of stockouts. As a result, the shopping behavior of customers is fully determined by assortments offered at the beginning of the sales period. Moreover, it also means that terms "demand" and "sales" can be used interchangeably. The assumption of unlimited inventories – which is common in the assortment planning literature – does not prevent our modeling framework from being applicable in a number of relevant and important practical situations. For example, this assumption is valid in the make-to-order setting, in which a company produces a product only after receiving an order and thus avoids carrying a lot of stock. It is also a valid assumption for companies which rarely have stockouts due to a high level of inventory.

1.3.2 Model Discussion

In this subsection, we provide the intuition behind the formulation of the MAM. Generally speaking, when it comes to omnichannel retailing, the product substitution behavior is not trivial, so our MAM requires a detailed description. At a high level, we assume that under the MAM, customers are subject to the following product substitution behavior. Suppose that if all products were available in all channels, a certain type-c customer would purchase product j in channel c. If this product is not offered in channel c, then the customer may either be determined to purchase product j anyway (potentially in channel $\bar{c}$), or decide to purchase another product $k \neq j$ instead, or leave without purchasing anything from this retailer. The outcomes of these three alternatives are summarized in Figure 1.1. In essence, if product j is not offered in channel c (i.e. $j \notin S_c$) and it is the first choice of a type-c customer, then this customer will either purchase another product $k \in S_c$, or a product $l \in S_{\bar{c}} \backslash S_c$, where $l \neq k$ but possibly $l = j$, or nothing at all.
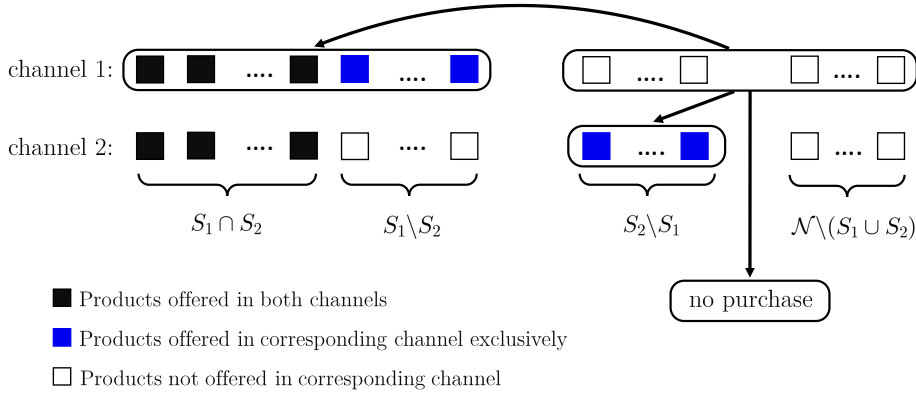
Figure 1.1: Product substitution pattern.

Let us consider the first two cases in more detail. First, suppose that the customer decides to stick with product j. In this case, the customer may be willing to search for product j in channel $\bar{c}$, or to go looking for this product somewhere else. To keep the model formulation tractable, we assume that if a customer decides to stick with a certain product and thus switches to another channel searching for it, then the product is always purchased if it is available in that channel, or else the customer leaves without purchasing anything from this retailer (e.g., goes to a competitor). Second, if the customer decides to purchase another product k, then there are two possibilities: If product k is offered in channel c, then the customer purchases it there; otherwise, the customer either continues the search in channel c, or looks for product k in channel $\bar{c}$, or looks for this product somewhere else. Similar to the previous case, we assume that if the customer is willing to switch to channel $\bar{c}$ in order to look for product k, then the product is always purchased if it is available there, otherwise the customer just leaves without purchasing anything.

Let us also provide additional insights into the structure of this product substitution behavior by showing the link between the MAM and Markov chain choice model (MCCM).

Proposition 1.1. *The MAM can be represented as a mixture of MCCMs where choices of type-c customers are characterized by the MCCM with the following parameters:*

$$
\begin{aligned}
&\lambda_{j_c}^{(c)} = v_j^{(c)}, \quad \lambda_{j_{\bar{c}}}^{(c)} = 0, \\
&\rho_{j_c i_c}^{(c)} = \frac{v_i^{(c)}(v_j^{(c)} - u_j^{(c)} - w_j^{(c)})}{v_j^{(c)} - v_j^{(c)}(v_j^{(c)} - u_j^{(c)} - w_j^{(c)})}, \\
&\rho_{j_c j_{\bar{c}}}^{(c)} = \frac{u_j^{(c)}}{v_j^{(c)}}, \quad \rho_{j_c i_{\bar{c}}}^{(c)} = 0, \\
&\rho_{j_c 0}^{(c)} = \frac{v_0^{(c)}(v_j^{(c)} - u_j^{(c)} - w_j^{(c)}) + u_j^{(c)} + w_j^{(c)}}{v_j^{(c)} - v_j^{(c)}(v_j^{(c)} + u_j^{(c)} + w_j^{(c)})} - \frac{u_j^{(c)}}{v_j^{(c)}},
\end{aligned}
\tag{1.7}
$$

$$\rho_{j_{\bar{c}}i_{\bar{c}}}^{(c)} = 0, \quad \rho_{j_{\bar{c}}j_c}^{(c)} = 0, \quad \rho_{j_{\bar{c}}i_c}^{(c)} = 0, \quad \rho_{j_{\bar{c}}0}^{(c)} = 1,$$

where j_c denotes product j in channel c, $\lambda^{(c)}$ is the vector of arrival probabilities, and $\rho^{(c)}$ is the matrix of transition probabilities.

The formal proof can be found in Appendix A.2. The intuition behind these expressions is as follows. We leverage the fact that the MAM restricted to type-c customers and products in channel c is equivalent to the GAM. Therefore, for such customers, the arrival probabilities of products in channel c as well as the transition probabilities between such products are defined by analogy to the probabilities that result in the GAM (see Blanchet et al., 2016). However, in contrast to the GAM, we split the transition probability from product j_c to the no-purchase alternative into two parts: one corresponds to choosing the no-purchase alternative directly, and the other one corresponds to first choosing product $j_{\bar{c}}$, and then – in case product j is not available in channel $\bar{c}$ – choosing the no-purchase alternative with probability 1. Note that we impose the constraint that type-c customers cannot purchase product $j_{\bar{c}}$ directly as they attempt to purchase product j_c first. This means, in turn, that type-c customers will never buy product j in channel $\bar{c}$ if this product is available in channel c. We believe that this assumption is not only reasonable, but also essential as it enables us to obtain simple analytical formulas (and thus achieve tractability and ensure interpretability of the model) for the choice probabilities, which would not be the case if we used a mixture of MCCMs in a general setting.

Let us also highlight the connection between the MAM and the random utility theory. First, note that the GAM is a random utility model (RUM), i.e., there exists a joint distribution of random utilities over a certain set of alternatives such that if each customer chooses an alternative with a maximum realization of utility, then the product choice probabilities are consistent with the GAM choice probabilities. Indeed, the GAM can be viewed as the nested logit model in the limit (see Gallego et al., 2014), and the nested logit model is a special case of the generalized extreme value (GEV) model, which is a RUM. Now since the GAM is a RUM, it is straightforward to show that for each customer type, there exists a joint distribution of random utilities over a certain set of alternatives that is consistent with the MAM choice probabilities. This can be done by using the fact that the condition of existence of a joint probability distribution of random utilities is equivalent to the condition of existence of a probability distribution over rankings of alternatives consistent with a given set of choice probabilities (see Block and Marschak, 1959). The MAM thus belongs to the class of RUMs and we can formulate the following proposition:

Proposition 1.2. *The MAM is a mixture of RUMs (one model per customer type), and as such it is also a RUM.*

The formal proof of this proposition can be found in Appendix A.3. However, note that our modelling approach is motivated by incorporating probabilistic cross-channel transitions into the GAM to account for the complex shopping behavior of customers in an omnichannel environment, rather than by proposing a new discrete choice model through specifying random utility of alternatives within a random utility maximization framework.

Finally, let us demonstrate the benefit of our formulation approach as opposed to a more traditional, utility-based way of generalizing the GAM to the omnichannel setting. As mentioned above, Gallego et al. (2014) showed that the GAM can be represented as a nested logit model in a limit. Following the same logic, one could have formulated the MAM so that it would emerge as a mixture of nested logit models (one model per customer type) in a limit, where each nest corresponds to a product and the dissimilarity parameter of each nest tends to zero. In this case, each nest would comprise three alternatives: purchasing the product in channel c, in channel $\bar{c}$, and from another source. Then, the choice probabilities would take the following form:

$$\pi_{cj}^{(c)}(S_c, S_{\bar{c}}) =$$

$$\begin{cases} \dfrac{v_j^{(c)}}{v_0^{(c)} + \sum\limits_{k \in S_c} v_k^{(c)} + \sum\limits_{i \in (\mathcal{N} \backslash S_c) \cap S_{\bar{c}}} \max\{u_i^{(c)}, w_i^{(c)}\} + \sum\limits_{l \in \mathcal{N} \backslash (S_c \cup S_{\bar{c}})} w_l^{(c)}} & \text{if } j \in S_c, \\ \\ 0 & \text{otherwise;} \end{cases}$$

and

$$\pi_{\bar{c}j}^{(c)}(S_c, S_{\bar{c}}) =$$

$$\begin{cases} \dfrac{u_j^{(c)}}{v_0^{(c)} + \sum\limits_{k \in S_c} v_k^{(c)} + \sum\limits_{i \in (\mathcal{N} \backslash S_c) \cap S_{\bar{c}}} \max\{u_i^{(c)}, w_i^{(c)}\} + \sum\limits_{l \in \mathcal{N} \backslash (S_c \cup S_{\bar{c}})} w_l^{(c)}} & \begin{array}{l} \text{if } j \in S_{\bar{c}} \backslash S_c \\ \text{and } u_j^{(c)} \geq w_j^{(c)}, \end{array} \\ \\ 0 & \text{otherwise.} \end{cases}$$

Importantly, in the above formulas the parameters $u_j^{(c)}$ and $w_j^{(c)}$ have a different interpretation compared to our model: for type-c customers, $u_j^{(c)}$ and $w_j^{(c)}$ here represent the shadow attractiveness of purchasing product j in channel $\bar{c}$ and elsewhere, respectively, whereas in our case $u_j^{(c)}$ is defined through the proportion of customers willing to purchase product j in channel $\bar{c}$ if it is not available in channel c (see Subsection 1.3.1).

However, this alternative formulation based on a mixture of nested logit models has important limitations. Indeed, it means that if $j \in S_{\bar{c}} \backslash S_c$, then either all type-$c$ customers

that choose nest j purchase product $j_{\bar{c}}$ (if $u_j^{(c)} > w_j^{(c)}$), or all of them leave the retailer (if $u_j^{(c)} < w_j^{(c)}$). Both these cases are rather extreme and hence not sufficiently realistic. There is also the special case when $u_j^{(c)} = w_j^{(c)}$, in which exactly half of the considered customers purchase product $j_{\bar{c}}$ and the other half leave the retailer, but this is also too restrictive. In contrast, our approach to formulating the MAM probabilities does not suffer from these limitations as it allows any partitioning of customers' choices between purchasing product $j_{\bar{c}}$ and leaving the retailer. We therefore believe that our approach is appealing because it is not only capable of more realistically representing omnichannel customer behaviour but it also provides for more flexibility without overcomplicating the choice model.

1.4 Assortment Optimization Problem

The assortment optimization problem under the MAM requires finding offer sets such that the total expected profit is maximized, that is determining the optimizers of the following problem:

$$\max_{S_c, S_{\bar{c}}} \sum_{c \in \mathcal{C}} \sum_{j \in \mathcal{N}} r_{cj} \pi_{cj}(S_c, S_{\bar{c}}), \tag{1.8}$$

where r_{cj} denotes the gross profit per unit of product j sold through channel c. This problem can be reformulated using binary variables. Let z_{cj} be a binary variable such that $z_{cj} = 1$ if $j \in S_c$, and 0 otherwise. Then, problem (1.8) can be written as follows:

$$\max_{S_c, S_{\bar{c}}} \sum_{c \in \mathcal{C}} \sum_{j \in \mathcal{N}} r_{cj} \pi_{cj}(S_c, S_{\bar{c}}) =$$

$$\max_{z_{cj}, z_{\bar{c}j}} \sum_{c \in \mathcal{C}} \sum_{j \in \mathcal{N}} \frac{r_{cj}}{\Lambda^{(c)} + \Lambda^{(\bar{c})}} \left(\frac{v_j^{(c)} \Lambda^{(c)} z_{cj}}{v_0^{(c)} + \sum_{k \in \mathcal{N}} v_k^{(c)} z_{ck} + \sum_{i \in \mathcal{N}} (u_i^{(c)} + w_i^{(c)})(1 - z_{ci})} + \right.$$

$$\left. \frac{u_j^{(\bar{c})} \Lambda^{(\bar{c})} z_{cj}(1 - z_{\bar{c}j})}{v_0^{(\bar{c})} + \sum_{k \in \mathcal{N}} v_k^{(\bar{c})} z_{\bar{c}k} + \sum_{i \in \mathcal{N}} (u_i^{(\bar{c})} + w_i^{(\bar{c})})(1 - z_{\bar{c}i})} \right).$$

This optimization problem is extremely difficult to solve as it includes binary decision variables and a nonlinear objective function. However, we can formulate an equivalent mixed-integer linear program (MILP) that can be solved using standard, off-the-shelf optimization software. Consider the following problem:

$$\max_x \sum_{c \in \mathcal{C}} \sum_{j \in \mathcal{N}} r_{cj} \left(x_{cj}^{(c)} + x_{cj}^{(\bar{c})} \right) \tag{1.9a}$$

$$\text{s.t.} \quad \frac{\tilde{v}_0^{(c)}}{v_0^{(c)}} x_{c0}^{(c)} + \sum_{j \in \mathcal{N}} \frac{\tilde{v}_j^{(c)}}{v_j^{(c)}} x_{cj}^{(c)} = \Lambda^{(c)} \qquad\qquad \forall c \in \mathcal{C}, \qquad\qquad (1.9b)$$

$$\frac{x_{cj}^{(c)}}{v_j^{(c)}} + \frac{x_{\bar{c}j}^{(c)}}{u_j^{(c)}} \leq \frac{x_{c0}^{(c)}}{v_0^{(c)}} \qquad\qquad \forall c \in \mathcal{C},\ j \in \mathcal{N}, \qquad (1.9c)$$

$$\frac{x_{c0}^{(c)}}{v_0^{(c)}} - \frac{x_{cj}^{(c)}}{v_j^{(c)}} \leq \frac{\Lambda^{(c)}}{\tilde{v}_0^{(c)}} (1 - z_{cj}) \qquad\qquad \forall c \in \mathcal{C},\ j \in \mathcal{N}, \qquad (1.9d)$$

$$x_{cj}^{(c)} \leq H_j^{(c)} z_{cj} \qquad\qquad \forall c \in \mathcal{C},\ j \in \mathcal{N}, \qquad (1.9e)$$

$$x_{\bar{c}j}^{(c)} \leq K_j^{(c)} x_{\bar{c}j}^{(\bar{c})} \qquad\qquad \forall c \in \mathcal{C},\ j \in \mathcal{N}, \qquad (1.9f)$$

$$x_{c0}^{(c)}, x_{cj}^{(c)}, x_{\bar{c}j}^{(c)} \in \mathbb{R}_{\geq 0} \qquad\qquad \forall c \in \mathcal{C},\ j \in \mathcal{N}, \qquad (1.9g)$$

$$z_{cj} \in \{0, 1\} \qquad\qquad \forall c \in \mathcal{C},\ j \in \mathcal{N}, \qquad (1.9h)$$

where $x = \left\{ x_{c0}^{(c)},\ x_{cj}^{(c)},\ x_{\bar{c}j}^{(c)}, z_{cj} \right\}_{c \in \mathcal{C}, j \in \mathcal{N}}$, $\tilde{v}_0^{(c)} = v_0^{(c)} + \sum_{i \in \mathcal{N}} \left(u_i^{(c)} + w_i^{(c)} \right)$, $\tilde{v}_j^{(c)} = v_j^{(c)} - \left(u_j^{(c)} + w_j^{(c)} \right)$, and constants $H_j^{(c)}$ and $K_j^{(c)}$ are given by

$$H_j^{(c)} = \frac{v_j^{(c)} \Lambda^{(c)}}{v_0^{(c)} + v_j^{(c)} + \sum_{i \in \mathcal{N} \backslash \{j\}} \left(u_i^{(c)} + w_i^{(c)} \right)}, \quad K_j^{(c)} = \frac{u_j^{(c)} \Lambda^{(c)}}{\tilde{v}_0^{(c)}} \bigg/ \frac{v_j^{(\bar{c})} \Lambda^{(\bar{c})}}{v_0^{(\bar{c})} + \sum_{k \in \mathcal{N}} v_k^{(\bar{c})}}.$$

We call this problem formulation the sales-based mixed-integer linear program (SBMILP) by analogy with the sales-based linear program presented by Gallego et al. (2014) for the GAM. Constraints (1.9b) are similar to the balance constraint in the SBLP, and constraints (1.9c) are modified scale constraints. However, due to the multichannel structure of the MAM and hence a more complex product substitution behavior of customers, we need additional constraints with binary variables. The meaning of each constraint as well as the equivalence of problems (1.8) and (1.9) becomes evident from the proof of the following Theorem (see Appendix A.4):

Theorem 1.1. *The SBMILP is a valid formulation of the assortment optimization problem under the MAM, that is, the optimal value of problem (1.9) is equal to the optimal value of problem (1.8) multiplied by the constant $(\Lambda^{(c)} + \Lambda^{(\bar{c})})$.*

Remark 1.1. *It is straightforward to verify that constraints (1.9d), (1.9e) and (1.9f) cannot be tightened, that is, the constant coefficients on the right-hand side of these constraints cannot be reduced.*

Remark 1.2. *The SBMILP can easily be modified to incorporate additional constraints. For example, if there is a cost a_{cj} associated with product j offered in channel c, and the total cost induced by products offered in this channel is limited by the upper bound L_c,*

then the following constraint has to be added to the SBMILP:

$$\sum_{j \in \mathcal{N}} a_{cj} z_{cj} \leq L_c. \tag{1.10}$$

If channel c is the physical channel, then constraint (1.10) can be viewed as a shelf-space constraint where a_{cj} represents the shelf space required for product j to be offered in channel c, and L_c is the total shelf space in that channel.

Adding the shelf space constraint described in Remark 1.2 to the SBMILP is of particular practical importance (typically omnichannel retailers cannot offer all products in the physical channel due to limited shelf space), but has implications in terms of computational complexity:

Proposition 1.3. *The assortment optimization problem represented by the shelf-space-constrained SBMILP, which is given by adding constraint (1.10) to the SBMILP formulation (1.9), is NP-hard.*

The proof can be found in Appendix A.5. It is important to note that there is another modification of the assortment optimization problem under the MAM that not only makes the modified problem NP-hard, but may also suggest that the original SBMILP is NP-hard too. In particular, if assortments in both channels have to be the same (i.e. $S_1 = S_2$), then this problem is equivalent to the assortment optimization problem under a mixture of two GAMs in a single channel. Such a problem is NP-hard as a generalization of the assortment optimization problem under a mixture of two MNL models, which has been shown by Rusmevichientong et al. (2014) to be NP-hard. This well-known result, formulated for the simplest illustrative case of the assortment optimization problem under a mixture of discrete choice models, strongly indicates that the original SBMILP formulation (1.9) is also NP-hard since the MAM itself is essentially a mixture of discrete choice models (one model per customer type). While an interesting problem, formally establishing NP-hardness of the original SBMILP is left for future research.

If the assortment in one of the channels is fixed and equals $\mathcal{N}$ (i.e., all products are offered), then we can build upon some of the results obtained for the SBLP in Gallego et al. (2014) and establish certain analytical properties of the optimal assortment in the other channel. Let $R^{(c)}(S_c, S_{\bar{c}})$ be the total profit generated by type-c customers given assortments S_c and $S_{\bar{c}}$. Suppose that the assortment in channel $\bar{c}$ is fixed so that $S_{\bar{c}} = \mathcal{N}$, thus no customers of type $\bar{c}$ will switch to channel c. We can then formulate the following proposition:

Proposition 1.4. *Let $S_{\bar{c}} = \mathcal{N}$ and, without loss of generality, suppose that all products*

are sorted in descending order of the ratio $(r_{cj}v_j^{(c)} - r_{\bar{c}j}u_j^{(c)})\Lambda^{(c)}/\tilde{v}_j^{(c)}$. Then, the optimal assortment in channel c is given by

$$
z_{cj} = \begin{cases} 1 & \text{if } j \leq m, \\ 0 & \text{otherwise,} \end{cases} \tag{1.11}
$$

where $m = \max\{j \in \mathcal{N} : R^{(c)}(\{1, \ldots, j\}, \mathcal{N}) < (r_{c,j+1}v_{j+1}^{(c)} - r_{\bar{c},j+1}u_{j+1}^{(c)})\Lambda^{(c)}/\tilde{v}_{j+1}^{(c)}\}.$

The proof can be found in Appendix A.6. The intuition behind this finding is clear. Even if the gross profit per unit of a certain product j in channel c is high and the product has a high attractiveness value (i.e., the demand for product j in channel c is high compared to other products), it may not be profitable to include this product in the channel c assortment.

Further developing the idea behind the proof of Proposition 1.4, we can formulate the following property of the optimal assortment in both channels:

Proposition 1.5. *Let $(S_c, S_{\bar{c}})$ be the optimal combination of assortments, and suppose that $k \in S_c$. Then $j \in S_c$ as well if $F_c(j, S_c, S_{\bar{c}}) \geq F_c(k, S_c, S_{\bar{c}})$, where*

$$
F_c(j, S_c, S_{\bar{c}}) = \frac{r_{cj}v_j^{(c)} - r_{\bar{c}j}u_j^{(c)}\mathbb{1}_{j \in S_{\bar{c}}}}{\tilde{v}_j^{(c)}}
$$
$$
+ \frac{\left(v_0^{(c)} + \sum_{k \in S_c \cup \{j\}} v_k^{(c)} + \sum_{i \in \mathcal{N} \setminus (S_c \cup \{j\})} (u_i^{(c)} + w_i^{(c)})\right)r_{cj}u_j^{(\bar{c})}\Lambda^{(\bar{c})}/\Lambda^{(c)}\mathbb{1}_{j \in \mathcal{N} \setminus S_{\bar{c}}}}{\tilde{v}_j^{(c)}\left(v_0^{(\bar{c})} + \sum_{k \in S_{\bar{c}}} v_k^{(\bar{c})} + \sum_{i \in \mathcal{N} \setminus S_{\bar{c}}} (u_i^{(\bar{c})} + w_i^{(\bar{c})})\right)}.
$$

$$\tag{1.12}$$

This property, the proof of which is provided in Appendix A.7, could serve as a base for a heuristic algorithm that allows to determine a high-quality solution if solving the SBMILP is computationally infeasible. In particular, $F_c(j, S_c, S_{\bar{c}})$ could be approximated with an expression that depends only on product j and channel-$\bar{c}$ assortment, which would allow for the approximate characterization of the optimal assortment in channel c given the assortment in channel $\bar{c}$. Then, the heuristic algorithm could take form of an iterative procedure, where in each iteration the assortment in one of the channels is determined given the assortment in the other channel. In this work, however, we focus on the development of a heuristic algorithm based on a relaxation approach for situations in which directly solving the SBMILP is computationally challenging.

1.5 Heuristic Method

The SBMILP formulation (1.9) has $2n$ binary variables and $4n$ constraints containing binary variables. For large values of n, this problem may become more challenging to solve. Let us therefore consider the LP relaxation of the SBMILP derived from formulation (1.9) by removing binary variables z_{cj} together with the corresponding constraints (1.9d) and (1.9e). Importantly, preliminary numerical experiments showed that a solution to the relaxed problem satisfies the removed constraints for almost all $c \in \mathcal{C}$, $j \in \mathcal{N}$, and this is a key observation underlying our heuristic. We propose the following two-step algorithm:

1. Solve the relaxed problem. Let $\big\{\hat{x}_{c0}^{(c)},\ \hat{x}_{cj}^{(c)},\ \hat{x}_{\bar{c}j}^{(c)}\big\}_{c \in \mathcal{C}, j \in \mathcal{N}}$ be its optimal solution, and $\mathcal{J}^{(c)}$ be the set of indexes $j \in \mathcal{N}$ such that either $\dfrac{\hat{x}_{cj}^{(c)}}{v_j^{(c)}} = \dfrac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}}$ or $\hat{x}_{cj}^{(c)} = 0$.

2. Solve problem (1.9) with the following additional constraints:

$$z_{cj} = \mathbb{1}_{\hat{x}_{cj}^{(c)} > 0} \quad \forall c \in \mathcal{C},\ j \in \mathcal{J}^{(c)}, \tag{1.13}$$

   where $\mathbb{1}$ is the indicator function. The obtained solution is the heuristic output.

To gain insights into the computational performance of the proposed heuristic method, we generate the parameters of the omnichannel assortment optimization problem in the following way:

(i) $r_{1j} = u(0,1) + \varepsilon$, $r_{2j} = r_{1j}(1 + u(0, 0.5))\ \forall j \in \mathcal{N}$;

(ii) $v_j^{(1)} = u(0,1) + \varepsilon$, $v_j^{(2)} = u(0,1) + \varepsilon\ \forall j \in \mathcal{N} \cup \{0\}$;

(iii) $u_j^{(1)} = u(0, 0.5)v_j^{(1)}$, $u_j^{(2)} = u(0, 0.5)v_j^{(2)}\ \forall j \in \mathcal{N}$;

(iv) $w_j^{(1)} = u(0, 0.5)v_j^{(1)}$, $w_j^{(2)} = u(0, 0.5)v_j^{(2)}\ \forall j \in \mathcal{N}$,

where $u(a, b)$ denotes a value sampled from the uniform distribution $\mathcal{U}(a,b)$, and $\varepsilon = 0.01$. Also, we normalize the attractiveness values so that $v_0^{(c)} + \sum_{j \in \mathcal{N}} v_j^{(c)} = 1\ \forall c \in \mathcal{C}$. While the generated values of the parameters may not necessarily be representative of a real-world example, they are well-suited for the purpose of evaluating the computational effort required to solve the problem instance. The only meaningful restriction we impose is that the gross profit per unit of a product is higher for the online channel than for the offline channel, which can be justified by the difference in holding costs. Lastly, we fix the values of parameters $\Lambda^{(1)}$ and $\Lambda^{(2)}$ at 10^4 and $3 \cdot 10^4$, respectively.

We use the SBMILP as a benchmark to evaluate the comparative performance of our
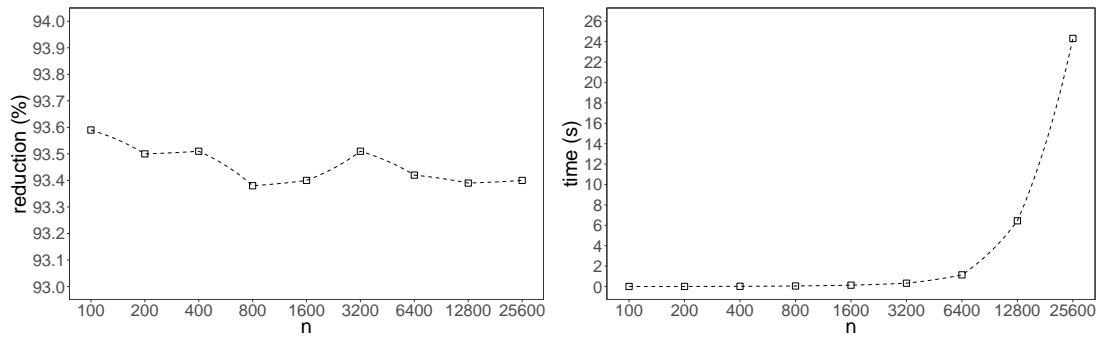
(a) Reduction in the number of binary variables.

(b) Solving time of the heuristic method.

Figure 1.2: Computational performance of the heuristic method as a function of the product set size (n).

heuristic method. The computational study was carried out on a laptop with Intel Core i7-8650U CPU (1.90 GHz), 8 GB RAM and 64-bit Windows 10 OS. To solve our MILPs we used Gurobi (version 8.0.1). We ran 100 experiments for each value of $n \in \{100, 200, 300\}$. The results showed that, on average, the SBMILP can be solved to optimality in 0.29 seconds if $n = 100$, in 1.48 seconds if $n = 200$, and in 10.72 seconds if $n = 300$. However, running 100 experiments for $n = 400$ turned out to be not particularly feasible, as some instances took up to 500 seconds to solve (even though some other instances were solved in less than 7 seconds). This highlights the strong need for a computationally efficient heuristic method to solve the assortment optimization problem.

Our proposed heuristic method allows to drastically reduce both the number of binary variables and the number of constraints containing binary variables in the SBMILP formulation. This is because each binary variable z_{cj} such that $j \in \mathcal{J}^{(c)}$ turns into a parameter. The numerical results in terms of reduction in the number of binary variables and solving time for different values of n are shown in Figure 1.2(a) and (b), respectively (note that the x-axis is log-scaled with base 2). The results are averaged over 100 generated instances. It can be seen that the reduction in the number of binary variables is more than 93% and this is independent of the actual value of n. As a consequence, the optimization problem (1.9) with additional constraints (1.13) is computationally much easier to solve than the original one. The heuristic still requires solving a MILP whose size grows linearly in n, so the solving time grows exponentially in n. However, even if n is of order 10^4, the heuristic solution can be found in a matter of seconds, making this method attractive for most practical applications.

For moderate values of n – in the range from 50 to 300 – we also compared the output and the solving time of the heuristic to those of the SBMILP. The results, which again represent averages over 100 generated instances, are given in Table 1.1. As can be

n	ratio of solving times	ratio of optimal values	proportion of mismatched decisions
50	8.95	0.999984	0.003400
100	39.76	0.999975	0.003950
150	86.17	0.999979	0.003400
200	138.61	0.999960	0.003925
250	218.32	0.999974	0.003540
300	595.31	0.999966	0.004050

Table 1.1: Comparative performance of the heuristic method.

seen, on average, the profit yielded by the heuristic is around 0.99997 of the optimal profit, and the proportion of mismatched assortment decisions (i.e., the proportion of z_{cj} values which are different for the heuristic output and for the optimal output) is almost always less than 0.004. At the same time, the ratio of the solving time of the original formulation to the solving time of the heuristic is around 9 if $n = 50$, and around 595 if $n = 300$. These results demonstrate that the developed heuristic method yields close-to-optimal assortments whilst generally being substantially superior in terms of solving time, especially as n grows.

Finally, for large values of n, we verified numerically that our heuristic method yields close-to-optimal solutions. For each $n \in \{5000, 10000, 15000, 20000, 25000\}$, we generated 100 random MAM instances. Then, we compared the heuristic outputs with the solutions to the linear relaxations of the corresponding SBMILPs, which provide upper bounds on the true optimal values. For each problem instance, we computed the value of $(obj_{rel} - obj_{heur})/obj_{heur}$, where obj_{rel} is the value of the linear relaxation of the SBMILP and obj_{heur} is the value of the heuristic. As can be seen in Figure 1.3, the value of these gaps never exceeded 0.2%, with the average gap between 0.15% and 0.16%, confirming the high solution accuracy of the proposed heuristic.

Note that since the MAM is a regular choice model – i.e., adding a product to an assortment in one of the channels cannot lead to an increase in the probability of customers choosing any other product – the revenue-ordered heuristic algorithm (see, e.g., Berbeglia and Joret, 2020) can be used to solve the assortment optimization problem under the MAM. In particular, if all products (in both channels simultaneously) are ranked in ascending order of the price, and the heuristic solution is determined by finding the best cutoff in this ranking, then this solution approximates the optimum revenue within a factor of $\frac{1}{1+ln(r_{max}/r_{min})}$, where r_{max} and r_{min} are the maximum and the minimum price, respectively. This method can prove to be very useful if prices are sufficiently close to
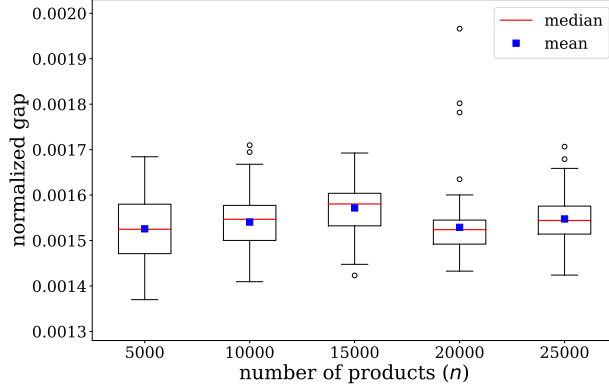
Figure 1.3: Heuristic performance bounds for large numbers of products.

each other. By contrast, the quality of our proposed heuristic method, which is very promising in terms of computational time and solution accuracy, does not rely on the closeness of prices of products, meaning it is readily applicable in a general setting.

1.6 Parameter Estimation

We start with describing a basic method to estimate the parameters of the MAM. Recall that $\pi_{cj}(S_c, S_{\bar{c}})$ is the probability that a customer buys product j in channel c given assortments S_c and $S_{\bar{c}}$ in channels c and $\bar{c}$, respectively. Suppose that for all $i, j \in \mathcal{N}$, $c \in \mathcal{C}$ the probabilities $\pi_{cj}(\mathcal{N}, \mathcal{N})$, $\pi_{cj}(\mathcal{N}\backslash\{i\}, \mathcal{N})$ and $\pi_{cj}(\mathcal{N}, \mathcal{N}\backslash\{i\})$ are known along with the respective values of $\Lambda^{(c)}$ and $\Lambda^{(\bar{c})}$. Such information can be obtained from data comprising the aggregate demand values (or, equivalently, the aggregate sales, assuming that there are no shortages) for the corresponding assortments together with the shares of customers of each type who chose the no-purchase alternative if all the products from $\mathcal{N}$ had been offered in both channels. The latter means that the ratios $v_0^{(c)}/\sum_{j \in \mathcal{N}} v_j^{(c)}$ $\forall c \in \mathcal{C}$ are estimated exogenously. While the assumption of having these exogenous estimates is fairly restrictive, it aligns with the existing literature (Vulcano et al., 2012).

Without loss of generality, we can assume $v_0^{(c)} + \sum_{j \in \mathcal{N}} v_j^{(c)} = 1$ $\forall c \in \mathcal{C}$. For all $c \in \mathcal{C}$, $j \in \mathcal{N}$, the MAM parameters can then be determined using the following expressions:

$$v_j^{(c)} = \pi_{cj}(\mathcal{N}, \mathcal{N}) \frac{\Lambda^{(c)} + \Lambda^{(\bar{c})}}{\Lambda^{(c)}}, \quad v_0^{(c)} = 1 - \sum_{j \in \mathcal{N}} v_j^{(c)}, \tag{1.14a}$$

$$u_j^{(c)} = \frac{\pi_{\bar{c}j}(\mathcal{N}\backslash\{j\}, \mathcal{N}) - \pi_{\bar{c}j}(\mathcal{N}, \mathcal{N})}{\pi_{ck}(\mathcal{N}\backslash\{j\}, \mathcal{N})} v_k^{(c)}, \tag{1.14b}$$

$$w_j^{(c)} = \frac{\pi_{ck}(\mathcal{N}, \mathcal{N})}{\pi_{ck}(\mathcal{N}\backslash\{j\}, \mathcal{N})} + v_j^{(c)} - u_j^{(c)} - 1, \tag{1.14c}$$

where $k \in \mathcal{N}$ is any product different from j. Expressions (1.14) can be verified by straightforward calculations. This method, however, requires specific information on the product demands for a certain set of assortments, which may be difficult to obtain in practice. Therefore, there is a need for more general parameter estimation methods for situations in which only limited data about product demands is available. The two most common parameter estimation techniques for discrete choice models are maximum likelihood estimation (MLE) and least squares estimation. In their recent empirical study, Berbeglia, Garassino, et al. (2021) compared the estimation results produced by these two standard estimation techniques and, considering a range of prominent discrete choice models, found that the quality of estimates of these two methods is very similar. Importantly, least squares estimation requires only aggregate sales data (i.e., how many units of each product were sold during each period with a fixed assortment), whereas MLE is typically used when the information about all individual sales transactions is available. In this work, we focus on the limited setting of aggregate sales transaction data (i.e. only sales quantities and product availability in each period are observed) in which direct maximization of the (log-)likelihood function is computationally unappealing (see Vulcano et al., 2012), so we resort to an estimation based on least squares which is readily applicable in incomplete data situations.

Suppose that product demands which arise from the MAM are observed in both channels for T periods. In other words, we implicitly assume customers make choices according to an MAM in a homogeneous market (i.e. preferences of customer are homogeneous across the selling horizon, meaning their choice behaviour can be modelled by a single MAM). Note that although this assumption is standard in the estimation of discrete choice models, it can be straightforwardly relaxed (see, for example, the discussion in Vulcano et al. (2012)). For each period t, we denote the assortment in channel c by S_{ct} and the observed demand for product j in channel c by d_{cjt}. We also assume that the demand rate is constant for both channels, i.e. $\Lambda_t^{(c)} = \Lambda^{(c)} \ \forall t \in \{1, \ldots, T\}$ and that the market size is sufficiently large (i.e. observed historical sales are representative of expected sales). If the available data is composed solely of the demand values and the corresponding assortments, we can obtain estimates of the MAM parameters by minimizing the sum of squared residuals, i.e., by solving the following optimization problem:

$$
\min_{\substack{v^{(c)}, \Lambda^{(c)} \\ u^{(c)}, w^{(c)}}} \sum_{t=1}^{T} \sum_{c \in \mathcal{C}} \sum_{j \in \mathcal{N}} \Bigg(\frac{v_j^{(c)} \Lambda^{(c)} \mathbb{1}_{j \in S_{ct}}}{v_0^{(c)} + \sum_{k \in S_{ct}} v_k^{(c)} + \sum_{i \in \mathcal{N} \setminus S_{ct}} (u_i^{(c)} + w_i^{(c)})} +
$$

$$
\frac{u_j^{(\bar{c})} \Lambda^{(\bar{c})} \mathbb{1}_{j \in S_{ct} \setminus S_{\bar{c}t}}}{v_0^{(\bar{c})} + \sum_{k \in S_{\bar{c}t}} v_k^{(\bar{c})} + \sum_{i \in \mathcal{N} \setminus S_{\bar{c}t}} (u_i^{(\bar{c})} + w_i^{(\bar{c})})} - d_{cjt} \Bigg)^2 \qquad (1.15a)
$$

$$\text{s.t. } u_j^{(c)} + w_j^{(c)} \leq v_j^{(c)} \qquad\qquad \forall c \in \mathcal{C}, \ j \in \mathcal{N}, \quad (1.15\text{b})$$

$$v_0^{(c)} + \sum_{j \in \mathcal{N}} v_j^{(c)} = 1 \qquad\qquad \forall c \in \mathcal{C}, \quad (1.15\text{c})$$

$$\Lambda^{(c)}, v_0^{(c)}, v_j^{(c)}, u_j^{(c)}, w_j^{(c)} \in \mathbb{R}_{\geq 0} \qquad\qquad \forall c \in \mathcal{C}, \ j \in \mathcal{N}. \quad (1.15\text{d})$$

Importantly, despite the fact that problem (1.15) is nonconvex, some off-the-shelf solvers are able to cope with this problem quite well. In particular, the IPOPT package developed by Wächter and Biegler (2006) for nonlinear optimization shows a surprisingly good performance. Following Vulcano et al. (2012) and Gallego et al. (2014), we illustrate the performance of the above parameter estimation method by considering an exemplary setup with $n = 5$ and $T = 15$. We simulate 100 instances of demand arising from the MAM with fixed parameters (which are randomly generated in the way described in Section 1.5), with each instance corresponding to a set of randomly simulated assortments S_{ct}, $c \in \mathcal{C}$, $t \in \{1, \ldots, T\}$. We consider two cases: in the first case, we assume that the values of the ratios $v_0^{(c)} / \sum_{j \in N} v_j^{(c)}$ $\forall c \in \mathcal{C}$ (or, equivalently, the values of $v_0^{(c)}$ if $\sum_{j \in \mathcal{N}} v_j^{(c)} = 1$) are given exogenously, whereas no such information is available in the second case. The least squares estimates obtained by solving problem (1.15) and averaged over 100 instances are presented in Table 1.2. It can be seen that these estimates are particularly close to the true parameter values if the values of $v_0^{(c)}$ are known, which highlights the importance of information availability. Indeed, if a firm has access to accurate exogenous estimates of the attractiveness of the no-purchase option (e.g., by keeping track of the no-purchase outcomes), then the accuracy of parameter estimation is shown to improve dramatically.

An alternative way to estimate the MAM parameters is to build upon the Expectation Maximization (EM) algorithm which was developed by Vulcano et al. (2012) for estimating the parameters of the BAM when only aggregate sales data are available, which makes the standard MLE approach extremely computationally challenging. Their algorithm was later adapted by Gallego et al. (2014) to estimate the parameters of the GAM. The idea behind these algorithms is to estimate the model parameters iteratively using estimates of the first-choice demand. For large-scale problems, such an approach can be more effective than solving the least squares problem. However, similar to the algorithm presented by Gallego et al. (2014), the method based on an adaptation of the EM algorithm to the MAM suffers from an important limitation: its convergence is not theoretically guaranteed. We provide a detailed description and performance examples of this method in Appendix A.8.

		$v_j^{(c)}$			$u_j^{(c)}$			$w_j^{(c)}$			$\Lambda^{(c)}$	
	j	True	Estimated ($v_0^{(c)}$ known)	Estimated ($v_0^{(c)}$ unknown)	True	Estimated ($v_0^{(c)}$ known)	Estimated ($v_0^{(c)}$ unknown)	True	Estimated ($v_0^{(c)}$ known)	Estimated ($v_0^{(c)}$ unknown)	True	Estimated ($v_0^{(c)}$ known)	Estimated ($v_0^{(c)}$ unknown)
channel 1	1	0.068	0.068	0.083	0.013	0.018	0.021	0.027	0.028	0.033			
	2	0.145	0.146	0.173	0.045	0.045	0.058	0.031	0.042	0.055			
	3	0.233	0.227	0.304	0.106	0.106	0.131	0.017	0.044	0.054	10000	10124	11991
	4	0.096	0.102	0.117	0.047	0.050	0.057	0.039	0.031	0.038			
	5	0.221	0.219	0.255	0.035	0.029	0.036	0.095	0.101	0.133			
channel 2	1	0.105	0.106	0.122	0.021	0.022	0.025	0.046	0.046	0.059			
	2	0.199	0.194	0.249	0.068	0.067	0.081	0.093	0.087	0.121			
	3	0.195	0.197	0.248	0.081	0.081	0.092	0.083	0.084	0.098	30000	29884	23254
	4	0.236	0.234	0.313	0.061	0.062	0.070	0.088	0.086	0.129			
	5	0.008	0.013	0.013	0.000	0.004	0.003	0.001	0.004	0.005			

Table 1.2: Least squares estimates of the MAM parameters.

1.7 Impact of Implementation of Omnichannel Initiatives and Sensitivity Analysis

1.7.1 General Effects of Omnichannel Initiatives

In this subsection, we consider widely used omnichannel initiatives and discuss general effects of their implementation on the MAM parameters. Importantly, we do not account for implementation and maintenance costs associated with these initiatives – our goal is to estimate and explore the demand evolution. In the following, we assume that channel 1 is the retailer's offline channel (i.e., a physical store or a chain of physical stores) and channel 2 is the online channel.

The most straightforward effect takes place when customers have in-store access to information about the availability of online inventory, e.g., through in-store digital help desks. In this case, values of $u_j^{(1)}$, $j \in \mathcal{N}$ have to increase as ratios $u_j^{(1)}/(u_j^{(1)} + w_j^{(1)})$ determine the probability that customers are willing to switch from the offline to online channel when looking for the desired product, and offline customers have an additional incentive to check the online assortment. In other words, if product j is not available in the offline channel but it is available in the online channel, then the demand for product j in the online channel is expected to increase due to the fact that more customers switch from the offline to online channel when looking for this product. At the same time, one can expect that the shadow attractiveness values of purchasing products from another source, $u_j^{(1)} + w_j^{(1)}$, should remain constant for all $j \in \mathcal{N}$ since there should be no impact on choices of customers who select products which are offered in the offline channel. That is, the increase in demand for products in the online channel caused by this omnichannel initiative takes place on account of type-1 customers who would otherwise go to a different retailer. However, note that parameters $u_j^{(1)}$ may not increase considerably because nowadays customers have an option to access the retailer's online store using their smartphones. It is also reasonable to assume that the effects on all other parameters is negligible as no additional customers are drawn to the store, and customers willing to buy something from the in-store assortment are not likely to switch to the online channel looking for other products. Therefore, the benefits of this initiative may be outweighed by its implementation costs. It would thus be interesting for future studies to compare the additional revenue generated by increased cross-channel demand with the implementation costs of this initiative through real-world case studies.

We now investigate the effect of BOPS, which is arguably the most prevalent omnichannel initiative. We assume that BOPS orders are fulfilled from the physical store inventory. Otherwise, if a customer is only allowed to pick up a product after it is delivered

from a warehouse to the store, there is no substantial difference between regular online transactions and BOPS transactions. The main distinction is in the delivery cost – if the delivery is carried out by the retailer and it is not paid separately for by customers, then the delivery cost in the latter case should at least not be higher than in the former because deliveries to the store can be organized in batches. Thus, the major issue for the retailer is to compare the benefits generated by attracting new online customers through introducing the BOPS functionality with the associated implementation costs (including those for adjusting the supply chain). This is, however, outside the scope of our research.

Within the framework of our research, it is more interesting to study the effect of BOPS if orders have to be fulfilled from the physical store inventory. This is also a common practice for omnichannel retailers, usually due to the need to have items ready for collection shortly after the order was placed (see Gallino and Moreno, 2014). In general, the overall traffic of customers should increase because customers who do not want to wait for a delivery have an additional convenient way to receive a product. However, even if we do not consider implementation costs, introducing BOPS can be unprofitable. Despite counting purchases made through BOPS as online transactions, the empirical analysis carried out by Gallino and Moreno (2014) revealed that the introduction of the BOPS functionality generally leads to a reduction in online sales and an increase in offline sales. They explained this phenomenon by the impact of sharing reliable inventory availability information on customers' decisions. If it is guaranteed that a certain product is available in-store, customers may choose to go to the store (even without reserving the product using BOPS) rather than order it online. A similar effect can take place if the retailer provides online information about the current stock level of each product in each store. As a consequence, the retailer may lose part of its online customers while attracting more in-store customers instead. This can lead to losses under the assumption that the gross profits per unit of each product in the online are higher than those in the offline channel.

Based on these assertions, we can describe the general effects of implementing the BOPS functionality on the MAM parameters as follows: Firstly, one can expect an increase in the expected number of customers visiting at least one of the retail channels, $\Lambda^{(1)} + \Lambda^{(2)}$, together with a decrease in the number of online customers $\Lambda^{(2)}$. Secondly, since online customers now have information about the in-store assortment, parameters $u_j^{(2)}$ should go up, while the shadow attractiveness values of purchasing products from another source, $u_j^{(2)} + w_j^{(2)}$, should remain constant (similar to the impact of in-store information about online inventory availability). Note that in this case, we count purchases made through BOPS as in-store purchases.

A more rigorous way to study the effect of BOPS on customer choices would be to introduce a separate channel for BOPS transactions, and suitably adjust the parameters

related to other channels. This approach would also allow us to formulate an optimization problem for finding optimal assortments in all three channels, thereby determining the products for which it is profitable to implement the BOPS functionality. Future research should focus on studying the impact of this and other omnichannel initiatives in order to gain a better understanding of, and novel insights into the profitability of adopting such initiatives.

1.7.2   Numerical Analysis and Managerial Insights

For the first part of our numerical analysis, we generate the MAM parameters as described in Section 1.5 and keep them fixed throughout the numerical experiments. More specifically, we consider the case $n = 30$ in order to be able to clearly visualize the optimal assortments. In the below figures, we indicate products that belong to the optimal assortment of both channels by black squares, products that belong to one optimal channel assortment exclusively by blue squares, and the remaining products – i.e. those not offered at all – by white squares (similar to Figure 1.1). Each figure consists of two plots, with the plot(s) on the left- and right-hand side corresponding to channel 1 (offline) and channel 2 (online), respectively.

It is important to recall that the generated parameter values are convenient for carrying out this analysis, but they are not necessarily realistic. For example, one can expect that if the gross profit per unit of a certain product j is high compared to other products, then the proportion of the demand for product j to the total demand in the corresponding channel is probably small, i.e., there is a negative correlation between the values of r_{cj} and $v_j^{(c)}$. Moreover, we assume that the values of r_{cj} and $v_j^{(c)}$ are uniformly distributed, which is unlikely to be the case in practice. However, our goal is to study the relationship between the parameters values and the optimal solution to the SBMILP, rather than to investigate a real-world case study, and using such generated data ideally fits this purpose.

First, let us explore the relation between the gross profit per unit of each product and the optimal assortments given by the optimal solution of the SBMILP (see Figure 1.4). We observe that the higher the gross profit per unit of a product, the more likely it is that this product belongs to the optimal assortment. However, such a relation is not always evident, i.e., a product with a relatively high unit profit in one of the channels may not be included in the corresponding assortment, whereas a product with a relatively low unit profit may be offered. Moreover, the relationship between unit profits and optimal assortments seems to be more pronounced in channel 2 than in channel 1. It can possibly be explained by the fact that we set the total number of type-2 customers to be
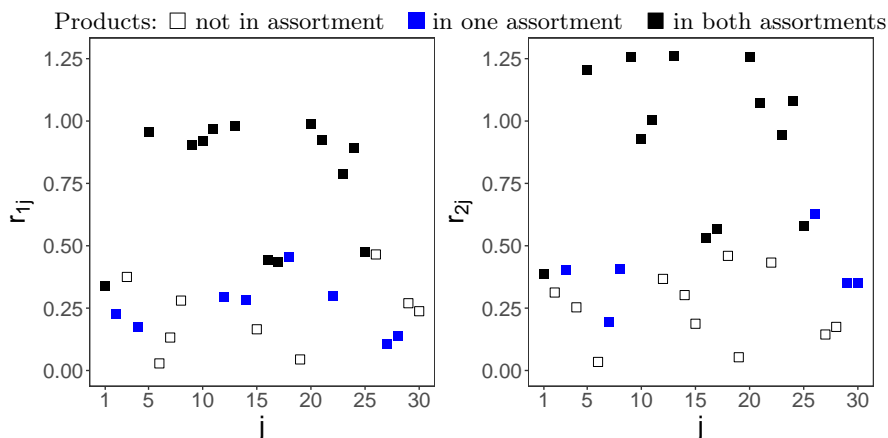
Figure 1.4: Relation between the optimal assortments and r_{cj} for channels 1 (left) and 2 (right).

considerably higher than that of type-1 customers ($\Lambda^{(1)} = 10^4$ and $\Lambda^{(2)} = 3 \cdot 10^4$), which could reflect a situation where channel 2 represents an online channel. Also, note that if a product is not included in the assortment in one channel, and the profit per unit of the product is low in both channels, it can still be profitable to offer this product in the other channel due to demand generated by customers who switch to that channel and those who shop there in the first place.

Next, we create a similar visualization but for the attractiveness values $v_j^{(c)}$ instead of the gross profit values, with the results shown by Figure 1.5. Unlike Figure 1.4, it can be observed that there is no apparent relation. To further analyze the structure of optimal assortments, recall the finding formulated in Proposition 1.5. Based on $F_c(j, S_c, S_{\bar{c}})$, we can derive a more tractable expression that approximately characterizes the optimal assortment in one channel given the assortment in the other channel. Indeed, consider the following expression:

$$
f_{cj}(S_{\bar{c}}) = \begin{cases} \dfrac{r_{cj}v_j^{(c)} - r_{\bar{c}j}u_j^{(c)}}{\tilde{v}_j^{(c)}} & \text{if } j \in S_{\bar{c}}, \\[4mm] \dfrac{r_{cj}v_j^{(c)} + r_{cj}u_j^{(\bar{c})}\Lambda^{(\bar{c})}/\Lambda^{(c)}}{\tilde{v}_j^{(c)}} & \text{otherwise.} \end{cases} \tag{1.16}
$$

The values of $f_{cj}(S_{\bar{c}})$ $\forall c \in \mathcal{C}$, $j \in \mathcal{N}$ for the optimal assortments are illustrated by Figure 1.6a. We observe that unlike the attractiveness values, these expressions do provide an approximate characterization of the optimal assortments. Indeed, as can be seen, if $f_{cj}(S_{\bar{c}}) > f_{ck}(S_{\bar{c}})$ and product k belongs to the optimal assortment in channel c, then product j tends to belong to the optimal assortment as well. In the following, we

Products: □ not in assortment ■ in one assortment ■ in both assortments
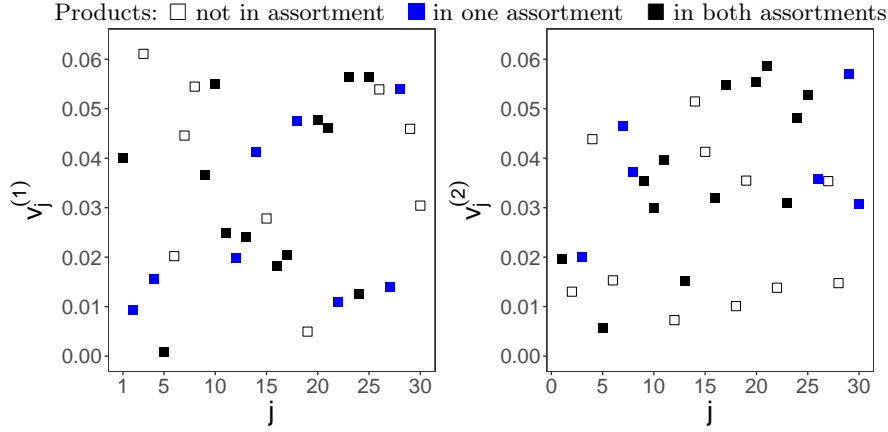
Figure 1.5: Relation between the optimal assortments and $v_j^{(c)}$ for channels 1 (left) and 2 (right).

use the values of f_{cj} to produce more illustrative plots.

To study how the optimal assortments are being affected by a change in the ratio $\Lambda^{(1)}/\Lambda^{(2)}$, we consider two additional cases: $\Lambda^{(1)} = \Lambda^{(2)} = 10^4$ and $\Lambda^{(1)} = 3 \cdot 10^4$, $\Lambda^{(2)} = 10^4$. The results are summarized in Figure 1.6b and 1.6c. Interestingly, it can be seen that with an increase in the ratio $\Lambda^{(1)}/\Lambda^{(2)}$ from 1/3 to 1, the optimal assortment in channel 1 becomes smaller in size, whereas the optimal assortment in channel 2 becomes larger. However, with a further increase in the ratio $\Lambda^{(1)}/\Lambda^{(2)}$ from 1 to 3, the optimal assortments in both channels do not change. To sum up, the smaller the ratio of customers whose primary choice is to shop online to those whose primary choice is to go to a retail store, the smaller the optimal assortment in the physical channel and the larger the assortment in the online channel (up to a certain limit). This is intuitively clear given that an online purchase is generally more profitable for the retailer than the in-store purchase of the same product, so the retailer is interested in a high online traffic and therefore increases the online assortment.

We now investigate the effect of the values of $u_i^{(c)}/(u_i^{(c)} + w_i^{(c)})$ on the optimal assortments. These ratios determine how willing customers are to switch to another channel when looking for the desired product. We therefore fix the values of all parameters except for the values of $u_j^{(c)}$ and $w_j^{(c)}$, and also set $u_j^{(c)} + w_j^{(c)} = 0.5v_j^{(c)}$. We then consider the following three cases: $u_j^{(c)} = 0.01v_j^{(c)}$, $u_j^{(c)} = 0.25v_j^{(c)}$ and $u_j^{(c)} = 0.49v_j^{(c)}$ $\forall c \in \mathcal{C}, j \in \mathcal{N}$. The optimal assortments are displayed in Figure 1.7. It is interesting to note that a similar effect to the one described above for different values of $\Lambda^{(1)}/\Lambda^{(2)}$ can be observed when increasing the ratio $u_j^{(c)}/(u_j^{(c)} + w_j^{(c)})$: given a fixed sum $u_j^{(c)} + w_j^{(c)}$ $\forall c \in \mathcal{C}, j \in \mathcal{N}$, the optimal assortment in channel 1 becomes smaller in size, whereas the optimal assortment in channel 2 becomes larger. In other words, the larger the proportion of customers willing

Products: □ not in assortment ■ in one assortment ■ in both assortments

(a) Optimal assortments if $\Lambda^{(1)} = 10^4$, $\Lambda^{(2)} = 3 \cdot 10^4$.

(b) Optimal assortments if $\Lambda^{(1)} = \Lambda^{(2)} = 10^4$.

(c) Optimal assortments if $\Lambda^{(1)} = 3 \cdot 10^4$, $\Lambda^{(2)} = 10^4$.
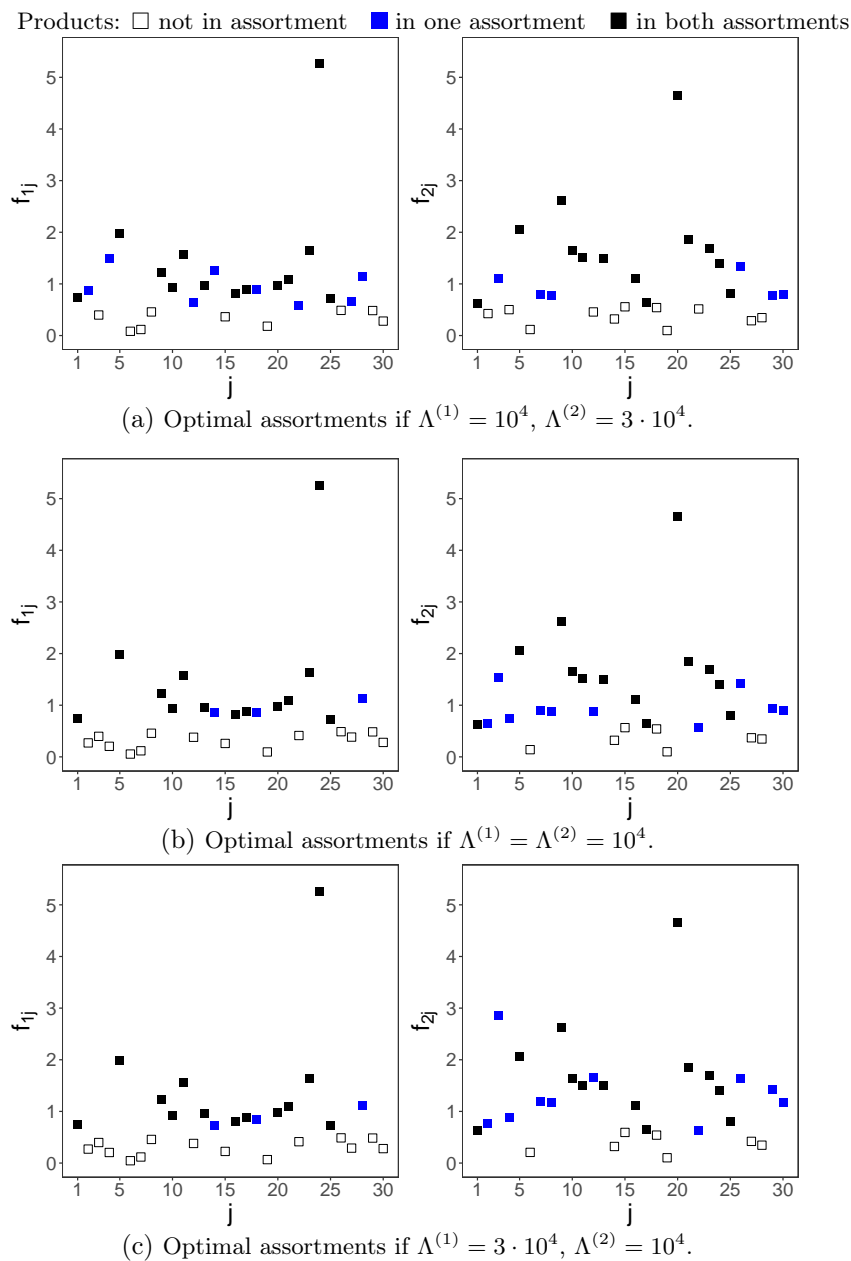
Figure 1.6: Optimal assortments for different values of $\Lambda^{(1)}/\Lambda^{(2)}$ for channels 1 (left) and 2 (right).

to switch from one channel to another in case of absence of a certain product, the larger the optimal assortment in the online channel and the smaller the optimal assortment in the physical channel. Trivially, in the case when 100% of in-store customers are willing to switch to the online channel if their primary-choice product is not available, the optimal assortment in the physical channel is the empty set.

Products: □ not in assortment ■ in one assortment ■ in both assortments

(a) The optimal assortments if $u_j^{(c)} = 0.01 v_j^{(c)}$, $w_j^{(c)} = 0.49 v_j^{(c)}$ $\forall c \in \mathcal{C}$, $j \in \mathcal{N}$.

(b) The optimal assortments if $u_j^{(c)} = 0.25 v_j^{(c)}$, $w_j^{(c)} = 0.25 v_j^{(c)}$ $\forall c \in \mathcal{C}$, $j \in \mathcal{N}$.

(c) The optimal assortments if $u_j^{(c)} = 0.49 v_j^{(c)}$, $w_j^{(c)} = 0.01 v_j^{(c)}$ $\forall c \in \mathcal{C}$, $j \in \mathcal{N}$.
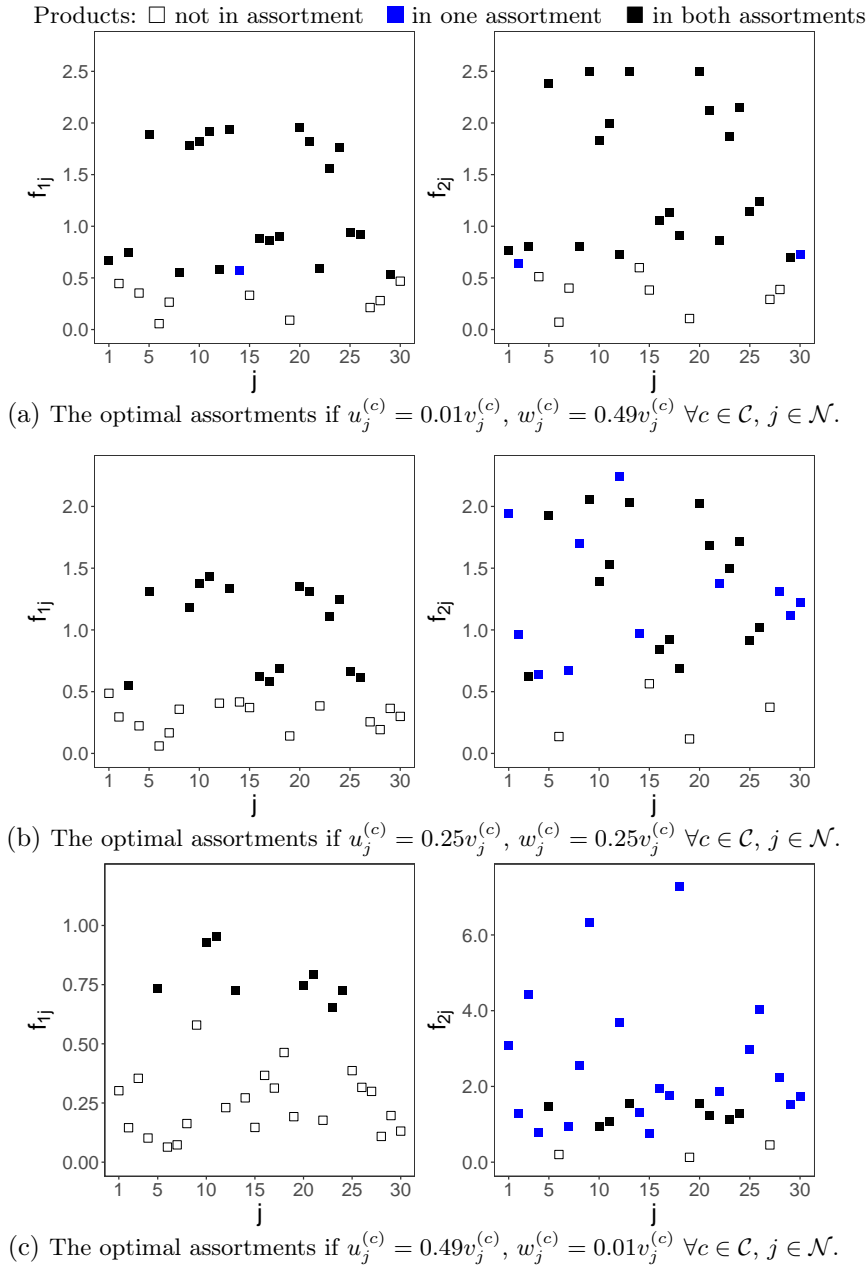
Figure 1.7: Optimal assortments for different values of $u_j^{(c)}$ and $w_j^{(c)}$ for channels 1 (left) and 2 (right).

Next, we investigate the extent to which the implementation of the BOPS functionality affects the total profit of the retailer. To this end, we use the initially generated values of parameters $u_j^{(c)}$ and $w_j^{(c)}$ $\forall c \in \mathcal{C}$, $j \in \mathcal{N}$. Recall that the impact of introducing the BOPS functionality on the MAM parameters can be described approximately as follows: the sum $\Lambda^{(1)} + \Lambda^{(2)}$ increases, $\Lambda^{(2)}$ decreases, and $u_j^{(2)}$ increases under the condition that the sum $u_j^{(2)} + w_j^{(2)}$ does not change $\forall j \in \mathcal{N}$. We carried out several numerical experiments

change in $\Lambda^{(1)}$	change in $\Lambda^{(2)}$	change in $u_j^{(2)}$	objective value
–	–	–	24025.16
7500	−6000	–	23489.87
9000	−6000	–	24133.95
10500	−9000	–	22362.52
12000	−9000	–	22931.03
13500	−9000	–	23499.54
13500	−9000	$0.2w_j^{(2)}$	23514.09
13500	−9000	$0.5w_j^{(2)}$	23582.26
13500	−9000	$0.8w_j^{(2)}$	23723.54
15000	−9000	–	24068.05

Table 1.3: Different impacts of the BOPS implementation on total profit.

considering different degrees of such impact and report results in Table 1.3. Based on these results, the following key observations can be made. The implementation of the BOPS functionality can be unprofitable if the proportion of online customers using this option is too large compared to the additional traffic attracted to the offline channel. Furthermore, the primary effect on total profit is due to the change in the parameters $\Lambda^{(1)}$ and $\Lambda^{(2)}$, whereas the effect of an increase of $u_j^{(2)}$-values is less pronounced. This is not surprising since $u_j^{(2)}$-values only reflect the cross-channel demand volume generated by customers switching from channel 2 to channel 1, while $\Lambda^{(1)} + \Lambda^{(2)}$ represents the total number of customers visiting the retailer. Ultimately, our results indicate that the profitability of adopting the BOPS functionality needs to be evaluated on a case-by-case basis, which is in line with the empirical findings of Gallino and Moreno (2014).

Unlike the above analysis, which was carried out for illustrative purposes for one problem instance, we now analyze the effect of the considered parameter changes on both the size of assortments and the degree of assortment overlap for a large number of problem instances (simulated in the same way). First, we simulate 1000 problem instances, and for each problem instance, we fix all the parameters except for $\Lambda^{(1)}$ and $\Lambda^{(2)}$. We then consider several cases with different values of the $\Lambda^{(1)}/\Lambda^{(2)}$-ratio. Note that the absolute values of $\Lambda^{(1)}$ and $\Lambda^{(2)}$ are irrelevant since the objective function in the SBMILP can be divided by the constant $\Lambda^{(2)}$ (leaving the optimal solution unchanged), which makes it a function of $\Lambda^{(1)}/\Lambda^{(2)}$. Subsequently, we solve the SBMILP in each considered case and record the size of assortments in both channels and the degree of assortment overlap. The aggregated results are shown in Figure 1.8(a) (note that the x-axis is log-scaled). We
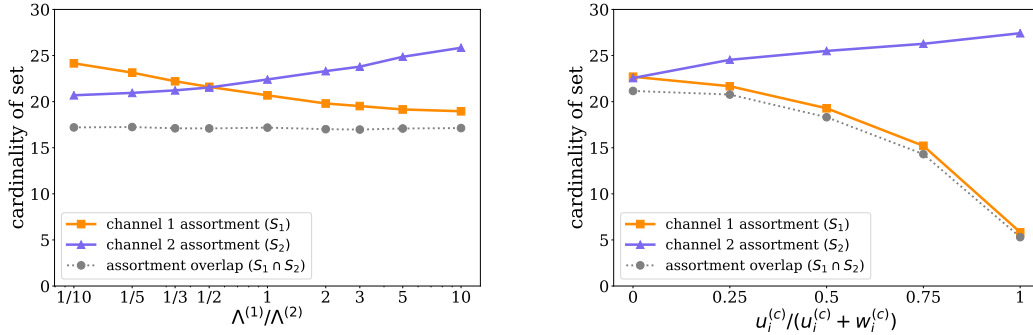
(a) Impact of ratio of numbers of different customer types.

(b) Impact of proportions of customers switching to another channel.

Figure 1.8: Effect of ratio of physical to online traffic (left) and of proportions of customers switching channels (right) on both the average size of assortments and the degree of assortment overlap.

observe that the average assortment sizes follow the same trends as the ones described earlier for one problem instance: with an increase in the ratio $\Lambda^{(1)}/\Lambda^{(2)}$, on average the size of assortment in channel 1 (physical channel) decreases while the size of assortment in channel 2 (online channel) increases. Interestingly, we also observe that on average, the degree of assortment overlap remains virtually constant. This can be explained by fact that the trend lines are almost symmetric, that is the absolute values of their slopes are very close to each other.

Next, we study the effect of the values of $u_i^{(c)}/(u_i^{(c)} + w_i^{(c)})$ on both the size of assortments and the degree of assortment overlap. As previously, we simulate 1000 problem instances. For each problem instance, we fix all parameters except for u and w. Similar to the analysis of one problem instance, we set $u_i^{(c)} + w_i^{(c)} = 0.5v_i^{(c)}$ and consider several cases with the ratio $u_i^{(c)}/(u_i^{(c)} + w_i^{(c)})$ taking values from 0 to 1. The results are summarized by Figure 1.8(b). We can see that if $u_i^{(c)}/(u_i^{(c)} + w_i^{(c)})$ tends zero – that is if there is no cross-channel demand – then the average sizes of assortments in both channels are the same. This is not surprising because if there is no cross-channel demand, then each assortment can be optimized independently. Since the parameters related to each channel are simulated randomly, on average the assortment sizes are virtually identical, i.e. $|S_1| \approx |S_2|$. Also, we can observe that on average, an increase in $u_i^{(c)}/(u_i^{(c)} + w_i^{(c)})$ leads to an increase in the size of channel 1 assortment ($|S_1|$) and a decrease in the size of channel 2 assortment ($|S_2|$). This is consistent with our previous observations for one problem instance. Lastly, we observe that the line representing the average overlap degree closely follows the line representing the average size of channel 1 assortment – in other words, there are hardly any products that belong to channel 1 assortment but
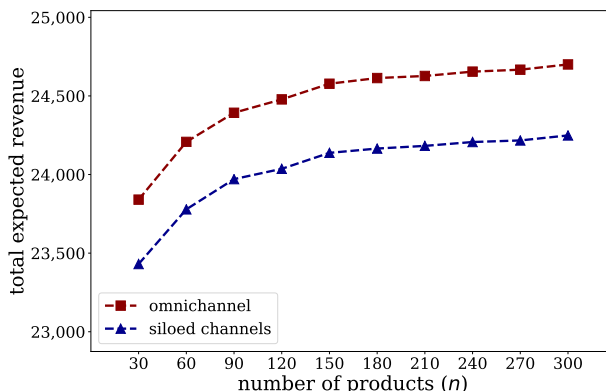
Figure 1.9: Profitability of omnichannel assortment optimization and assortment optimization in siloed channels.

not to channel 2 assortment. This might be explained by the fact that since channel 2 assortment comprises almost all products, the few products that are not part of it are characterized by very low unit gross profits. In this case, their unit gross profits are also low in channel 1 (due to the parameter simulation procedure), making them also less likely to be part of the optimal assortment in that channel.

Finally, let us evaluate the revenue benefits of solving the omnichannel assortment problem as opposed to optimizing the two assortments in siloed channels. To this end, we consider several problem sizes with the number of products, n, ranging from 30 to 300. For each n, we simulate 1000 problem instances in the way described at the beginning of Section 1.5. Then, for each problem instance, we compute two revenues: one that corresponds to omnichannel assortment optimization; and one that corresponds to assortment optimization in siloed channels. The former is obtained by solving the SBMILP, whereas the latter is determined in the following way: First, we solve the assortment optimization problem under the GAM for each channel separately, using the fact that the MAM restricted to choices of type-c customers in channel c is equivalent to the GAM. Once assortments in the two channels are identified, we compute the corresponding total expected revenue assuming that the customers make their choices according to the MAM. Importantly, by doing this we do not neglect the cross-channel demand – even though the assortments are optimized in siloed channels, the underlying demand model is the same as in the omnichannel case. Figure 1.9 shows that there is a clear benefit of solving the omnichannel assortment optimization problem as opposed to optimizing assortments in siloed channels. In fact, for the n-values under consideration, the omnichannel solution turned out to between 1.8% and 1.9% more profitable on average than the solution obtained for siloed channels, with maximum improvements of up to

6.7%, highlighting the considerable revenue gains that may be achieved by omnichannel assortment planning.

1.8 Conclusions and Future Work

In this paper, we have developed an analytical framework for both modeling product demand and making assortment decisions in an omnichannel environment. In particular, we have introduced a discrete choice model referred to as the multichannel attraction model (MAM) that specifically accounts for the complex nature of omnichannel shopping behavior. Compared to the single-channel setup – which corresponds to the general attraction model (GAM) – the multichannel structure of our choice model substantially increases the complexity of associated problems. For example, the assortment optimization problem under the GAM can be formulated as a linear program, and the optimal assortment can actually be found analytically. For the MAM, on the other hand, we have formulated the sales-based mixed-integer linear program (SBMILP) – a tight MILP formulation of the corresponding assortment optimization problem – and proved that the optimal assortment in one channel can be found analytically if all products are available in the remaining channel. We have proposed a computationally efficient heuristic method to approximately solve the SBMILP, and showed numerically that its output is extremely close to the optimal solution. We have also presented three different methods to estimate the parameters of the MAM, and demonstrated that if product demands are only known for a limited number of assortments in each channel, the MAM parameters can be estimated fairly accurately by simply solving a least squares problem.

We have analyzed general effects of the implementation of widely-used omnichannel initiatives on the MAM parameters, and have carried out numerical experiments to investigate the structure of optimal assortments. We demonstrated that in an omnichannel environment, optimal assortments cannot be characterized by a single factor alone as they are affected by a combination of several factors. In fact, in our experiments we identified a relation between the sizes of optimal assortments and the following two factors: the ratio of customers whose primary choice is to shop online to those whose primary choice is to go to a retail store, and the proportion of customers willing to switch from one channel to another in case of absence of a certain product. We also showed numerically that implementing the buy-online-and-pick-up-in-store (BOPS) initiative is not always profitable, which supports previous findings in the omnichannel literature. Finally, we evaluated the benefits of omnichannel assortment optimization as opposed to optimizing siloed assortments and showed that the former can result in substantial revenue gains for omnichannel retailers. Our numerical analysis indicated that the omnichannel solution

is 1.8%-1.9% more profitable on average than the solution obtained for siloed channels, with maximum gains of up to 6.7%. These findings are encouraging and demonstrate that our framework can be beneficially used for omnichannel assortment planning as well as for exploring the profitability of implementing different omnichannel initiatives, which should be of great interest to decision-makers in the retailing industry.

The proposed framework can be the basis for a range of important further developments. Firstly, a better understanding of the structure of optimal assortments in each channel could be obtained through an extensive numerical analysis based on values of the MAM parameters estimated using a real-world dataset. Secondly, additional theoretical results related to the assortment optimization problem under the MAM could be derived, with the question of NP-hardness of the SBMILP being of particular interest. In a similar vein, the possibility of developing an FPTAS (e.g. along the lines of the recent work of Désir et al. (2022)) may be explored. Also, the MAM could be subject to multiple extensions. For example, the MAM could be calibrated on the product features level. In this case, it would be very interesting to compare the results obtained using such a modified version of the MAM to those obtained by Dzyabura and Jagabathula (2018) and Lo and Topaloglu (2022). Another promising research direction would be to formulate and explore a stochastic version of the SBMILP for the case of uncertain demand in each channel. It would also be beneficial to study the effects of BOPS on customer choices by introducing a separate channel for BOPS transactions. Such an approach would allow to investigate not only whether but also for which specific products it is profitable to implement the BOPS initiative. Finally, deriving analytical properties of solutions to the SBMILP should lead to additional managerial insights and result in a more efficient SBMILP formulation.

2 Optimizing Omnichannel Assortments and Inventory Provisions Using an Attraction Model

2.1 Introduction

Omnichannel retail has emerged as the new norm in today's commerce landscape. However, few retailers have adopted the omnichannel approach from their inception. Instead, many businesses have undergone a transition from traditional single-channel retailing to modern omnichannel retailing. Prominent examples include Walmart, a traditional brick-and-mortar (B&M) retailer, and Amazon, an e-commerce giant. Initially, Walmart focused on B&M stores while Amazon sold exclusively online; both companies have since expanded their presence in the digital and physical retail realms (Rooderkerk and Kök, 2019). In particular, Walmart has heavily invested in IT-related technology and digital infrastructure to create a seamless customer shopping experience (Banker, 2021). Likewise, Amazon has opened physical bookstores and acquired Whole Foods, including its hundreds of physical stores, in a massive $13.7 billion deal (Debter, 2017). As a result of gradual implementation of omnichannel practices, few retailers have fully exploited the potential of the omnichannel approach, which necessitates further development of methodologies that can facilitate decision-making in omnichannel environments.

While omnichannel retail enhances customer shopping experience, the necessary channel integration makes decision-making extremely complex. Optimizing assortments in digital and physical channels under stochastic demand, while simultaneously managing inventory across a large and possibly diverse network of B&M stores, poses significant challenges for omnichannel retailers. This complexity stems from the interconnected nature of operational decisions related to inventory management and assortment optimization, particularly in the context of demand pooling across different sales channels. In this paper, we address this challenge by building upon the multichannel attraction model (MAM) introduced by Vasilyev et al. (2023), which is a discrete choice model specifically tailored

for omnichannel environments. Although the MAM was designed primarily for settings with just two retail channels, we expand its applicability to a setting with an online store and a network of physical stores, thereby enhancing its practical relevance. Furthermore, we integrate assortment optimization under the MAM with demand stochasticity and inventory management considerations, factors not taken into account by Vasilyev et al. (2023). These extensions of the modeling framework have significant implications. We demonstrate that accounting for demand variability is crucial due to demand pooling effects, which may impact a retailer's assortment decisions when inventory considerations are factored in. Additionally, our generalized framework makes it possible to analyze the impact of various relevant factors, such as the size or density of the physical store network, on properties of optimal assortments.

To incorporate inventory provisions into our modeling framework, we employ the newsvendor model, which effectively balances inventory costs and stockouts. We derive a profit function that accounts for optimal inventory decisions while assuming a static product substitution behavior of customers. We then formulate the assortment optimization problem as a mixed-integer second-order cone program. This problem can be solved exactly using off-the-shelf solvers for small to medium-sized instances. For larger problems, we propose two heuristic methods based on two different relaxations of the formulated assortment optimization problem. We derive the conditions under which the two relaxations are equivalent to each other and to the initial problem, and provide insights into the performance of our heuristic methods. Moreover, we conduct an extensive numerical analysis to derive managerial insights. Our findings reveal that an increasing coefficient of variation of demand has a dual effect on optimal assortment sizes: initially causing a decrease in online assortment size due to rising costs, followed by an increase in online assortment size as a result of the demand pooling effect. We also study the impact of other model parameters on optimal assortments, highlighting the importance of considering the structure of the physical store network and confirming some of the findings of Vasilyev et al. (2023). Lastly, we assess the potential benefits of omnichannel assortment optimization compared to assortment optimization in siloed channels, demonstrating that the former approach yields a noticeable increase in expected profit.

The remainder of this paper is organized as follows. Section 2.2 presents a review of two streams of literature closely related to our research. In Section 2.3, we provide a concise introduction to the MAM and adapt this model to the case of a retailer managing both an online store and a network of physical stores. In Section 2.4, we utilize the newsvendor model to derive a function for computing expected profits given assortments in all channels, and provide an illustrative example highlighting the significance of accounting for demand variability and inventory costs in omnichannel assortment decision-making. In

Section 2.5, we formulate the assortment problem, examine its computational complexity, and reformulate it as a mixed-integer second-order cone program. Section 2.6 is devoted to heuristic methods that can be used to solve this optimization problem. In Section 2.7, we perform a sensitivity analysis to gain insights into the properties of optimal assortments, and evaluate the benefits of omnichannel assortment optimization as opposed to optimizing assortments in isolated channels. Finally, Section 2.8 summarizes our findings and outlines potential avenues for future research.

2.2 Literature Review

In this section, we review two primary streams of literature closely related to our research: the integration of inventory decisions with choice modeling and assortment optimization, and assortment planning in the context of omnichannel retailing.

2.2.1 Combining Inventory Decisions with Choice Modeling

In this paper, we utilize the multichannel attraction model (MAM) developed by Vasilyev et al. (2023), a discrete choice model specifically tailored for omnichannel environments. The MAM builds on the general attraction model (GAM) introduced by Gallego et al. (2014), which, in turn, is a natural extension of the multinomial logit model (MNL) formulated by Luce (1959) and McFadden (1973). Arising from the random utility theory, the MNL is often regarded as the most prominent discrete choice model and is extensively applied in practice due to its simple structure. We refer the reader to Vasilyev et al. (2023) for an in-depth review of the theoretical background behind the MAM. One of the key contributions of our paper is incorporating inventory decisions into the MAM framework. There are two primary approaches to integrating choice models with inventory decisions: assuming either static or dynamic product substitution behavior of customers. The static approach suggests that customer shopping behavior is determined solely by the retailer's assortment decisions, and if a customer is willing to buy a product that is out of stock, it results in a lost sale. In contrast, the dynamic approach suggests that customers may switch to other items when their desired products are out of stock.

The static approach has been widely explored in academic literature. Smith and Agrawal (2000) studied the problem of joint assortment and inventory decisions under the exogenous demand model. In this model, base product demands are assumed to be pairwise independent, with fixed parameters determining the redistribution of product demand to other products when a product is unavailable. If the second-choice product is also unavailable, a lost sale is generally assumed to occur, because modeling more complex

product substitution behavior would require defining numerous additional parameters that are difficult to estimate in practice. The authors derived optimal inventory quantities using the newsvendor model and formulated the assortment optimization as a nonlinear integer program. Ryzin and Mahajan (1999) analyzed the same problem using the MNL model for consumer choice and the newsvendor model for optimal inventory levels. They imposed the restrictive assumption that products have identical price-to-cost ratios, which led to an analytical characterization of optimal assortments. Several studies have built upon the model of Ryzin and Mahajan (1999), such as Cachon et al. (2005), who incorporated consumer search behavior into their model. This behavior suggests that customers may find a suitable product at one retailer but choose not to purchase it, preferring to explore other retailers' offers for better options. The authors showed that neglecting consumer search can result in suboptimal assortment decisions. Li (2007) relaxed the assumption of identical price-to-cost ratios and demonstrated that the optimal assortments have a simple structure that can be explicitly derived under the assumption that store traffic is a continuous random variable. Maddah and Bish (2007) extended the model of Ryzin and Mahajan (1999) by incorporating pricing decisions into the modeling framework.

Gaur and Honhon (2006) addressed the problem of joint assortment and inventory decisions in a setting similar to Ryzin and Mahajan (1999), but employed the locational choice model instead of the MNL. The locational choice model suggests that each customer has an ideal product configuration in the attribute space, and they either purchase the product closest to their ideal configuration or leave the retailer. The authors showed that under the locational choice model, products in the optimal assortment tend to be equally spaced out in the attribute space. They also used their results for the static model to develop two heuristic methods for the problem under dynamic substitution. Fisher and Vaidyanathan (2014) addressed a similar problem under a custom-built demand model. They considered each SKU as a collection of attribute levels, estimated the demand share of each attribute level based on historical sales data, and derived the demand share of each SKU as the product of its attribute levels' demand shares. The authors developed several heuristic methods to solve the assortment optimization problem. Topaloglu (2013) extended the work of Ryzin and Mahajan (1999) by making not only assortment and inventory decisions, but also selling horizon decisions. They demonstrated that including these decisions makes it possible to relax the assumption of products having the same price-to-cost ratios, while still avoiding the combinatorial aspects of the assortment optimization problem. Geunes and Su (2020) devised an analytical framework for making joint assortment, inventory, and pricing decisions in online and physical retail channels, where customer choices are modeled using a mixture of MNL models. All these decisions are represented by variables in a two-stage stochastic optimization problem that can be

solved using a simulation-based approximation algorithm.

The alternative approach to assuming the static product substitution behavior is considering dynamic product substitution behavior, under which customers may switch to other products when faced with a stockout. Although this assumption is arguably more realistic, it leads to extremely challenging problems even in the simplest settings due to the stochastic nature of stockout events. Mahajan and Ryzin (2001) analyzed a newsvendor-style inventory problem under dynamic substitution, where a retailer selects initial inventory levels for products in a given assortment to maximize expected profits. They employed a utility maximization framework to model customer choices. The authors demonstrated that for a continuous relaxation of the problem, the expected profit is not even quasiconcave in the initial inventory levels. They proposed a sample path gradient method for the relaxed problem and showed that it converges to a stationary point of the expected profit function under some mild conditions. Honhon, Gaur, et al. (2010) addressed the problem of joint assortment and inventory decisions under dynamic substitution, or dynamic assortment planning for short. In their model, customers can be separated into different segments based on their preference lists, with each customer purchasing their highest-ranked product available at the time of their visit. The total customer demand is assumed to be random and comprises fixed proportions of customers from different segments. The authors developed a dynamic programming-based algorithm with a computational complexity of $O(8^n)$, where n represents the number of products. Honhon and Seshadri (2013) built upon the model of Honhon, Gaur, et al. (2010) by considering random proportions of customers from different segments, and providing a performance guarantee for the fixed proportions solution in the random proportions setting. Aouad, Levi, et al. (2018) studied the dynamic assortment planning problem under the MNL. Importantly, they disregarded inventory costs, instead imposing a capacity constraint on the total number of units stocked. The authors presented an approximation algorithm that yields expected revenue of at least 0.139 times the optimum given that the total demand has an increasing-failure rate distribution. Aouad and Segev (2019) developed an approximation scheme for the same problem, with the running time of their randomized algorithm of $O(\frac{1}{\delta} \cdot (nC)^{O(\log \Delta/\epsilon^2)})$, where ϵ is the accuracy level, δ is the confidence level, n is the number of products, C is the total number of units that can be stocked, and Δ is the ratio between the maximal and minimal product weights. Unlike Aouad, Levi, et al. (2018), the algorithm proposed by Aouad and Segev (2019) does not rely on the assumption of a specific distribution of the total number of customers.

Several papers have examined the dynamic assortment planning problem while making simplifying assumptions to avoid complexities induced by the relevance of stockout event sequences. Goyal et al. (2016), Segev (2019), Aouad, Levi, et al. (2019), Transchel et al.

(2022) considered dynamic assortment planning models for which they devised efficient solution algorithms, providing either approximation guarantees or numerical performance validation. However, the choice models in these papers impose a very specific order in which products are consumed. Finally, some studies assume that if the on-hand inventory cannot satisfy the demand for a product during a sales period, fixed proportions of the unmet demand are redirected to other products (see, e.g., Netessine and Rudi (2003), Kök and Fisher (2007), Schlapp and Fleischmann (2018)). Under this assumption, product substitution behavior is not truly dynamic since it is not affected by the order of stockout events. In this paper, we follow the more conventional static approach, as the complexity of the omnichannel environment makes it exceedingly difficult to employ the dynamic approach. Moreover, as many authors have noted, the static approach provides a reasonable approximation of customer shopping behavior in various real-world scenarios (see, e.g., Ryzin and Mahajan (1999), Topaloglu (2013)).

2.2.2 Assortment and Inventory Decision-Making in an Omnichannel Context

The topic of assortment optimization in an omnichannel environment has received considerable attention in the past decade. In one of the earliest works on the subject, Bhatnagar and Syam (2014) considered the problem of finding the optimal item allocation for a hybrid retailer that operates an online store and a chain of B&M stores. They developed an integer programming framework, assuming fixed product demands with no product substitution. Dzyabura and Jagabathula (2018) investigated the problem of selecting a subset of the online assortment to offer in the physical channel to maximize profits across both channels. In their modeling approach, product substitution is taken into account via an MNL-based consumer choice model where utilities depend on product features (or attributes) but customer preferences may change following in-store product inspections. Lo and Topaloglu (2022) built upon the work of Dzyabura and Jagabathula (2018) by characterizing the product portfolio with a features tree, which allowed them to relax the assumption made by Dzyabura and Jagabathula (2018) that a product exists for every combination of features. Chen, Liang, et al. (2022) used a mixture of MNLs, or the mixed MNL, to model customer preferences. They addressed the problem of determining both optimal locations and assortments of B&M stores in the presence of an online sales channel that maximize the total profit. Most recently, Vasilyev et al. (2023) studied the assortment optimization problem in a dual-channel setting, utilizing the previously mentioned MAM. The authors formulated this problem as a mixed-integer linear program and provided an efficient heuristic method for solving it. They also described general effects of the implementation of widely-used omnichannel initiatives on the MAM parameters and

explored the properties of optimal assortments through numerical experiments.

Several studies have explored inventory management in the context of omnichannel decision-making. One of the first contributions in this area was made by Seifert et al. (2006), who analyzed a company that serves both online and in-store customers. The authors studied an integration strategy where a firm can utilize excess inventory at B&M stores to fulfill online orders, a strategy known as inventory pooling. They quantified the benefits of channel integration over maintaining dedicated channels under centralized and decentralized decision-making. They also established the conditions under which using local stock for online order fulfillment can be beneficial. Hu et al. (2022) investigated the impact of the buy-online-pick-up-in-store (BOPS) initiative on B&M store operations from an inventory management perspective, using a stylized model with two customer types: store-only and omnichannel. The authors demonstrated that the economic viability of this initiative depends on the relative and absolute magnitude of costs related to store visiting and online waiting. In particular, if the latter cost is relatively low and the former cost is even lower, introducing the BOPS functionality benefits the retailer due to the demand pooling effect. Conversely, if both costs are relatively high, with the online waiting cost being even higher, then the BOPS initiative may negatively impact the retailer because of demand depooling. Govindarajan et al. (2021) studied joint inventory and fulfillment decisions of an omnichannel retailer operating a network of B&M stores and online fulfillment centers. Their network-based approach made it possible to account for the demand that "spills over" to other inventory locations in case of stockouts, as well as for the synergies arising from pooling in-store and online demands within each inventory location. The authors found that the financial benefits of centralized inventory planning increase with network size and are highest when the retailer experiences a moderate mix of in-store and online demands. Abouelrous et al. (2022) addressed a similar problem as Govindarajan et al. (2021) but employed a different methodology. The authors developed an algorithm based on solving a two-stage stochastic optimization problem on a reduced number of scenarios. They showed that their proposed algorithm results in an average cost reduction of 7.56% compared to the algorithm of Govindarajan et al. (2021). Gabor et al. (2022) analyzed a two-echelon inventory model for an omnichannel retailer, where a warehouse follows an (R, Q) inventory policy and uses its stock to serve online customers and replenish B&M stores. They showed that offering customers a discount for purchasing online can lead to substantial financial benefits. Other relevant papers studied the effect of inventory availability information (Gao and Su, 2017) and store return options (He, Xu, et al., 2020) on optimal inventory decisions.

Apart from the previously mentioned paper by Geunes and Su (2020), there have been very few works on joint assortment and inventory optimization in omnichannel environments.

Recently, Hense and Hübner (2022) proposed an approach based on the exogenous demand (ED) model to investigate the omnichannel assortment planning problem, taking into account limited shelf space and space elasticity of the demand. While the ED model-based approach has some limitations as outlined in Section 2.2.1, it allowed the authors to determine optimal assortment composition along with optimal space allocation and inventory provisions across channels. To solve the profit maximization problem, which was formulated as a multi-knapsack problem, Hense and Hübner (2022) proposed a heuristic algorithm. Building upon the work of Hense and Hübner (2022), Schäfer et al. (2023) incorporated various additional demand effects into the modeling framework, including product positioning on store shelves. The authors extended the algorithm presented by Hense and Hübner (2022) to solve the resulting nonlinear integer program.

In this paper, we go beyond the existing literature by addressing the problem of joint assortment and inventory optimization under a utility-based choice model specifically designed for omnichannel environments. We leverage the MAM framework developed by Vasilyev et al. (2023), and broaden its scope in multiple key directions. Notably, we incorporate demand stochasticity and inventory management considerations into the assortment optimization problem under the MAM. We highlight the importance of accounting for these aspects by demonstrating how neglecting them can result in suboptimal assortment decisions. By examining a setting with both an online store and a chain of physical stores, as opposed to a simple dual-channel setup, we are able to uncover the crucial role the demand pooling effect plays in determining optimal assortments. Our work offers various theoretical and practical insights, deepening the understanding of factors affecting omnichannel assortment decision-making.

2.3 Multichannel Attraction Model with a Network of Physical Stores

This section provides a brief overview of the multichannel attraction model (MAM) introduced by Vasilyev et al. (2023) and adapts it to the context of our modeling framework. We consider a setting in which an omnichannel retailer manages an online store and a chain of physical (brick-and-mortar) stores. We assume that the physical stores can be represented by a network, where nodes correspond to stores, and two nodes are connected via an edge if and only if customers can switch between them. For tractability, we consider the case where customers can move between connected stores in both directions (i.e., edges are not directed), but this assumption can easily be relaxed. We also assume that customers can always switch from a physical store to the online store and vice versa. In essence, the online store can be viewed as an additional node in
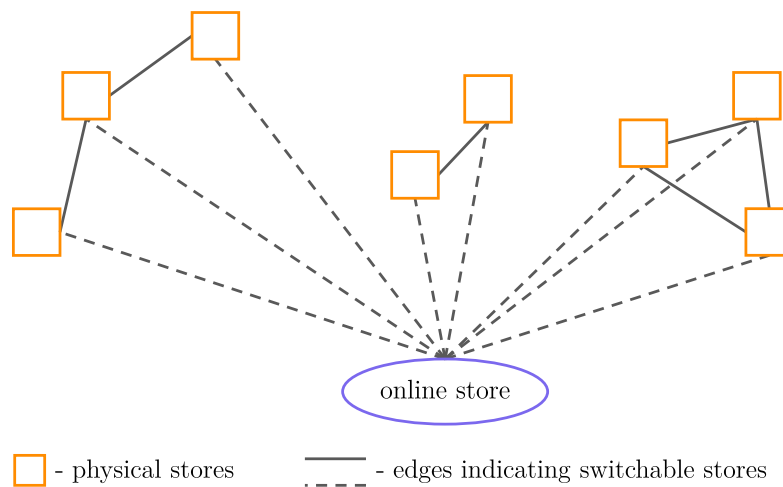
Figure 2.1: Store network illustration.

the network that is connected to all other nodes (see Figure 2.1). To maintain clarity, we focus on the setting with a single online store, but our modeling approach can be generalized to the case with a larger number of online stores – for example, if a retailer has a website and a page on a social media platform – by considering additional nodes that are connected to all other nodes in the network. Throughout this paper, we will sometimes refer to stores as sales channels.

Vasilyev et al. (2023) primarily studied the MAM in a dual-channel setting. However, the MAM can be extended to accommodate a network of physical stores rather than a single physical channel. Such an extended version of the MAM can be formally defined in the following way. Suppose that $\mathcal{N} = \{1, 2, \dots, n\}$ is the set of products that can be offered in all channels, and let 0 represent the no-purchase alternative. Furthermore, let $\mathcal{C}$ denote the set of channels (stores), and $c \in \mathcal{C}$ denote a channel index. Lastly, let $S_c \subseteq \mathcal{N}$ denote the product assortment in channel c, and $S = \{S_c\}_{c \in \mathcal{C}}$ denote the collection of assortments in all channels. Following Vasilyev et al. (2023), we identify different types of customers (one customer type per channel) depending on customers' preferred channels, with type-c customers being the ones who would shop in channel c if all alternatives were offered in all channels. For type-c customers, we adopt the following modified version of the notation used by Vasilyev et al. (2023):

- $v_j^{(c)}$, $v_0^{(c)}$: attractiveness values of purchasing product $j \in S_c$ and selecting the no-purchase alternative, respectively;

- $\phi_j^{(c)}$: shadow attractiveness value of purchasing product $j \in \mathcal{N} \backslash S_c$ from another source;

- $\mathcal{C}_{cj}$: subset of channels a customer can switch to in case $j \notin S_c$;

- $\delta_{\bar{c}j}^{(c)}$: proportion of customers switching to channel $\bar{c} \in \mathcal{C}_{cj}$ out of those willing to purchase product $j \notin S_c$ from another source.

Parameters $v_j^{(c)}$ are assumed to be strictly positive for all $c \in \mathcal{C}$, $j \in \mathcal{N} \cup \{0\}$. Furthermore, each shadow attractiveness value $\phi_i^{(c)}$ has to belong to the interval $[0, v_i^{(c)}]$. This is because the MAM is defined as a regular choice model, i.e., only product substitution effects (and not complementarity effects) are present in the product portfolio. In other words, adding a product to the assortment cannot increase the probability of customers selecting other products. Finally, in order for the parameters $\delta_{\bar{c}j}^{(c)}$ to be correctly defined, we must ensure that $\sum_{\bar{c} \in \mathcal{C}_{cj}} \delta_{\bar{c}j}^{(c)} \leq 1$ for all $j \in \mathcal{N}$, $c \in \mathcal{C}$.

Given assortments in all channels, the choice probabilities under the MAM are as follows. The probability of a type-c customer purchasing product j in channel c (store c) is:

$$\pi_{cj}^{(c)}(S_c) = \frac{v_j^{(c)} \mathbb{1}_{j \in S_c}}{v_0^{(c)} + \sum_{k \in S_c} v_k^{(c)} + \sum_{i \in \mathcal{N} \setminus S_c} \phi_i^{(c)}}, \tag{2.1}$$

and the probability that such a customer purchases product j in channel $\bar{c}$ is:

$$\pi_{\bar{c}j}^{(c)}(S) = \frac{\phi_j^{(c)} \delta_{\bar{c}j}^{(c)} \mathbb{1}_{j \in S_{\bar{c}} \setminus S_c}}{v_0^{(c)} + \sum_{k \in S_c} v_k^{(c)} + \sum_{i \in \mathcal{N} \setminus S_c} \phi_i^{(c)}}, \tag{2.2}$$

where $\mathbb{1}$ represents the indicator function.

The MAM is exceptionally well-suited to model the shopping behavior of customers in an omnichannel environment. First, the MAM is a very intuitive model with parameters that admit a natural interpretation. The concept of shadow attractiveness values, initially introduced by Gallego et al. (2014), is particularly fitting for an omnichannel setting (see Vasilyev et al. (2023) for more details). Second, this model has a sound theoretical justification. Vasilyev et al. (2023) showed that this model belongs to the family of random utility models as there exists a distribution over rankings of alternatives consistent with the MAM choice probabilities. They also proved that the MAM can be represented as a mixture of Markov chain choice models introduced by Blanchet et al. (2016). Finally, the model parameters can be estimated using conventional techniques such as least squares or maximum likelihood (Vasilyev et al., 2023).

In addition to the limitation of focusing on a dual-channel setting, the modeling framework presented by Vasilyev et al. (2023) does not incorporate demand stochasticity and inventory

management considerations. In our paper, we aim to fill this gap by assuming that the retailer is subject to inventory costs, and the profit associated with each product is computed using a variant of the single-period newsvendor model rather than a linear function of the expected demand quantity.

2.4 Profit Function Under Stochastic Demand and Inventory Costs

In this section, we derive the function for computing the expected profit given assortments in all channels, and provide an example demonstrating the importance of accounting for demand stochasticity and inventory costs when making assortment decisions in an omnichannel environment.

2.4.1 Profit Function Derivation

First, we assume that the number of customers of a given type visiting the retailer during the considered period is a normally distributed random variable. The assumption of having normally distributed demands is common in the academic literature, especially if it is important to maintain flexibility when modeling the variability of the demand (see, e.g., Geunes and Su (2020)). Note that the support of a normally distributed random variable is the whole real line, which presents a challenge as demand cannot be negative. However, if the standard deviation of such a random variable is substantially less than its expected value, then the probability of attaining negative values becomes negligible. For example, if the value of the standard deviation is 0.3 of the expected value (equivalently, if the corresponding coefficient of variation equals 0.3), then the probability of such a normally distributed random variable attaining negative values is 0.0004. In this paper, we assume that the coefficients of variation of random variables representing the numbers of customers of each type are upper-bounded by $CV_{max} = 0.3$. Such an upper bound on the coefficients of variation of normally distributed demands can also be observed in the literature (see, e.g., Geunes and Su (2020)).

Let $\xi^{(c)} \sim \text{NORMAL}(\mu^{(c)}, (\sigma^{(c)})^2)$ be the number of type-c customers, We assume that $\xi^{(c)}$ and $\xi^{(\bar{c})}$ are independent for any channels $c \neq \bar{c}$. We also assume that the choice probabilities of customers are given by the multichannel attraction model. Then, the demand D_{cj} for product j in channel c is as follows:

$$D_{cj} = \xi^{(c)}\pi_{cj}^{(c)}(S_c) + \sum_{\bar{c}\in\mathcal{C}_{cj}} \xi^{(\bar{c})}\pi_{cj}^{(\bar{c})}(S). \tag{2.3}$$

Importantly, we assume that customers exhibit static product substitution behavior, meaning that the distributions of product demands are determined solely by assortment decisions and not by inventory levels. Under this assumption, stockouts are considered abnormal situations that result in lost sales (e.g., due to customer dissatisfaction with the service). As discussed in Section 2.2.1, the static modeling approach provides a reasonable approximation of real-life customer shopping behavior in common practical settings.

Using properties of the normal distribution, we obtain that the individual product demands (2.3) are normally distributed with the following parameters:

$$
D_{cj} \sim \text{NORMAL}\!\left(\mu^{(c)}\pi_{cj}^{(c)}(S_c) + \sum_{\bar{c}\in\mathcal{C}_{cj}} \mu^{(\bar{c})}\pi_{cj}^{(\bar{c})}(S), (\sigma^{(c)}\pi_{cj}^{(c)}(S_c))^2 + \sum_{\bar{c}\in\mathcal{C}_{cj}} (\sigma^{(\bar{c})}\pi_{cj}^{(\bar{c})}(S))^2 \right).
$$

(2.4)

Note that if the coefficients of variation of aggregate demands generated by customers of each type are upper bounded by a certain constant, then the coefficients of variation of individual product demands are upper bounded by the same constant. Formally, consider the following proposition:

Proposition 2.1. *If the coefficient of variation $CV(\xi^{(c)})$ of the number of type-c customers is less than CV_{max} for each $c \in \mathcal{C}$, then the coefficient of variation $CV(D_{cj})$ of any individual product demand D_{cj} is also less than CV_{max}.*

The proof of this proposition can be found in Appendix B.1. Now, given product demand distributions, we can compute the profit associated with each product using a variant of the single-period newsvendor model. For the sake of clarity, let us slightly abuse the notation and omit channel and product indices in the profit function derivation part. Suppose that for a certain product, its demand D follows distribution $\text{NORMAL}(\mu, \sigma^2)$ with probability density function $f(\cdot)$ and cumulative distribution function $F(\cdot)$. Let us introduce the following additional notation:

- r, b: unit price and unit ordering cost of the product, respectively;
- h: holding cost per unit of the product left over at the end of the period;
- s: base-stock quantity of the product at the beginning of the period.

Note that we assume that each unit of the product left over at the end of the period is penalized with the holding cost h. Alternatively, it could be assumed that leftover units are sold back to the supplier. In this case, the holding cost h would have to be replaced with $-\alpha b$, where $\alpha \in [0,1]$ is a sell-back discount factor.

Then, the expected revenue is:

$$r\mathbb{E}(\min(s, D)) = r\left(\int_0^s tf(t)dt + \int_s^\infty sf(t)dt\right) = r\mu - r\int_s^\infty (q-s)f(q)dq,$$

and the expected cost is:

$$bs + h\mathbb{E}(\max(0, s - D)) = bs + h(s - \mu) + h\int_s^\infty (q-s)f(q)dq.$$

Thus, the expected profit associated with the product can be computed as follows:

$$\Pi(s, \mu, \sigma) = (r+h)\mu - (b+h)s - (r+h)\int_s^\infty (q-s)f(q)dq. \tag{2.5}$$

It is straightforward to verify that the optimal base-stock quantity s^* maximizing the expected profit is given by $s^* = F^{-1}\left(\dfrac{r-b}{r+h}\right)$. Let $\varphi(\cdot)$ and $\Phi(\cdot)$ denote the probability density function and the cumulative distribution function of the standard normal distribution NORMAL$(0, 1)$, respectively. Then, the loss function $\int_s^\infty (q-s)f(q)dq$ and the optimal base-stock quantity s^* can be rewritten in the following way:

$$\int_s^\infty (q-s)f(q)dq = \sigma\left(\varphi\left(\frac{s-\mu}{\sigma}\right) - \frac{s-\mu}{\sigma}\left(1 - \Phi\left(\frac{s-\mu}{\sigma}\right)\right)\right),$$

$$s^* = \mu + \Phi^{-1}\left(\frac{r-b}{r+h}\right)\sigma. \tag{2.6}$$

Substituting (2.6) into (2.5) yields the following expression for the expected profit:

$$\Pi(\mu, \sigma) = (r-b)\mu - (r+h)\varphi\left(\frac{r-b}{r+h}\right)\sigma.$$

Thus, by reintroducing channel and product indices, we arrive at the following expression for the profit associated with product j in channel c given that the product demand is distributed as NORMAL(μ, σ^2):

$$\Pi_{cj}(\mu, \sigma) = (r_{cj} - b_{cj})\mu - (r_{cj} + h_{cj})\varphi\left(\frac{r_{cj} - b_{cj}}{r_{cj} + h_{cj}}\right)\sigma. \tag{2.7}$$

Finally, since each individual product demand follows distribution (2.4), we obtain that

the total expected profit given assortments in all channels is as follows:

$$\sum_{c \in \mathcal{C}} \sum_{j \in \mathcal{N}} \Pi_{cj} \left(\mu^{(c)} \pi_{cj}^{(c)}(S_c) + \sum_{\bar{c} \in \mathcal{C}_{cj}} \mu^{(\bar{c})} \pi_{cj}^{(\bar{c})}(S), \sqrt{(\sigma^{(c)} \pi_{cj}^{(c)}(S_c))^2 + \sum_{\bar{c} \in \mathcal{C}_{cj}} (\sigma^{(\bar{c})} \pi_{cj}^{(\bar{c})}(S))^2} \right). \quad (2.8)$$

2.4.2 Importance of Inventory Considerations Given Stochastic Demand in an Omnichannel Environment

Let us consider an illustrative example highlighting the importance of incorporating inventory considerations into the modeling framework when addressing assortment optimization in an omnichannel environment. Suppose that a retailer operates an online store and m_p identical geographically distributed physical stores, meaning that the total number of stores (physical and online) is $m = m_p + 1$. The assumption that the physical stores are geographically distributed implies that customers do not switch from one physical store to another, i.e., there are no edges between physical stores in the retail store network. Also, suppose that this retailer sells only one product (alternatively, we can consider a product such that the demand for this product is independent of demands for all other products). We will now demonstrate that although the deterministic approach to assortment optimization under the MAM without inventory considerations may suggest that the most profitable strategy is to offer the product in all stores (channels), taking the demand stochasticity and inventory considerations into account may reveal that it is more profitable for the retailer to offer the product exclusively online.

Suppose that the product price is the same in online and physical channels. If we disregard inventory-related costs, then the optimal retailer strategy under the MAM is to offer the product everywhere as it maximizes the total expected number of customers, and both online and physical customers are equally profitable. Now, let us show that if we include inventory management considerations and use expression (2.8) for profit calculation, then the optimal strategy may change due to the demand pooling effect. Suppose that in addition to the same price, the holding cost and the unit ordering cost are also the same for all channels (with the purpose of isolating the demand pooling effect from the costs difference effect). Let $\mu^{(p)}$ and $\sigma^{(p)}$ be the expected value and the standard deviation of the product demand in each of the identical physical retail stores, and $\mu^{(o)}$ and $\sigma^{(o)}$ be the expected value and the standard deviation of the product demand in the online store. Furthermore, let r, b, and h be the unit price, ordering cost, and holding cost of the product, respectively. Also, let v, ϕ, and v_0 be the product attractiveness value, product shadow attractiveness value, and the attractiveness value of the no-purchase alternative in either of the channels, respectively, and suppose that these values satisfy $v = \phi = v_0$. Finally, let δ be the proportion of customers switching from a physical store to the online

store out of those willing to purchase the product from another source if this product is not offered in the physical store.

In this case, if the product is offered in all stores, then the expected profit is:

$$\Pi_{all} = 0.5\left((r-b)(\mu^{(o)} + m_p\mu^{(p)}) - (r+h)\varphi\left(\frac{r-b}{r+h}\right)(\sigma^{(o)} + m_p\sigma^{(p)})\right), \qquad (2.9)$$

and if the product is offered exclusively online, then the expected profit is:

$$\Pi_{online} = 0.5\left((r-b)(\mu^{(o)} + m_p\delta\mu^{(p)}) - (r+h)\varphi\left(\frac{r-b}{r+h}\right)\sqrt{\left(\sigma^{(o)}\right)^2 + m_p\left(\delta\sigma^{(p)}\right)^2}\right). \quad (2.10)$$

It can be observed that Π_{all} can be less than Π_{online} depending on the values of the parameters. Let us consider an example in which $m = 10$ (hence $m_p = 9$), $r = 1$, $b = 0.8$, $h = 0.2$, $\delta = 0.7$, $\mu^{(o)} = 1000$, $\sigma^{(o)} = 150$, $\mu^{(p)} = 50$, and $\sigma^{(p)} = 7.5$. Then, the ratio Π_{all}/Π_{online} is approximately 0.975, meaning that it is more profitable for the retailer to offer the product exclusively online compared to offering it in all stores.

A key factor affecting how the ratio Π_{all}/Π_{online} compares to 1 is the value of the coefficient of variation $CV^{(p)} = \sigma^{(p)}/\mu^{(p)}$. Indeed, let $\mu^{(p)}$ be fixed (so that the total number of customers remains constant), and let $CV^{(p)}$ take values from 0.05 to 0.3. The relation between $CV^{(p)}$ and Π_{all}/Π_{online} is shown in Figure 2.2a. Ultimately, there is a profitability cutoff such that if the coefficient of variation exceeds a certain threshold, then it becomes more profitable for the retailer to offer such a product exclusively online. To gain additional managerial insights, let us consider a stylized example of a portfolio of non-substitutable (independent) products sorted in the descending order of the coefficient of variation of product demand in a physical store. In practice, the values of the coefficients of variation of product demands often follow an L-shaped curve, with a few products having a high coefficient of variation and the majority of products having a lower coefficient of variation. Similar to the single-product setting, suppose that all physical stores are identical and geographically distributed and that for each product, its price, purchasing cost, and holding cost are the same for all stores. Furthermore, suppose that all products are offered in all stores and that purchasing and inventory costs divided by corresponding prices do not depend on the product index. Then, from the reasoning for a single product it follows that the considered product portfolio can be separated into two parts, such that it is more profitable for the retailer to offer products with a high coefficient of variation of demand exclusively online, and to offer the rest of the product portfolio both online and offline (see Figure 2.2b). While this is just a stylized example based on several strong assumptions, it effectively demonstrates the crucial role of the coefficient of variation of product demand in determining the optimal allocation of

(a) Relation between the coefficient of variation of the demand in a physical store and the profit ratio.

(b) Stylized example of a product portfolio sorted in the descending order of the coefficient of variation of product demand in a physical store.
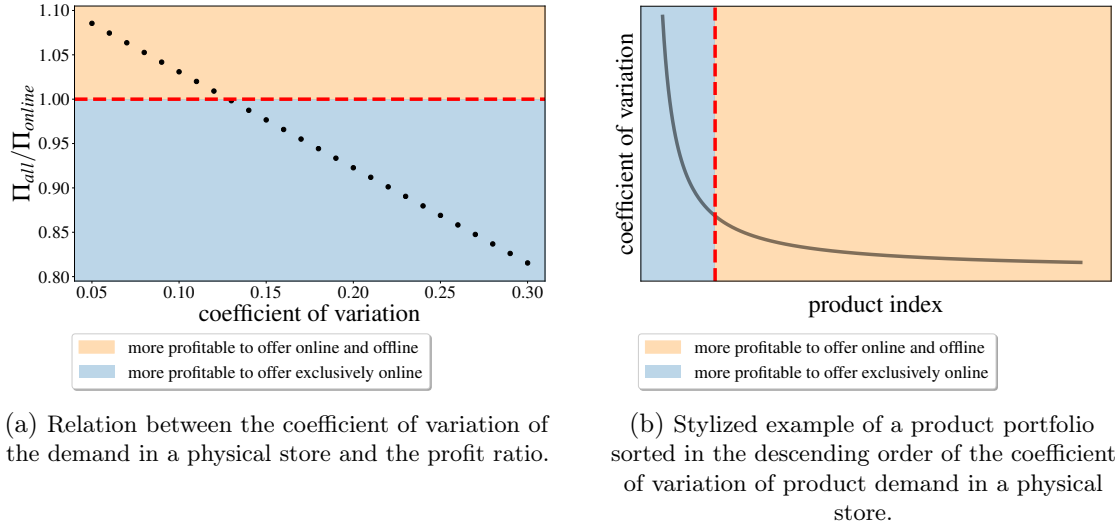
Figure 2.2: Profitability cutoff with respect to the coefficient of variation of product demand.

products between online and physical stores.

In the single-product setting, the connection between the coefficient of variation of the demand in a physical store and the profitability of different product allocation strategies becomes even more evident if we assume that the demand generated by online customers is zero, i.e., if $\mu^{(o)} = 0$ and $\sigma^{(o)} = 0$. In this case:

$$\Pi_{all} - \Pi_{online} = 0.5\left(m_p(1 - \delta)(r - b)\mu^{(p)} - (m_p - \delta\sqrt{m_p})(r + h)\varphi\left(\frac{r - b}{r + h}\right)\sigma^{(p)}\right).$$

This equation shows that the difference between the expected profit obtained by offering the product everywhere and the one obtained by offering the product exclusively online is a linear function of $\mu^{(p)}$ and $\sigma^{(p)}$. Therefore, the profit difference decreases linearly if the coefficient of variation $CV^{(p)} = \sigma^{(p)}/\mu^{(p)}$ increases due to either a linear decrease in $\mu^{(p)}$ or a linear increase in $\sigma^{(p)}$.

2.5 Assortment Problem

The goal of the assortment optimization problem under the MAM with inventory considerations is to maximize the total expected profit (2.8) over all possible assortments $S = \{S_c\}_{c \in \mathcal{C}}$, i.e.:

$$\max_S \sum_{c \in \mathcal{C}} \sum_{j \in \mathcal{N}} \Pi_{cj}\left(\mu^{(c)}\pi_{cj}^{(c)}(S_c) + \sum_{\bar{c} \in \mathcal{C}_{cj}} \mu^{(\bar{c})}\pi_{cj}^{(\bar{c})}(S), \sqrt{(\sigma^{(c)}\pi_{cj}^{(c)}(S_c))^2 + \sum_{\bar{c} \in \mathcal{C}_{cj}} (\sigma^{(\bar{c})}\pi_{cj}^{(\bar{c})}(S))^2}\right).$$

Such a nonlinear discrete optimization problem is extremely difficult to solve. First, this problem is NP-hard in the number of stores even in the setting with a single product and deterministic demand (i.e., if $\sigma^{(c)} = 0$ for all $c \in \mathcal{C}$) as demonstrated by the following theorem:

Theorem 2.1. *The assortment optimization problem under the MAM is NP-hard in the number of stores.*

The proof of this theorem is based on the fact that a special case of our assortment optimization problem is equivalent to the assortment optimization problem under the 2-product nonparametric choice model presented by Feldman et al. (2019), which is in turn NP-hard. The full proof of this theorem is provided in Appendix B.2.

Remark 2.1. *Proof of Theorem 2.1 implies that the considered assortment optimization problem is not just NP-hard, but also APX-hard in the number of stores, meaning that it does not admit any polynomial-time approximation scheme (PTAS) unless $P = NP$. Indeed, Feldman et al. (2019) proved the NP-hardness of the assortment optimization problem under the 2-product nonparametric choice model by constructing a reduction from the minimum vertex cover problem on cubic graphs, which is APX-hard as was shown by Alimonti and Kann (2000).*

Furthermore, it follows from the findings of Vasilyev et al. (2023) that if at least one of the channels has limited shelf space, or assortments in at least two channels have to be the same, then the assortment optimization problem is NP-hard in the number of products (even if we disregard inventory considerations and assume deterministic demand). Proving that the assortment optimization problem without such constraints is NP-hard in the number of products remains an interesting problem for future research.

Despite having such a complex structure of the objective function, our assortment optimization problem can be reformulated as the following mixed-integer second-order cone program that can be handled by off-the-shelf solvers:

$$\max_{x} \sum_{c \in \mathcal{C}} \sum_{j \in \mathcal{N}} \Pi_{cj}\big(x_{cj}^{(c)} + \sum_{\bar{c} \in \mathcal{C}_{cj}} x_{cj}^{(\bar{c})}, t_{cj}\big) \tag{2.11a}$$

$$\text{s.t. } \big\|(y_{cj}^{(c)}, \{y_{cj}^{(\bar{c})}\}_{\bar{c} \in \mathcal{C}_{cj}})\big\|_2 \leq t_{cj} \qquad\qquad \forall c \in \mathcal{C}, \ j \in \mathcal{N}, \tag{2.11b}$$

$$\frac{\tilde{v}_0^{(c)}}{v_0^{(c)}} x_{c0}^{(c)} + \sum_{j \in \mathcal{N}} \frac{\tilde{v}_j^{(c)}}{v_j^{(c)}} x_{cj}^{(c)} = \mu^{(c)} \qquad\qquad \forall c \in \mathcal{C}, \tag{2.11c}$$

$$\frac{x_{cj}^{(c)}}{v_j^{(c)}} + \frac{x_{\bar{c}j}^{(c)}}{\phi_j^{(c)} \delta_{\bar{c}j}^{(c)}} \leq \frac{x_{c0}^{(c)}}{v_0^{(c)}} \qquad\qquad \forall c \in \mathcal{C}, \ \bar{c} \in \mathcal{C}_{cj}, \ j \in \mathcal{N}, \tag{2.11d}$$

$$\frac{x_{cj}^{(c)}}{v_j^{(c)}} \leq \frac{x_{c0}^{(c)}}{v_0^{(c)}} \qquad\qquad \forall c \in \mathcal{C},\ \bar{c} \in \mathcal{C}\backslash\mathcal{C}_{cj},\ j \in \mathcal{N}, \qquad (2.11\text{e})$$

$$\frac{x_{c0}^{(c)}}{v_0^{(c)}} - \frac{x_{cj}^{(c)}}{v_j^{(c)}} \leq \frac{\mu^{(c)}}{\tilde{v}_0^{(c)}}(1 - z_{cj}) \qquad\qquad \forall c \in \mathcal{C},\ j \in \mathcal{N}, \qquad (2.11\text{f})$$

$$\frac{x_{c0}^{(c)}}{v_0^{(c)}} - \frac{x_{\bar{c}j}^{(c)}}{\phi_j^{(c)}\delta_{\bar{c}j}^{(c)}} \leq \frac{\mu^{(c)}}{\tilde{v}_0^{(c)}}(1 + z_{cj} - z_{\bar{c}j}) \qquad \forall c \in \mathcal{C},\ \bar{c} \in \mathcal{C}_{cj},\ j \in \mathcal{N}, \qquad (2.11\text{g})$$

$$\mu^{(c)}y_{cj}^{(c)} \geq \sigma^{(c)}x_{cj}^{(c)} \qquad\qquad \forall c \in \mathcal{C},\ j \in \mathcal{N}, \qquad (2.11\text{h})$$

$$\mu^{(c)}y_{\bar{c}j}^{(c)} \geq \sigma^{(c)}x_{\bar{c}j}^{(c)} \qquad\qquad \forall c \in \mathcal{C},\ \bar{c} \in \mathcal{C}_{cj},\ j \in \mathcal{N}, \qquad (2.11\text{i})$$

$$x_{cj}^{(c)} \leq H_j^{(c)}z_{cj} \qquad\qquad \forall c \in \mathcal{C},\ j \in \mathcal{N}, \qquad (2.11\text{j})$$

$$x_{\bar{c}j}^{(c)} \leq K_{\bar{c}j}^{(c)}x_{\bar{c}j}^{(\bar{c})} \qquad\qquad \forall c \in \mathcal{C},\ \bar{c} \in \mathcal{C}_{cj},\ j \in \mathcal{N}, \qquad (2.11\text{k})$$

$$x_{c0}^{(c)}, x_{cj}^{(c)}, x_{\bar{c}j}^{(c)}, y_{cj}^{(c)}, y_{\bar{c}j}^{(c)} \in \mathbb{R}_{\geq 0} \qquad \forall c \in \mathcal{C},\ \bar{c} \in \mathcal{C}_{cj},\ j \in \mathcal{N}, \qquad (2.11\text{l})$$

$$z_{cj} \in \{0, 1\} \qquad\qquad \forall c \in \mathcal{C},\ j \in \mathcal{N}, \qquad (2.11\text{m})$$

where $x = \big\{x_{c0}^{(c)}, x_{cj}^{(c)}, x_{\bar{c}j}^{(c)}, y_{cj}^{(c)}, y_{\bar{c}j}^{(c)}, t_{cj}, z_{cj}\big\}_{c\in\mathcal{C},\bar{c}\in\mathcal{C}_{cj},j\in\mathcal{N}}$, $\tilde{v}_0^{(c)} = v_0^{(c)} + \sum_{i\in\mathcal{N}} \phi_i^{(c)}$, $\tilde{v}_j^{(c)} = v_j^{(c)} - \phi_j^{(c)}$, and constants $H_j^{(c)}$ and $K_{\bar{c}j}^{(c)}$ are as follows:

$$H_j^{(c)} = \frac{v_j^{(c)}\mu^{(c)}}{\tilde{v}_0^{(c)} + \tilde{v}_j^{(c)}}, \quad K_{\bar{c}j}^{(c)} = \frac{\phi_j^{(c)}\delta_{\bar{c}j}^{(c)}\mu^{(c)}}{\tilde{v}_0^{(c)}} \bigg/ \frac{v_j^{(\bar{c})}\mu^{(\bar{c})}}{v_0^{(\bar{c})} + \sum_{k\in\mathcal{N}} v_k^{(\bar{c})}}. \qquad (2.12)$$

We call problem (2.11) a demand-based mixed-integer second-order cone program, or DB-MISOCP. The name is derived from the interpretation of its decision variables. Specifically, $x_{cj}^{(c)}$ represents the expected demand for product j in channel c generated by type-c customers, and $x_{cj}^{(\bar{c})}$ represents the expected demand for product j in channel c generated by type-$\bar{c}$ customers. Therefore, the sum $x_{cj}^{(c)} + \sum_{\bar{c}\in\mathcal{C}_{cj}} x_{cj}^{(\bar{c})}$ is the total expected demand for product j in channel c. Additionally, t_{cj} corresponds to the standard deviation of such demand. Note that due to the limited inventory, the actual sales may be lower than the demand. The equivalence of the considered assortment optimization problem and the DB-MISOCP is established by the following theorem:

Theorem 2.2. *The DB-MISOCP is a valid formulation of the assortment optimization problem under the MAM with inventory considerations.*

The DB-MISOCP can be viewed as an extension of the sales-based mixed-integer linear program (SBMILP) presented by Vasilyev et al. (2023). Compared to the SBMILP, the DB-MISOCP has a more complex formulation with a number of additional constraints and variables. For example, constraints (2.11g) are not present in the SBMILP, despite

the fact that they do not contain newly introduced variables that are unique for the DB-MISOCP. These constraints play an important role in ensuring that the expected demand variables are correctly defined given the intricate structure of our model. There are also a number of other constraints that are not present in the SBMILP, including the second-order cone constraints (2.11b), which are needed for the correct definition of the variables that correspond to the standard deviations of the demands. Ultimately, the significance of each constraint becomes clear through the course of the proof of Theorem 2.2 (see Appendix B.3). In the following section, we will present heuristic algorithms that make it possible to obtain high-quality solutions to the DB-MISOCP in cases where finding the exact solutions is not computationally feasible.

2.6   Heuristic Methods

2.6.1   Formulation and Theoretical Results

Although the DB-MISOCP can be handled by off-the-shelf solvers, solving this problem becomes very computationally demanding in the context of a large-scale omnichannel environment. For this reason, we have implemented two basic heuristic approaches that make it possible to obtain high-quality solutions to the DB-MISOCP when finding exact solutions is not computationally feasible.

These heuristic methods are based on two types of relaxations of problem (2.11). Consider the following two relaxations:

> **Relaxation (R1):** Derived from formulation (2.11) by removing binary variables z_{cj} together with the corresponding constraints (2.11f), (2.11g), (2.11j), and (2.11m);
>
> **Relaxation (R2):** Derived from formulation (2.11) by relaxing binary variables z_{cj}, i.e., by replacing constraints (2.11m) with the following constraints:
>
> $$0 \leq z_{cj} \leq 1 \qquad \forall c \in \mathcal{C}, \ j \in \mathcal{N}.$$

The first heuristic method is a generalization of the algorithm proposed by Vasilyev et al. (2023) for solving the SBMILP. This approach involves solving the relaxed problem (R1) and using the obtained solution to fix some of the assortment decisions. Subsequently, the remaining assortment decisions are derived by solving problem (2.11), in which some of the binary variables are replaced by fixed parameters determined in the previous step. The second heuristic method is based on the standard continuous relaxation of problem (2.11). The heuristic solution is obtained by rounding the values of relaxed variables z_{cj} to the nearest integers. A formal description of these two heuristic methods is provided below.

<div style="text-align:center">

Heuristic Method 1.   **Heuristic Method 2.**

</div>

1. Solve the relaxed problem (R1). Let $\left\{\hat{x}_{c0}^{(c)}, \hat{x}_{cj}^{(c)}\right\}_{c \in \mathcal{C}, j \in \mathcal{N}, \bar{c} \in \mathcal{C}_{cj}}$ be part of the obtained optimal solution, and $\mathcal{J}^{(c)}$ be the set of indices $j \in \mathcal{N}$ such that either $\dfrac{\hat{x}_{cj}^{(c)}}{v_j^{(c)}} = \dfrac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}}$ or $\hat{x}_{cj}^{(c)} = 0$.

2. Solve problem (2.11) with the following additional constraints:

$$z_{cj} = \mathbb{1}_{\hat{x}_{cj}^{(c)} > 0} \quad \forall c \in \mathcal{C}, \ j \in \mathcal{J}^{(c)},$$

where $\mathbb{1}$ represents the indicator function. The obtained solution is the heuristic output.

1. Solve the relaxed problem (R2). Let $\left\{\hat{z}_{cj}\right\}_{c \in \mathcal{C}, j \in \mathcal{N}}$ be part of the obtained optimal solution.

2. Round each $\hat{z}_{cj}$ to the nearest integer. The obtained binary values represent the heuristic assortment decisions for different channels. The values of all other variables can be unambiguously determined given the assortment decisions.

Interestingly, the two relaxations of the DB-MISOCP used in the heuristic methods are equivalent if the price and cost parameters satisfy certain conditions. First, we require the following theoretical statement:

Lemma 2.1. *For any $c \in \mathcal{C}$, $\bar{c} \in \mathcal{C}_{cj}$, $j \in \mathcal{N}$, if the following condition is satisfied:*

$$\frac{r_{cj} - b_{cj}}{r_{cj} + h_{cj}} \geq \frac{\sigma^{(c)}}{\mu^{(c)}} \varphi\left(\frac{r_{cj} - b_{cj}}{r_{cj} + h_{cj}}\right), \tag{2.13}$$

then either constraint of type (2.11d) or constraint of type (2.11k) in the DB-MISOCP relaxation (R1) is binding.

The proof of this lemma can be found in Appendix B.4. In practice, condition (2.13) is not overly restrictive. For $\dfrac{\sigma^{(c)}}{\mu^{(c)}} = CV_{max} = 0.3$, this condition translates into $\dfrac{r_{cj} - b_{cj}}{r_{cj} + h_{cj}} \gtrapprox 0.119$. Such an inequality is satisfied if, for example, $b_{cj} = 0.8 r_{cj}$ and $h_{cj} = 0.65 r_{cj}$, in which case $\dfrac{r_{cj} - b_{cj}}{r_{cj} + h_{cj}} \approx 0.121$. For smaller values of the coefficient of variation, condition (2.13) is even milder. Now, we can prove the following theorem:

Theorem 2.3. *If condition (2.13) is satisfied for all $c \in \mathcal{C}$, $\bar{c} \in \mathcal{C}_{cj}$, $j \in \mathcal{N}$, then problem (R1) is equivalent to problem (R2).*

A complete proof of this theorem is given in Appendix B.5. The theorem statement is further developed by the following proposition:

Proposition 2.2. *The DB-MISOCP is equivalent to the relaxed problems (R1) and (R2) if the following conditions are satisfied:*

- *condition (2.13) holds for all $c \in \mathcal{C}$, $\bar{c} \in \mathcal{C}_{cj}$, $j \in \mathcal{N}$;*
- *customers do not switch between channels, i.e., if for any $c \in \mathcal{C}$, $j \in \mathcal{N}$, either $\phi_j^{(c)} = 0$ or $\delta_{\bar{c}j}^{(c)} = 0$ $\forall \bar{c} \in \mathcal{C} \backslash \{c\}$.*

The equivalence of the DB-MISOCP and relaxed problems (R1) and (R2) implies that heuristic methods 1 and 2 are exact.

The proof of this proposition is available in Appendix B.6. In addition to offering clear theoretical insights, Proposition 2.2 also provides some intuition about the performance of our heuristic methods in certain scenarios. Specifically, the two heuristic methods are expected to perform especially well if the number of customers willing to switch to other channels in case of absence of their desired product is low. In the next subsection, we conduct a numerical analysis to evaluate the performance of the heuristic methods on a number of generated problem instances.

2.6.2 Heuristics Performance Analysis

In this subsection, we examine how the two heuristic methods perform in terms of both solution quality and computational time. Let us first describe the parameter generation procedure for our numerical experiments. Suppose that $u(0,1)$ denotes a random value drawn from the standard uniform distribution $\mathcal{U}(0,1)$ and $u(0,1,q)$ denotes a vector of q random values drawn from the same distribution. Furthermore, let ε be a small positive value, e.g., $\varepsilon = 0.01$. For all $c \in \mathcal{C}$, $j \in \mathcal{N}$, we generate the MAM parameters in the following way:

- $v_j^{(c)} = u(0,1) + \varepsilon$;
- $\phi_j^{(c)} = u(0,1)v_j^{(1)}$;
- $\delta_{cj}^{(\bar{c})} = \dfrac{\rho_{cj}^{(\bar{c})}}{\sum\limits_{d \in \mathcal{C}_{cj}} \rho_{cj}^{(d)} + u(0,1)}$, where $\left(\rho_{cj}^{(d)}\right)_{d \in \mathcal{C}_{cj}} = u(0,1,|\mathcal{C}_{cj}|)$.

Next, we assume that for any product, prices are the same across all channels, i.e., $r_{cj} = r_j$ $\forall c \in \mathcal{C}$, $j \in \mathcal{N}$. This assumption is highly appropriate for an omnichannel environment since maintaining consistent prices across channels can be viewed as one of the potential omnichannel initiatives that enhance the overall customer experience. We then generate

problem size	average solving time (s)			average objective value		
	exact	heuristic 1	heuristic 2	exact	heuristic 1	heuristic 2
$n = 20, m = 5$	0.809	0.457	0.025	5403.8	5403.4	5356.4
$n = 30, m = 5$	2.110	1.356	0.035	5484.1	5484.0	5430.1
$n = 40, m = 5$	5.598	3.555	0.046	5549.0	5548.6	5493.6
$n = 50, m = 5$	11.681	6.106	0.057	5613.9	5613.3	5548.7
$n = 60, m = 5$	26.199	8.274	0.067	5707.2	5706.9	5645.3
$n = 20, m = 3$	0.149	0.114	0.014	2706.4	2706.4	2690.0
$n = 20, m = 6$	1.318	0.902	0.045	6644.8	6644.2	6574.4
$n = 20, m = 9$	5.743	3.395	0.086	10698	10697	10569
$n = 20, m = 12$	10.196	7.793	0.151	14992	14991	14766
$n = 20, m = 15$	16.010	10.288	0.203	18978	18976	18684

Table 2.1: Performance of the heuristic methods compared to the exact method.

prices as $r_j = u(0,1) + \varepsilon$. We also set the values of the purchasing costs to $b_{cj} = 0.8r_j$ $\forall c \in \mathcal{C}$, and the values of the holding costs to $h_{1j} = 0.4r_j$ and $h_{cj} = 0.2r_j$ $\forall c \in \mathcal{C}\backslash\{1\}$. Furthermore, we assume that the numbers of customers are given by random variables $\xi^{(c)} \sim \text{NORMAL}(\mu^{(c)}, (\sigma^{(c)})^2)$, where $\mu^{(1)} = 10^4(m - 1)$, $\mu^{(c)} = 10^4$ $\forall c \in \mathcal{C}\backslash\{1\}$, and $CV(\xi^{(c)}) = \sigma^{(c)}/\mu^{(c)} = 0.15$ $\forall c \in \mathcal{C}$. Note that this assumption implies that the total expected number of online customers is the same as the total number of customers in physical stores, which helps us to avoid biases in our estimates. Finally, while we have described a generalized procedure for generating parameters $\delta_{cj}^{(\bar{c})}$, we will focus on the case with geographically distributed physical stores for this part of our numerical experiments. Similar to Section 2.4.2, in this setting customers can switch between a physical store and the online store, but not between different physical stores. We will utilize the generalized procedure in the next section where we analyze the impact of the density of the physical store network on properties of optimal assortments.

We start by considering relatively small problems that can be effectively solved using the DB-MISOCP formulation. We first set the number of channels $m = |\mathcal{C}|$ to 5 and modify the number of products n within the range of 20 to 60 with a step of 10. As a result, the total number of assortment decisions that the retailer has to make varies from 100 to 300, which translates into 2^{100} to 2^{300} potential assortments. We then fix the number of products n at 20 and change the value of m from 3 to 15 with a step of 3, meaning that the total number of retailer's assortment decisions changes from 60 to 300, and the number of potential assortments changes from 2^{60} to 2^{300}. For each pair (n, m), we generate 100 problem instances, and solve them using the exact method and the two heuristic methods.

(a) Gaps for different numbers of products with the number of stores fixed at $m = 5$.

(b) Gaps for different numbers of stores with the number of products fixed to $n = 50$.
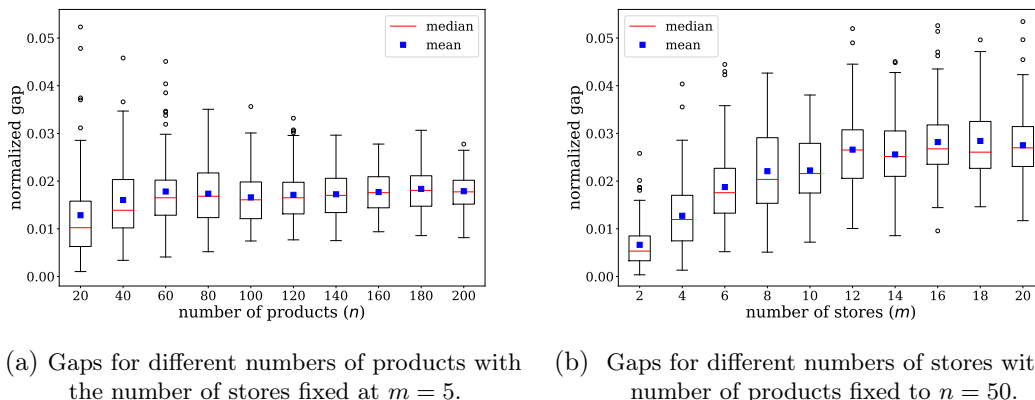
Figure 2.3: Normalized gaps between profits yielded by heuristic method 2 and the upper bounds obtained by solving a linear programming relaxation of the DB-MISOCP.

For all generated problem instances, we record the objective values yielded by the three methods as well as their solving times. All numerical experiments were carried out on a laptop with AMD Ryzen 5 4600H CPU, 16 GB RAM and 64-bit Windows 10 OS. The optimization problems were solved using Gurobi version 10.0 (Gurobi Optimization, LLC, 2023). The results averaged over all simulated instances are provided in Table 2.1.

We can observe that the first heuristic method consistently yields near-optimal solutions and reduces the average solving time nearly twofold. However, this gain in computational efficiency is not sufficient for solving large problem instances. On the other hand, the second heuristic method is substantially more computationally efficient as it only requires solving a continuous second-order cone program. Although it falls slightly short of the performance of the first heuristic method in terms of solution quality, it still shows excellent performance. Overall, if the assortment optimization problem is relatively small, then the best solution approach is to first try to solve this problem exactly using the DB-MISOCP formulation. If the problem turns out to be too large and the solving time exceeds a reasonable limit, then the first heuristic method should be applied. If this method still cannot handle the given problem, then the second heuristic method should be used.

Let us now explore the performance of heuristic method 2 on a set of larger problems. In particular, we assess how the method performs as the problem size increases. First, we set the value of m to 5 and vary the value of n from 20 to 200 with the step of 20. Next, we set the value of n to 50 and vary the value of m from 2 to 20 with a step of 2. Using the procedure described above, we generate 100 problem instances for each pair (n, m). Following the heuristics performance analysis carried out by Vasilyev et al. (2023), we

compute the value of the normalized gap $(obj_{R2} - obj_{H2})/obj_{H2}$ for each problem instance, where obj_{R2} is the value of relaxation (R2) and obj_{H2} is the value of heuristic method 2. This gap upper bounds the relative optimality gap between the optimal objective value and the value of the heuristic, and serves as a proxy for the heuristic's performance. The results of our numerical experiments are presented in Figure 2.3. Figure 2.3a shows that as the number of products increases, the normalized gaps initially increase on average but quickly stabilize once the number of products reaches $n = 60$. Interestingly, the spread of gaps appears to decrease as the number of products increases, which demonstrates the robustness of the performance of heuristic method 2. On average, the gap values are below 0.02, which is an excellent result given that the value of relaxation (R2) is only an upper bound to the optimal objective value. As for the performance of the heuristic method as the number of stores increases, Figure 2.3b indicates that both the average value and spread of the gaps initially grow, but reach a plateau when the number of stores becomes $m = 12$. On average, the normalized gaps between the values of heuristic method 2 and relaxation (R2) are below 0.03. These findings, along with the results obtained for the fixed value of m and a growing number of products n, confirm the potential of the second heuristic method in generating high-quality solutions to the assortment optimization problem under MAM with inventory considerations.

2.7 Numerical Study and Managerial Insights

In this section, we carry out numerical experiments to extract key managerial insights from our modeling framework. We start by exploring how optimal assortment sizes are impacted by changes in certain model parameters. We consider problems with the number of products $n = 20$ and the number of stores $m = 5$. Although these problems are of a relatively small size, they are sufficient to provide insights into general trends. We generate the model parameters using the procedure described at the beginning of Section 2.6.2, with a few exceptions that will be explicitly specified.

We begin our analysis by investigating the impact of the coefficient of variation of the demand on the sizes of optimal assortments. Instead of using fixed values $CV(\xi^{(c)}) = 0.15$ $\forall c \in \mathcal{C}$ as outlined at the beginning of Section 2.6.2, we examine how optimal assortment sizes change when the coefficients of variation range from $CV(\xi^{(c)}) = 0$ $\forall c \in \mathcal{C}$ to $CV(\xi^{(c)}) = 0.3$ $\forall c \in \mathcal{C}$, assuming that parameters $\mu^{(c)}$ $\forall c \in \mathcal{C}$ remain constant (i.e., the total expected number of customers does not change). We generate 100 problem instances for each parameter configuration and solve them exactly using the DB-MISOCP formulation. Figure 2.4 illustrates the results of these numerical experiments. Specifically, the left-hand side plot shows the evolution of assortment sizes averaged over all simulated
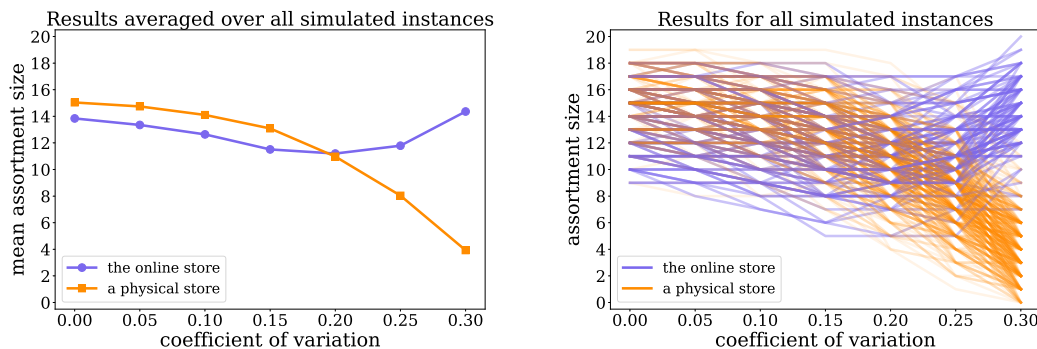
Figure 2.4: Effect of values of coefficients of variation $CV(\xi^{(c)})$ $\forall c \in \mathcal{C}$ on optimal assortment sizes.
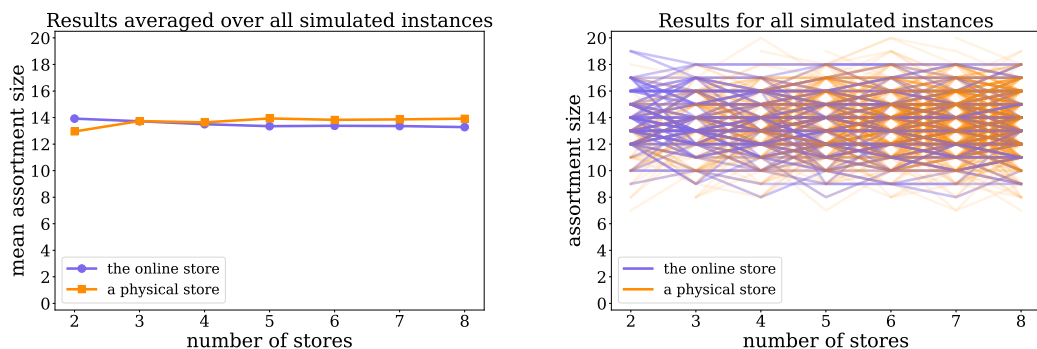
Figure 2.5: Effect of the number of physical stores on optimal assortment sizes.

instances, while the right-hand side plot displays the evolution of assortment sizes for all individual problem instances (with each line representing the assortment size evolution in one store for one generated set of parameters). Note that all figures in the first part of this section will have the same layout. We observe that as the coefficient of variation of the demand increases, the sizes of the optimal assortments in physical stores decrease. This outcome is expected since an increase in the coefficient of variation of the demand may make some products less appealing to offer due to inventory costs, resulting in smaller optimal assortments. As for the optimal assortment sizes in online stores, we notice an intriguing pattern. When the coefficient of variation grows, the online assortment size initially decreases but then starts to increase, sometimes even surpassing the initial assortment size derived for a zero coefficient of variation. This phenomenon can be attributed to two different effects. Initially, online assortment sizes shrink since some products become less appealing to offer due to inventory costs, which mirrors the behavior of assortments in physical stores. However, the demand pooling effect eventually comes into play. As the retailer offers fewer products in physical stores, it becomes profitable to make some products available online, capturing a portion of the demand from physical

stores. The adverse effects of a higher coefficient of variation are thus mitigated by the demand pooling effect. These findings are consistent with the conclusions drawn from a simple illustrative example provided in Section 2.4.2. It is also worth noting that the observed trends apply not only to average values but also to all individual problem instances. Geunes and Su (2020) studied a similar question concerning the impact of the coefficient of variation of demand on the optimal assortment sizes for a dual-channel retailer. Interestingly, their work showed that both online and physical assortment sizes decrease as the coefficient of variation increases, meaning that the authors observed only the inventory cost effect and not the demand pooling effect. This can be explained by the fact that the authors considered a setting with just two channels, which makes the demand pooling effect significantly less impactful.

Next, we examine the effect of the number of physical stores on the optimal assortment sizes. As previously, we generate the parameters using the procedure described at the beginning of Section 2.4.2. We initially assume that $n = 20$ and $m = 2$, and subsequently vary the value of m from 2 to 8. Importantly, as we add physical stores one by one, we retain the parameters for previously added stores and only generate parameters for the newly added stores. The evolution of assortment sizes can be seen in Figure 2.5. We do not observe any apparent upward or downward trends. This becomes particularly evident when analyzing the results for individual problem instances (see the plot on the right-hand side of Figure 2.5). At the same time, we find that adding a physical store can have a substantial effect (either positive or negative) on the sizes of optimal assortments in other stores.

Furthermore, we investigate the effect of the density of the physical store network on the optimal assortment sizes. We return to the setting where the number of stores is $m = 5$ and consider four network density values: 0, 1/3, 2/3, and 1. If the density is 0, the physical stores are assumed to be connected to the online store but not to each other. On the other hand, if the network density is 1, all stores are assumed to be interconnected, meaning that customers can switch between any two stores. For each density value, we simulate 100 problem instances. When the density is 1/3 or 2/3, for each problem instance we randomly generate edges in the physical store network until the target density is reached. Once the network is formed, we generate parameters $\delta_{cj}^{(\bar{c})}$ in the way described at the beginning of the previous section. Our numerical experiment results are presented in Figure 2.6. Similar to the case with varying numbers of stores, no apparent trends emerge for the optimal assortment sizes. We also observe substantial variation in optimal assortment sizes when examining the results for individual problem instances. This is not surprising, because modifying the network structure requires us to generate new parameters $\delta_{cj}^{(\bar{c})}$, which has a considerable impact on the optimal solution.
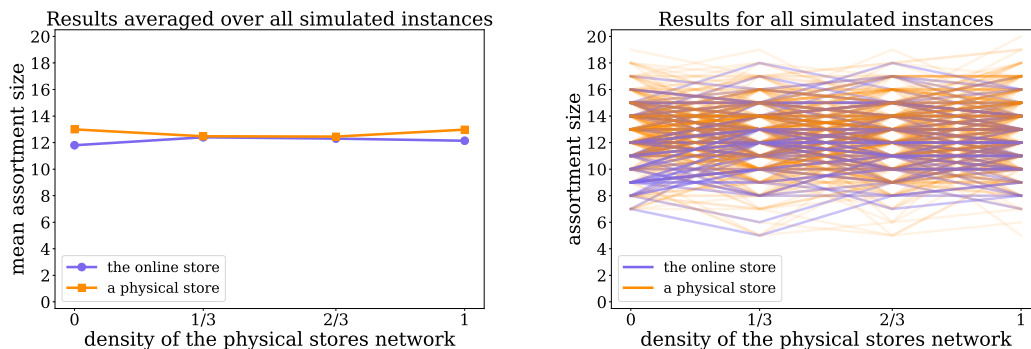
Figure 2.6: Effect of the density of the physical store network on optimal assortment sizes.
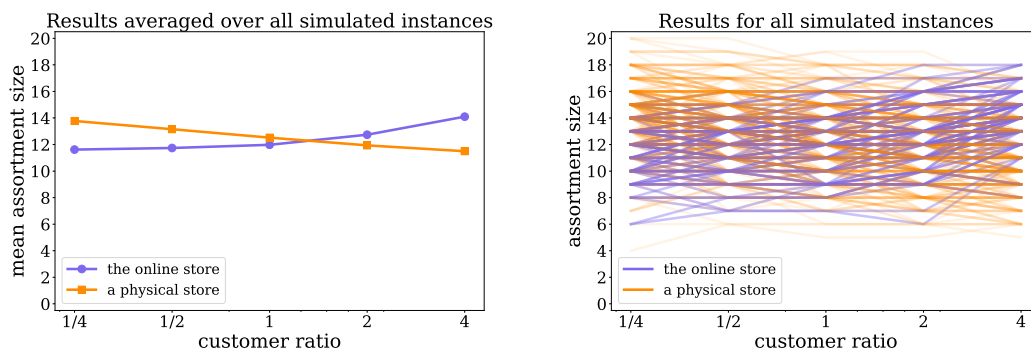
Figure 2.7: Effect of the ratio of the number of physical customers to the number of online customers on optimal assortment sizes.

Lastly, we explore what happens with the optimal assortment sizes when the ratio of the expected number of physical customers to the expected number of online customers increases, i.e., when $\sum_{c \in \mathcal{C} \setminus \{1\}} \mu^{(c)}/\mu^{(1)}$ grows. In our previous numerical experiments, we generated parameters in a way that this ratio equaled 1. We now examine the effects on optimal assortment sizes when this ratio varies from 1/4 to 4, while keeping the total expected number of customers $\sum_{c \in \mathcal{C}} \mu^{(c)}$ constant. We continue to assume that physical stores are identical in size, i.e., $\mu^{(c)} = \mu^{(d)} \; \forall c, d \in \mathcal{C} \setminus \{1\}$. The results are displayed in Figure 2.7 (note that the x-axes are log-scaled in both plots). We observe that as the considered ratio grows, the online assortment size expands while the physical assortment sizes shrink. This is an intuitively clear result, as inventory costs are assumed to be lower in the online store, hence the retailer aims to attract more online traffic by increasing the online assortment. These findings align with those of Vasilyev et al. (2023), who also explored the effect of the ratio of the number of physical customers to the number of online customers on optimal assortment sizes. However, as mentioned previously, Vasilyev et al. (2023) only considered a setting with two channels and deterministic demands, and
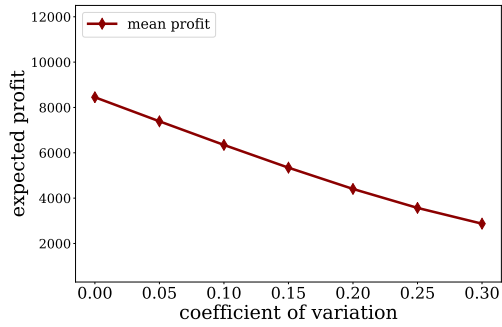
(a) Profit sensitivity with respect to the values of coefficients of variation $\dfrac{\sigma^{(c)}}{\mu^{(c)}} \ \forall c \in \mathcal{C}$.

(b) Profit sensitivity with respect to the number of stores.
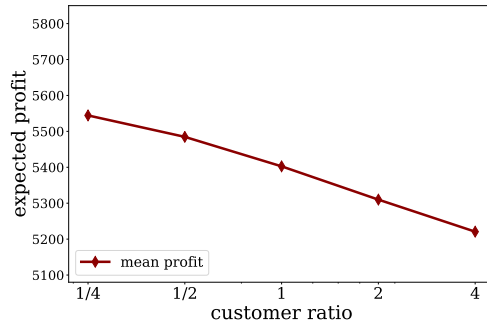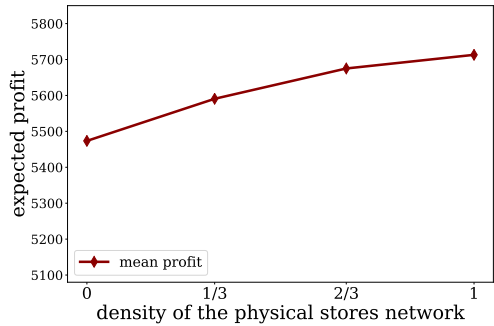
(c) Profit sensitivity with respect to the density of physical store network.

(d) Profit sensitivity with respect to the ratio of the number of physical customers to the number of online customers.

Figure 2.8: Expected profit sensitivity.

did not explicitly account for inventory costs.

In addition to analyzing the evolution of optimal assortment sizes, we have also examined the effects of the considered model parameter changes on the average objective values (i.e., on the average profits of the retailer). These effects are illustrated in Figure 2.8. Figure 2.8a demonstrates that the average expected profit decreases as the coefficients of variation of the demands generated by customers of each type increase. Naturally, a higher coefficient of variation of the demand leads to a lower expected profit per product due to an increased probability of incurring lost sales or carrying excessive inventory. Next, when the retailer opens additional physical stores, the expected profit grows as a result of an expanding customer base (see Figure 2.8b). Remarkably, the magnitude of changes in expected profit resulting from modifying the coefficient of variation is comparable to the magnitude of changes caused by adding more physical stores, even though the expected number of customers remains constant in the former case. This emphasizes the importance of accounting for the coefficient of variation of the demand. Figure 2.8c reveals that on
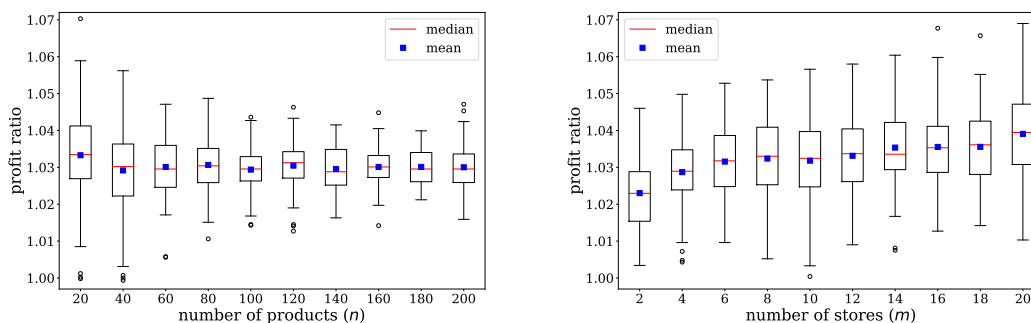
(a) Profit ratios for different numbers of products.   (b) Profit ratios for different numbers of stores.

Figure 2.9: Comparison of omnichannel solutions yielded by heuristic method 2 to the siloed-channel solutions.

average, the expected profit slightly increases as the density of the physical store network grows. This is also not surprising since customers' willingness to switch between different physical stores implies a higher probability of retaining those customers who cannot find desired products in their preferred stores. However, this effect on the total expected profit is less pronounced than the previously considered effects (note that the y-axis scale in Figures 2.8a and 2.8b is different from the y-axis scale in Figures 2.8c and 2.8d), because the expected number of customers in each channel remains stable. On average, increasing the network density from 0 to 1 results in a profit gain of approximately 4.7%. Lastly, as shown in Figure 2.8d, the average expected profit declines as the ratio of the number of physical customers to the number of online customers decreases. Similar to the case with the varying density of the physical store network, this effect is relatively minor, with a profit loss of about 5.8% as the ratio changes from 1/4 to 4. The observed trend is in line with our expectations, as a product offered online generally results in higher profit compared to a product offered in a physical store, all else being equal (due to the assumption of lower inventory costs in the online channel).

Finally, we analyze the potential benefits of optimizing assortments in an omnichannel environment as opposed to optimizing assortments in siloed channels. To obtain more realistic estimates, we focus on a set of larger problems considered in the end of Section 2.6.2. We compare the solutions yielded by heuristic method 2 to those derived from solving a series of assortment optimization problems formulated for individual channels. Following Vasilyev et al. (2023), we obtain the latter solutions by leveraging the fact that the MAM restricted to choices of type-c customers in channel c is equivalent to the GAM. Therefore, for each channel, we can solve the assortment optimization problem under the GAM independently from other channels. Once omnichannel and siloed-channel solutions (assortments) are obtained, we compute the expected profits assuming the same

underlying demand model. The ratios of the profits yielded by omnichannel solutions to those yielded by siloed-channel solutions for different problem sizes are presented in Figure 2.9. First, we observe that as the number of products n increases, the average profit ratio remains stable, whereas the spread of profit ratios decreases up to a certain limit. Meanwhile, as the number of stores m increases, the average profit ratio grows (although the growth rate diminishes with larger values of m), while the spread of profit ratios remains stable. In some rare cases, these ratios take values slightly below 1, which indicates certain suboptimality of the chosen heuristic method. The average profit ratio across all our numerical experiments is slightly above 3%, and it can be as high as 7%. These estimates are at the lower end of the spectrum of potential profit gains. As shown in Section 2.6.2, heuristic method 2 – despite showing good performance – still yields suboptimal decisions. Moreover, we consider the setting with isolated physical stores, and one could expect even more substantial profit gains in the case of a densely connected physical store network.

2.8   Conclusions and Future Work

In this paper, we addressed the omnichannel assortment optimization problem under stochastic demand and inventory considerations. Our approach builds upon the multichannel attraction model (MAM) proposed by Vasilyev et al. (2023), which allows us to model customer choices in an omnichannel environment. In their work, the authors essentially assumed deterministic demand and focused on the case with a dual-channel retailer. We extended this modeling framework by including stochastic demand, accounting for the retailer's optimal inventory decisions, and considering a generalized setting where an omnichannel retailer manages an online store and a chain of physical stores. Our study highlighted the importance of incorporating demand stochasticity into the modeling framework. We showed that ignoring the coefficient of variation of the demand can lead to suboptimal assortment decisions due to the demand pooling effect. We also proved that the assortment optimization problem is NP-hard (and, moreover, APX-hard) in the number of stores even in the setting with a single product and deterministic demand. We then formulated the assortment optimization problem as a mixed-integer second-order cone program that we called the DB-MISOCP. Additionally, we presented two heuristic algorithms based on two different relaxations of the DB-MISOCP. We derived the conditions under which the two relaxations are equivalent to each other, and the conditions under which they are also equivalent to the DB-MISOCP. We showed numerically that our first heuristic method consistently yields near-optimal solutions while reducing the average solving times nearly twofold. We also observed that our second heuristic method falls slightly short compared to the first heuristic method in terms of solution quality,

but it demonstrates great computational efficiency, making it suitable for solving the assortment optimization problem in large-scale omnichannel environments. Furthermore, we performed an extensive numerical analysis to gather managerial insights. In particular, we analyzed the complex effect of the coefficient of variation of the demand on optimal assortment sizes. We demonstrated that an increasing coefficient of variation initially leads to a decrease in the online assortment size due to increasing costs, followed by an increase in the online assortment size due to the demand pooling effect. Last but not least, we showed that omnichannel assortment optimization brings a clear gain in expected profit compared to assortment optimization in siloed channels.

Our work gives rise to several interesting research questions. First, both heuristic methods show great performance, with the first method being particularly accurate. It would thus be interesting to theoretically justify why the first heuristic method yields such high-quality solutions. This could lead to the development of a more advanced heuristic method based on the DB-MISOCP formulation, which could further enhance solution quality and reduce solving time. Another promising avenue would be to expand the modeling framework to include additional dimensions, such as capacity constraints and pricing decisions. This would allow the framework to better capture the complexities of real-world scenarios and improve its applicability to practical situations. Finally, since demand variability has a substantial impact on the structure of optimal assortments, it would be interesting to explore a robust formulation of the assortment optimization problem for cases with highly uncertain demand. Ultimately, our study contributes to the literature on omnichannel assortment optimization and lays a foundation for future research in the field of omnichannel decision-making.

3 Omnichannel Assortment Optimization Given Basket Shopping Behavior

3.1 Introduction

Optimizing product assortment can play a crucial role in increasing a retailer's earnings by both reducing operational costs and redirecting the demand to more profitable products. The critical aspect lies in deciding what products to eliminate from an existing assortment because removing a single product can have a substantial impact on demand allocation for the remaining products. For example, if a retailer decides to discard a low-demand or low-margin product, it may accidentally eliminate an item that attracts customers to shop at this retailer and also buy other products (Timonina-Farkas et al., 2020). The importance of this phenomenon cannot be understated. In 2009, Walmart cut 15% of its inventory, with dire consequences: "Sales declined for seven consecutive quarters as shoppers took their entire shopping lists elsewhere. By April 2011, Walmart added 8,500 SKUs back to its mix, an average of 11% of its products" (Pearson, 2015). This example clearly shows how ignoring basket shopping behavior can result in suboptimal assortment decisions.

The majority of assortment optimization models assume that each customer purchases at most one product. Such an approach generally allows researchers to account only for the product substitution effect, but not for the complementarity effect. In other words, excluding a product from the assortment can only increase the probability of customers purchasing other products instead. Discrete choice models satisfying this property are often referred to as regular choice models (see, e.g., Berbeglia and Joret, 2020). However, in many industries where customers tend to buy baskets of products, the complementarity effect cannot be ignored: Eliminating a product from the assortment can reduce the probability of customers purchasing other products. There is an evident need for suitable methodologies designed to support assortment decision-making in such complex environments. This is clearly reflected in recent interviews with practitioners

that indicated that most retailers are willing to adopt more advanced analytics in the area of assortment planning (Rooderkerk, DeHoratius, et al., 2022).

In this paper, we utilize Markov random fields (MRFs) to model basket shopping behavior. In particular, we provide a fresh perspective on the Ising model – which is a special case of an MRF – by showing that it can be viewed as a multi-purchase choice model, i.e., a choice model under which each customer can select a set of products instead of just one product. We consider the Ising model such that each node represents a random variable indicating whether a product belongs to the basket, and an edge between two nodes represents the dependency between the two random variables. The joint probability distribution of such random variables determines the choice probabilities for all possible baskets from a given assortment. Note that we assume that each basket consists of unique products, which is a standard assumption for multi-purchase choice models (see Section 3.3.2 for more details). Under this assumption, the term "basket" is sometimes replaced with other terms such as "bundle" (see, e.g., Tulabandhula et al., 2020) or "subset" (see, e.g., Benson et al., 2018). However, in this paper we refer to customer choices as baskets since this is intuitive terminology that improves the clarity of exposition.

One of the contributions of this paper is that it highlights the connection between the Ising model and the multivariate logit model (MVL), which is one of the most prominent techniques used for modeling basket shopping behavior. In the classic formulation of the Ising model, the random variables take spin values, i.e., -1 or 1. However, here we assume that these random variables are binary. As shown in Section 3.3.2, for a given offered set of products, these two assumptions on the domain of the random variables result in the same choice probabilities up to a parameter transformation. At the same time, the binary values assumption turns the Ising model formulation into the MVL formulation (see Section 3.3.2 for more details). The purpose of this paper is thus not to introduce a novel multi-purchase choice model, but rather to establish the link between MRFs and multi-purhcase choice models. It ultimately allows us to leverage the vast methodology developed for the Ising model for omnichannel assortment optimization taking into account basket shopping behavior.

We first consider a single-channel setting. One of the biggest challenges in modeling basket shopping behavior is obtaining a valid probability distribution over all possible shopping baskets. In particular, computing the normalization coefficient that makes all the probabilities sum up to one is computationally demanding even for relatively small problems. This makes the task of estimating the parameters of such models extremely challenging. In the case of the Ising model, however, we can invoke existing methods that help to alleviate the computational complexity issue and obtain parameter estimates from historical sales data. Next, we formulate the assortment optimization problem under

the Ising model. Using the fact that the problem of computing the partition function of the Ising model is NP-hard (see, e.g., Istrail, 2000), we show that the decision version of the assortment optimization problem is NP-hard as well. We then establish some theoretical results on the structure of the optimal assortments based on the graphical representation of the Ising model which can be used to reduce the dimensionality of the assortment optimization problem. Furthermore, when it comes to assortment optimization, being able to estimate the expected revenue for a given assortment is essential. Directly computing such revenues might not be computationally feasible assuming that the choice probabilities of customers are defined over all possible baskets. However, these values can be approximated, and the methodology developed for the Ising model proves to be beneficial for this task as well. The expected revenues can be estimated by simulating a large number of customer choices, which is equivalent to drawing samples from the Ising model. There are several Markov chain Monte Carlo (MCMC) methods that can be applied to efficiently generate such samples. In this paper, we use systematic scan Gibbs sampling to this end. Lastly, we develop a customized simulated annealing algorithm for finding high-quality solutions to the assortment optimization problem, where each candidate solution is evaluated using the aforementioned simulation procedure.

This paper equally contributes to the literature on omnichannel commerce. Even though the tasks of modeling basket shopping behavior and solving the corresponding assortment optimization problem have previously been addressed in the academic literature, they are currently underrepresented in the context of omnichannel retailing. At the same time, the need to incorporate multiple channels into the modeling framework cannot be understated (Guo and Keskin, 2022). COVID-19 forced traditional brick-and-mortar retailers to develop their online capabilities, and it has become clear that there is no going back once customers become used to a seamless shopping experience (Barr, 2021). Moreover, some of the originally pure online players such as Amazon and the Alibaba Group have been opening physical stores to provide customers with additional options for their shopping journey (Rooderkerk and Kök, 2019). From the methodology perspective, accounting for several retail channels is a challenging task, and often methods developed for a single-channel case cannot be readily generalized to a setting with multiple channels. In this paper, we extend our modeling framework to the omnichannel environment by considering customers of different types. We focus on a two-channel setting (with a physical and an online channel) and differentiate four types of customers depending on their consideration sets. In particular, each customer either shops in one channel exclusively or purchases part of the basket in their primary-choice channel and the remaining part in the other channel. We thus formulate the omnichannel assortment optimization problem as a generalization of the single-channel assortment optimization problem. Furthermore, we generalize some of our theoretical results formulated for a

single-channel setting to the omnichannel setting, and develop a simulated annealing algorithm that can be used to solve the omnichannel assortment optimization problem when finding the exact solution is not computationally feasible. Finally, we carry out a numerical study to obtain insights into the properties of optimal assortments and to evaluate the benefits of accounting for omnichannel shopping behavior of customers as opposed to considering isolated channels.

The remainder of this paper is organized as follows. In Section 3.2, we review the relevant literature with a primary focus on modeling techniques related to multi-purchase shopping behavior. In Section 3.3, we provide the background behind our modeling framework and highlight the connection between the Ising model and the MVL. In Section 3.4, we describe a method that can be used to estimate the model parameters and provide an illustrative example. Section 3.5 is devoted to assortment optimization under the Ising model in a single-channel setting. In Section 3.6, we extend our modeling framework (including the theoretical results obtained for the single-channel setting) to an omnichannel setting and carry out an extensive numerical study. In Section 3.7, we summarize our findings and discuss future research directions.

3.2   Literature Review

Discrete choice modeling is a vast research area that aims to predict customer choices given different sets of alternatives. A common approach to discrete choice modeling implies that each customer selects at most one alternative. This applies to the multinomial logit model (MNL) introduced by Luce (1959), which is arguably the most prominent discrete choice model. Despite being widely adopted in practice, models that assume a single-item customer shopping behavior are less suitable for industries where customers purchase products primarily in baskets. This led to the need to develop methods for modeling the basket shopping behavior of customers. Such models can be separated into three categories: multi-purchase choice models (sometimes referred to as menu or subset selection choice models), multiple-discrete choice models, and multiple discrete-continuous choice models. In multi-purchase choice models, several alternatives can be chosen at the same time, but at most one unit of each alternative can be selected. Multiple-discrete choice models allow an integer number of units of each of the selected alternatives to be chosen. Finally, in multiple discrete-continuous choice models, noninteger amounts of several alternatives can be selected simultaneously. Our focus is on the setting in which customers purchase products in sets, i.e., where each shopping basket consists of unique products. This assumption is fairly justified for many retail industries such as consumer electronics, fashion, toys, etc. It also applies if the goal is to study product categories

rather than products themselves. Therefore, theory on multiple-discrete and multiple discrete-continuous choice models is less relevant to our research. We only refer the reader to the notable works of Hendel (1999), Dubé (2004), Kim et al. (2002), Bhat (2005) and Bhat (2008) in this field.

With regard to multi-purchase choice modeling, one of the most prevalent models is the multivariate logit model (MVL). Hruschka et al. (1999) used the MVL to analyze how cross-category sales promotion effects influence purchase probabilities. The theoretical justification for the MVL was developed by Russell and Petersen (2000), who derived basket choice probabilities from conditional probabilities of purchasing each product given the purchase decisions related to all other products. The authors showed that the only joint distribution that is consistent with the specified conditional choice probabilities is the multivariate logistic distribution (Cox, 1972). Russell and Petersen (2000) applied the MVL to analyze basket purchases in a fairly small setting with four product categories. Later, Boztuğ and Reutterer (2008) proposed a way of applying the MVL in settings of a larger scale. In their paper, the authors first determine prototypical baskets that are used for segmentation of the customer base and then estimate a separate MVL for each customer segment. This approach also makes it possible to account for customer heterogeneity by considering a mixture of the MVL models. Song and Chintagunta (2006) presented the idea of leveraging the MVL framework to model customer choices across different product categories assuming that customers choose at most one product within each category (i.e., products in one category are considered to be strong substitutes). Such an extension of the MVL is often referred to as the multivariate MNL, or MVMNL (see, e.g., Lyu et al., 2022 and Chen, Li, et al., 2022). The MVL can thus be viewed as a special case of the MVMNL where each product category contains a single product. The MVL, however, can model an arbitrarily strong substitution effect between any two products by assigning a sufficiently large negative value to the parameter representing the pairwise demand dependency between these products (see Section 3.3 for parameter definition). An interesting model similar to the MVL was proposed by Benson et al. (2018). In their work, the utility of each basket equals the sum of utilities of individual products in that basket plus an optional correction term. The authors proved that the problem of determining the optimal set of baskets receiving corrective utilities is NP-hard and developed several heuristic algorithms for finding such sets. Overall, the MVL and MVMNL models have been applied in various contexts such as recommendation systems (Moon and Russell, 2008) and pricing under competition (Richards et al., 2018) (See Lyu et al. (2022) for a brief overview of different application areas of these models.) Another classic approach to multi-purchase choice modeling is based on the multivariate probit model (MVP) introduced by Manchanda et al. (1999). In the MVP framework, vectors of unobserved parts of product utilities are assumed to follow the multivariate normal

distribution. The MVP makes it possible to capture the correlation of customer product preferences across different baskets. However, this modeling approach does not consider the complementarity and substitution effects occurring within one purchased basket. In other words, unlike the MVL, the MVP does not account for the fact that purchasing one product may change the marginal value of adding other products to the same basket (Kamakura and Kwak, 2020).

To the best of our knowledge, there are only three papers that address assortment optimization under the MVL model and its variations. Tulabandhula et al. (2020) studied the assortment optimization problem under the BundleMVL-K model – a version of the MVL in which the size of each basket is restricted by an exogenously given constant K. The authors focused on a setting in which each basket consists of at most two products, i.e., $K = 2$. They showed that the decision version of the assortment problem under this model is NP-complete. This powerful result might seem somewhat counterintuitive. Indeed, the BundleMVL-2 specified for n products can be viewed as the MNL defined over the set of n^2 basket alternatives, meaning that the optimal set of baskets (which is generally not translated into the optimal set of products) can be found in polynomial time. The authors proposed a binary search-based heuristic algorithm to solve the assortment problem and compared it against a mixed-integer programming benchmark, as well as two other heuristic algorithms: a greedy approach and the revenue-ordered heuristic (see, e.g., Rusmevichientong et al., 2014). Lyu et al. (2022) addressed the assortment optimization problem under the MVMNL assuming that the parameters representing cross-category demand dependencies are the same for all products in one category. The authors showed that the decision version of the assortment optimization problem is NP-hard even if each category comprises no more than two products and there are no cross-category interaction terms (i.e., the utility of each basket is the sum of utilities of the corresponding product categories). They developed a fully polynomial-time approximation scheme (FPTAS) that can be used to solve this optimization problem. The authors also proposed the generalization of the FPTAS for the case with nonzero cross-category interactions assuming that the maximum number of interacting categories is two. Lastly, they considered a setting with capacity constraints and developed an FPTAS for the capacitated assortment problem by building on the ideas presented for the uncapacitated case. Chen, Li, et al. (2022) studied the assortment optimization problem under the MVMNL with two product categories, i.e., where the size of each basket is at most two. However, in contrast to the work of Tulabandhula et al. (2020), Chen, Li, et al. (2022) explicitly separated the product portfolio into two disjoint product categories. The authors proved that the assortment optimization problem in this setting is strongly NP-hard. They proposed the concept of adjusted-revenue-ordered assortments and showed that the assortment with the highest revenue provides a 0.5-approximation. They also

developed a 0.74-approximation algorithm based on a linear programming relaxation of the assortment optimization problem. Finally, they considered three extensions of the original setting: with capacity constraints, with generally defined basket prices, and with three product categories. They proved that the assortment optimization problems in all these settings do not admit constant-factor approximation algorithms assuming the Exponential Time Hypothesis.

Another stream of literature closely related to our work concerns assortment optimization in omnichannel retailing. An extensive overview of the existing literature on this topic can be found in Vasilyev et al. (2023). Therefore, in this paper we only briefly mention the most relevant works published in the field. Bhatnagar and Syam (2014) presented an integer programming framework for determining the optimal product allocation for a hybrid retailer. The authors imposed the strong assumption of having product demands as fixed model parameters. Dzyabura and Jagabathula (2018) developed a modeling approach to determine the subset of products from the online channel to offer in the physical channel to maximize the aggregate profit. The authors incorporated the impact of in-store product evaluation on customer preferences into the consumer demand model based on the MNL where product utilities depend on their features. Lo and Topaloglu (2022) addressed the same problem as Dzyabura and Jagabathula (2018). However, Lo and Topaloglu (2022) relaxed the assumption made by Dzyabura and Jagabathula (2018) that for every combination of product features there exists a product. Instead, they considered a setting in which the product portfolio can be characterized by a features tree in which each leaf corresponds to a product. Geunes and Su (2020) developed an analytical framework for making joint assortment, inventory, and pricing decisions in online and physical retail channels. The authors modeled customer choices using a mixture of MNLs for different customer segments determined by their consideration sets. The resulting decision problem is a two-stage stochastic optimization problem that can be solved using a simulation-based approximation algorithm. Hense and Hübner (2022) studied the omnichannel assortment, space, and inventory problem under an exogenous demand (ED) model. The ED framework implies that base product demands are pairwise independent, and if one product is not available, then the proportion of the demand redirected to another product is given by a fixed parameter. The authors also incorporated space-elasticity and limited shelf space considerations into their model. They formulated the retailer's profit maximization problem as a multi-knapsack problem with a nonlinear objective function and developed a heuristic algorithm to solve it. Schäfer et al. (2023) addressed a similar problem while accounting for a range of additional demand effects such as product positioning on store shelves. The authors developed a specialized heuristic to solve the resulting nonlinear integer program by generalizing the algorithm presented by Hense and Hübner (2022). Lastly, Vasilyev et al. (2023) presented a discrete choice model

called the multichannel attraction model (MAM) designed specifically for an omnichannel setting. The MAM – which is an extension of the general attraction model introduced by Gallego et al. (2014) – accounts for both the product substitution behavior of customers within each channel and the switching behavior between channels. The authors formulated the corresponding assortment optimization problem as a mixed-integer linear program and developed an efficient heuristic method based on a relaxation of the original problem.

As detailed in Section 3.1, we go beyond the extant research on assortment optimization by explicitly considering basket shopping behavior in an omnichannel setting. To tackle this challenging problem, we adopt methodologies and theoretical results developed for the Ising model. The necessary theoretical background related to the Ising model is given in Sections 3.3.1 and 3.4.

3.3 Model Formulation

In this section, we recall the definition of a Markov random field (MRF) and show that the Ising model can be viewed as a multi-purchase choice model.

3.3.1 Background on Markov Random Fields

The Ising model was introduced by Lenz (1920) and studied in a one-dimensional case in the work of Ising (1925). It marked the beginning of the development of graphical models, including MRFs. For clarity, we start by providing a formal definition of an MRF.

Definition 3.1. *Let $\xi = \{\xi_i\}_{i \in \{1,...,n\}}$ be a random field represented by an undirected graph $G = (V, E)$, where $V = \{v_i\}_{i \in \{1,...,n\}}$ is the set of nodes such that each node v_i is associated with a random variable ξ_i, and E is the set of edges. Then, the random field ξ is called a Markov random field if the joint probability distribution $p(\xi)$ is positive and satisfies the conditional independence property, i.e., each random variable ξ_i is independent of all other variables given its neighbors.*

The MRF model received major attention in the 1970s when the equivalence relation between this model and the Gibbs distribution was established. To define the Gibbs distribution, recall that a subset of nodes of a graph is called a clique if the corresponding subgraph is complete. Let $\mathfrak{L}$ be the set of all cliques of graph G. Suppose that for each clique $l \in \mathfrak{L}$, there exists a strictly positive function ψ_l referred to as the potential function of this clique. Then, the Gibbs distribution can be defined in the following way (see, e.g., Murphy, 2012):

Definition 3.2. *A joint probability distribution p defined over the nodes of graph G is*

called a Gibbs distribution if it can be expressed in terms of potential functions of cliques of graph G in the following way:

$$p(x) = \frac{1}{Z} \prod_{l \in \mathfrak{L}} \psi_r(x),$$

where $Z = \sum_x \prod_{l \in \mathfrak{L}} \psi_l(x)$ is the normalization coefficient (also known as the partition function).

The connection between MRFs and the Gibbs distribution was first studied in the works of Dobruschin (1968) and Spitzer (1971). Subsequently, the equivalence relation between these two concepts was proved in an unpublished paper by Hammersley and Clifford (1971). This important result was later revisited in the work of Besag (1974), where the author restated the Hammersley-Clifford theorem and gave an alternative proof of it. This theorem can be formulated as follows (Hristopulos, 2020):

Theorem (Hammersley-Clifford). *A random field ξ is a Markov random field if and only if the joint probability distribution $p(\xi)$ is a Gibbs distribution.*

Note that the Hammersley-Clifford theorem establishes the equivalence between local properties of a random field (the conditional independence properties) and its global property (being defined by the Gibbs distribution).

The Ising model is a prototypical example of an MRF. It originates from statistical mechanics and is historically defined on a lattice rather than a general graph. However in our case, we do not assume any lattice structure in the graph representation. The Ising model can be defined by the joint distribution of random variables $\xi = \{\xi_i\}_{i \in \{1,\ldots,n\}}$. Let us now specify the probability mass function of this distribution.

Let $\mathcal{N}$ denote the set of indices $\{1,\ldots,n\}$. Suppose that each ξ_i, $i \in \mathcal{N}$ is a binary random variable. For the Ising model, the probability mass function is defined in the following way:

$$p_\theta(x|\mathcal{N}) = \exp\left(\sum_{i \in \mathcal{N}} \theta_{ii}x_i + \sum_{i \neq j,\ i,j \in \mathcal{N}} x_i\theta_{ij}x_j - A_\theta(\mathcal{N})\right), \tag{3.1}$$

where $x \in \mathcal{X}(\mathcal{N}) = \{0,1\}^{|\mathcal{N}|}$ is a realization of the random vector ξ, $\theta \in \mathbb{S}^n$ is a symmetric matrix of parameters, and $A_\theta(\mathcal{N})$ is the logarithm of the partition function given by

$$A_\theta(\mathcal{N}) = \log\left(\sum_{x \in \mathcal{X}(\mathcal{N})} \exp\left(\sum_{i \in \mathcal{N}} \theta_{ii}x_i + \sum_{i \neq j,\ i,j \in \mathcal{N}} x_i\theta_{ij}x_j\right)\right). \tag{3.2}$$

Clearly, distribution (3.1) is a special case of the Gibbs distribution. Therefore, from the Hammersley-Clifford theorem, it follows that the Ising model is an MRF. The conditional independence property can also be verified directly, thereby confirming that the Ising model is an MRF. Importantly, the Ising model can define very complex probability distributions. In fact, any pairwise MRF – i.e., a graphical model defined by expressions similar to (3.1) and (3.2) but with the sums $\sum_i \theta_{ii} x_i$ and $\sum_{i \neq j} x_i \theta_{ij} x_j$ replaced by $\sum_i E_i(x_i)$ and $\sum_{i \neq j} E_{ij}(x_i, x_j)$, respectively, where E_i and E_{ij} are arbitrary real-valued functions – is equivalent to the Ising model with one extra node (see Globerson and Jaakkola, 2006 and Schraudolph and Kamenetsky, 2008).

Now, having provided all the necessary theoretical background, we can offer a novel perspective on the Ising model as a multi-purchase choice model.

3.3.2   Ising Model as a Multi-Purchase Choice Model

As mentioned in Section 3.2, we consider the setting in which each basket consists of unique products, i.e., it is assumed that customers do not purchase more than one unit of each product. Then, basket $\{i_1, \ldots, i_k\} \subseteq \mathcal{N}$ can be represented by vector $x = \{x_1, \ldots, x_n\}$ such that

$$
x_i = \begin{cases} 1 & \text{if } i \in \{i_1, \ldots, i_k\}, \\ 0 & \text{otherwise.} \end{cases}
$$

We can thus consider the Ising model such that each node represents a binary random variable corresponding to a certain product, and an edge between two nodes represents the dependency between the corresponding random variables. The binary value of each random variable indicates whether the corresponding product belongs to the basket. Then, each parameter θ_{ij}, $i \neq j$ represents the pairwise demand dependency between products i and j, and parameter θ_{ii} determines the individual attractiveness of product i. Consequently, the joint probability distribution of these random variables defines the probability of a random customer selecting any basket, i.e., any set of products from $\mathcal{N}$. Finally, formulas (3.1) and (3.2) can be straightforwardly adjusted for the case when only a subset $S \subseteq \mathcal{N}$ of products is offered, thereby defining a multi-purchase choice model:

$$
p_\theta(x|S) = \exp\left(\sum_{i \in S} \theta_{ii} x_i + \sum_{i \neq j, \ i,j \in S} x_i \theta_{ij} x_j - A_\theta(S) \right), \tag{3.3a}
$$

$$
A_\theta(S) = \log\left(\sum_{x \in \mathcal{X}(S)} \exp\left(\sum_{i \in S} \theta_{ii} x_i + \sum_{i \neq j, \ i,j \in S} x_i \theta_{ij} x_j \right) \right), \tag{3.3b}
$$

where x represents a basket, and $\mathcal{X}(S)$ is the set of all possible baskets of products from S.

Importantly, in the classic formulation of the Ising model, random variables are assumed to be spin variables rather than binary variables, i.e., they are assumed to take values -1 or 1. However, for a given offered set S, these two assumptions on the domain of the random variables result in models that are equivalent to each other up to a parameter transformation. In other words, one can easily construct a parameter transformation that maps the choice probabilities yielded by one model onto the choice probabilities yielded by another model.

Proposition 3.1. *Suppose that $\mathcal{M}_1$ is the Ising model with parameters θ defined over binary variables, and $\mathcal{M}_2$ is the Ising model with parameters $\tilde{\theta}$ defined over spin variables. Let $\mathcal{M}_1$ and $\mathcal{M}_2$ define the choice probabilities of customers selecting baskets from a fixed assortment S. If θ and $\tilde{\theta}$ satisfy the following conditions:*

$$
\begin{aligned}
\theta_{ii} &= 2\tilde{\theta}_{ii} - 4 \sum_{j \in S : \, j \neq i} \tilde{\theta}_{ij} &&\forall i \in S, \\
\theta_{ij} &= 4\tilde{\theta}_{ij} &&\forall i, j \in S : \, i \neq j,
\end{aligned}
\tag{3.4}
$$

then the choice probabilities yielded by models $\mathcal{M}_1$ and $\mathcal{M}_2$ are identical.

The formal proof of this proposition can be found in Appendix C.1. If the variables in the Ising model take binary values instead of spin values, then for a given choice set the Ising model takes exactly the same functional form as the multivariate logit model (MVL). Proposition 3.1 allows us to extend this equivalence relation to the Ising model defined over spin variables. Formally, we can formulate the following corollary of Proposition 3.1:

Corollary 3.1. *The following relation between the Ising model and the MVL holds:*

- *For a given choice set, the binary Ising model is equivalent to the MVL;*

- *For a given choice set, the spin Ising model is equivalent to the MVL up to a parameter transformation.*

As previously stated, the purpose of this paper is not to introduce a novel multi-purchase choice model, but to underline the connection (in a certain sense, the equivalence relation) between the Ising model and the MVL. In Section 3.2, we emphasized that the MVL model is one of the most prominent multi-purchase choice models. This model can be theoretically justified in two different ways based on the random utility theory. First, one can specify the conditional utility of purchasing each product given the purchase decisions related to all other products in the assortment and derive the corresponding

conditional choice probabilities. Then, the factorization theorem of Besag (1974) can be used to obtain the unique joint distribution consistent with this conditional probability distribution. We refer the reader to Russell and Petersen (2000), who developed this theoretical justification for the MVL model, and to Tulabandhula et al. (2020), who adapted this approach to the BundleMVL-K model and provided a detailed description of it. Alternatively, one can assume that the utility of a basket represented by vector x is given by the following expression:

$$U_\theta(x|S) = \sum_{i \in S} \theta_{ii} x_i + \sum_{i \neq j, \ i,j \in S} x_i \theta_{ij} x_j + \epsilon_x,$$

where ϵ_x is a random variable representing the unobserved part of the utility. If variables ϵ_x are independent and identically distributed (i.i.d.) for all possible baskets and normalized so that their mean is zero and the variance is $\pi^2/6$, then the MVL can be viewed as an extended version of the MNL, where the set of alternatives is given by the set of all possible baskets and the resulting choice probabilities are given by expressions (3.3). This theoretical justification approach was pursued by, for example, Lyu et al. (2022) for the MVMNL model. The independence assumption might be criticized as being too restrictive since it must hold even for baskets that share some of the products (see, e.g., Tulabandhula et al., 2020). However, this assumption is milder than it might look. Note that the independence assumption applies to the MNL model, despite the fact that some alternatives (i.e., individual products) can share similar attributes. In fact, as Train (2002) pointed out with respect to the MNL model, this assumption can be viewed as a natural consequence of a well-specified model. It implies that for each basket, the constant part of utility is specified sufficiently so that the unobserved part of utility does not provide any information about the unobserved parts of utilities of other alternatives. Therefore, this is an assumption on, above all, the quality of model specification. To summarize, each of the two described theoretical justifications is well grounded and can be viewed as the utility theory foundation underlying the considered model.

Since the Ising model was introduced long before the MVL, we will primarily use the former term when referring to customer choice probabilities. Using such terminology also highlights the fact that most of the methodology utilized in this paper was originally developed for the Ising model, which includes parameter estimation.

3.4 Parameter Estimation and Illustrative Example

For tractability, let us first consider a single-channel setting. The estimation approach described in this section can be easily generalized to the omnichannel environment if we

impose some additional assumptions on the available data (see Section 3.6).

Since we utilize the methodology developed for the classic version of the Ising model – with variables taking spin values – we first describe how to obtain estimates of parameters $\tilde{\theta}$. Then, the estimates of parameters θ for the binary Ising model can be obtained by applying transformation (3.4) to $\tilde{\theta}$. We assume that the only sales history information that is available is the list of purchased baskets assuming that all products are offered, i.e., $S = \mathcal{N}$. Such an assumption makes the parameter estimation method described in this section readily applicable in a range of practical situations. Note that the assortment optimization problem studied later in this paper can thus be viewed as one of finding a subset of the current product portfolio that maximizes the expected profit. Finally, since the assortment is fixed throughout this section (and equals the whole product portfolio), we will slightly abuse the notation by omitting the argument $\mathcal{N}$ of the log-partition function $\tilde{A}_{\tilde{\theta}}$.

Suppose that the available historical data sample comprises a list of purchased baskets $\{a^k\}_{k=1,\dots,m}$, where each basket a^k is given in the form of a spin vector with $a_i^k = 1$ if product $i \in \mathcal{N}$ belongs to basket a_k, otherwise $a_i^k = -1$. Then, under the Ising model with parameters $\tilde{\theta}$, the negative mean log-likelihood of selecting baskets from the given sample is as follows:

$$-LL_{mean}(\tilde{\theta}) = \tilde{A}_{\tilde{\theta}} - \left(\sum_{i \in \mathcal{N}} \tilde{\theta}_{ii} \mu_i + \sum_{i \neq j,\ i,j \in \mathcal{N}} \tilde{\theta}_{ij} s_{ij} \right), \tag{3.5}$$

where $\mu = \frac{1}{m} \sum_{k=1}^{m} a^k$ is the sample mean vector, and $s = \frac{1}{m} \sum_{k=1}^{m} a^k (a^k)^T$ is the sample second-order moment matrix. The model parameters can be learned by minimizing the right-hand side of (3.5) over $\tilde{\theta}$. Moreover, the matrix of off-diagonal entries of $\tilde{\theta}$ can be sparsified by using ℓ_1-regularization, i.e., by adding a ℓ_1-norm penalty term to the corresponding optimization problem. This allows us to focus on the most important pairwise dependencies while removing nonmeaningful interactions from consideration. Formally, sparse estimates of parameters $\tilde{\theta}$ can be obtained by solving the following problem:

$$\min_{\tilde{\theta} \in \mathbb{S}^n} \tilde{A}_{\tilde{\theta}} - \left(\sum_{i \in \mathcal{N}} \tilde{\theta}_{ii} \mu_i + \sum_{i \neq j,\ i,j \in \mathcal{N}} \tilde{\theta}_{ij} s_{ij} \right) + \rho ||\tilde{\theta} - \text{diag}(\tilde{\theta})||_1, \tag{3.6}$$

where $\rho \geq 0$ is the penalty weight. The greater the value of ρ, the fewer nonzero off-diagonal entries in matrix $\tilde{\theta}$ and the fewer edges in the corresponding network.

The log-partition function $\tilde{A}_{\tilde{\theta}}$ contains an exponential number of terms, making optimization problem (3.6) extremely computationally challenging. In fact, even the problem of

computing the partition function for a given $\tilde{\theta}$ is NP-hard in general, which was proved by Barahona (1982). Istrail (2000) extended the work of Barahona (1982) and showed that the computational complexity of this problem arises from the topology of the underlying graphical model: If the corresponding graph is nonplanar, then the problem is NP-hard. Since we do not make any restrictive assumptions on the topological structure of the Ising model, the presence of the (log-) partition function poses a challenging problem. It can be overcome by the following result obtained for the Ising model. Wainwright and Jordan (2006) showed that the log-partition function can be upper bounded by the solution to a certain variational problem, and Banerjee et al. (2008) proved that this upper bound can be rewritten as follows:

$$\widehat{\widetilde{A}}_{\tilde{\theta}} = \frac{n}{2} \log\left(\frac{e\pi}{2}\right) - \frac{1}{2}(n+1) - \frac{1}{2}\left(\max_{v \in \mathbb{R}^{n+1}} v^T q + \log \det\left(-Q(\tilde{\theta}) - \mathrm{diag}(v)\right)\right), \quad (3.7)$$

where $q = (1, 4/3, \ldots, 4/3)^T \in \mathbb{R}^{n+1}$ and

$$Q(\tilde{\theta}) = \begin{pmatrix} 0 & \tilde{\theta}_{11} & \tilde{\theta}_{22} & \ldots & \tilde{\theta}_{nn} \\ \tilde{\theta}_{11} & 0 & 2\tilde{\theta}_{12} & \ldots & 2\tilde{\theta}_{1n} \\ \tilde{\theta}_{22} & 2\tilde{\theta}_{21} & 0 & \ldots & 2\tilde{\theta}_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \tilde{\theta}_{nn} & 2\tilde{\theta}_{n1} & 2\tilde{\theta}_{n2} & \ldots & 0 \end{pmatrix}.$$

Thus, the model parameters can be estimated by solving the following optimization problem:

$$\min_{\tilde{\theta} \in \mathbb{S}^n} \widehat{\widetilde{A}}_{\tilde{\theta}} - \left(\sum_{i \in \mathcal{N}} \tilde{\theta}_{ii} \mu_i + \sum_{i \neq j, \ i,j \in \mathcal{N}} \tilde{\theta}_{ij} s_{ij}\right) + \rho ||\tilde{\theta} - \mathrm{diag}(\tilde{\theta})||_1. \quad (3.8)$$

Optimization problem (3.8) is convex and can be solved using one of the commercial or open-source solvers. We used SDPT3 v4.0 – a MATLAB package for semidefinite-quadratic-linear programming (see Toh et al., 1999 and Tutuncu et al., 2003). Once optimal values of parameters $\tilde{\theta}$ are identified, we can obtain the values of parameters θ for the binary Ising model by applying transformation (3.4) to $\tilde{\theta}$.

Now, for illustration purposes, let us apply this estimation method to an open-source dataset. We chose the BAKERY dataset – utilized by Benson et al. (2018) to evaluate the performance of their model – on account of its moderate size, which makes it very well suited to visualizing results. The dataset comprises a list of selected baskets obtained from the receipts of purchases by customers of a bakery. It was preprocessed so that it contains baskets with less than 6 products and only those products that were selected at least 25 times, leaving 67,488 baskets and 50 products in total. We randomly split this
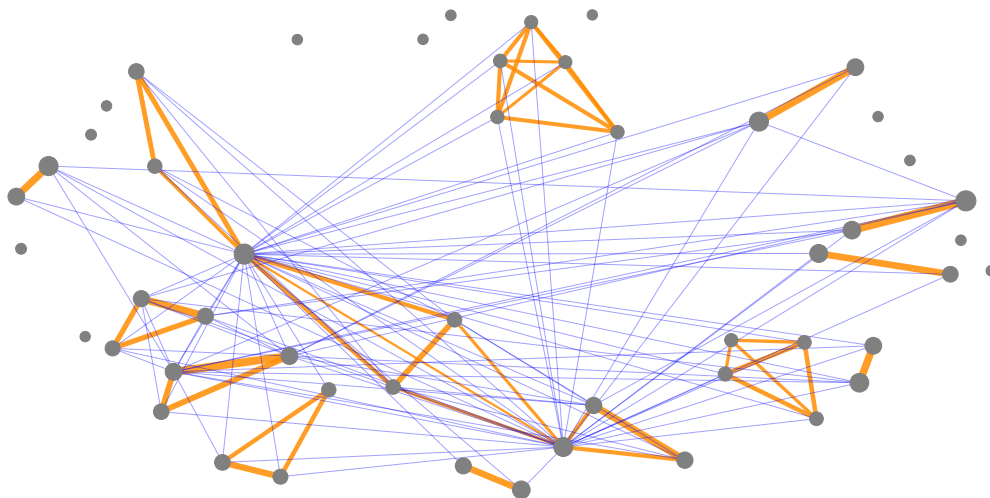
Figure 3.1: Example of the Ising model visualization.

dataset into training and test samples at a ratio of 80:20. Using the training sample, we first obtain estimates of parameters $\tilde{\theta}$ by solving problem (3.8) with the penalty weight $\rho = 0.015$. The penalty weight was chosen empirically with the goal of obtaining a representative graph with a clear structure (i.e., with the goal of removing all negligible interactions but without oversimplifying the model). Lastly, we convert $\tilde{\theta}$ to θ by applying transformation (3.4) and construct a graph based on the values of θ. We visualize this graph in the following way (see Figure 3.1):

- The greater the value of θ_{ii}, the larger the size of node i;
- The greater the absolute value of θ_{ij}, the thicker the edge (i, j);
- If $\theta_{ij} < 0$, then the corresponding edge is blue (meaning a negative direct dependency between i and j), and if $\theta_{ij} > 0$, then the corresponding edge is orange (meaning a positive direct dependency between i and j).

Such an illustration provides valuable insights into the structure of the product portfolio. For example, one can easily identify "established" baskets of products (represented by groups of nodes mutually connected by edges with positive weights), or independent-demand products (represented by isolated nodes).

We investigate the performance of the Ising model as a multi-purchase choice model on the considered dataset (in particular, on the aforementioned test sample) in Appendix C.2. Nevertheless, as detailed in Section 3.2, the MVL (which is equivalent to the binary Ising model) is a well-established multi-purchase choice model, which is why we will not delve into a deep exploration of its modeling strength.

3.5 Assortment Optimization in a Single-Channel Setting

Let us first consider a single-channel setting in which customer choices correspond to the Ising model. Let r_j denote the gross profit per unit of product $j \in \mathcal{N}$. Then, the expected profit generated by a random customer given assortment $S \subseteq \mathcal{N}$ is as follows:

$$R(S) = \sum_{x \in \mathcal{X}(S)} \left(p_\theta(x|S) \sum_{j \in S} r_j x_j \right).$$

The assortment optimization problem under the Ising model can be formulated as the problem of maximizing $R(S)$ over all possible assortments S, i.e.,

$$\max_S R(S). \tag{3.9}$$

Note that this formulation does not depend on the expected number of customers since the optimal solution to an optimization problem does not change if the objective function is multiplied by a positive constant.

Solving optimization problem (3.9) is an extremely difficult task. In fact, the decision version of this problem is NP-hard as demonstrated by the following theorem:

Theorem 3.1. *The decision version of the assortment optimization problem under the Ising model is NP-hard.*

Remark 3.1. *If a capacity constraint is introduced, then the decision version of the capacitated assortment optimization problem under the Ising model is also NP-hard since the uncapacitated problem is a special case of the capacitated one (in the case of infinite capacity).*

The theorem statement follows from the fact that the problem of computing the partition function of the Ising model is NP-hard (see Appendix C.3 for the full proof). Furthermore, the proof of Theorem 3.1 implies that the problem in question does not admit any polynomial-time approximation scheme (PTAS) unless P = NP. Indeed, Istrail (2000) proved NP-hardness of the problem of computing the partition function of the Ising model using a reduction from the Max-Cut problem for 3-regular (cubic) graphs. At the same time, the latter problem was shown to be APX-hard by Alimonti and Kann (2000). Therefore, the decision version of our assortment optimization problem is APX-hard as well, meaning that this problem does not admit PTAS unless P = NP.

Despite the complexity of the considered assortment optimization problem, we can formulate some theoretical insights into the structure of its optimal solutions based on the graphical representation of the Ising model.

Proposition 3.2. *If a node in the graphical representation of the Ising model is isolated, then the corresponding product belongs to the optimal assortment.*

The proof of this proposition is provided in Appendix C.4. This is an intuitively clear result – the fact that $\theta_{kj} = 0 \; \forall j \in \mathcal{N}\backslash\{k\}$ means that the probability of any product $j \in \mathcal{N}\backslash\{k\}$ belonging to a random basket is not affected by the presence of product k in the assortment. However, if a node in the graphical representation of the Ising model is such that there are no edges with negative weights incident on it, then the corresponding product does not necessarily belong to the optimal assortment. Consider the following example:

Example 3.1. *Let $\mathcal{N} = \{1, 2, 3\}$, $r_1 = 10$, $r_2 = 10$, and $r_3 = 100$. Suppose that the customer choices follow the Ising model with the following parameters (see Figure 3.2 for the graphical representation of this model):*

$$\theta = \begin{pmatrix} 1 & 5 & 2 \\ 5 & 5 & -5 \\ 2 & -5 & 5 \end{pmatrix}.$$

Then, it can be checked that $P(\{1, 2, 3\}) \approx 47$ and $P(\{2, 3\}) \approx 55$, meaning that it is more profitable to offer products $\{2, 3\}$ than products $\{1, 2, 3\}$ despite the fact that all edges connected to node 1 have positive weights.
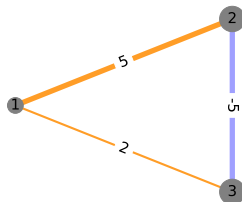
Figure 3.2: Example of when it is profitable to exclude a product from the assortment even though there are only positive edges connected to it.

The above example demonstrates the complexity of the structure of the optimal assortment under the Ising model. A product added to a basket can have only positive *direct* impacts on the probabilities of other products belonging to this basket. However, when all effects are taken into account jointly, it might happen that having such a product in the assortment reduces the marginal probabilities of customers buying some other products, meaning that this product does not necessarily belong to the optimal assortment.

Next, we can further exploit the fact that the Ising model is an MRF and formulate the following proposition:

Proposition 3.3. *Suppose that $\mathcal{N} = \mathcal{H} \sqcup \mathcal{K}$. If $\theta_{ij} = 0 \ \forall i \in \mathcal{H}, \ j \in \mathcal{K}$, then the assortment optimization problem (3.9) can be separated into assortment optimization problems for product sets $\mathcal{H}$ and $\mathcal{K}$.*

This proposition, the proof of which is provided in Appendix C.5, means that the properties of optimal assortments can be formulated on the level of isolated subgraphs of the graphical representation of the Ising model rather than on the level of the whole graph. In Theorem 3.2, we provide a sufficient condition under which all products in an isolated subgraph belong to the optimal assortment. We do it by showing that removing a product from such a subgraph can only reduce the marginal probabilities of customers choosing other products. When it comes to marginal probabilities, one can think of removing product k from the assortment as considering conditional marginal probabilities under condition $x_k = 0$. Formally,

$$p_\theta(x_l = 1 | x_k = 0, S) = p_\theta(x_l = 1 | S \backslash \{k\}).$$

Indeed, one can easily see that

$$p_\theta(x_l = 1 | x_k = 0, S) = \frac{p_\theta(x_l = 1, x_k = 0 | S)}{p_\theta(x_k = 0 | S)}$$

$$= \frac{\sum\limits_{x \in \mathcal{X}(S): \ x_l=1, x_k=0} \exp\Big(\sum\limits_{i \in S} \theta_{ii} x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i \theta_{ij} x_j\Big)}{\sum\limits_{x \in \mathcal{X}(S): \ x_k=0} \exp\Big(\sum\limits_{i \in S} \theta_{ii} x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i \theta_{ij} x_j\Big)}$$

$$= \frac{\sum\limits_{x \in \mathcal{X}(S \backslash \{k\}): \ x_l=1} \exp\Big(\sum\limits_{i \in S \backslash \{k\}} \theta_{ii} x_i + \sum\limits_{i \neq j, \ i,j \in S \backslash \{k\}} x_i \theta_{ij} x_j\Big)}{\sum\limits_{x \in \mathcal{X}(S \backslash \{k\})} \exp\Big(\sum\limits_{i \in S \backslash \{k\}} \theta_{ii} x_i + \sum\limits_{i \neq j, \ i,j \in S \backslash \{k\}} x_i \theta_{ij} x_j\Big)} = p_\theta(x_l = 1 | S \backslash \{k\}).$$

This observation plays an important role in the proof of Theorem 3.2.

Theorem 3.2. *If an isolated subgraph in the graphical representation of the Ising model does not contain edges with negative weights, then all products from this subgraph belong to the optimal assortment.*

Although the theorem statement is intuitively clear, its formal proof is rather involved and requires a combination of different techniques and supporting statements. The full proof can be found in Appendix C.6. Besides explicit theoretical insights, Theorem 3.2 also provides intuition on the sizes of optimal assortments in certain cases. In particular, if there are few pairwise negative dependencies, or if the absolute values of negative parameters θ_{ij} are small compared to the rest of the Ising model parameters, then one can expect that it is optimal to offer almost all products.

We can also identify a specific condition on the parameters associated with a product under which this product has to belong to the optimal assortment. Consider the following proposition (see Appendix C.7 for the proof):

Proposition 3.4. *Let $\mathcal{H} \subseteq \mathcal{N}$ induce an isolated subgraph in the graphical representation of the Ising model. If product $k \in \mathcal{H}$ is such that $\theta_{ki} = \alpha r_i \ \forall i \in \mathcal{H} \backslash \{k\}$, where α is a positive constant, then product k belongs to the optimal assortment.*

The theoretical results obtained can be used towards assortment optimization, in particular, to reduce the dimensionality of problem (3.9). After producing the graphical representation of the Ising model as discussed in Section 3.4, the following preprocessing procedure can be applied:

1. Add isolated nodes to the optimal assortment, thereby removing them from consideration;

2. Separate the graph into connected components (maximal connected subgraphs);

3. If some connected components do not contain edges with negative weights, add all nodes from these components to the optimal assortment;

4. For each remaining connected component, formulate the assortment optimization problem considering nodes from this component as a separate product portfolio.

Next, we need a computationally tractable method that will allow us to solve the assortment optimization problem after completing the preprocessing stage. As mentioned earlier, the decision version of problem (3.9) is NP-hard and, moreover, it does not admit any polynomial-time approximation scheme (PTAS) unless P = NP. Thus, we developed a metaheuristic algorithm that can be used to obtain high-quality solutions to the assortment optimization problem. Before providing a detailed description, let us first discuss the integral part of this algorithm, namely, how to evaluate potential solutions to

Algorithm 1 Systematic scan Gibbs generator	**Algorithm 2** Random scan Gibbs generator
1: **Input**	1: **Input**
2: $x^{(l-1)}$: previous sample	2: $x^{(l-1)}$: previous sample
3: **Output**	3: **Output**
4: $x^{(l)}$: new sample	4: $x^{(l)}$: new sample
5: **Initialization**	5: **Initialization**
6: $n \leftarrow \texttt{length}(x^{(l-1)})$	6: $n \leftarrow \texttt{length}(x^{(l-1)})$
7: $x^{(l)} \leftarrow x^{(l-1)}$	7: $x^{(l)} \leftarrow x^{(l-1)}$
8: **for** k **in** $[1,\dots,n]$ **do**:	8: Draw a random index k from $[1,\dots,n]$
9: Draw a random number $r \sim U(0,1)$	9: Draw a random number $r \sim U(0,1)$
10: **if** $r \le g(1,x^{(l)}_{-k})/\big(g(0,x^{(l)}_{-k}) + g(1,x^{(l)}_{-k})\big)$ where	10: **if** $r \le g(1,x^{(l)}_{-k})/\big(g(0,x^{(l)}_{-k}) + g(1,x^{(l)}_{-k})\big)$ where
11: $g(x_k, x_{-k}) = \exp\big(\sum_i \theta_{ii}x_i + \sum_{i \ne j} x_i \theta_{ij} x_j\big)$ **then**:	11: $g(x_k, x_{-k}) = \exp\big(\sum_i \theta_{ii}x_i + \sum_{i \ne j} x_i \theta_{ij} x_j\big)$ **then**:
12: $a \leftarrow 1$	12: $a \leftarrow 1$
13: **else**:	13: **else**:
14: $a \leftarrow 0$	14: $a \leftarrow 0$
15: $x^{(l)}_k \leftarrow a$	15: $x^{(l)}_k \leftarrow a$
16: **return** $x^{(l)}$	16: **return** $x^{(l)}$
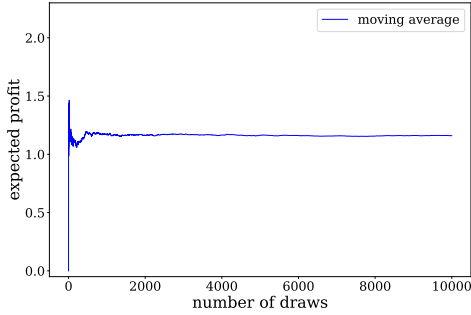
(a) Systematic scan Gibbs sampling.
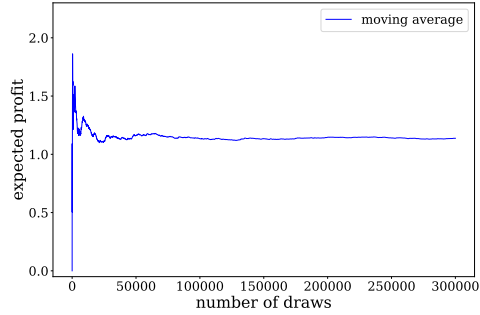
(b) Random scan Gibbs sampling.

Figure 3.3: Example of expected profit estimation using Gibbs sampling for Ising model.

the considered problem. Finding the exact value of the expected profit $R(S)$ might not be computationally feasible even for moderately sized assortments S. This is because of the need to compute the normalizing coefficient of $p_\theta(x|S)$, which contains an exponential number of terms. We thus need an efficient way of approximating the expected profit. This can be done by generating a large number of samples from the Ising model, i.e., generating a large number of purchased baskets. We employ Gibbs sampling for this purpose, which is a special case of the Metropolis-Hastings algorithm. There are two well-known variations of Gibbs sampling: systematic scan Gibbs sampling and random scan Gibbs sampling (see, e.g., He, De Sa, et al., 2016). In the former method, each new sample $x^{(l)}$ is obtained by iteratively sampling vector components $x^{(l)}_k$ from $p_\theta(\cdot|x^{(l)}_{-k})$ for $k = 1,\dots,n$, whereas in the latter method, each new sample is obtained by randomly selecting index k and sampling $x^{(l)}_k$ from $p_\theta(\cdot|x^{(l)}_{-k})$. Algorithms 1 and 2 describe these procedures adapted to

---

Algorithm 3 Simulated annealing procedure to solve problem (3.9)

1: **Input**
2:     $\mathcal{N}$: whole product set
3:     S_{start}: assortment representing the starting point of the algorithm
4:     d_{obj}: typical increase of the objective function
5:     k_{temps}: number of temperatures
6:     p_{min}, p_{max}: minimum and maximum target acceptance probabilities
7: **Output**
8:     S_{heur}: optimized assortment
9: **Initialization**
10:     $S_{cur} \leftarrow S_{start}, \ R_{cur} \leftarrow \texttt{estimateProfit}(S_{cur})$ ▷ Initialize the current solution and its value
11:     $S_{heur} \leftarrow S_{cur}, \ R_{heur} \leftarrow R_{cur}$ ▷ Initialize the heuristic solution and its value
12:     $T \leftarrow -d_{obj}/\log(p_{max})$ ▷ Initialize the temperature
13: **for** i **in** $[1, \ldots, k_{temps}]$ **do**:   ▷ Main loop
14:     Draw a random product $j \in \mathcal{N}$
15:     **if** $j \in S_{cur}$ **then**:   ▷ Generate the candidate solution
16:         $S_{can} \leftarrow S_{cur} \backslash \{j\}$
17:     **else**:
18:         $S_{can} \leftarrow S_{cur} \cup \{j\}$
19:     $R_{can} \leftarrow \texttt{estimateProfit}(S_{can})$ ▷ Estimate the value of the candidate solution
20:     **if** $R_{can} > R_{cur}$ **then**:   ▷ Update the current solution
21:         $S_{cur} \leftarrow S_{can}, \ R_{cur} \leftarrow R_{can}$
22:     **else**:
23:         Draw a random number $r \sim U(0, 1)$
24:         **if** $r < \exp\big((R_{can} - R_{cur})/T\big)$ **then**:
25:             $S_{cur} \leftarrow S_{can}, \ R_{cur} \leftarrow R_{can}$
26:     **if** $R_{cur} > R_{heur}$ **then**:   ▷ Update the heuristic solution
27:         $S_{heur} \leftarrow S_{cur}, \ R_{heur} \leftarrow R_{cur}$
28:     $T \leftarrow -d_{obj}/\log(p_{max} + (p_{min} - p_{max})i/k_{temps})$ ▷ Update the temperature
29: **return** S_{heur}, R_{heur}

---

sampling from the Ising model. In order to select the sampling procedure, we conducted a preliminary numerical analysis using the Ising model parameters estimated in Section 3.4 together with randomly generated profit margins $r_j \sim U(0, 1)$. Based on this preliminary analysis, we decided to use systematic scan Gibbs sampling as it appeared to converge faster than random scan Gibbs sampling for a similar amount of time (see Figure 3.3 for an illustration).

Being able to approximate the expected profit for any assortment allows us to implement a metaheuristic algorithm for finding promising solutions to problem (3.9). One of the most prominent metaheuristic algorithms is simulated annealing (SA). There are numerous variations of this procedure. In this work, we adopt the SA design described by Bierlaire (2015). We set the number of algorithm iterations per temperature to one. At each iteration, a candidate solution is generated and evaluated using simulations. Each candidate solution is selected randomly from the neighborhood of the current solution. We chose a fairly basic neighborhood structure: The neighborhood of an assortment comprises all possible assortments that can be obtained from the assortment in question

by either removing or adding a product. Next, if the candidate solution is better than the current solution, then the candidate solution is accepted and becomes the current solution. Otherwise, the candidate solution is accepted with a certain probability depending on the annealing temperature at this iteration. The temperature decreases with each new iteration. The maximum and minimum annealing temperatures are calculated based on the target maximum and minimum acceptance probabilities, respectively. It is done using parameter d_{obj} representing the typical increase of the objective function in the given neighborhood structure, which can be estimated empirically. In this algorithm variation, the acceptance probability decreases linearly with respect to the iteration counter. Algorithm 3 provides a detailed description of this SA procedure.

One of the greatest advantages of using SA for our purposes is its flexibility. First, various additional constraints – such as capacity or shelf-space constraints – can be easily incorporated into the algorithm by imposing additional restrictions on the neighborhood structure. Moreover, Algorithm 3 can be generalized to solve the assortment optimization problem in much more complex settings, including the omnichannel setting. In the next section, we discuss the omnichannel extension and provide the results of our numerical experiments.

3.6   Omnichannel Setting

In this section, we generalize our modeling framework to an omnichannel setting. We focus on the case where a retailer has two sales channels: an online channel and a physical channel. One of the key features of an omnichannel environment is that a customer might purchase subsets of a single shopping basket in two different channels. To account for this behavior, we consider four customer types: online-only, physical-only, omnichannel-online, and omnichannel-physical. Online-only or physical-only customers are assumed to shop exclusively in their respective channel. Omnichannel customers, however, might purchase the first part of the basket in their preferred channel (e.g., online in the case of omnichannel-online customers) and the rest of the basket in the other channel. We also assume that an omnichannel customer never purchases a product in the second-choice channel if the same product is available in their preferred channel. Let $\Omega = \{o, o_p, p_o, p\}$ be the set of customer types, where o-customers are online-only customers, o_p-customers are omnichannel-online customers, p_o-customers are omnichannel-physical customers, and p-customers are physical-only customers. Figure 3.4 illustrates the shopping behavior of customers of different types. In Figure 3.4(a), we provide a representative example of baskets that can be purchased by customers of each type. Note that it is also possible that an omnichannel customer purchases products only in one

(a) Example of purchased baskets.

(b) Summary of choice alternatives.
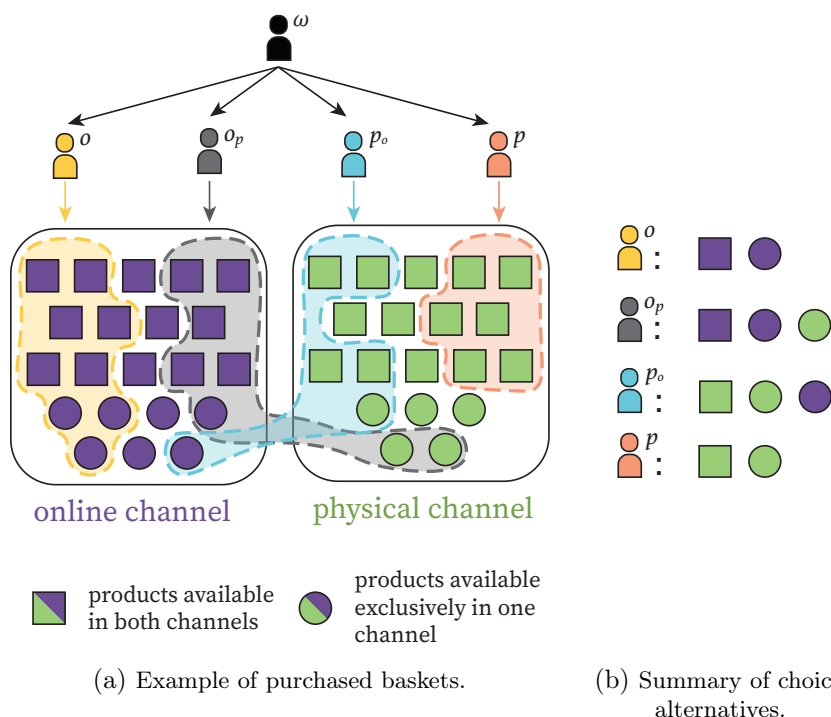
Figure 3.4: Basket shopping behavior of customers in an omnichannel environment.

channel, or that a customer (of any type) selects an empty basket, which represents a no-purchase alternative. Figure 3.4(b) summarizes what kinds of products are considered for purchase by customers of each type.

Suppose that for all customer types $\omega \in \Omega$, choices of ω-customers correspond to the Ising model with parameters θ^ω. Parameter estimation can be performed for each customer type separately as described in Section 3.4 if we assume that customers possess loyalty cards. Since loyalty programs have been widely adopted by retailers worldwide, the assumption that loyalty card data is available is fairly justified. In turn, such data would allow us to identify two parts of the same basket purchased in different channels by an omnichannel customer as a single basket. In the numerical analysis part of this section, we will highlight the importance of having such visibility across channels by evaluating the benefits of accounting for omnichannel purchases while optimizing assortments.

We can now formulate the omnichannel assortment optimization problem. Let $S^o, S^p \subseteq \mathcal{N}$ denote assortments in the online and physical channels, respectively. Since omnichannel customers can shop in both channels, the product set available for purchase by omnichannel customers (either o_p-customers or p_o-customers) is the union of the two assortments, i.e., $S^{o_p} = S^{p_o} = S^o \cup S^p \subseteq \mathcal{N}$. Importantly, even though both types of omnichannel customers select from the same set of products $S^o \cup S^p$, they purchase products that are

offered in both channels only in their preferred channel, which generally results in different gross profits per unit of each product depending on where it was bought. Suppose that r_j^o is the gross profit per unit of product $j \in \mathcal{N}$ offered in the online channel, and r_j^p is the gross profit per unit of the same product offered in the physical channel. Let us also introduce auxiliary parameters $r_j^{o_p}$ and $r_j^{p_o}$ which can be expressed through r_j^o and r_j^p in the following way:

$$r_j^{o_p} = \begin{cases} r_j^o & \text{if } j \in S^o, \\ r_j^p & \text{if } j \in S^p \backslash S^o, \end{cases} \qquad r_j^{p_o} = \begin{cases} r_j^p & \text{if } j \in S^p, \\ r_j^o & \text{if } j \in S^o \backslash S^p. \end{cases}$$

Parameters $r_j^{o_p}$ and $r_j^{p_o}$ defined above represent gross profits per unit of product $j \in \mathcal{N}$ bought by an o_p-customer and a p_o-customer, respectively. These parameters are required to formulate the standardized functions for computing the expected profits generated by customers of different types. The expected profit generated by an ω-customer can be written in the following way:

$$R^\omega(S^\omega) = \sum_{x \in \mathcal{X}(S^\omega)} \left(p_{\theta^\omega}(x|S^\omega) \sum_{j \in S^\omega} r_j^\omega x_j \right).$$

Finally, let Λ^ω be the expected number of ω-customers. Then, the omnichannel assortment optimization problem is as follows:

$$\max_{S^o, S^p} \sum_{\omega \in \Omega} \Lambda^\omega R^\omega(S^\omega). \tag{3.10}$$

Clearly, problem (3.10) can be reduced to problem (3.9) – the assortment optimization problem in a single-channel setting – by setting all but one parameter Λ^ω to zero. Therefore, we can formulate the following corollary of Theorem 3.1:

Corollary 3.2. *The decision version of the omnichannel assortment optimization problem (3.10) is NP-hard.*

Note that if adding a product to the assortment in either of the channels increases the expected profits generated by customers of each type (or leaves some of the profits unchanged), then adding such a product leads to an increase in the total expected profit. Thus, we can generalize some of our theoretical results formulated for a single-channel setting to the omnichannel setting. Consider the following corollaries of Propositions 3.2, 3.3 and Theorem 3.2, respectively:

Corollary 3.3. *If for each customer type ω, node $j \in \mathcal{N}$ in the graphical representation of the corresponding Ising model is isolated, then product j belongs to the optimal assortment.*

Corollary 3.4. *Suppose that $\mathcal{N} = \mathcal{H} \sqcup \mathcal{K}$. If for each customer type ω, $\theta_{ij}^{\omega} = 0 \; \forall i \in \mathcal{H}$, $j \in \mathcal{K}$, then the omnichannel assortment optimization problem (3.10) can be separated into omnichannel assortment optimization problems for product sets $\mathcal{H}$ and $\mathcal{K}$.*

Corollary 3.5. *If for each customer type ω, $\mathcal{H} \subseteq \mathcal{N}$ induces an isolated subgraph in the graphical representation of the corresponding Ising model, and all these subgraphs do not contain edges with negative weights, then all products from $\mathcal{H}$ belong to the optimal assortment.*

Similar to the single-channel setting, we can leverage the obtained theoretical results for preprocessing aimed at reducing the dimensionality of problem (3.10). Let G^{ω} denote the graph corresponding to the Ising model that defines the shopping behavior of ω-customers (in other words, G^{ω} represents the graph that can be created using θ^{ω} as adjacency matrix). Our preprocessing procedure becomes:

1. Add nodes that are isolated in G^{ω} for all $\omega \in \Omega$ to the optimal assortment;

2. Separate the product portfolio into subsets of maximum size such that any two subsets induce disconnected subgraphs of G^{ω} for all $\omega \in \Omega$;

3. If some of the subsets are such that the induced subgraphs of G^{ω} do not contain negative edges for all $\omega \in \Omega$, add all nodes from these subsets to the optimal assortment;

4. For each remaining subset, formulate the omnichannel assortment optimization problem considering elements from this subset as a separate product portfolio.

After preprocessing, the remaining omnichannel assortment optimization problems can be solved using Algorithm 4, which is the omnichannel extension of the SA procedure described in Section 3.5. At its core, Algorithm 4 is nearly identical to Algorithm 3 apart from a few key distinctions. In the omnichannel case, we assume that $\mathcal{N} = \mathcal{N}^o \cup \mathcal{N}^p$, where $\mathcal{N}^o$ and $\mathcal{N}^p$ represent the maximum sets of products that can be offered in the online and physical channels, respectively. Depending on the problem formulation, sets $\mathcal{N}^o$ and $\mathcal{N}^p$ may coincide (i.e., $\mathcal{N}^o = \mathcal{N}^p = \mathcal{N}$). In practice, however, there may be some products that cannot be offered in one of the channels – for example for marketing considerations, or if the goal is to optimize assortments without introducing new products in either channel. We thus provide a general algorithm formulation that can handle such limitations. Next, at each iteration of Algorithm 4, a candidate solution is generated by first randomly choosing a channel (with the probability calculated based on the sizes of $\mathcal{N}^o$ and $\mathcal{N}^p$) and then changing the current assortment in that channel. Finally, once the candidate solution is generated, it is evaluated by summing up the expected profits

---

Algorithm 4 Simulated annealing procedure to solve problem (3.10)

1: **Input**
2:     $\mathcal{N}^o, \mathcal{N}^p$: whole product sets that can be offered in the online and physical channels, respectively
3:     S^o_{start}, S^p_{start} : assortments representing the starting point of the algorithm
4:     d_{obj}: typical increase of the objective function
5:     k_{temps}: number of temperatures
6:     p_{min}, p_{max}: minimum and maximum target acceptance probabilities
7: **Output**
8:     S^o_{heur}, S^p_{heur}: optimized assortments in the online and physical channels
9: **Initialization**
10:     $S^o_{cur} \leftarrow S^o_{start}, \ S^p_{cur} \leftarrow S^p_{start}$ ▷ Initialize the current solution
11:     $R_{cur} \leftarrow \texttt{estimateOmnichannelProfit}(S^o_{cur}, S^p_{cur})$ ▷ Initialize the value of the current solution
12:     $S^o_{heur} \leftarrow S^o_{cur}, \ S^p_{heur} \leftarrow S^p_{cur}, \ R_{heur} \leftarrow R_{cur}$ ▷ Initialize the heuristic solution and its value
13:     $T \leftarrow -d_{obj}/\log(p_{max})$ ▷ Initialize the temperature
14: **for** i **in** $[1, \ldots, k_{temps}]$ **do:**   ▷ Main loop
15:     Draw a random number $r \sim U(0,1)$
16:     **if** $r < |\mathcal{N}^o|/(|\mathcal{N}^o| + |\mathcal{N}^p|)$ **then:**   ▷ Randomly choose channel
17:         Draw a random product $j \in \mathcal{N}^o$
18:         **if** $j \in S^o_{cur}$ **then:**   ▷ Generate the candidate solution
19:             $S^o_{can} \leftarrow S^o_{cur} \backslash \{j\}$
20:         **else:**
21:             $S^o_{can} \leftarrow S^o_{cur} \cup \{j\}$
22:     **else:**
23:         Draw a random product $j \in \mathcal{N}^p$
24:         **if** $j \in S^p_{cur}$ **then:**   ▷ Generate the candidate solution
25:             $S^p_{can} \leftarrow S^p_{cur} \backslash \{j\}$
26:         **else:**
27:             $S^p_{can} \leftarrow S^p_{cur} \cup \{j\}$
28:     $R_{can} \leftarrow \texttt{estimateOmnichannelProfit}(S^o_{can}, S^p_{can})$ ▷ Estimate the value of the candidate solution
29:     **if** $R_{can} > R_{cur}$ **then:**   ▷ Update the current solution
30:         $S^o_{cur} \leftarrow S^o_{can}, \ S^p_{cur} \leftarrow S^p_{can}, \ R_{cur} \leftarrow R_{can}$
31:     **else:**
32:         Draw a random number $r \sim U(0,1)$
33:         **if** $r < \exp\big((R_{can} - R_{cur})/T\big)$ **then:**
34:             $S^o_{cur} \leftarrow S^o_{can}, \ S^p_{cur} \leftarrow S^p_{can}, \ R_{cur} \leftarrow R_{can}$
35:     **if** $R_{cur} > R_{heur}$ **then:**   ▷ Update the heuristic solution
36:         $S^o_{heur} \leftarrow S^o_{cur}, \ S^p_{heur} \leftarrow S^p_{cur}, \ R_{heur} \leftarrow R_{cur}$
37:     $T \leftarrow -d_{obj}/\log(p_{max} + (p_{min} - p_{max})i/k_{temps})$ ▷ Update the temperature
38: **return** $S^o_{heur}, S^p_{heur}, R_{heur}$

---

generated by each of the four customer types estimated using the systematic scan Gibbs sampling. The explicit pseudocode description of this algorithm is provided below.

We now perform numerical analysis to obtain insights into the properties of heuristic and optimal solutions to the omnichannel assortment optimization problem. We start by describing the parameter generation procedure for our numerical experiments. We consider two sizes of the product portfolio $\mathcal{N}$: $n = 8$ and $n = 40$. In the former case, the size of the assortment optimization problem allows us to solve it by brute force (i.e., using a complete enumeration of all possible solutions). In the latter case, finding the exact solutions is not computationally feasible, which means it is necessary to use the

developed heuristic approach. For each product portfolio size, we generated 100 problem instances in the following way. First, in order to study a more general setting, we assume that some products can only be offered in one channel (in other words, $\mathcal{N}^o$ and $\mathcal{N}^p$ do not coincide). We generate sets $\mathcal{N}^o$ and $\mathcal{N}^p$ by taking random subsets of $\mathcal{N}$ of size $0.75n$ while ensuring that $\mathcal{N}^o \cup \mathcal{N}^p = \mathcal{N}$. Next, the generation process for parameters θ^ω is easier to describe in terms of the corresponding graphs G^ω. Consider the following graph generation procedure:

1. Create nodes with weights θ_{ii}^ω $\forall i \in \mathcal{N}$ sampled randomly from the interval $[2, 4]$;

2. For all unordered pairs $i, j \in \mathcal{N}$, create an edge between nodes i and j with probability p_{edge};

3. For each edge, sample its absolute weight θ_{ij}^ω from the interval $[1, 2]$ and multiply the obtained value by -1 with probability p_{neg}.

We apply the above procedure to first generate 100 graphs G^ω for one customer type $\omega = o_p$. In the case $n = 40$, we set the value of the sparsity-controlling parameter p_{edge} to 0.2. If $n = 8$, we use p_{edge} equal to 1 (i.e., we generate complete graphs) because for such a small product portfolio, sparse graphs with low values of p_{edge} would not be representative. As for parameter p_{neg}, we set its value to 0.8 for both $n = 8$ and $n = 40$. This means that direct product substitution effects between pairs of adjacent products (i.e., products connected via an edge) are somewhat prevalent. For $n = 40$, direct product substitution effects thus exist with probability $p_{edge} \cdot p_{neg} = 0.16$, which seems to be fair value for illustrative purposes. Once graphs G^{o_p} are generated, we use them to obtain graphs for the remaining customer types. It is reasonable to assume that graphs G^ω for different types ω have a similar structure. To this end, we obtain each graph G^ω, $\omega \in \Omega \backslash \{o_p\}$ by first adding random noise $\zeta_{ij} \sim U(-0.3 \cdot \theta_{ij}^{o_p}, 0.3 \cdot \theta_{ij}^{o_p})$ to all weights $\theta_{ij}^{o_p}$ and then swapping the absolute values of one-third of all edges. Furthermore, following Vasilyev et al. (2023), we assume that the gross profit per unit of each product is higher for the online channel than for the physical channel (e.g., due to the difference in holding costs) and thus generate gross profits per unit of each product $j \in \mathcal{N}$ in the following way:

$$r_j^p \sim U(0, 1), \quad r_j^o \sim U(r_j^p, 1.5 \cdot r_j^p).$$

Finally, for each simulated problem instance, we consider several customer portfolio compositions, i.e., several sets of values of parameters Λ^ω, $\omega \in \Omega$. Our primary goal is to investigate the relationship between different customer portfolio compositions and the properties of optimized assortments. To this end, we consider the following realizations

of vector $\mathbf{\Lambda} = (\Lambda^o, \Lambda^{o_p}, \Lambda^{p_o}, \Lambda^p)$:

$$
\begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \end{pmatrix} = 10^4 \cdot \begin{pmatrix} 5 & 1 & 1 & 5 \\ 4 & 2 & 2 & 4 \\ 3 & 3 & 3 & 3 \\ 2 & 4 & 4 & 2 \\ 1 & 5 & 5 & 1 \end{pmatrix}, \quad \text{and} \quad \begin{pmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \\ b_5 \end{pmatrix} = 10^4 \cdot \begin{pmatrix} 1 & 1 & 5 & 5 \\ 2 & 2 & 4 & 4 \\ 3 & 3 & 3 & 3 \\ 4 & 4 & 2 & 2 \\ 5 & 5 & 1 & 1 \end{pmatrix}.
$$

We start by exploring what happens with the sizes of optimal (for $n = 8$) and heuristic (for $n = 40$) solutions as the customer portfolio composition $\mathbf{\Lambda}$ evolves. To obtain heuristic solutions, we apply Agorithm 4 with the following input: $p_{min} = 0.001$, $p_{max} = 0.999$, $k_{temps} = 300$, $d_{obj} = 10000$, $S_{start}^o = \mathcal{N}^o$, and $S_{start}^p = \mathcal{N}^p$. In Figure 3.5, we visualize the sizes of assortments relative to the maximum number of products that can be offered in the corresponding channels. For each customer portfolio composition, we display values averaged over 100 simulated problem instances. First, note that the general trends are similar for both exact and heuristic solutions (i.e., the plots on the left-hand side and on the right-hand side of the figure are largely alike), which validates the results obtained using our heuristic algorithm. In Figure 3.5(a), we can observe that as $\mathbf{\Lambda}$ changes from a_1 to a_5 – in other words, as the proportion of omnichannel customers (of both types) grows – the average size of the online assortment remains stable, whereas the sizes of the physical assortment and assortment overlap decrease. This is in line with our expectations – indeed, in the limit with only omnichannel customers and no purely online and physical customers, offering a product exclusively online is always more profitable than offering this product in both channels simultaneously. This is because the gross profit per unit of each product is assumed to be higher in the online channel than in the physical channel. Therefore, in the considered extreme case, one can expect to see a "large" online assortment and a "small" physical assortment, with the latter assortment comprising only those products that are not offered in the online channel. Conversely, in the limit where all customers either shop exclusively online or in person (i.e., there are no omnichannel customers), the two channels are independent and their average assortment sizes should be approximately the same due to a similar parameter generation procedure. We also study what happens with average assortment sizes if the proportion of customers who prefer the online channel (including both o-customers and o_p-customers) grows compared to the proportion of customers who prefer the physical channel. The results of our numerical experiments visualized in Figure 3.5(b) do not indicate any apparent trends. Note that for $n = 40$, the difference between the average sizes of online and physical assortments is less pronounced in the case $\mathbf{\Lambda} = b_1$ than for other customer portfolio compositions. However, this observation is not substantiated by the results obtained for exact solutions, so this
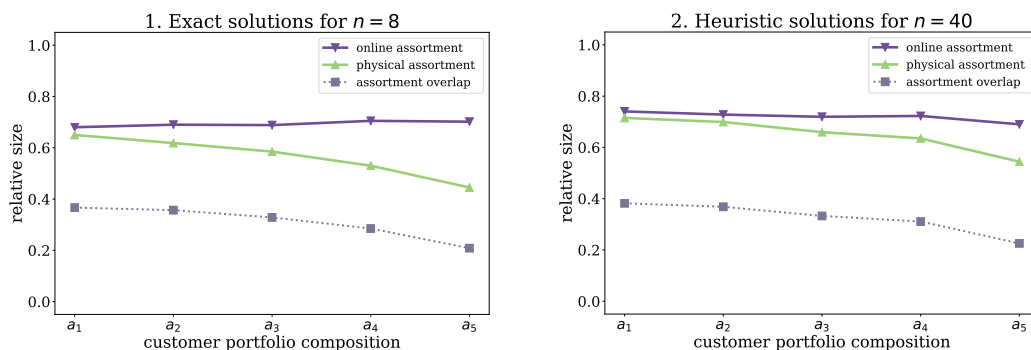
(a) Sensitivity with respect to $\dfrac{\Lambda^{o_p} + \Lambda^{p_o}}{\Lambda^o + \Lambda^p}$ (as the proportion of omnichannel customers grows).
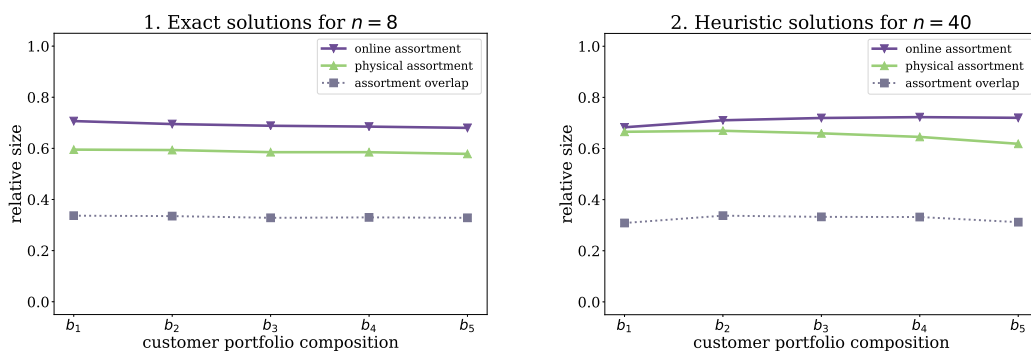
(b) Sensitivity with respect to $\dfrac{\Lambda^o + \Lambda^{o_p}}{\Lambda^p + \Lambda^{p_o}}$ (as the proportion of customers preferring the online channel grows).

Figure 3.5: Average assortment sizes.

aberration might arise because of some suboptimal heuristic solutions. Nevertheless, we can still see that, on average, the size of the online assortment is greater than the size of the physical assortment, which is intuitively not surprising given that the unit gross profits are assumed to be higher in the online channel than in the physical channel.

We also explore the benefits of omnichannel assortment optimization as opposed to optimizing assortments in siloed channels. Let us start by describing how we obtain assortments optimized in the two channels separately for a fixed set of generated parameters. First, we apply Gibbs sampling to simulate a large number of baskets bought by customers of each type. Then, assuming that there is no visibility across different retail channels, we separate each omnichannel basket into "online" and "physical" baskets and add them to the lists of generated purchases made by online-only customers and physical-only customers, respectively. Hence, we obtain two lists of baskets purchased in the online and physical channels. Based on these lists, we estimate two sets of Ising model parameters as described in Section 3.4 and formulate the corresponding assortment optimization problems (one problem per channel). Consequently, these optimization

(a) Sensitivity with respect to $\dfrac{\Lambda^{o_p} + \Lambda^{p_o}}{\Lambda^o + \Lambda^p}$ (as the proportion of omnichannel customers grows).
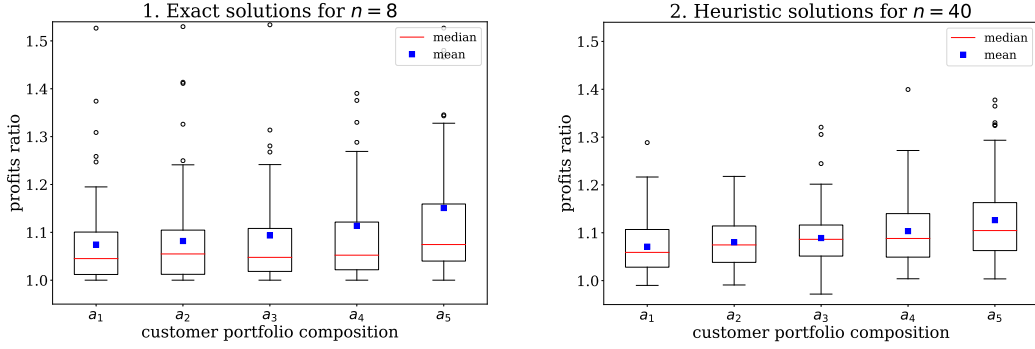

(b) Sensitivity with respect to $\dfrac{\Lambda^o + \Lambda^{o_p}}{\Lambda^p + \Lambda^{p_o}}$ (as the proportion of customers preferring the online channel grows).
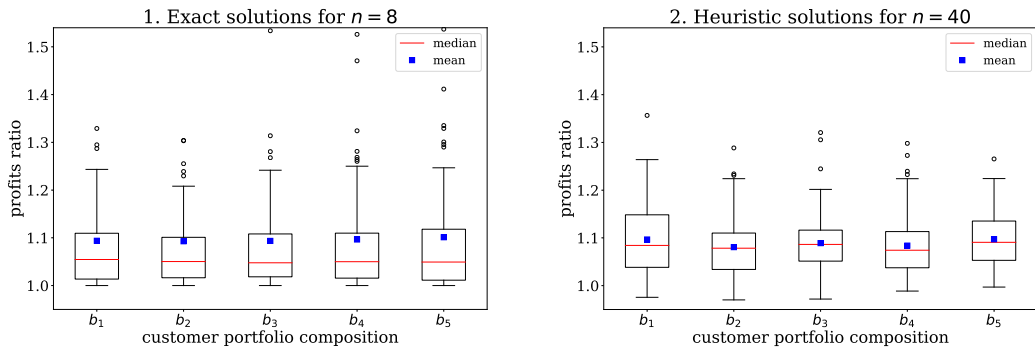
Figure 3.6: Benefits of omnichannel assortment optimization as opposed to optimizing assortments in siloed channels.

problems are solved either by brute force (if $n = 8$) or by using Algorithm 3 (if $n = 40$) with the following input: $p_{min} = 0.001$, $p_{max} = 0.999$, $k_{temps} = 300$, $d_{obj} = 2500$, and $S_{start} = \mathcal{N}^c$, where $c = o$ in the case of the online channel and $c = p$ in the case of the physical channel. We then compare the omnichannel solution (i.e., solution to the omnichannel assortment optimization problem) to the solution obtained for the siloed channels. In particular, we compute the ratio of the profit yielded by the omnichannel solution to the profit yielded by the siloed-channel solution. Importantly, even though the latter overall product assortment is obtained assuming independent demands in the two channels, we always consider the same omnichannel demand when comparing the expected profits of the two models. We repeat the described procedure for each generated set of parameters. The results of our numerical experiments are provided in Figure 3.6. As can be seen in Figure 3.6(a), the larger the proportion of omnichannel customers (of both types), the greater the average profit gain yielded by optimizing product assortment decisions explicitly taking into account the omnichannel setting. This shows for both cases ($n = 8$ and $n = 40$). Meanwhile, the difference between the mean and

the median ratio values computed over different problem instances is more pronounced in the case of exact solutions. This could be explained by the noticeable presence of outliers, i.e., instances in which omnichannel solutions greatly outperform siloed-channel solutions, which occurs more rarely for heuristic solutions than for exact solutions. Next, as previously, Figure 3.6(b) does not indicate any apparent trends as the proportion of customers preferring the online channel grows. We do notice that in some rare cases, the heuristic solutions obtained for the omnichannel case may be slightly worse than the heuristic solutions obtained for siloed channels, which indicates a certain suboptimality of our heuristics. On average, however, the omnichannel solution yields a substantially higher profit than the siloed-channel solution. The gross profit gain averaged over all our numerical experiments is close to 10% (more specifically, 9.9% in the case of exact solutions and 9.2% in the case of heuristic solutions), which underlines the importance of having visibility of customer shopping behavior across retail channels.

3.7 Conclusions and Future Work

In this paper, we addressed the question of omnichannel assortment optimization given basket shopping behavior of customers. To tackle this problem, we developed a comprehensive methodology at the interface of operations research and computer science. We brought into the spotlight the equivalence relation between the Ising model and the multivariate logit model. This allowed us to utilize classic theoretical results devised for the Ising model for various purposes, including parameter estimation and deriving complexity results for the assortment optimization problem. Furthermore, we introduced a preprocessing procedure that can be used to reduce the dimensionality of the assortment optimization problem based on the graphical representation of the Ising model. We developed a customized simulated annealing algorithm for solving the assortment optimization problem in cases when finding the exact solution is not computationally tractable. Finally, we carried out an extensive numerical study that allowed us, inter alia, to quantify the value of having visibility across different retail channels, i.e., having information about purchases made by omnichannel customers. We observed that, on average, omnichannel assortment optimization results in significant gross profit gains compared to assortment optimization in siloed channels.

Our work motivates a number of interesting research questions. We believe that the graphical representation of the Ising model can be further exploited to boost assortment optimization capabilities through network analytics techniques. For example, one can identify "important" products using different centrality measures, or "established" baskets using community detection algorithms. This information can be incorporated into the

developed simulated annealing procedure, for example, in such a way that at each iteration, high centrality products are added to the assortment with a higher probability and removed from the assortment with a lower probability. Another interesting avenue is to develop a heuristic procedure whereby all products are first ranked with respect to a selected centrality measure (potentially multiplied by the corresponding unit gross profit), and then the heuristic solution is determined by finding a cutoff in this ranking. Last but not least, more general MRFs with other joint distribution functions could be used for more accurate modeling of basket shopping behavior (e.g., by enabling our model to account for product quantities in each basket). Ultimately, our work takes the first step toward a comprehensive framework for omnichannel assortment optimization that takes into account basket shopping behavior, and it provides a base for future studies on omnichannel decision-making.

Conclusion

In this thesis, we addressed the complex problem of omnichannel assortment optimization, presenting three chapters that tackle different aspects of this topic. In the first chapter, we introduced the multichannel attraction model (MAM), a discrete choice model specifically designed for omnichannel environments. We focused on a dual-channel setting and formulated the assortment optimization problem under the MAM as a mixed-integer linear program. We proved that the optimal assortment in one channel can be found analytically if all products are available in the remaining channel. This result also allowed us to derive an approximate characterization of optimal assortments, thereby providing insights into their structure. We proposed a computationally efficient heuristic method to solve the assortment optimization problem, and showed numerically that it steadily yields near-optimal solutions. We also showed that implementing the buy-online-and-pick-up-in-store (BOPS) functionality can be unprofitable if the proportion of online customers selecting this option is too large compared to the additional traffic in the physical channel. Lastly, we demonstrated that omnichannel assortment optimization results in a 1.8%-1.9% average profit gain compared to assortment optimization in siloed channels, with maximum gains reaching up to 6.7%.

In the second chapter, we extended our modeling framework to the case of a retailer managing an online store and a network of physical stores. We also incorporated demand stochasticity and inventory management considerations into the assortment optimization problem under the MAM, which we formulated as a mixed-integer second-order cone program. This problem can be solved exactly using off-the-shelf solvers for small to medium-sized instances. For larger problems, we proposed two heuristic methods based on different relaxations of the problem. We identified the conditions under which the two relaxations are equivalent to each other, and the conditions under which they are also equivalent to the initial assortment problem. Our numerical analysis indicated that demand variability has a dual effect on optimal assortment sizes: a combination of the increasing costs effect and the demand pooling effect. Similar to the first chapter, the numerical experiments showed noticeable financial benefits as a result of omnichannel optimization compared to assortment optimization in siloed channels. On average, the

profit gain yielded by the omnichannel solution was slightly above 3%, reaching up to 7%. These estimates, obtained using a basic heuristic method in a simplified setting where customers do not switch between physical stores, were at the lower end of potential profit gains. Nonetheless, our results surpassed those from the first chapter, emphasizing the importance of incorporating inventory decisions and demand variability into the modeling framework.

In the third chapter, we tackled omnichannel assortment optimization while taking into account the basket shopping behavior of customers. We established the equivalence relation between the Ising model and the multivariate logit model, which made it possible to leverage well-known theorems and methodologies for parameter estimation and derivation of complexity results for the assortment optimization problem. We developed a preprocessing procedure based on the graphical representation of the Ising model for dimensionality reduction of the assortment optimization problem, as well as a customized simulated annealing algorithm for solving this problem when finding the exact solution is not computationally feasible. Our numerical study allowed us to quantify the value of having visibility of basket purchases across retail channels. The average gross profit gain was close to 10%, exceeding the potential profit gains estimated in the first two chapters and highlighting the crucial role such visibility plays in omnichannel assortment optimization.

This thesis suggests several interesting avenues for future research. First, the modeling framework presented in Chapters 1 and 2 could be extended by incorporating additional dimensions, such as pricing and promotion decisions, capacity constraints, and product return policies. These factors are closely intertwined with inventory management and assortment planning. Addressing all these decisions simultaneously presents a considerable challenge, but overcoming it could lead to significant financial benefits. It would also be valuable to integrate the time component into our model, which would allow it to span across numerous sales periods, as opposed to being limited to a single period. Furthermore, incorporating retailer competition into the framework would provide a more accurate representation of real-world complexities and enhance the model's practical applicability. Another promising research direction involves deriving additional theoretical properties of the assortment optimization problem and its optimal solutions. This would enable the development of more advanced heuristic methods for solving this problem in large-scale omnichannel environments, potentially improving solution quality and reducing solving time. Additionally, since demand variability significantly affects the structure of optimal assortments, exploring a robust formulation of the assortment optimization problem for cases with highly uncertain demand would be of interest. Lastly, a more in-depth understanding of the characteristics of optimal assortments in each channel could be

obtained through an extensive numerical analysis based on values of the MAM parameters estimated using a real-world dataset.

For the third chapter, it would be interesting to explore the possibility of tightening the upper bound of the partition function to improve the estimation quality of the Ising model parameters. Another potential avenue would be to leverage the graphical representation of the Ising model to enhance assortment optimization capabilities through network analytics techniques. For example, one could identify "important" products (those whose removal would substantially impact the demand for other products) using various centrality measures, or "established" baskets (sets of products that are often purchased together) using community detection algorithms. This information could then be incorporated into a customized assortment optimization algorithm. Furthermore, the Ising model – yet a powerful tool – is a rather basic example of an MRF. Considering more complex MRFs could enable our modeling framework to account for product quantities in each basket, marking a transition from the multi-purchase to the multiple-discrete choice modeling paradigm. This could be achieved by replacing binary variables with integer variables when defining basket choice probabilities, or alternatively, by creating a separate binary variable for each possible quantity of each product. Finally, the current configuration of our model does not allow for an explicit specification of the distribution of basket sizes. Finding a way to incorporate this distribution into the model would be a significant step forward. Overall, the MRF-based approach presented in this chapter holds considerable potential to improve omnichannel decision-making and advance the field of assortment optimization in general.

Acknowledgments

A Appendices for Chapter 1

A.1 The MAM Formulation in the General Setup

Suppose that $\mathcal{C} = \{1, \ldots, K\}$ is the set of channels with $K > 2$. We adjust and extend the notation presented in Section 1.3 in the following way. Let $u_{di}^{(c)}/v_i^{(c)}$ be the proportion of type-c customers willing to purchase product $i \in \mathcal{N}\backslash S_c$ in channel $d \in \mathcal{C}\backslash\{c\}$ if it is not available in channel c, and let $u_i^{(c)}$ denote the sum $\sum_{d \in \mathcal{C}\backslash\{c\}} u_{di}^{(c)}$. In other words, $u_i^{(c)}/v_i^{(c)}$ is the proportion of type-c customers willing to purchase product i in any of the retailer's channels if it is not available in channel c.

Then, the choice probabilities under the MAM given assortments in all channels can be defined by analogy to the two-channel case. The probability that a type-c customer buys product j in channel c is

$$
\pi_{cj}^{(c)}(S_c) = \begin{cases} \dfrac{v_j^{(c)}}{v_0^{(c)} + \sum\limits_{k \in S_c} v_k^{(c)} + \sum\limits_{i \in \mathcal{N}\backslash S_c} (u_i^{(c)} + w_i^{(c)})} & \text{if } j \in S_c, \\ 0 & \text{otherwise;} \end{cases}
$$

and the probability that a type-c customer buys product j in channel d is

$$
\pi_{dj}^{(c)}(S_c, S_d) = \begin{cases} \dfrac{u_{dj}^{(c)}}{v_0^{(c)} + \sum\limits_{k \in S_c} v_k^{(c)} + \sum\limits_{i \in \mathcal{N}\backslash S_c} (u_i^{(c)} + w_i^{(c)})} & \text{if } j \in S_d\backslash S_c, \\ 0 & \text{otherwise.} \end{cases}
$$

A.2 Proof of Proposition 1.1

First, note that the MAM restricted to type-c customers and products in channel c is equivalent to the GAM for products in this channel. At the same time, Blanchet et al. (2016) provided the set of MCCM parameters under which the choice probabilities can be expressed as the GAM probabilities. Suppose that we only consider choices of type-c customers in channel c. Then, the MCCM parameters that result in the GAM probabilities are as follows (see Blanchet et al., 2016):

$$
\lambda_{j_c}^{(c)} = v_j^{(c)}, \quad \rho_{j_c i_c}^{(c)} = \frac{v_i^{(c)}(v_j^{(c)} - u_j^{(c)} - w_j^{(c)})}{v_j^{(c)} - v_j^{(c)}(v_j^{(c)} - u_j^{(c)} - w_j^{(c)})},
$$
$$
\rho_{j_c 0}^{(c)} = \frac{v_0^{(c)}(v_j^{(c)} - u_j^{(c)} - w_j^{(c)}) + u_j^{(c)} + w_j^{(c)}}{v_j^{(c)} - v_j^{(c)}(v_j^{(c)} + u_j^{(c)} + w_j^{(c)})}.
$$
(A.1)

Now, let us also take into consideration products in channel $S_{\bar{c}}$. We therefore split the transition probability from product j_c to the no-purchase alternative into two parts: one corresponds to switching to the no-purchase alternative directly, and the other one corresponds to first purchasing product $j_{\bar{c}}$, and then – in case this product is not available – selecting the no-purchase option with probability 1. Formally, let us define

$$
\rho_{j_c 0}^{(c)} = \frac{v_0^{(c)}(v_j^{(c)} - u_j^{(c)} - w_j^{(c)}) + u_j^{(c)} + w_j^{(c)}}{v_j^{(c)} - v_j^{(c)}(v_j^{(c)} + u_j^{(c)} + w_j^{(c)})} - \frac{u_j^{(c)}}{v_j^{(c)}}, \quad \rho_{j_c j_{\bar{c}}}^{(c)} = \frac{u_j^{(c)}}{v_j^{(c)}}, \quad \rho_{j_{\bar{c}} 0}^{(c)} = 1,
$$

and leave $\lambda_{j_c}^{(c)} = v_j^{(c)}$ and $\rho_{j_c i_c}^{(c)}$ unchanged as in (A.1).

Finally, let us set the remaining parameters $\lambda_{j_{\bar{c}}}^{(c)}$, $\rho_{j_c i_{\bar{c}}}^{(c)}$, $\rho_{j_{\bar{c}} i_{\bar{c}}}^{(c)}$, $\rho_{j_{\bar{c}} j_c}^{(c)}$ and $\rho_{j_{\bar{c}} i_c}^{(c)}$ to be zero, which gives us exactly the set of parameters shown in (1.7). By construction, if type-c customers make their choices according to the MCCM with this set of parameters, then the probability that such a customer buys product j in channel c is exactly the MAM probability (1.5). What remains to be proven is that the probability of such a customer purchasing product j in channel $\bar{c}$ is the MAM probability (1.6).

Suppose that product j is offered in channel $\bar{c}$. Note that type-c customers interested in this product consider alternatives only in channel c before switching to channel $\bar{c}$. Therefore, in order to obtain probability $\pi_{\bar{c}j}^{(c)}(S_c, S_{\bar{c}})$, we can consider a subchain of our constructed Markov chain comprising all products in channel c, product j in channel $\bar{c}$, and the no-purchase alternative. Following Blanchet et al. (2016), let B denote the transition probability submatrix from alternatives that are not in the assortment to alternatives from the assortment (or to the no-purchase alternative), and let C denote the

transition probability submatrix among the alternatives that are not in the assortment. Furthermore, let $\hat{\lambda}$ be the vector of arrival probabilities to alternatives that are not in the assortments. Finally, let $e_{j_{\bar{c}}}$ be the standard unit vector such that the product $Be_{j_{\bar{c}}}$ corresponds to the vector of transition probabilities from alternatives that are not in the assortment to product j in channel $\bar{c}$. Then, using the formula for Markov chain choice probabilities (see Blanchet et al., 2016), we obtain the following:

$$
\begin{aligned}
\pi_{\bar{c}j}^{(c)}(S_c, S_{\bar{c}}) &= 0 + \hat{\lambda}^T(I - C)^{-1}Be_{j_{\bar{c}}} \\
&= \hat{\lambda}^T\left(\sum_{q=0}^{\infty} C^q\right)Be_{j_{\bar{c}}} = \left(\sum_{q=0}^{\infty}\Big(\sum_{i\in\mathcal{N}\backslash S_c} v_i^{(c)} - \sum_{i\in\mathcal{N}\backslash S_c}(u_i^{(c)} + w_i^{(c)})\Big)^q\right)\hat{\lambda}^T Be_{j_{\bar{c}}} \\
&= \frac{\hat{\lambda}^T Be_{j_{\bar{c}}}}{1 - \big(\sum_{i\in\mathcal{N}\backslash S_c} v_i^{(c)} - \sum_{i\in\mathcal{N}\backslash S_c}(u_i^{(c)} + w_i^{(c)})\big)} = \frac{v_j^{(c)}\rho_{j_c j_{\bar{c}}}^{(c)}}{v_0^{(c)} + \sum_{k\in S_c} v_k^{(c)} + \sum_{i\in\mathcal{N}\backslash S_c}(u_i^{(c)} + w_i^{(c)})} \\
&= v_0^{(c)} + \sum_{k\in S_c} v_k^{(c)} + \sum_{i\in\mathcal{N}\backslash S_c}(u_i^{(c)} + w_i^{(c)}),
\end{aligned}
$$

which concludes the proof. $\qquad\square$

A.3 Proof of Proposition 1.2

Recall that the GAM can be viewed as a limited case of the nested logit model where the dissimilarity parameter of each nest goes to zero (see Gallego et al., 2014). In this model, each nest corresponds to a product, and alternatives within the nest correspond to different sources where this product can be bought. Since the dissimilarity parameter of each nest tends to zero, it can be assumed that each nest comprises only two sources: The retailer itself (with the highest utility overall) and the outside source (with the highest utility among other available sources). Since the nested logit model is a RUM, the GAM is a RUM as well and it can be represented through a distribution over rankings (permutations) of alternatives. Readers are referred to Block and Marschak (1959), who showed how to construct the distribution over rankings from the joint distribution of random utilities, and vice versa. Importantly, if the number of products is n, then each ranking comprises $2n + 1$ alternatives since each product can be bought either from the retailer or from the outside source (and there is also the no-purchase alternative).

We can use the distribution over rankings that is consistent with the GAM to construct the distribution over rankings that is consistent with the MAM choice probabilities for one customer type. First, note that the choice probabilities of type-c customers selecting alternatives from channel c are exactly the same as the GAM choice probabilities defined over this set of products. Let us consider the corresponding distribution over rankings.

Appendices

One can insert the alternative of buying a certain product j from channel $\bar{c}$ into the rankings and modify the distribution in such a way that the probabilities of choosing products from channel c do not change while the probability of choosing product j from channel $\bar{c}$ is exactly the same as the MAM choice probability. Indeed, let us consider a ranking $r = (r_1, r_2, \ldots r_{2n+1})$ with $\mathbb{P}(r) = p$. Suppose that r_k corresponds to the option of buying product j from the outside source (which is always available). Let s denote the alternative of buying a certain product j from channel $\bar{c}$. Then, let us replace ranking r with two rankings $r' = (r_1, r_2, \ldots, r_{k-1}, s, r_k, r_{k+1} \ldots, r_{2n+1})$ and $r'' = (r_1, r_2, \ldots, r_{k-1}, r_k, s, r_{k+1} \ldots, r_{2n+1})$, such that $\mathbb{P}(r') = pu_j^{(c)}/(u_j^{(c)} + w_j^{(c)})$ and $\mathbb{P}(r'') = pw_j^{(c)}/(u_j^{(c)} + w_j^{(c)})$. Note that $u_j^{(c)}/(u_j^{(c)} + w_j^{(c)})$ is exactly the probability of type-c customers switching to channel $\bar{c}$ given that they are willing to purchase product j outside of channel c. If we do this for each ranking, then the probabilities of choosing products from channel c will not change by construction of rankings r' and r'' and since $\mathbb{P}(r') + \mathbb{P}(r'') = \mathbb{P}(r)$ for all rankings r. At the same time, the probability of choosing product $j \in S_{\bar{c}}$ will be defined as the probability of choosing the outside option according to the GAM multiplied by the coefficient $u_j^{(c)}/(u_j^{(c)} + w_j^{(c)})$, that is

$$\frac{(u_j^{(c)} + w_j^{(c)})\mathbb{1}_{j \in \mathcal{N} \setminus S_c}}{v_0^{(c)} + \sum_{k \in S_c} v_k^{(c)} + \sum_{i \in \mathcal{N} \setminus S_c}(u_i^{(c)} + w_i^{(c)})} \cdot \frac{u_j^{(c)} \mathbb{1}_{j \in S_{\bar{c}}}}{u_j^{(c)} + w_j^{(c)}} = \frac{u_j^{(c)} \mathbb{1}_{j \in S_{\bar{c}} \setminus S_c}}{v_0^{(c)} + \sum_{k \in S_c} v_k^{(c)} + \sum_{i \in \mathcal{N} \setminus S_c}(u_i^{(c)} + w_i^{(c)})},$$

which is exactly the MAM choice probability $\pi_{\bar{c}j}^{(c)}(S_c, S_{\bar{c}})$. If we repeat this procedure n times for all products in channel $\bar{c}$, then we obtain the distribution over rankings that is consistent with the MAM choice probabilities. The MAM is thus a mixture of RUMs.

Finally, note that a mixture of RUMs is also a RUM. Indeed, from the distributions over rankings of alternatives consistent with each individual model in the mixture, one can straightforwardly construct a distribution over rankings corresponding to the mixture model. In the case of the MAM, suppose that the distribution over rankings that corresponds to choice probabilities of type-c customers is as follows:

$$D^{(c)} = \{(r^1, p_1^{(c)}), (r^2, p_2^{(c)}), \ldots, (r^h, p_h^{(c)})\},$$

where each r^i is a ranking and $p_i^{(c)}$ is the associated probability. Similarly, let

$$D^{(\bar{c})} = \{(r^1, p_1^{(\bar{c})}), (r^2, p_2^{(\bar{c})}), \ldots, (r^h, p_h^{(\bar{c})})\}$$

be the distribution over rankings of alternatives that corresponds to choice probabilities of type-$\bar{c}$ customers. Note that both distributions $D^{(c)}$ and $D^{(\bar{c})}$ are defined over the same set of rankings of alternatives which is composed of all permutations of all possible

purchase outcomes (including the no-purchase option). In particular, each ranking r^i contains $3n + 1$ alternatives (the no-purchase option together with products in channel c, channel $\bar{c}$, and the outside source) and h equals the total number of permutations of alternatives, i.e., $h = (3n + 1)!$. Then, it is easy to see that the following distribution corresponds to the MAM choice probabilities:

$$D = \{(r^1, \omega p_1^{(c)} + (1 - \omega)p_1^{(\bar{c})}), (r^2, \omega p_2^{(c)} + (1 - \omega)p_2^{(\bar{c})}), \ldots, (r^h, \omega p_h^{(c)} + (1 - \omega)p_h^{(\bar{c})})\},$$

where $\omega = \Lambda^{(c)}/(\Lambda^{(c)} + \Lambda^{(\bar{c})})$. Thus, the MAM is also a RUM. $\qquad\square$

A.4 Proof of Theorem 1.1

In order to prove the theorem, we need to show the following:

1) Let $S_c, S_{\bar{c}}$ be arbitrary subsets of $\mathcal{N}$. If $x_{cj}^{(c)}$, $x_{\bar{c}j}^{(c)}$ correspond to probabilities (1.5), (1.6) multiplied by $\Lambda^{(c)}$, z_{cj} take the corresponding binary values, and $x_{c0}^{(c)}$ satisfy

$$x_{c0}^{(c)} = \frac{v_0^{(c)}\Lambda^{(c)}}{v_0^{(c)} + \sum\limits_{k \in S_c} v_k^{(c)} + \sum\limits_{i \in \mathcal{N} \setminus S_c} (u_i^{(c)} + w_i^{(c)})},$$

then $\left\{x_{c0}^{(c)}, x_{cj}^{(c)}, x_{\bar{c}j}^{(c)}, z_{cj}\right\}_{c \in \mathcal{C}, j \in \mathcal{N}}$ is a feasible solution to the SBMILP.

2) If $\left\{x_{c0}^{(c)}, x_{cj}^{(c)}, x_{\bar{c}j}^{(c)}, z_{cj}\right\}_{c \in \mathcal{C}, j \in \mathcal{N}}$ is an optimal solution to the SBMILP, then there exist offer sets $S_c, S_{\bar{c}} \subseteq \mathcal{N}$ such that $x_{cj}^{(c)}$, $x_{\bar{c}j}^{(c)}$ correspond to probabilities (1.5), (1.6) multiplied by $\Lambda^{(c)}$, and z_{cj} take the corresponding binary values.

The first part can be shown straightforwardly by substituting $\left\{x_{c0}^{(c)}, x_{cj}^{(c)}, x_{\bar{c}j}^{(c)}, z_{cj}\right\}_{c \in \mathcal{C}, j \in \mathcal{N}}$ into the SBMILP constraints. For the second part, note that from constraints (1.9e) and the form of the objective function it follows that $z_{cj} = 1$ if and only if $x_{cj}^{(c)} > 0$ (since z_{cj} are binary and $x_{cj}^{(c)}$ are nonnegative). Furthermore, from constraints (1.9c) and (1.9d) we can see that

$$\frac{x_{cj}^{(c)}}{v_j^{(c)}} = \begin{cases} \dfrac{x_{c0}^{(c)}}{v_0^{(c)}} & \text{if } z_{cj} = 1, \\[2ex] 0 & \text{otherwise.} \end{cases} \tag{A.2}$$

Let us set $S_c = \{j \in \mathcal{N} : z_{cj} = 1\}$. Substituting (A.2) into constraints (1.9b), we obtain

the following:

$$\frac{\tilde{v}_0^{(c)}}{v_0^{(c)}}x_{c0}^{(c)} + \sum_{j \in \mathcal{N}} \frac{\tilde{v}_j^{(c)}}{v_j^{(c)}}x_{cj}^{(c)} = \frac{\tilde{v}_0^{(c)}}{v_0^{(c)}}x_{c0}^{(c)} + \sum_{k \in S_c} \frac{\tilde{v}_k^{(c)}}{v_0^{(c)}}x_{c0}^{(c)} = \Lambda^{(c)}.$$

Hence,

$$x_{c0}^{(c)} = \frac{v_0^{(c)}\Lambda^{(c)}}{\tilde{v}_0^{(c)} + \sum\limits_{k \in S_c} \tilde{v}_k^{(c)}} = \frac{v_0^{(c)}\Lambda^{(c)}}{v_0^{(c)} + \sum\limits_{k \in S_c} v_k^{(c)} + \sum\limits_{i \in \mathcal{N} \setminus S_c} (u_i^{(c)} + w_i^{(c)})}. \tag{A.3}$$

Subsequently, by combining (A.2) and (A.3), we obtain the desired values of $x_{cj}^{(c)}$. Finally, note that constraints (1.9c) and (1.9f) imply that $x_{\bar{c}j}^{(c)} \neq 0$ if and only if $x_{cj}^{(c)} = 0$ and $x_{\bar{c}j}^{(\bar{c})} \neq 0$. In that case, $x_{\bar{c}j}^{(c)}$ attains its maximum value at

$$x_{\bar{c}j}^{(c)} = \frac{u_j^{(c)}\Lambda^{(c)}}{v_0^{(c)} + \sum\limits_{k \in S_c} v_k^{(c)} + \sum\limits_{i \in \mathcal{N} \setminus S_c} (u_i^{(c)} + w_i^{(c)})},$$

which concludes the proof. $\qquad \square$

A.5 Proof of Proposition 1.3

Let all shadow attractiveness values together with all parameters related to channel $\bar{c}$ be zero. Then, the assortment optimization problem under the MAM with shelf space constraint (1.10) takes the following form:

$$\max_{z_{cj} \in \{0,1\}^n} \frac{\sum\limits_{j \in \mathcal{N}} r_{cj} v_j^{(c)} z_{cj}}{v_0^{(c)} + \sum\limits_{k \in \mathcal{N}} v_k^{(c)} z_{ck}} \quad \text{s.t.} \quad \sum_{j \in \mathcal{N}} a_{cj} z_{cj} \leq L_c. \tag{A.4}$$

Problem (A.4) is essentially the shelf-space-constrained assortment optimization problem under the MNL, which has been shown by Désir et al. (2022) to be NP-hard. Indeed, suppose that $r_{cj} = 1$ for all $j \in \mathcal{N}$ and, following Désir et al. (2022), note that function $f(x) = \dfrac{x}{v_0^{(c)} + x}$ is increasing in x, meaning that the objective function of problem (A.4) can be replaced by $\sum_{j \in \mathcal{N}} v_j^{(c)} z_{cj}$. The resulting problem is equivalent to the knapsack problem. Therefore, the shelf-space constrained SBMILP – i.e. the SMBILP formulation (1.9) with additional constraint (1.10) – is NP-hard by reduction from the knapsack problem. $\qquad \square$

A.6   Proof of Proposition 1.4

First, let us show that $R^{(c)}(S_c \cup \{j\}, \mathcal{N})$ can be expressed as a convex combination of $R^{(c)}(S_c, \mathcal{N})$ and $(r_{cj}v_j^{(c)} - r_{\bar{c}j}u_j^{(c)})\Lambda^{(c)}/\tilde{v}_j^{(c)}$. Consider the following:

$$\alpha = \frac{\tilde{v}_j^{(c)}}{v_0^{(c)} + \sum\limits_{k \in S_c \cup \{j\}} v_k^{(c)} + \sum\limits_{i \in \mathcal{N} \setminus (S_c \cup \{j\})} (u_i^{(c)} + w_i^{(c)})}.$$

Note that $\alpha \in (0, 1)$. Then, it is easy to verify that

$$\alpha \frac{(r_{cj}v_j^{(c)} - r_{\bar{c}j}u_j^{(c)})\Lambda^{(c)}}{\tilde{v}_j^{(c)}} + (1 - \alpha)R^{(c)}(S_c, \mathcal{N}) = R^{(c)}(S_c \cup \{j\}, \mathcal{N}).$$

This fact is sufficient for showing that the assortment given by (1.11) is optimal for channel c. Indeed, suppose that there is an optimal assortment S_c such that $q \in S_c$ and $p \notin S_c$ for some $p < q$. Then, since S_c is optimal, $R^{(c)}(S_c \setminus \{q\}, \mathcal{N}) \leq R^{(c)}(S_c, \mathcal{N}) \leq (r_{cq}v_q^{(c)} - r_{\bar{c}q}u_q^{(c)})\Lambda^{(c)}/\tilde{v}_q^{(c)}$. At the same time, since $p < q$ and hence $(r_{cq}v_q^{(c)} - r_{\bar{c}q}u_q^{(c)})\Lambda^{(c)}/\tilde{v}_q^{(c)} < (r_{cp}v_p^{(c)} - r_{\bar{c}p}u_p^{(c)})\Lambda^{(c)}/\tilde{v}_p^{(c)}$, it follows that $R^{(c)}(S_c, \mathcal{N}) < (r_{cp}v_p^{(c)} - r_{\bar{c}p}u_p^{(c)})\Lambda^{(c)}/\tilde{v}_p^{(c)}$. Therefore, $R^{(c)}(S_c \cup \{p\}, \mathcal{N})$ is greater than $R^{(c)}(S_c, \mathcal{N})$ as a convex combination of $R^{(c)}(S_c, \mathcal{N})$ and $(r_{cp}v_p^{(c)} - r_{\bar{c}p}u_p^{(c)})\Lambda^{(c)}/\tilde{v}_p^{(c)}$, which contradicts the assumption that S_c is optimal. It means that the optimal allocation has to be in descending order of $(r_{cj}v_j^{(c)} - r_{\bar{c}j}u_j^{(c)})\Lambda^{(c)}/\tilde{v}_j^{(c)}$. Finally, the fact that the optimal assortment has to be of the form (1.11) with index m specified in Proposition 1.4 follows from the same convex combination observation. $\qquad \square$

A.7   Proof of Proposition 1.5

Let $(S_c, S_{\bar{c}})$ be the optimal combination of assortments. Consider channel c and suppose that all products in that channel are sorted in descending order of the value of expression (1.5). Suppose that the proposition does not hold, i.e., for some $p < q$ we have that $q \in S_c$ and $p \notin S_c$. First, note that the total revenue generated by type-c customers is as follows:

$$R^{(c)}(S_c, S_{\bar{c}}) = \frac{\left(\sum\limits_{k \in S_c} r_{ck}v_k^{(c)} + \sum\limits_{i \in S_{\bar{c}} \setminus S_c} r_{\bar{c}i}u_i^{(c)}\right)\Lambda^{(c)}}{v_0^{(c)} + \sum\limits_{k \in S_c} v_k^{(c)} + \sum\limits_{i \in \mathcal{N} \setminus S_c} (u_i^{(c)} + w_i^{(c)})},$$

and the revenue generated by type-$\bar{c}$ customers purchasing product j in channel c is:

$$R_{cj}^{(\bar{c})}(S_c, S_{\bar{c}}) = \frac{r_{cj} u_j^{(\bar{c})} \Lambda^{(\bar{c})} \mathbb{1}_{j \in S_c \setminus S_{\bar{c}}}}{v_0^{(\bar{c})} + \sum\limits_{k \in S_{\bar{c}}} v_k^{(\bar{c})} + \sum\limits_{i \in \mathcal{N} \setminus S_{\bar{c}}} (u_i^{(\bar{c})} + w_i^{(\bar{c})})}.$$

Given that $q \in S_c$ and $p \notin S_c$, one can check that

$$\beta_1 \Lambda^{(c)} F_c(p, S_c, S_{\bar{c}}) + (1 - \beta_1) R^{(c)}(S_c, S_{\bar{c}}) = R^{(c)}(S_c \cup \{p\}, S_{\bar{c}}) + R_{cp}^{(\bar{c})}(S_c \cup \{p\}, S_{\bar{c}}) \quad \text{(A.5)}$$

and

$$\beta_2 \Lambda^{(c)} F_c(q, S_c, S_{\bar{c}}) + (1 - \beta_2) R^{(c)}(S_c \setminus \{q\}, S_{\bar{c}}) = R^{(c)}(S_c, S_{\bar{c}}) + R_{cq}^{(\bar{c})}(S_c, S_{\bar{c}}), \quad \text{(A.6)}$$

where

$$\beta_1 = \frac{\tilde{v}_p^{(c)}}{v_0^{(c)} + \sum\limits_{k \in S_c \cup \{p\}} v_k^{(c)} + \sum\limits_{i \in \mathcal{N} \setminus (S_c \cup \{p\})} (u_i^{(c)} + w_i^{(c)})}$$

and

$$\beta_2 = \frac{\tilde{v}_q^{(c)}}{v_0^{(c)} + \sum\limits_{k \in S_c} v_k^{(c)} + \sum\limits_{i \in \mathcal{N} \setminus S_c} (u_i^{(c)} + w_i^{(c)})}.$$

Since $(S_c, S_{\bar{c}})$ is assumed to be the optimal combination of assortments, it holds that $R^{(c)}(S_c \setminus \{q\}, S_{\bar{c}}) < R^{(c)}(S_c, S_{\bar{c}}) + R_{cq}^{(\bar{c})}(S_c, S_{\bar{c}})$. Therefore, from relation (A.6) it follows that $\Lambda^{(c)} F_c(q, S_c, S_{\bar{c}}) > R^{(c)}(S_c, S_{\bar{c}}) + R_{cq}^{(\bar{c})}(S_c, S_{\bar{c}}) \geq R^{(c)}(S_c, S_{\bar{c}})$. Since $p < q$ and products are sorted in descending order of the value of expression (1.5), we obtain that $\Lambda^{(c)} F_c(p, S_c, S_{\bar{c}}) > R^{(c)}(S_c, S_{\bar{c}})$. Thus, from relation (A.5) it follows that $R^{(c)}(S_c \cup \{p\}, S_{\bar{c}}) + R_{cp}^{(\bar{c})}(S_c \cup \{p\}, S_{\bar{c}}) > R^{(c)}(S_c, S_{\bar{c}})$. This means that the combination of assortments $(S_c \cup \{p\}, S_{\bar{c}})$ is more profitable than $(S_c, S_{\bar{c}})$, which contradicts the initial assumption and thereby concludes the proof. $\qquad\square$

A.8 MAM Estimation based on an Adaptation of the EM Algorithm

We provide an alternative method to estimate the MAM parameters by building upon the Expectation Maximization (EM) algorithm which was developed by Vulcano et al. (2012)

for estimating the parameters of the BAM and then adapted to the GAM by Gallego et al. (2014). We first give a brief description of the algorithm proposed by Gallego et al. (2014) and discuss its limitations.

Their algorithm allows to estimate the GAM parameters given product demands in T periods assuming that for each period $t = 1, \ldots, T$, the offered set S_t is constant and the arrival rate is $\Lambda_t = \Lambda$. The parameters are estimated through an iterative procedure in which the observed demand is viewed as an incomplete observation of the first-choice demand (the demand occurring when all the products belong to the offered set). More specifically, the estimation procedure is based on the following expressions. For each period t, let d_{jt} and X_{jt} be the observed demand and the first-choice demand for product j, respectively. If the GAM parameters are assumed to be known, then an estimate $\hat{X}_{jt}$ of the first-choice demand given observed demands in period t can be found in the following way (see Gallego et al. (2014)):

$$
\hat{X}_{jt} = \begin{cases} \dfrac{\pi_j(\mathcal{N}) d_{jt}}{\pi_j(S_t)} & \text{if } j \in S_t, \\[2ex] \dfrac{\pi_j(\mathcal{N}) \sum_{j \in S_t} d_{jt}}{\sum_{k \in S_t} \pi_k(S_t)} & \text{otherwise.} \end{cases} \tag{A.7}
$$

Aggregating over all the periods, the estimate of the first-choice demand can be obtained as $\hat{X}_j = \sum_{t=1}^{T} \hat{X}_{jt}/T$. Gallego et al. (2014) also proposed to use the following weighting scheme to compute $\hat{X}_j$, which allows to reduce the variance of the estimators: $\hat{X}_j = \sum_{t=1}^{T} a_{jt} \hat{X}_{jt}/(\sum_{t=1}^{T} a_{jt})$, where $a_{jt} = \pi_j(S_t)$ if $j \in S_t$; and $a_{jt} = \sum_{k \in S_t} \pi_k(S_t)$ otherwise.

Without loss of generality, it can be assumed that $v_0 + \sum_{j \in \mathcal{N}} v_j = 1$. Let $r = \dfrac{v_0}{\sum_{j \in \mathcal{N}} v_j}$. Then, each iteration of the estimation procedure comprises the following two steps:

E-step (GAM). Using current estimates of the parameters, compute the values of the first-choice demands and update the estimates of v in the following way:

$$
\hat{v}_j = \frac{\hat{X}_j}{r \sum_{k \in \mathcal{N}} \hat{X}_k}. \tag{A.8}
$$

M-step (GAM). Substitute the estimates of v obtained in the *E-step* into the following least squares minimization problem:

$$
\min_{w, \Lambda} \sum_{t=1}^{T} \sum_{j \in S_t} \left(\frac{v_j \Lambda}{v_0 + \sum_{k \in S_t} v_k + \sum_{i \in \mathcal{N} \setminus S_t} w_i} - d_{jt} \right)^2 \tag{A.9}
$$

subject to the constraints $0 \le w_j \le v_j \; \forall j \in \mathcal{N}$ and $\Lambda \ge 0$. Solve this optimization problem to update the values of w and Λ.

Appendices

In their paper, Gallego et al. (2014) also suggested to update the value of $r = \dfrac{v_0}{\sum_{j \in \mathcal{N}} v_j}$ with each iteration. However, from (A.8) it follows that $\sum_{j \in \mathcal{N}} \hat{v}_j$ remains constant over iterations, which means that the ratio $\dfrac{\hat{v}_0}{\sum_{j \in \mathcal{N}} \hat{v}_j} = \dfrac{1 - \sum_{j \in \mathcal{N}} \hat{v}_j}{\sum_{j \in \mathcal{N}} \hat{v}_j}$ does not change either. It is therefore reasonable to treat r as a given parameter, that is to assume that r (or, equivalently, v_0) is given exogenously, rather than being determined within the estimation procedure.

To set initial values of v and w before starting the iteration procedure, Gallego et al. (2014) proposed to solve the least squares minimization problem of the form (A.9) but with respect to v, w and Λ. However, considering v as a vector of decision variables rather than a vector of parameters makes such an optimization problem quite challenging to solve. An alternative way to initialize v and w is to choose their values from the corresponding domain, e.g., by setting $v_j = (1 - v_0)/n$ and $w_j = 0 \; \forall j \in \mathcal{N}$.

The described algorithm to estimate the GAM parameters suffers from two major issues. Firstly and most importantly, its convergence is not theoretically guaranteed. The reason is that problem (A.9) – which has to be solved at each iteration – is not convex and, as such, it may have several optimal solutions. Secondly, solving a nonconvex minimization problem at each iteration is computationally hard, making this approach increasingly less attractive as the number of iterations needed to obtain accurate parameter estimates grows. However, running time issues emerging from solving problem (A.9) become substantially less pronounced if the additional assumption of having exogenous knowledge of the true value of Λ is imposed. Such an assumption is rather weak, because in the case where the offered set is $S_t = \mathcal{N}$ for at least one period t, it is equivalent to the previously mentioned assumption that v_0 is given exogenously. In addition, note that if the value of Λ is known, then the expressions for the first-choice demand (A.7) can be simplified in the following way:

$$
\hat{X}_{jt} = \begin{cases} \dfrac{\pi_j(\mathcal{N}) d_{jt}}{\pi_j(S_t)} & \text{if } j \in S_t, \\[2ex] \pi_j(\mathcal{N}) \Lambda & \text{otherwise.} \end{cases}
$$

Nevertheless, problem (A.9) remains nonconvex even if we consider Λ as a parameter rather than a decision variable, and hence the convergence of the algorithm is still not guaranteed.

We now adapt this estimation algorithm to the MAM. Without loss of generality, let $v_0^{(c)} + \sum_{j \in \mathcal{N}} v_j^{(c)} = 1 \; \forall c \in \mathcal{C}$. Similar to the case of the GAM, we assume that for each channel c, the value of $v_0^{(c)}$ is given exogenously and $\Lambda_t^{(c)} = \Lambda^{(c)} \; \forall t = 1, \ldots, T$. In

addition, we also assume that we have knowledge of the true values of $\Lambda^{(c)}$ $\forall c \in \mathcal{C}$. Since the values of $v_0^{(c)}$ $\forall c \in \mathcal{C}$ are known, the latter assumption is rather weak (similar to the single-channel case discussed above). If it does not hold, however, our numerical experiments have shown that there is no advantage of our proposed algorithm over the least-squares estimation of the MAM parameters described in Section 1.6 in terms of solution quality.

For each period t, let d_{cjt} and X_{cjt} be the observed demand and the first-choice demand for product j in channel c, respectively. If the MAM parameters are assumed to be known, then an estimate $\hat{X}_{cjt}$ of the first-choice demand can be obtained in the following way:

$$
\hat{X}_{cjt} = \begin{cases} (d_{cjt} - \pi_{cj}^{(\bar{c})} \Lambda^{\bar{c}}) \dfrac{\pi_{cj}^{(c)}(\mathcal{N})}{\pi_{cj}^{(c)}(S_t)} & \text{if } j \in S_t, \\[12pt] \pi_{cj}^{(c)}(\mathcal{N})\Lambda^{(c)} & \text{otherwise.} \end{cases}
$$

We define $\hat{X}_{cj}$ – the estimate of the first-choice demand for product j in channel c aggregated over all the periods – in the following way: $\hat{X}_{cj} = \sum_{t=1}^{T} \hat{X}_{cjt}/T$. Lastly, let $r^{(c)} = \dfrac{v_0^{(c)}}{\sum_{j \in \mathcal{N}} v_j^{(c)}}$ $\forall c \in \mathcal{C}$. Note that $r^{(c)}$ is a constant under the previously mentioned assumption that $v_0^{(c)}$ is given exogenously.

For the parameter initialization stage, we set $v_j^{(c)} = (1 - v_0^{(c)})/n$ and $u_j^{(c)} = w_j^{(c)} = 0$ $\forall c \in \mathcal{C}$, $j \in \mathcal{N}$. Then, each iteration of our algorithm to estimate the MAM parameters comprises the following two steps:

E-step (MAM). Using current estimates of the parameters, compute the values of the first-choice demands for each channel c and update the estimates of $v^{(c)}$ in the following way:

$$
\hat{v}_j^{(c)} = \frac{\hat{X}_{cj}}{r^{(c)} \sum_{k \in \mathcal{N}} \hat{X}_{ck}}.
$$

M-step (MAM). Substitute the estimates of $v^{(c)}$ obtained in the *E-step* into the following least squares minimization problem:

$$
\min_{u^{(c)}, w^{(c)}} \sum_{t=1}^{T} \sum_{c \in \mathcal{C}} \sum_{j \in \mathcal{N}} \Bigg(\frac{v_j^{(c)} \Lambda^{(c)} \mathbb{1}_{j \in S_{ct}}}{v_0^{(c)} + \sum_{k \in S_{ct}} v_k^{(c)} + \sum_{i \in \mathcal{N} \setminus S_{ct}} (u_i^{(c)} + w_i^{(c)})} +
$$
$$
\frac{u_j^{(\bar{c})} \Lambda^{(\bar{c})} \mathbb{1}_{j \in S_{ct} \setminus S_{\bar{c}t}}}{v_0^{(\bar{c})} + \sum_{k \in S_{\bar{c}t}} v_k^{(\bar{c})} + \sum_{i \in \mathcal{N} \setminus S_{\bar{c}t}} (u_i^{(\bar{c})} + w_i^{(\bar{c})})} - d_{cjt} \Bigg)^2
\tag{A.10}
$$

123

subject to the constraints $0 \leq u_j^{(c)} + w_j^{(c)} \leq v_j^{(c)} \ \forall c \in \mathcal{C}, \ j \in \mathcal{N}$. Solve this optimization problem to update the values of $\{u^{(c)}, w^{(c)}\}_{c \in \mathcal{C}}$.

To illustrate the performance of the just-described estimation method, we use the same setup as in Section 1.6. More precisely, we set $n = 5$ and $T = 15$, and we simulate 100 instances of demand arising from the MAM with fixed parameters, with each instance corresponding to a set of randomly generated assortments S_{ct}, $c \in \mathcal{C}$, $t \in \{1, \ldots, T\}$. The estimation results are provided in Table A.1. It can be seen that the estimated values are close to the true values of the parameters, especially in the case of parameters $v_j^{(c)}$ and $u_j^{(c)}$. Even though these estimates are more accurate compared to those obtained in Section 1.6, here we impose the additional assumption that the values of $\Lambda^{(c)} \ \forall c \in \mathcal{C}$ are given exogenously, which clearly improves the estimation accuracy. Moreover, similar to the algorithm developed by Gallego et al. (2014), the convergence of our proposed EM-based algorithm is not theoretically guaranteed since problem (A.10) is also nonconvex and, as such, it may also have several optimal solutions. However, in our extensive numerical experiments the proposed algorithm has always converged for randomly simulated sets of the MAM parameters, meaning that our estimation procedure can prove useful and effective in practice.

	j	$v_j^{(c)}$ True	$v_j^{(c)}$ Estimated	$u_j^{(c)}$ True	$u_j^{(c)}$ Estimated	$w_j^{(c)}$ True	$w_j^{(c)}$ Estimated
channel 1	1	0.068	0.068	0.013	0.013	0.027	0.027
	2	0.145	0.146	0.045	0.044	0.031	0.032
	3	0.233	0.232	0.106	0.105	0.017	0.019
	4	0.096	0.096	0.047	0.047	0.039	0.039
	5	0.221	0.220	0.035	0.034	0.095	0.096
channel 2	1	0.105	0.105	0.021	0.021	0.046	0.028
	2	0.199	0.199	0.068	0.068	0.093	0.035
	3	0.195	0.194	0.081	0.081	0.083	0.021
	4	0.236	0.236	0.061	0.061	0.088	0.042
	5	0.008	0.008	0.000	0.001	0.001	0.096

Table A.1: Estimates of the MAM parameters based on an adaptation of the EM algorithm.

B Appendices for Chapter 2

B.1 Proof of Proposition 2.1

If $\mu^{(c)} > \sigma^{(c)}/CV_{max}$ for each $c \in \mathcal{C}$, then:

$$\left(\mu^{(c)} \pi_{cj}^{(c)}(S_c) + \sum_{\bar{c} \in \mathcal{C}_{cj}} \mu^{(\bar{c})} \pi_{cj}^{(\bar{c})}(S) \right)^2 \geq (\mu^{(c)} \pi_{cj}^{(c)}(S_c))^2 + \sum_{\bar{c} \in \mathcal{C}_{cj}} (\mu^{(\bar{c})} \pi_{cj}^{(\bar{c})}(S))^2 >$$

$$> \left((\sigma^{(c)} \pi_{cj}^{(c)}(S_c))^2 + \sum_{\bar{c} \in \mathcal{C}_{cj}} (\sigma^{(\bar{c})} \pi_{cj}^{(\bar{c})}(S))^2 \right)/CV_{max}^2.$$

Therefore, we obtain:

$$CV(D_{cj}) = \frac{\left((\sigma^{(c)} \pi_{cj}^{(c)}(S_c))^2 + \sum_{\bar{c} \in \mathcal{C}_{cj}} (\sigma^{(\bar{c})} \pi_{cj}^{(\bar{c})}(S))^2 \right)^{1/2}}{\mu^{(c)} \pi_{cj}^{(c)}(S_c) + \sum_{\bar{c} \in \mathcal{C}_{cj}} \mu^{(\bar{c})} \pi_{cj}^{(\bar{c})}(S)} < CV_{max}. \qquad \square$$

B.2 Proof of Theorem 2.1

We show that a special case of our assortment optimization problem is equivalent to the assortment optimization problem under the 2-product nonparametric choice model presented by Feldman et al. (2019), which in turn has been proven to be NP-hard. Feldman et al. (2019) studied a model where each customer is characterized by a ranked preference list comprising at most k products. Consequently, customers can be separated into different classes based on their preference lists, and customers in each class purchase the highest ranking product available from the corresponding list. Every customer belongs to a class with a given probability. Feldman et al. (2019) named this model the k-product nonparametric choice model, and they showed that the assortment optimization problem under this model is NP-hard even if $k = 2$ (through a reduction from the minimum vertex

cover problem on cubic graphs).

Let us recall how the choice probabilities are defined under the 2-product nonparametric choice model (2-NPCM). We adjust the notation used by Feldman et al. (2019) to enhance the clarity of our proof. Let $\mathcal{N}$ denote the set of choice alternatives. Also, let λ_{jk} denote the probability of a customer belonging to the class with the preference list $\{j, k\}$, where j is the most preferred alternative. Probability λ_{jk} can be expressed as the product $\lambda_{jk} = \lambda_j p(k|j)$, where λ_j represents the probability of a customer belonging to a class in which j is the most preferred alternative, and $p(k|j)$ denotes the conditional probability of a customer belonging to the class in which k is the second most preferred alternative given that j is the most preferred one. Then, the probability of a customer purchasing alternative j is as follows:

$$Pr_j(z) = \sum_{k \in \mathcal{N} \setminus \{j\}} \lambda_{jk} z_j + \sum_{i \in \mathcal{N} \setminus \{j\}} \lambda_{ij}(1 - z_i)z_j = \lambda_j z_j + \sum_{i \in \mathcal{N} \setminus \{j\}} \lambda_i p(j|i)(1 - z_i)z_j, \quad \text{(B.1)}$$

where z_j denotes a binary variable representing whether product j is offered, and $z = \{z_j\}_{j \in \mathcal{N}}$.

We now construct an instance of the MAM that is equivalent to the 2-NPCM. Suppose that in the network representation, all channels (both physical and online) are interconnected. Additionally, suppose that the retailer sells a single product, and all costs are zero, meaning that unit product prices are equal to the unit profit margins in each channel. Since the retailer sells only one product, we omit product indices in our notation. Let the attractiveness value of the no-purchase alternative $v_0^{(c)}$ be zero, the product attractiveness value $v^{(c)}$ be equal to the product shadow attractiveness value $\phi^{(c)}$ for all channels $c \in \mathcal{C}$. Moreover, let the standard deviation $\sigma^{(c)}$ of the number of type-c customers be zero for all channels $c \in \mathcal{C}$. Lastly, let $\hat{\mu}^{(c)} = \mu^{(c)}/\left(\sum_{c \in \mathcal{C}} \mu^{(c)}\right)$ denote the ratio of the expected number of type-c customers to the expected total number of customers. Then, the probability that a customer of any type purchases the product in channel c is as follows:

$$\pi_c(z) = \hat{\mu}^{(c)} z_c + \sum_{\bar{c} \in \mathcal{C} \setminus \{c\}} \hat{\mu}^{(\bar{c})} \delta_c^{(\bar{c})}(1 - z_{\bar{c}})z_c. \quad \text{(B.2)}$$

Clearly, probabilities (B.1) and (B.2) are equivalent if the set of channels under the MAM is viewed as the set of alternatives under the 2-NPCM (i.e. $\mathcal{N} = \mathcal{C}$), $\hat{\mu}^{(j)} = \lambda_j$ and $\delta_j^{(i)} = p(j|i)$. Therefore, the assortment optimization problems under the described special case of the MAM and the 2-NPCM are equivalent to each other, which concludes the proof. $\square$

B.3 Proof of Theorem 2.2

Let us call a vector of decision variables $x = \left\{ x_{c0}^{(c)}, x_{cj}^{(c)}, x_{\bar{c}j}^{(c)}, y_{cj}^{(c)}, y_{\bar{c}j}^{(c)}, t_{cj}, z_{cj} \right\}_{c \in \mathcal{C}, j \in \mathcal{N}, \bar{c} \in \mathcal{C}_{cj}}$ "consistent" if it corresponds to the demand under the MAM, i.e, if all decision variables satisfy the following expressions for all $c \in \mathcal{C}$, $\bar{c} \in \mathcal{C}_{cj}$, $j \in \mathcal{N}$:

$$x_{cj}^{(c)} = \frac{\mu^{(c)} v_j^{(c)} z_{cj}}{v_0^{(c)} + \sum_{k \in \mathcal{N}} v_k^{(c)} z_{ck} + \sum_{i \in \mathcal{N}} \phi_i^{(c)}(1 - z_{ci})}, \tag{B.3a}$$

$$y_{cj}^{(c)} = \frac{\sigma^{(c)} v_j^{(c)} z_{cj}}{v_0^{(c)} + \sum_{k \in \mathcal{N}} v_k^{(c)} z_{ck} + \sum_{i \in \mathcal{N}} \phi_i^{(c)}(1 - z_{ci})}, \tag{B.3b}$$

$$x_{\bar{c}j}^{(c)} = \frac{\mu^{(c)} \phi_j^{(c)} \delta_{\bar{c}j}^{(c)}(1 - z_{cj}) z_{\bar{c}j}}{v_0^{(c)} + \sum_{k \in \mathcal{N}} v_k^{(c)} z_{ck} + \sum_{i \in \mathcal{N}} \phi_i^{(c)}(1 - z_{ci})}, \tag{B.3c}$$

$$y_{\bar{c}j}^{(c)} = \frac{\sigma^{(c)} \phi_j^{(c)} \delta_{\bar{c}j}^{(c)}(1 - z_{cj}) z_{\bar{c}j}}{v_0^{(c)} + \sum_{k \in \mathcal{N}} v_k^{(c)} z_{ck} + \sum_{i \in \mathcal{N}} \phi_i^{(c)}(1 - z_{ci})}, \tag{B.3d}$$

$$t_{cj} = \sqrt{\left(y_{cj}^{(c)}\right)^2 + \sum_{\bar{c} \in \mathcal{C}_{cj}} \left(y_{cj}^{(\bar{c})}\right)^2} = \left\| \left(y_{cj}^{(c)}, \left\{ y_{cj}^{(\bar{c})} \right\}_{\bar{c} \in \mathcal{C}_{cj}} \right) \right\|_2. \tag{B.3e}$$

To prove the theorem, it is sufficient to demonstrate the following two points:

1) For any set of binary values of z_{cj} defining the channel assortments, the corresponding consistent vector of decision variables (B.3) is a feasible solution to the DB-MISOCP. In other words, each assortment corresponds to a feasible solution to the DB-MISOCP, ensuring that no assortment is excluded from consideration when solving the DB-MISOCP.

2) The optimal solution to the DB-MISOCP is a consistent vector of variables for a certain set of binary values of z_{cj}. This means that the optimal solution corresponds to product demands yielded by the MAM for a certain combination of channel assortments.

The first point is straightforward to verify by substituting expressions (B.3) into the constraints of problem (2.11). Similar to the SBMILP formulation presented by Vasilyev et al. (2023), constants (2.12) are chosen to be large enough so that constraints (2.11j) and (2.11k) are always satisfied. However, it can be checked that in extreme cases when all products are offered in some channels while no products are offered in other channels, these constraints become tight.

Let us prove the second point. First, note that each profit function Π_{cj} is decreasing in t_{cj}. Since the decision variables t_{cj} are only present in constraints (2.11b), it means that these constraints have to be tight. Therefore, equalities (B.3e) hold true. Next, suppose that $z_{cj} = 1$ for a certain $c \in \mathcal{C}$, $j \in \mathcal{N}$. From constraints (2.11d), (2.11e), and (2.11f), it follows that $\dfrac{x_{cj}^{(c)}}{v_j^{(c)}} = \dfrac{x_{c0}^{(c)}}{v_0^{(c)}}$ and $x_{\bar{c}j}^{(c)} = 0$ for all $\bar{c} \in \mathcal{C}_{cj}$. At the same time, constraints (2.11j) imply that if $z_{cj} = 0$, then $x_{cj}^{(c)} = 0$. Hence, from (2.11c) it follows that equalities (B.3a) must be satisfied.

Constraints (2.11k) imply that variable $x_{\bar{c}j}^{(c)}$ is nonzero only if $x_{\bar{c}j}^{(\bar{c})} \neq 0$, meaning that $z_{\bar{c}j} \neq 0$. Furthermore, as previously demonstrated, variable $x_{cj}^{(c)}$ is nonzero only if $z_{cj} = 0$. Therefore, $x_{\bar{c}j}^{(c)}$ is nonzero only if $z_{\bar{c}j} = 1$ and $z_{cj} = 0$. Meanwhile, if $z_{\bar{c}j} = 1$ and $z_{cj} = 0$, then from constraints (2.11d) and (2.11g) it follows that $\dfrac{x_{\bar{c}j}^{(c)}}{\phi_j^{(c)}\delta_{\bar{c}j}^{(c)}} = \dfrac{x_{c0}^{(c)}}{v_0^{(c)}}$, which means that equalities (B.3c) must be satisfied.

Finally, since each profit function Π_{cj} decreases if either $y_{cj}^{(c)}$ or $y_{\bar{c}j}^{(c)}$ grows, the DB-MISOCP constraints (2.11h) and (2.11i) must be tight. Using our previous findings on the values of $x_{cj}^{(c)}$ and $x_{\bar{c}j}^{(c)}$ in the optimal solution to the DB-MISOCP, we deduce that equalities (B.3b) and (B.3d) hold true, which concludes the proof. $\qquad\square$

B.4 Proof of Lemma 2.1

Let us consider fixed indices $c, \bar{c}$, and j. In problem (R1), variable $x_{\bar{c}j}^{(c)}$ is only present in three constraints, namely:

$$
\frac{x_{cj}^{(c)}}{v_j^{(c)}} + \frac{x_{\bar{c}j}^{(c)}}{\phi_j^{(c)}\delta_{\bar{c}j}^{(c)}} \leq \frac{x_{c0}^{(c)}}{v_0^{(c)}},
$$
$$
\mu^{(c)} y_{\bar{c}j}^{(c)} \geq \sigma^{(c)} x_{\bar{c}j}^{(c)},
$$
$$
x_{\bar{c}j}^{(c)} \leq K_{\bar{c}j}^{(c)} x_{\bar{c}j}^{(\bar{c})}.
$$

The second constraint has to be binding, as it is the only constraint on variable $y_{\bar{c}j}^{(c)}$ and the objective function negatively depends on this variable. Suppose that the other two constraints are not binding. In this case, variable $x_{\bar{c}j}^{(c)}$ can be increased by some $\epsilon > 0$ while keeping all other variables except for $y_{\bar{c}j}^{(c)}$ unchanged. If $x_{\bar{c}j}^{(c)}$ is increased by ϵ, then, since the second constraint is binding, $y_{\bar{c}j}^{(c)}$ will be increased by $\dfrac{\sigma^{(c)}}{\mu^{(c)}}\epsilon$. To prove the lemma, it is sufficient to show that this change has a positive effect on the objective function.

This can be done by proving that the following function is increasing in ϵ at $\epsilon = 0$:

$$\Pi_{cj}\left(x_{cj}^{(c)} + \sum_{d \in \mathcal{C}_{cj}} x_{cj}^{(d)} + \epsilon, \sqrt{(y_{cj}^{(c)})^2 + \left(y_{cj}^{(\bar{c})} + \frac{\sigma^{(c)}}{\mu^{(c)}}\epsilon\right)^2 + \sum_{d \in \mathcal{C}_{cj} \setminus \{\bar{c}\}} (y_{cj}^{(d)})^2}\right).$$

The derivative of this function with respect to ϵ is as follows:

$$(r_{cj} - b_{cj}) - \frac{\frac{\sigma^{(c)}}{\mu^{(c)}}(r_{cj} + h_{cj})\varphi\left(\frac{r_{cj} - b_{cj}}{r_{cj} + h_{cj}}\right)\left(y_{cj}^{(\bar{c})} + \frac{\sigma^{(c)}}{\mu^{(c)}}\epsilon\right)}{\sqrt{(y_{cj}^{(c)})^2 + \left(y_{cj}^{(\bar{c})} + \frac{\sigma^{(c)}}{\mu^{(c)}}\epsilon\right)^2 + \sum_{d \in \mathcal{C}_{cj} \setminus \{\bar{c}\}} (y_{cj}^{(d)})^2}}. \tag{B.4}$$

Since all y-variables are nonnegative, it is straightforward to verify that expression (B.4) is positive if condition (2.13) is satisfied. $\qquad\square$

B.5 Proof of Theorem 2.3

First, note that problem (R1) is a relaxation of problem (R2) with the same objective function that does not depend on variables z_{cj}. Therefore, to show the equivalence of these two problems, it is sufficient to show that if $\left\{\hat{x}_{c0}^{(c)}, \hat{x}_{cj}^{(c)}, \hat{x}_{\bar{c}j}^{(c)}, \hat{y}_{cj}^{(c)}, \hat{y}_{\bar{c}j}^{(c)}, \hat{t}_{cj}\right\}_{c \in \mathcal{C}, j \in \mathcal{N}, \bar{c} \in \mathcal{C}_{cj}}$ is a solution to problem (R1), then there exist a set of variables z_{cj} such that $\left\{\hat{x}_{c0}^{(c)}, \hat{x}_{cj}^{(c)}, \hat{x}_{\bar{c}j}^{(c)}, \hat{y}_{cj}^{(c)}, \hat{y}_{\bar{c}j}^{(c)}, \hat{t}_{cj}, z_{cj}\right\}_{c \in \mathcal{C}, j \in \mathcal{N}, \bar{c} \in \mathcal{C}_{cj}}$ is a solution to problem (R2). In essence, we aim to prove that for all $c \in \mathcal{C}$, $j \in \mathcal{N}$, the following constraints on z_{cj} are not mutually contradictory:

$$\frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}} - \frac{\hat{x}_{cj}^{(c)}}{v_j^{(c)}} \leq \frac{\mu^{(c)}}{\tilde{v}_0^{(c)}}(1 - z_{cj}),$$

$$\hat{x}_{cj}^{(c)} \leq H_j^{(c)} z_{cj},$$

$$\frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}} - \frac{\hat{x}_{\bar{c}j}^{(c)}}{\phi_j^{(c)}\delta_{\bar{c}j}^{(c)}} \leq \frac{\mu^{(c)}}{\tilde{v}_0^{(c)}}(1 + z_{cj} - z_{\bar{c}j}) \qquad \forall \bar{c} \in \mathcal{C}_{cj},$$

$$0 \leq z_{cj} \leq 1,$$

or, equivalently:

$$\frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} z_{cj} \leq \frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} + \frac{\hat{x}_{cj}^{(c)}}{v_j^{(c)}} - \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}},$$

$$\frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} z_{cj} \geq \frac{\tilde{v}_0^{(c)} + \tilde{v}_j^{(c)}}{v_j^{(c)}\tilde{v}_0^{(c)}} \hat{x}_{cj}^{(c)},$$

$$\frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} z_{cj} \geq \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}} - \frac{\hat{x}_{\bar{c}j}^{(c)}}{\phi_j^{(c)} \delta_{\bar{c}j}^{(c)}} + \frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} (z_{\bar{c}j} - 1) \qquad \forall \bar{c} \in \mathcal{C}_{cj},$$

$$0 \leq z_{cj} \leq 1.$$

Note that the latter double inequality is redundant if other inequalities are satisfied. Indeed, from the first inequality, it follows that $z_{cj} \leq 1$ (since $\frac{\hat{x}_{cj}^{(c)}}{v_j^{(c)}} \leq \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}}$, which holds because constraint (2.11d) must be satisfied), and from the second inequality, it follows that $z_{cj} \geq 0$. Thus, to prove that the above inequalities are not mutually contradictory, it is sufficient to show that the upper bound on $\frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} z_{cj}$ is greater than or equal to its lower bounds, i.e.:

$$\frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} + \frac{\hat{x}_{cj}^{(c)}}{v_j^{(c)}} - \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}} \geq \frac{\tilde{v}_0^{(c)} + \tilde{v}_j^{(c)}}{v_j^{(c)} \tilde{v}_0^{(c)}} \hat{x}_{cj}^{(c)}, \tag{B.5}$$

$$\frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} + \frac{\hat{x}_{cj}^{(c)}}{v_j^{(c)}} - \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}} \geq \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}} - \frac{\hat{x}_{\bar{c}j}^{(c)}}{\phi_j^{(c)} \delta_{\bar{c}j}^{(c)}} + \frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} (z_{\bar{c}j} - 1) \qquad \forall \bar{c} \in \mathcal{C}_{cj}. \tag{B.6}$$

Inequality (B.5) can be rewritten in the following way:

$$\frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} \geq \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}} + \frac{\tilde{v}_j^{(c)}}{v_j^{(c)} \tilde{v}_0^{(c)}} \hat{x}_{cj}^{(c)}, \tag{B.7}$$

which holds because constraint (2.11c) of the DB-MISOCP has to be satisfied.

Now, let us prove inequality (B.6). From Lemma 2.1 it follows that for any given $\bar{c} \in \mathcal{C}_{cj}$, at least one of the following two constraints has to be binding:

$$\hat{x}_{\bar{c}j}^{(c)} \leq K_{\bar{c}j}^{(c)} \hat{x}_{\bar{c}j}^{(\bar{c})}, \tag{B.8}$$

$$\frac{\hat{x}_{cj}^{(c)}}{v_j^{(c)}} + \frac{\hat{x}_{\bar{c}j}^{(c)}}{\phi_j^{(c)} \delta_{\bar{c}j}^{(c)}} \leq \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}}. \tag{B.9}$$

First, suppose that inequality (B.8) is binding, i.e.:

$$\hat{x}_{\bar{c}j}^{(c)} = K_{\bar{c}j}^{(c)} \hat{x}_{\bar{c}j}^{(\bar{c})},$$

or, equivalently:

$$\hat{x}_{\bar{c}j}^{(c)} = \frac{\phi_j^{(c)} \delta_{\bar{c}j}^{(c)} \mu^{(c)}}{\tilde{v}_0^{(c)}} \Big/ \frac{v_j^{(\bar{c})} \mu^{(\bar{c})}}{v_0^{(\bar{c})} + \sum_{k \in \mathcal{N}} v_k^{(\bar{c})}} \hat{x}_{\bar{c}j}^{(\bar{c})}.$$

By substituting this expression into inequality (B.6), we obtain:

$$\frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} + \frac{\hat{x}_{cj}^{(c)}}{v_j^{(c)}} - \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}} \geq \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}} - \frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} \frac{v_0^{(\bar{c})} + \sum_{k \in \mathcal{N}} v_k^{(\bar{c})}}{v_j^{(\bar{c})} \mu^{(\bar{c})}} \hat{x}_{\bar{c}j}^{(\bar{c})} + \frac{\mu^{(c)}}{\tilde{v}_0^{(c)}}(z_{\bar{c}j} - 1).$$

This inequality can be rewritten as:

$$\frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} \left(1 + \frac{v_0^{(\bar{c})} + \sum_{k \in \mathcal{N}} v_k^{(\bar{c})}}{v_j^{(\bar{c})} \mu^{(\bar{c})}} \hat{x}_{\bar{c}j}^{(\bar{c})} \right) + \frac{\hat{x}_{cj}^{(c)}}{v_j^{(c)}} \geq 2 \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}} + \frac{\mu^{(c)}}{\tilde{v}_0^{(c)}}(z_{\bar{c}j} - 1).$$

Since $\dfrac{\mu^{(c)}}{\tilde{v}_0^{(c)}} \geq \dfrac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}}$, it is sufficient to show that:

$$\frac{v_0^{(\bar{c})} + \sum_{k \in \mathcal{N}} v_k^{(\bar{c})}}{v_j^{(\bar{c})} \mu^{(\bar{c})}} \hat{x}_{\bar{c}j}^{(\bar{c})} \geq z_{\bar{c}j}. \tag{B.10}$$

We can safely assume that $z_{\bar{c}j}$ satisfies the following constraint of the DB-MISOCP:

$$\frac{\hat{x}_{\bar{c}0}^{(\bar{c})}}{v_0^{(\bar{c})}} - \frac{\hat{x}_{\bar{c}j}^{(\bar{c})}}{v_j^{(\bar{c})}} \leq \frac{\mu^{(\bar{c})}}{\tilde{v}_0^{(\bar{c})}}(1 - z_{\bar{c}j}).$$

Therefore, since $\dfrac{\mu^{(\bar{c})}}{\tilde{v}_0^{(\bar{c})}} \geq \dfrac{\hat{x}_{\bar{c}0}^{(\bar{c})}}{v_0^{(\bar{c})}}$, the following inequality holds:

$$\frac{\tilde{v}_0^{(\bar{c})} \hat{x}_{\bar{c}j}^{(\bar{c})}}{v_j^{(\bar{c})} \mu^{(\bar{c})}} \geq z_{\bar{c}j}. \tag{B.11}$$

Since inequality (B.11) is stronger than inequality (B.10), we obtain that inequality (B.10) holds as well.

Next, suppose that inequality (B.9) is binding, i.e.:

$$\frac{\hat{x}_{cj}^{(c)}}{v_j^{(c)}} + \frac{\hat{x}_{\bar{c}j}^{(c)}}{\phi_j^{(c)} \delta_{\bar{c}j}^{(c)}} = \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}}. \tag{B.12}$$

Recall that $z_{\bar{c}j} \leq 1$ for all $\bar{c} \in \mathcal{C}_{cj}$. Therefore, to prove inequality (B.6), it is sufficient to

show that:

$$\frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} + \frac{\hat{x}_{cj}^{(c)}}{v_j^{(c)}} - \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}} \geq \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}} - \frac{\hat{x}_{\bar{c}j}^{(c)}}{\phi_j^{(c)}\delta_{\bar{c}j}^{(c)}} \qquad \forall \bar{c} \in \mathcal{C}_{cj}. \tag{B.13}$$

By substituting expression (B.12) into inequality (B.13), we obtain:

$$\frac{\mu^{(c)}}{\tilde{v}_0^{(c)}} \geq \frac{\hat{x}_{c0}^{(c)}}{v_0^{(c)}},$$

which holds as a special case of inequality (B.7). Therefore, inequality (B.6) holds as well, which concludes the proof. $\qquad\square$

B.6 Proof of Proposition 2.2

If customers do not switch between channels, then the omnichannel assortment optimization problem can be separated into assortment optimization problems for individual channels. In this case, for each channel c, type-c customers make choices according to the GAM with product set $\mathcal{N}$, product attractiveness values $v_j^{(c)}$, shadow attractiveness values $\phi_j^{(c)}$, and the attractiveness value of the no-purchase option $v_0^{(c)}$.

Next, it is easy to verify that the DB-MISOCP constraints (2.11b) and (2.11i) are binding. Therefore, at the optimum $y_{cj}^{(c)} = t_{cj}^{(c)} = \frac{\sigma^{(c)}}{\mu^{(c)}}x_{cj}^{(c)}$ for each $c \in \mathcal{C}$, $j \in \mathcal{N}$. It means that the objective function of the assortment optimization problem for a given channel c can be rewritten in the following way:

$$\sum_{j \in \mathcal{N}}(r_{cj} - b_{cj})x_{cj}^{(c)} - (r_{cj} + h_{cj})\varphi\Big(\frac{r_{cj} - b_{cj}}{r_{cj} + h_{cj}}\Big)\frac{\sigma^{(c)}}{\mu^{(c)}}x_{cj}^{(c)} =$$

$$\sum_{j \in \mathcal{N}}\Big((r_{cj} - b_{cj}) - (r_{cj} + h_{cj})\varphi\Big(\frac{r_{cj} - b_{cj}}{r_{cj} + h_{cj}}\Big)\frac{\sigma^{(c)}}{\mu^{(c)}}\Big)x_{cj}^{(c)}.$$

Since variables $x_{cj}^{(\bar{c})}$ are not present in the DB-MISOCP formulation with $\phi_j^{(c)}\delta_{\bar{c}j}^{(c)} = 0$ $\forall c \in \mathcal{C}$, $\bar{c} \in \mathcal{C}\backslash\{c\}$, $j \in \mathcal{N}$, and variables t_{cj}, $y_{cj}^{(c)}$, and $y_{cj}^{(\bar{c})}$ have been shown to be redundant, relaxation (R1) of the DB-MISOCP is reduced to a series of the following

optimization problems (with one problem per channel c):

$$\max_x \sum_{j \in \mathcal{N}} \left((r_{cj} - b_{cj}) - (r_{cj} + h_{cj})\varphi\left(\frac{r_{cj} - b_{cj}}{r_{cj} + h_{cj}}\right)\frac{\sigma^{(c)}}{\mu^{(c)}} \right) x_{cj}^{(c)}$$

$$\text{s.t.} \quad \frac{\tilde{v}_0^{(c)}}{v_0^{(c)}} x_{c0}^{(c)} + \sum_{j \in \mathcal{N}} \frac{\tilde{v}_j^{(c)}}{v_j^{(c)}} x_{cj}^{(c)} = \mu^{(c)},$$

$$\frac{x_{cj}^{(c)}}{v_j^{(c)}} \leq \frac{x_{c0}^{(c)}}{v_0^{(c)}} \qquad \forall j \in \mathcal{N},$$

$$x_{c0}^{(c)}, x_{cj}^{(c)} \in \mathbb{R}_{\geq 0} \qquad \forall j \in \mathcal{N}.$$

The above linear program is essentially the sales-based linear program (SBLP) formulated for an instance of the GAM, where the price of product j in channel c is given by $(r_{cj} - b_{cj}) - (r_{cj} + h_{cj})\varphi\left(\frac{r_{cj} - b_{cj}}{r_{cj} + h_{cj}}\right)\frac{\sigma^{(c)}}{\mu^{(c)}}$. Gallego et al. (2014) proved that the SBLP is a valid formulation of the assortment optimization problem under the GAM. Therefore, the optimal solution to problem (R1) – which is equivalent to problem (R2) as per Theorem 2.3 – yields the optimal assortments in all channels in the considered setting. $\square$

C Appendices for Chapter 3

C.1 Proof of Proposition 3.1

Let x be a binary vector representing a basket of products from S. Furthermore, let $\tilde{x}$ be the corresponding spin vector, i.e. the vector such that its entries satisfy

$$\tilde{x}_i = \begin{cases} 1 & \text{if } x_i = 1, \\ -1 & \text{otherwise.} \end{cases}$$

Note that $x = (\tilde{x} + 1)/2$. Since θ and $\tilde{\theta}$ satisfy conditions (3.4), we obtain that

$$\sum_{i \in S} \theta_{ii} x_i + \sum_{i,j \in S: i \neq j} x_i \theta_{ij} x_j = \sum_{i \in S} \Big(2\tilde{\theta}_{ii} - 4 \sum_{j \in S: j \neq i} \tilde{\theta}_{ij} \Big) \frac{\tilde{x}_i + 1}{2} + \sum_{i,j \in S: i \neq j} \frac{\tilde{x}_i + 1}{2} 4\tilde{\theta}_{ij} \frac{\tilde{x}_j + 1}{2},$$

$$= \sum_{i \in S} \Big(\tilde{\theta}_{ii} - 2 \sum_{j \in S: j \neq i} \tilde{\theta}_{ij} \Big) (\tilde{x}_i + 1) + \sum_{i,j \in S: i \neq j} \tilde{\theta}_{ij} (\tilde{x}_i \tilde{x}_j + \tilde{x}_i + \tilde{x}_j + 1)$$

$$= \sum_{i \in S} \tilde{\theta}_{ii} \tilde{x}_i + \sum_{i,j \in S: i \neq j} \tilde{x}_i \tilde{\theta}_{ij} \tilde{x}_j + \sum_{i \in S} \Big(\tilde{\theta}_{ii} - \sum_{j \in S: j \neq i} \tilde{\theta}_{ij} \Big).$$

Therefore,

$$\exp\Big(\sum_{i \in S} \theta_{ii} x_i + \sum_{i,j \in S: i \neq j} x_i \theta_{ij} x_j \Big) = \exp\Big(\sum_{i \in S} \tilde{\theta}_{ii} \tilde{x}_i + \sum_{i,j \in S: i \neq j} \tilde{x}_i \tilde{\theta}_{ij} \tilde{x}_j \Big) \cdot const,$$

from which it follows that $p_\theta(x|S) = p_{\tilde{\theta}}(\tilde{x}|S)$, which concludes the proof. $\qquad \square$

C.2 Performance of the Ising model as a Multi-Purchase Choice Model

To evaluate the performance of the Ising model as a multi-purchase choice model on the BAKERY dataset (see Benson et al., 2018), we require a benchmark. Following Benson et al. (2018), we use the separable model – in which each product has a certain utility and a basket's utility is equal to the sum of the utilities of its elements assuming a fixed basket size – as a benchmark. The key feature of the separable model is that demands for different products are independent of each other. In this model, the probability of choosing basket $\{i_1, i_2, \ldots, i_k\}$ from the given choice set $\mathcal{N}$ is given by

$$p_{sep}(\{i_1, i_2, \ldots, i_k\}) = p_{size}(k) \frac{\prod_{i=i_1}^{i_k} p_i}{\sum_{\{j_1, j_2, \ldots, j_k\} \subseteq \mathcal{N}} \prod_{j=j_1}^{j_k} p_j},$$

where $p_{size}(k)$ denotes the empirical probability of a customer selecting a basket of size k, and p_i is the empirical probability of a customer selecting item $i \in \mathcal{N}$. These auxiliary probabilities can be empirically estimated in the following way. Let B be the total number of baskets in the training sample, and B_k be the number of baskets of size k in the training sample. Furthermore, let H be the total number of units of all products in the training sample, and H_i be the number of units of product i in the training sample. Then, the empirical values of $p_{size}(k)$ and p_i are as follows:

$$p_{size}(k) = \frac{B_k}{B}, \quad p_i = \frac{H_i}{H}.$$

Let LL_{ising} and LL_{base} be the mean log-likelihood yielded by the Ising model and the benchmark model on the test sample, respectively. Then, the log-likelihood values are $LL_{ising} = -10.802$ and $LL_{base} = -11.538$, and the relative improvement is $e^{(LL_{ising} - LL_{base})} = 2.09$. Furthermore, if we set the regularization weight ρ to zero, then $LL_{ising} = -10.567$ and the relative improvement becomes 2.64. We see that the Ising model outperforms the subset selection choice model presented by Benson et al. (2018) when the set of corrections is determined using the (normalized) lift heuristic and the determinantal point process (DPP) model, and underperforms compared to their model when the set of corrections is determined using the frequency heuristic. The latter occurs due to the fact that, unlike Benson et al. (2018), we do not specifically account for the distribution of basket sizes and focus on pairwise product interactions. On the other hand, assortment optimization under the general version of the subset selection choice model presented by Benson et al. (2018) is extremely challenging and, to the best of our knowledge, this problem has not been addressed in the literature. We also note that the

Ising model is a basic example of an MRF, and considering more complex MRFs is a promising research direction that could lead to significant gains in performance.

C.3   Proof of Theorem 3.1

The decision version of the assortment optimization problem under the Ising model is as follows:

$$\textit{For any given } K, \textit{ is there assortment } S \textit{ such that } R(S) \geq K? \tag{D-AO}$$

Recall that $R(S)$ is defined by the following expression:

$$R(S) = \frac{\sum\limits_{x \in \mathcal{X}(S)} \exp\Big(\sum\limits_{i \in S} \theta_{ii} x_i + \sum\limits_{i \neq j, \; i,j \in S} x_i \theta_{ij} x_j\Big) \sum\limits_{j \in S} r_j x_j}{\sum\limits_{x \in \mathcal{X}(S)} \exp\Big(\sum\limits_{i \in S} \theta_{ii} x_i + \sum\limits_{i \neq j, \; i,j \in S} x_i \theta_{ij} x_j\Big)}.$$

Therefore, the inequality in the decision problem (D-AO) can be rewritten in the following way:

$$\sum\limits_{\substack{x \in \mathcal{X}(S), \\ x \neq 0}} \exp\Big(\sum\limits_{i \in S} \theta_{ii} x_i + \sum\limits_{i \neq j, \; i,j \in S} x_i \theta_{ij} x_j\Big) \Big(\sum\limits_{j \in S} r_j x_j - K\Big) \geq K. \tag{C.1}$$

Let $r_1 = K + 1$, $\theta_{11} = 0$, and $r_j = 0 \; \forall j \in \mathcal{N} \backslash \{1\}$. Since product 1 is the only product with a nonzero profit margin, it has to belong to S in order for inequality (C.1) to be satisfied assuming that $K > 0$. Then, inequality (C.1) takes the following form:

$$\sum\limits_{\substack{x \in \mathcal{X}(S), \\ x_1 = 1}} \exp\Big(\sum\limits_{i \in S \backslash \{1\}} (\theta_{ii} + 2\theta_{1i}) x_i + \sum\limits_{\substack{i \neq j, \\ i,j \in S \backslash \{1\}}} x_i \theta_{ij} x_j\Big) \geq K.$$

Lastly, let $\theta'_{ii} = \theta_{ii} + 2\theta_{1i} \; \forall i \in \mathcal{N} \backslash \{1\}$ and $\theta'_{ij} = \theta_{ij} \; \forall i, j \in \mathcal{N} \backslash \{1\}$. We can see that solving problem (D-AO) requires answering the question of whether the partition function of the Ising model with parameters θ' defined over nodes $\mathcal{N} \backslash \{1\}$ is greater than or equal to a given constant. Such a problem is NP-hard as shown by Istrail (2000) for nonplanar graphs, meaning that problem (D-AO) is NP-hard as well. □

C.4   Proof of Proposition 3.2

Let product k be such that $\theta_{kj} = 0 \; \forall j \in \mathcal{N} \backslash \{k\}$. Then, our goal is to show that $R(S \cup \{k\}) \geq R(S)$ for any assortment $S \subseteq \mathcal{N}$ such that $k \notin S$. This can be verified

directly:

$$R(S \cup \{k\}) =$$

$$= \frac{\sum\limits_{x \in \mathcal{X}(S \cup \{k\})} \exp\Big(\sum\limits_{i \in S \cup \{k\}} \theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S \cup \{k\}} x_i\theta_{ij}x_j\Big) \sum\limits_{j \in S \cup \{k\}} r_jx_j}{\sum\limits_{x \in \mathcal{X}(S \cup \{k\})} \exp\Big(\sum\limits_{i \in S \cup \{k\}} \theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S \cup \{k\}} x_i\theta_{ij}x_j\Big)}$$

$$\geq \frac{\sum\limits_{x \in \mathcal{X}(S \cup \{k\})} \exp\Big(\sum\limits_{i \in S \cup \{k\}} \theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S \cup \{k\}} x_i\theta_{ij}x_j\Big) \sum\limits_{j \in S} r_jx_j}{\sum\limits_{x \in \mathcal{X}(S \cup \{k\})} \exp\Big(\sum\limits_{i \in S \cup \{k\}} \theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S \cup \{k\}} x_i\theta_{ij}x_j\Big)}$$

$$= \frac{\sum\limits_{x \in \mathcal{X}(S)} \big(1 + \exp(\theta_{kk} + 2\sum\limits_{i \in S} x_i\theta_{ki})\big) \exp\Big(\sum_{i \in S} \theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i\theta_{ij}x_j\Big) \sum\limits_{j \in S} r_jx_j}{\sum\limits_{x \in \mathcal{X}(S)} \big(1 + \exp(\theta_{kk} + 2\sum\limits_{i \in S} x_i\theta_{ki})\big) \exp\Big(\sum_{i \in S} \theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i\theta_{ij}x_j\Big)}$$

$$= \frac{\sum\limits_{x \in \mathcal{X}(S)} \big(1 + \exp(\theta_{kk})\big) \exp\Big(\sum_{i \in S} \theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i\theta_{ij}x_j\Big) \sum\limits_{j \in S} r_jx_j}{\sum\limits_{x \in \mathcal{X}(S)} \big(1 + \exp(\theta_{kk})\big) \exp\Big(\sum_{i \in S} \theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i\theta_{ij}x_j\Big)}$$

$$= \frac{\sum\limits_{x \in \mathcal{X}(S)} \exp\Big(\sum_{i \in S} \theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i\theta_{ij}x_j\Big) \sum\limits_{j \in S} r_jx_j}{\sum\limits_{x \in \mathcal{X}(S)} \exp\Big(\sum_{i \in S} \theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i\theta_{ij}x_j\Big)} = R(S). \quad \square$$

C.5 Proof of Proposition 3.3

Let $S_{\mathcal{H}} = S \cap \mathcal{H}$ and $S_{\mathcal{K}} = S \cap \mathcal{K}$, and let us fix product $l \in S_{\mathcal{H}}$. To prove the proposition, it is sufficient to show that the marginal probability of a customer choosing product $l \in S_{\mathcal{H}}$ does not depend on $S_{\mathcal{K}}$. This is a direct implication of the fact that the Ising model satisfies global Markov properties. This can also be verified directly:

$$p_\theta(x_l = 1 | S) = \frac{\sum\limits_{x \in \mathcal{X}(S): \ x_l = 1} \exp\Big(\sum\limits_{i \in S} \theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i\theta_{ij}x_j\Big)}{\sum\limits_{x \in \mathcal{X}(S)} \exp\Big(\sum\limits_{i \in S} \theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i\theta_{ij}x_j\Big)}$$

$$= \frac{\sum\limits_{\substack{x \in \mathcal{X}(S_{\mathcal{H}}): \\ x_l = 1}} \bigg(\exp\Big(\sum\limits_{i \in S_{\mathcal{H}}} \theta_{ii}x_i + \sum\limits_{\substack{i \neq j, \\ i,j \in S_{\mathcal{H}}}} x_i\theta_{ij}x_j\Big) \sum\limits_{x' \in \mathcal{X}(S_{\mathcal{K}})} \exp\Big(\sum\limits_{i \in S_{\mathcal{K}}} \theta_{ii}x_i' + \sum\limits_{\substack{i \neq j, \\ i,j \in S_{\mathcal{K}}}} x_i'\theta_{ij}x_j'\Big)\bigg)}{\sum\limits_{x \in \mathcal{X}(S_{\mathcal{H}})} \bigg(\exp\Big(\sum\limits_{i \in S_{\mathcal{H}}} \theta_{ii}x_i + \sum\limits_{\substack{i \neq j, \\ i,j \in S_{\mathcal{H}}}} x_i\theta_{ij}x_j\Big) \sum\limits_{x' \in \mathcal{X}(S_{\mathcal{K}})} \exp\Big(\sum\limits_{i \in S_{\mathcal{K}}} \theta_{ii}x_i' + \sum\limits_{\substack{i \neq j, \\ i,j \in S_{\mathcal{K}}}} x_i'\theta_{ij}x_j'\Big)\bigg)}$$

$$
= \frac{\sum\limits_{x \in \mathcal{X}(S_{\mathcal{H}}):\ x_l = 1} \exp\Big(\sum\limits_{i \in S_{\mathcal{H}}} \theta_{ii} x_i + \sum\limits_{i \neq j,\ i,j \in S_{\mathcal{H}}} x_i \theta_{ij} x_j\Big)}{\sum\limits_{x \in \mathcal{X}(S_{\mathcal{H}})} \exp\Big(\sum\limits_{i \in S_{\mathcal{H}}} \theta_{ii} x_i + \sum\limits_{i \neq j,\ i,j \in S_{\mathcal{H}}} x_i \theta_{ij} x_j\Big)} = p_\theta(x_l = 1 | S_{\mathcal{H}}). \quad \square
$$

C.6 Proof of Theorem 3.2

Let us first formulate the following auxiliary lemma:

Lemma C.1.

$$
p_\theta(x_l = 1 | S) \geq p_\theta(x_l = 1 | x_k = 0, S) \iff p_\theta(x_l = 1 | x_k = 1, S) \geq p_\theta(x_l = 1 | x_k = 0, S).
$$

Proof of Lemma C.1. Suppose that

$$
p_\theta(x_l = 1 | S) \geq p_\theta(x_l = 1 | x_k = 0, S). \tag{C.2}
$$

By the law of total probability, we obtain that:

$$
p_\theta(x_l = 1 | x_k = 0, S) p_\theta(x_k = 0 | S) + p_\theta(x_l = 1 | x_k = 1, S) p_\theta(x_k = 1 | S) \geq
$$
$$
p_\theta(x_l = 1 | x_k = 0, S),
$$

or

$$
p_\theta(x_l = 1 | x_k = 1, S) p_\theta(x_k = 1 | S) \geq p_\theta(x_l = 1 | x_k = 0, S)(1 - p_\theta(x_k = 0 | S)).
$$

Since $p_\theta(x_k = 1 | S) = 1 - p_\theta(x_k = 0 | S)$ and $p_\theta(x_k = 1 | S) > 0$, we can divide both sides of the inequality by $p_\theta(x_k = 1 | S)$ while preserving the inequality sign, thus obtaining the desired inequality:

$$
p_\theta(x_l = j | x_k = 1, S) \geq p_\theta(x_l = 1 | x_k = 0, S). \tag{C.3}
$$

Finally, one can prove that inequality (C.2) follows from inequality (C.3) by repeating the above steps from the bottom up. $\square$

Proof of Theorem 3.2. From Proposition 3.3 it follows that we can consider products in an isolated subgraph as a separate product portfolio $\mathcal{N}$. Suppose that $\theta_{ij} \geq 0 \ \forall i, j \in \mathcal{N}$, $i \neq j$. To prove the theorem, it is sufficient to show that removing any product from the assortment can only reduce the marginal probabilities of customers choosing other

products, i.e., for any $k, l \in \mathcal{N}$:

$$p_\theta(x_l = 1|S) \geq p_\theta(x_l = 1|x_k = 0, S).$$

Without loss of generality, suppose that $S = \{1, \ldots, m\}$, $l = 1$, and $k = 2$. As shown in Lemma C.1, the above inequality is equivalent to the following one:

$$p_\theta(x_1 = 1|x_2 = 1, S) \geq p_\theta(x_1 = 1|x_2 = 0, S). \tag{C.4}$$

Let $f(a, b) = \displaystyle\sum_{x \in \mathcal{X}(S): \ x_1 = a, x_2 = b} \exp\Big(\sum_{i \in S} \theta_{ii} x_i + \sum_{i \neq j, \ i,j \in S} x_i \theta_{ij} x_j\Big)$. Then, inequality (C.4) can be rewritten as

$$\frac{f(1, 1)}{f(0, 1) + f(1, 1)} \geq \frac{f(1, 0)}{f(0, 0) + f(1, 0)},$$

or

$$f(0, 0) f(1, 1) \geq f(1, 0) f(0, 1). \tag{C.5}$$

Let $g(x|S) = \exp\Big(\sum_{i \in S} \theta_{ii} x_i + \sum_{i \neq j, \ i,j \in S} x_i \theta_{ij} x_j\Big)$, and let $S_m^- = \{3, \ldots, m\}$. Then, inequality (C.5) takes the following form:

$$\sum_{x \in \mathcal{X}(S_m^-)} g\big[(0, 0, x)|S\big] \sum_{x \in \mathcal{X}(S_m^-)} g\big[(1, 1, x)|S\big] \geq \sum_{x \in \mathcal{X}(S_m^-)} g\big[(1, 0, x)|S\big] \sum_{x \in \mathcal{X}(S_m^-)} g\big[(0, 1, x)|S\big].$$
$$\tag{C.6}$$

We prove inequality (C.6) by induction over m.

Induction base. If $m = 2$, then inequality (C.6) can be written as:

$$1 \cdot \exp(\theta_{11} + \theta_{22} + 2\theta_{12}) \geq \exp(\theta_{11}) \exp(\theta_{22}),$$

which holds true since $\theta_{12} \geq 0$. We also check that inequality (C.6) holds if $m = 3$. In this case, this inequality becomes:

$$\big(1 + \exp(\theta_{33})\big)\big(\exp(\theta_{11} + \theta_{22} + 2\theta_{12}) + \exp(\theta_{11} + \theta_{22} + \theta_{33} + 2\theta_{12} + 2\theta_{13} + 2\theta_{23})\big) \geq$$
$$\big(\exp(\theta_{11}) + \exp(\theta_{11} + \theta_{33} + 2\theta_{13})\big)\big(\exp(\theta_{22}) + \exp(\theta_{22} + \theta_{33} + 2\theta_{23})\big).$$

It is easy to see that the above inequality holds if:

$$\exp(2\theta_{12}) + \exp(2\theta_{12} + 2\theta_{13} + 2\theta_{23}) \geq \exp(2\theta_{23}) + \exp(2\theta_{13}). \tag{C.7}$$

Note that it is sufficient to show that the latter inequality holds for $\theta_{12} = 0$, in which case this inequality is equivalent to the following one:

$$(\exp(2\theta_{13}) - 1)(\exp(2\theta_{23}) - 1) \geq 0.$$

The above inequality is true for any $\theta_{13}, \theta_{23} > 0$, meaning that inequality (C.6) holds for $m = 3$.

Induction step. Assuming that inequality (C.6) holds for dimension $m - 1$, our goal is to show that it also holds for dimension m. This inequality can be rewritten in the following way:

$$\sum_{x,x'\in\mathcal{X}(S_m^-)} g\big[(0,0,x)|S\big] \cdot g\big[(1,1,x')|S\big] \geq \sum_{x,x'\in\mathcal{X}(S_m^-)} g\big[(1,0,x)|S\big] \cdot g\big[(1,0,x)|S\big],$$

which is equivalent to:

$$\sum_{x,x'\in\mathcal{X}(S_m^-)} g(x|S_m^-)g(x'|S_m^-)\exp(\theta_{11} + \theta_{22} + 2\theta_{12} + 2\sum_{i\in S_m^-}(\theta_{1i} + \theta_{2i})x_i') \geq$$
$$\sum_{x,x'\in\mathcal{X}(S_m^-)} g(x|S_m^-)g(x'|S_m^-)\exp(\theta_{11} + \theta_{22} + 2\sum_{i\in S_m^-}\theta_{1i}x_i + 2\sum_{i\in S_m^-}\theta_{2i}x_i').$$

Since $\theta_{12} > 0$ and $\exp(\theta_{12})$ is a multiplier that is only present on the left-hand side of the above inequality, it is sufficient to show that this inequality holds for $\theta_{12} = 0$. Therefore, we want to prove the following inequality:

$$\sum_{x,x'\in\mathcal{X}(S_m^-)} g(x|S_m^-)g(x'|S_m^-)\exp(2\sum_{i\in S_m^-}\theta_{2i}x_i')\Big(\exp(2\sum_{i\in S_m^-}\theta_{1i}x_i') - \exp(2\sum_{i\in S_m^-}\theta_{1i}x_i)\Big) \geq 0.$$
$$\tag{C.8}$$

It is easy to see that if $\theta_{2i} = 0 \ \forall i \in S_m^-$, then inequality (C.8) is satisfied as it turns into an equality. Therefore, one can prove inequality (C.8) by showing that its left-hand side is increasing in θ_{2i} for any $i \in S_m^-$. Without loss of generality, let us fix $i = m$ and show that the left-hand side of inequality (C.8) is increasing in θ_{2m}. We do this by taking a derivative of this function with respect to θ_{2m} and checking that it is always nonnegative

under the given assumptions. In other words, we prove the following inequality:

$$\sum_{\substack{x,x' \in \mathcal{X}(S_m^-) \\ x_m'=1}} g(x|S_m^-)g(x'|S_m^-) \exp(2 \sum_{i \in S_m^-} \theta_{2i}x_i') \Big(\exp(2 \sum_{i \in S_m^-} \theta_{1i}x_i') - \exp(2 \sum_{i \in S_m^-} \theta_{1i}x_i) \Big) \geq 0.$$

Clearly, the above inequality holds if the following inequalities are true:

$$\sum_{\substack{x,x' \in \mathcal{X}(S_m^-) \\ x_m=0, \, x_m'=1}} g(x|S_m^-)g(x'|S_m^-) \exp(2 \sum_{i \in S_m^-} \theta_{2i}x_i') \Big(\exp(2 \sum_{i \in S_m^-} \theta_{1i}x_i') - \exp(2 \sum_{i \in S_m^-} \theta_{1i}x_i) \Big) \geq 0,$$

(C.9)

$$\sum_{\substack{x,x' \in \mathcal{X}(S_m^-) \\ x_m=1, \, x_m'=1}} g(x|S_m^-)g(x'|S_m^-) \exp(2 \sum_{i \in S_m^-} \theta_{2i}x_i') \Big(\exp(2 \sum_{i \in S_m^-} \theta_{1i}x_i') - \exp(2 \sum_{i \in S_m^-} \theta_{1i}x_i) \Big) \geq 0.$$

(C.10)

Let $S_{m-1}^- = \{3, \ldots, m-1\}$. Then, inequality (C.9) can be rewritten as:

$$\sum_{x,x' \in \mathcal{X}(S_{m-1}^-)} g(x|S_{m-1}^-)g(x', |S_{m-1}^-) \exp(2 \sum_{i \in S_{m-1}^-} \theta_{mi}x_i' + \theta_{mm}) \cdot$$
$$\exp(2 \sum_{i \in S_{m-1}^-} \theta_{2i}x_i' + \theta_{2m}) \Big(\exp(2 \sum_{i \in S_{m-1}^-} \theta_{1i}x_i' + \theta_{1m}) - \exp(2 \sum_{i \in S_{m-1}^-} \theta_{1i}x_i) \Big) \geq 0.$$

It is sufficient to show that the above inequality holds for $\theta_{1m} = 0$, in which case it takes the following form:

$$\sum_{x,x' \in \mathcal{X}(S_{m-1}^-)} g(x|S_{m-1}^-)g(x'|S_{m-1}^-) \exp(2 \sum_{i \in S_{m-1}^-} (\theta_{mi} + \theta_{2i})x_i') \cdot$$
$$\Big(\exp(2 \sum_{i \in S_{m-1}^-} \theta_{1i}x_i') - \exp(2 \sum_{i \in S_{m-1}^-} \theta_{1i}x_i) \Big) \geq 0.$$

(C.11)

Inequality (C.11) is equivalent to inequality (C.8) where assortment S_m^- is replaced with assortment S_{m-1}^- and parameters θ_{2i} are replaced with parameters $\theta_{2i} + \theta_{mi} \ \forall i \in S_{m-1}^-$. Therefore, inequality (C.11) holds by induction hypothesis.

Similarly, inequality (C.10) is equivalent to the following one:

$$
\sum_{x,x' \in \mathcal{X}(S_{m-1}^-)} g(x|S_{m-1}^-)g(x',|S_{m-1}^-)\exp(2\sum_{i \in S_{m-1}^-}\theta_{mi}(x_i + x_i'))\cdot
$$
$$
\exp(2\sum_{i \in S_{m-1}^-}\theta_{2i}x_i')\Big(\exp(2\sum_{i \in S_{m-1}^-}\theta_{1i}x_i') - \exp(2\sum_{i \in S_{m-1}^-}\theta_{1i}x_i)\Big) \geq 0. \tag{C.12}
$$

Inequality (C.12) holds true since it is equivalent to inequality (C.8) where assortment S_m^- is replaced with assortment S_{m-1}^- and parameters θ_{ii} are replaced with parameters $\theta_{ii} + \theta_{mi} \ \forall i \in S_{m-1}^-$. Thus, inequality (C.8) holds for dimension m, which concludes our proof by induction. $\qquad\square$

C.7   Proof of Proposition 3.4

First, from Proposition 3.3 it follows that we can consider products from the isolated subgraph as a separate product portfolio, i.e., we can assume that $\mathcal{H} = \mathcal{N}$. Then, our goal is to show that $R(S \cup \{k\}) \geq R(S)$ for any assortment $S \subseteq \mathcal{N}$ such that $k \notin S$. Note that

$$
R(S \cup \{k\}) = \frac{\sum\limits_{x \in \mathcal{X}(S\cup\{k\})} \exp\Big(\sum\limits_{i \in S\cup\{k\}}\theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S\cup\{k\}} x_i\theta_{ij}x_j\Big)\sum\limits_{j \in S\cup\{k\}} r_j x_j}{\sum\limits_{x \in \mathcal{X}(S\cup\{k\})} \exp\Big(\sum\limits_{i \in S\cup\{k\}}\theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S\cup\{k\}} x_i\theta_{ij}x_j\Big)}
$$
$$
\geq \frac{\sum\limits_{x \in \mathcal{X}(S\cup\{k\})} \exp\Big(\sum\limits_{i \in S\cup\{k\}}\theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S\cup\{k\}} x_i\theta_{ij}x_j\Big)\sum\limits_{j \in S} r_j x_j}{\sum\limits_{x \in \mathcal{X}(S\cup\{k\})} \exp\Big(\sum\limits_{i \in S\cup\{k\}}\theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S\cup\{k\}} x_i\theta_{ij}x_j\Big)}
$$
$$
= \frac{\sum\limits_{x \in \mathcal{X}(S)} \big(1 + \exp(\theta_{kk} + 2\sum\limits_{i \in S} x_i\theta_{ki})\big)\exp\Big(\sum_{i \in S}\theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i\theta_{ij}x_j\Big)\sum\limits_{j \in S} r_j x_j}{\sum\limits_{x \in \mathcal{X}(S)} \big(1 + \exp(\theta_{kk} + 2\sum\limits_{i \in S} x_i\theta_{ki})\big)\exp\Big(\sum_{i \in S}\theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i\theta_{ij}x_j\Big)}
$$

and

$$
R(S) = \frac{\sum\limits_{x \in \mathcal{X}(S)} \exp\Big(\sum_{i \in S}\theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i\theta_{ij}x_j\Big)\sum\limits_{j \in S} r_j x_j}{\sum\limits_{x \in \mathcal{X}(S)} \exp\Big(\sum_{i \in S}\theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i\theta_{ij}x_j\Big)}.
$$

Let us denote $\exp\Big(\sum_{i \in S}\theta_{ii}x_i + \sum\limits_{i \neq j, \ i,j \in S} x_i\theta_{ij}x_j\Big)$ by $a(x)$, $\sum\limits_{j \in S} r_j x_j$ by $b(x)$, and $\big(1 + \exp(\theta_{kk} + 2\sum\limits_{i \in S} x_i\theta_{ki})\big)$ by $c(x)$. For readability, let us slightly abuse the notation and

write just the summation over x instead of the summation over $x \in \mathcal{X}(S)$. Then, our goal is to prove that the following inequality holds for any assortment $S \subseteq \mathcal{N}$:

$$\frac{\sum_x a(x)b(x)c(x)}{\sum_x a(x)c(x)} \geq \frac{\sum_x a(x)b(x)}{\sum_x a(x)},$$

or

$$\sum_x a(x)b(x)c(x) \sum_x a(x) \geq \sum_x a(x)b(x) \sum_x a(x)c(x).$$

The above inequality can be rewritten as:

$$\sum_{x,x'} a(x)b(x)c(x)a(x') \geq \sum_{x,x'} a(x)b(x)a(x')c(x'),$$

which is equivalent to

$$\sum_{x,x'} a(x)b(x)a(x')(c(x) - c(x')) \geq 0.$$

Let us multiply the latter inequality by 2 and split the sum on the left-hand side into pairs in the following way:

$$\sum_{x,x'} \Big(a(x)b(x)a(x')(c(x) - c(x')) + a(x')b(x')a(x)(c(x') - c(x)) \Big) \geq 0,$$

which can be rewritten as

$$\sum_{x,x'} a(x)a(x')(b(x) - b(x'))(c(x) - c(x')) \geq 0. \tag{C.13}$$

Finally, suppose that $b(x) > b(x')$, i.e., $\sum_{j \in S} r_j x_j > \sum_{j \in S} r_j x'_j$. Then, $\exp(\theta_{kk} + 2 \sum_{j \in S} \alpha r_j x_j) > \exp(\theta_{kk} + 2 \sum_{j \in S} \alpha r_j x'_j)$, meaning that $c(x) > c(x')$. Similarly, if $b(x) < b(x')$, then $c(x) < c(x')$ as well. Therefore, inequality (C.13) holds true, which concludes the proof. $\qquad \square$

Bibliography

Abouelrous, A., Gabor, A. F., and Zhang, Y. (2022). "Optimizing the inventory and fulfillment of an omnichannel retailer: a stochastic approach with scenario clustering". *Computers & Industrial Engineering*, 173, p. 108723.

Alimonti, P. and Kann, V. (2000). "Some APX-completeness results for cubic graphs". *Theoretical Computer Science*, 237(1), pp. 123–134.

Aouad, A., Levi, R., and Segev, D. (2018). "Greedy-Like Algorithms for Dynamic Assortment Planning Under Multinomial Logit Preferences". *Operations Research*, 66(5), pp. 1321–1345.

Aouad, A., Levi, R., and Segev, D. (2019). "Approximation Algorithms for Dynamic Assortment Optimization Models". *Mathematics of Operations Research*, 44(2), pp. 487–511.

Aouad, A. and Segev, D. (2019). *The Stability of MNL-Based Demand under Dynamic Customer Substitution and its Algorithmic Implications*. Available at SSRN: https://ssrn.com/abstract=3498325.

Banerjee, O., Ghaoui, L. E., and d'Aspremont, A. (2008). "Model Selection Through Sparse Maximum Likelihood Estimation for Multivariate Gaussian or Binary Data". *Journal of Machine Learning Research*, 9, pp. 485–516.

Banker, S. (2021). *Walmart's Massive Investment In A Supply Chain Transformation*. Forbes. Accessed on 05/05/2023.

Barahona, F. (1982). "On the computational complexity of Ising spin glass models". *Journal of Physics A: Mathematical and General*, 15(10), p. 3241.

Barr, J. (2021). *Why optimizing your omnichannel experience is critical after Covid-19*. Forbes. Accessed on 05/05/2023.

Bayram, A. and Cesaret, B. (2021). "Order fulfillment policies for ship-from-store implementation in omni-channel retailing". *European Journal of Operational Research*, 294(3), pp. 987–1002.

Bibliography

Ben-Akiva, M. E. and Lerman, S. R. (1985). *Discrete choice analysis: theory and application to travel demand*. MIT Press Cambridge (Mass.)

Benson, A., Kumar, R., and Tomkins, A. (2018). "A Discrete Choice Model for Subset Selection". In: *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*. ACM Press, pp. 37–45.

Berbeglia, G., Garassino, A., and Vulcano, G. (2021). "A Comparative Empirical Study of Discrete Choice Models in Retail Operations". *Management Science*.

Berbeglia, G. and Joret, G. (2020). "Assortment Optimisation Under a General Discrete Choice Model: A Tight Analysis of Revenue-Ordered Assortments". *Algorithmica*, 82(4), pp. 681–720.

Besag, J. (1974). "Spatial Interaction and the Statistical Analysis of Lattice Systems". *Journal of the Royal Statistical Society. Series B (Methodological)*, 36(2), pp. 192–236.

Bhat, C. R. (2005). "A multiple discrete–continuous extreme value model: Formulation and application to discretionary time-use decisions". *Transportation Research Part B: Methodological*, 39(8), pp. 679–707.

Bhat, C. R. (2008). "The multiple discrete-continuous extreme value (MDCEV) model: Role of utility function parameters, identification considerations, and model extensions". *Transportation Research Part B: Methodological*, 42(3), pp. 274–303.

Bhatnagar, A. and Syam, S. S. (2014). "Allocating a hybrid retailer's assortment across retail stores: Bricks-and-mortar vs online". *Journal of Business Research*, 67(6), pp. 1293–1302.

Bierlaire, M. (2015). *Optimization: Principles and Algorithms*. Lausanne: EPFL Press.

Blanchet, J., Gallego, G., and Goyal, V. (2016). "A Markov Chain Approximation to Choice Modeling". *Operations Research*, 64(4), pp. 886–905.

Block, H. and Marschak, J. (1959). *Random Orderings and Stochastic Theories of Response*. Cowles Foundation Discussion Papers 66. Cowles Foundation for Research in Economics, Yale.

Boztuğ, Y. and Reutterer, T. (2008). "A combined approach for segment-specific market basket analysis". *European Journal of Operational Research*, 187(1), pp. 294–312.

Brightpearl (2017). *The State of Omnichannel Retail – Survey of Leading Retailers*. White paper, Accessed on 05/05/2023.

Cachon, G. P., Terwiesch, C., and Xu, Y. (2005). "Retail Assortment Planning in the Presence of Consumer Search". *Manufacturing & Service Operations Management*, 7(4), pp. 330–346.

Cao, J., So, K. C., and Yin, S. (2016). "Impact of an "online-to-store" channel on demand allocation, pricing and profitability". *European Journal of Operational Research*, 248(1), pp. 234–245.

Chen, J., Liang, Y., Shen, H., Shen, Z.-J. M., and Xue, M. (2022). "Offline-Channel Planning in Smart Omnichannel Retailing". *Manufacturing & Service Operations Management*, 24(5), pp. 2444–2462.

Chen, X., Li, J., Li, M., Zhao, T., and Zhou, Y. (2022). *Assortment Optimization Under the Multivariate MNL Model*. Available at SSRN: https://ssrn.com/abstract=4233712.

Cox, D. R. (1972). "The Analysis of Multivariate Binary Data". *Journal of the Royal Statistical Society: Series C*, 21(2), pp. 113–120.

Debter, L. (2017). *Amazon Is Buying Whole Foods For $13.7 Billion*. Forbes. Accessed on 05/05/2023.

Désir, A., Goyal, V., and Zhang, J. (2022). "Technical Note—Capacitated Assortment Optimization: Hardness and Approximation". *Operations Research*, 70(2), pp. 893–904.

Dobruschin, P. (1968). "The Description of a Random Field by Means of Conditional Probabilities and Conditions of Its Regularity". *Theory of Probability and Its Applications*, 13(2), pp. 197–224.

Domencich, T. and McFadden, D. L. (1975). *Urban Travel Demand: A Behavioral Analysis*. North-Holland Pub. Co.

Dubé, J.-P. (2004). "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks". *Marketing Science*, 23(1), pp. 66–81.

Dzyabura, D. and Jagabathula, S. (2018). "Offline Assortment Optimization in the Presence of an Online Channel". *Management Science*, 64(6), pp. 2767–2786.

Feldman, J., Paul, A., and Topaloglu, H. (2019). "Technical Note—Assortment Optimization with Small Consideration Sets". *Operations Research*, 67(5), pp. 1283–1299.

Fisher, M. and Vaidyanathan, R. (2014). "A Demand Estimation Procedure for Retail Assortment Optimization with Results from Implementations". *Management Science*, 60(10), pp. 2401–2415.

Gabor, A. F., van Ommeren, J.-K., and Sleptchenko, A. (2022). "An inventory model with discounts for omnichannel retailers of slow moving items". *European Journal of Operational Research*, 300(1), pp. 58–72.

Gallego, G., Ratliff, R., and Shebalov, S. (2014). "A General Attraction Model and Sales-Based Linear Program for Network Revenue Management Under Customer Choice". *Operations Research*, 63(1), pp. 212–232.

Bibliography

Gallino, S. and Moreno, A. (2014). "Integration of Online and Offline Channels in Retail: The Impact of Sharing Reliable Inventory Availability Information". *Management Science*, 60(6), pp. 1434–1451.

Gao, F. and Su, X. (2017). "Online and Offline Information for Omnichannel Retailing". *Manufacturing & Service Operations Management*, 19(1), pp. 84–98.

Gao, F. and Su, X. (2016). "Omnichannel Retail Operations with Buy-Online-and-Pick-up-in-Store". *Management Science*, 63(8), pp. 2478–2492.

Gaur, V. and Honhon, D. (2006). "Assortment Planning and Inventory Decisions Under a Locational Choice Model". *Management Science*, 52(10), pp. 1528–1543.

Geunes, J. and Su, Y. (2020). "Single-period assortment and stock-level decisions for dual sales channels with capacity limits and uncertain demand". *International Journal of Production Research*, 58(18), pp. 5579–5600.

Globerson, A. and Jaakkola, T. (2006). "Approximate inference using planar graph decomposition". In: *Advances in Neural Information Processing Systems*. Ed. by B. Schölkopf, J. Platt, and T. Hoffman. Vol. 19.

Govindarajan, A., Sinha, A., and Uichanco, J. (2021). "Joint inventory and fulfillment decisions for omnichannel retail networks". *Naval Research Logistics (NRL)*, 68(6), pp. 779–794.

Goyal, V., Levi, R., and Segev, D. (2016). "Near-Optimal Algorithms for the Assortment Planning Problem Under Dynamic Substitution and Stochastic Demand". *Operations Research*, 64(1), pp. 219–235.

Guo, J. and Keskin, B. B. (2022). "Designing a centralized distribution system for omni-channel retailing". *Production and Operations Management*. (In Press).

Gurobi Optimization, LLC (2023). *Gurobi Optimizer Reference Manual*.

Hammersley, J. M. and Clifford, P. E. (1971). "Markov random fields on finite graphs and lattices". Working paper.

Harsha, P., Subramanian, S., and Ettl, M. (2019). "A Practical Price Optimization Approach for Omnichannel Retailing". *INFORMS Journal on Optimization*, 1(3), pp. 241–264.

Harsha, P., Subramanian, S., and Uichanco, J. (2019). "Dynamic Pricing of Omnichannel Inventories". *Manufacturing & Service Operations Management*, 21(1), pp. 47–65.

He, B., De Sa, C., Mitliagkas, I., and Ré, C. (2016). "Scan Order in Gibbs Sampling: Models in Which it Matters and Bounds on How Much". In: *Advances in Neural*

Information Processing Systems. Ed. by D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, and R. Garnett. Vol. 29.

He, Y., Xu, Q., and Wu, P. (2020). "Omnichannel retail operations with refurbished consumer returns". *International Journal of Production Research*, 58(1), pp. 271–290.

Hendel, I. (1999). "Estimating Multiple-Discrete Choice Models: An Application to Computerization Returns". *The Review of Economic Studies*, 66(2), pp. 423–446.

Hense, J. and Hübner, A. (2022). "Assortment optimization in omni-channel retailing". *European Journal of Operational Research*, 301(1), pp. 124–140.

Honhon, D., Gaur, V., and Seshadri, S. (2010). "Assortment Planning and Inventory Decisions Under Stockout-Based Substitution". *Operations Research*, 58(5), pp. 1364–1379.

Honhon, D. and Seshadri, S. (2013). "Fixed vs. Random Proportions Demand Models for the Assortment Planning Problem Under Stockout-Based Substitution". *Manufacturing & Service Operations Management*, 15(3), pp. 378–386.

Hristopulos, D. T. (2020). *Random Fields for Spatial Data Modeling.* Springer Dordrecht.

Hruschka, H., Lukanowicz, M., and Buchta, C. (1999). "Cross-Category Sales Promotion Effects". *Journal of Retailing and Consumer Services*, 6, pp. 99–105.

Hu, M., Xu, X., Xue, W., and Yang, Y. (2022). "Demand Pooling in Omnichannel Operations". *Management Science*, 68(2), pp. 883–894.

Ising, E. (1925). "Beitrag zur Theorie des Ferromagnetismus". *Zeitschrift fur Physik*, 31, pp. 253–258.

Istrail, S. (2000). "Statistical Mechanics, Three-Dimensionality and NP-Completeness: I. Universality of Intractability for the Partition Function of the Ising Model across Non-Planar Lattices". In: *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing.* New York, NY, USA, pp. 87–96.

iVend Retail (2019). *Global Shopper Trends Report.* White paper. Accessed on 05/05/2023.

Kamakura, W. A. and Kwak, K. (2020). "Menu-choice modeling with interactions and heterogeneous correlated preferences". *Journal of Choice Modelling*, 37, p. 100214.

Kim, J., Allenby, G. M., and Rossi, P. E. (2002). "Modeling Consumer Demand for Variety". *Marketing Science*, 21(3), pp. 229–250.

Kök, A. G. and Fisher, M. L. (2007). "Demand Estimation and Assortment Optimization Under Substitution: Methodology and Application". *Operations Research*, 55(6), pp. 1001–1021.

Bibliography

Lenz, W. (1920). "Beiträge zum Verständnis der magnetischen Eigenschaften in festen Körpern". *Physikalische Zeitschrift*, 21, pp. 613–615.

Li, Z. (2007). "A Single-Period Assortment Optimization Model". *Production and Operations Management*, 16(3), pp. 369–380.

Lo, V. and Topaloglu, H. (2019). "Assortment optimization under the multinomial logit model with product synergies". *Operations Research Letters*, 47(6), pp. 546–552.

Lo, V. and Topaloglu, H. (2022). "Omnichannel Assortment Optimization Under the Multinomial Logit Model with a Features Tree". *Manufacturing & Service Operations Management*, 24(2), pp. 1220–1240.

Luce, R. D. (1959). "Individual Choice Behavior: A Theoretical Analysis". In: Wiley New York.

Lyu, C., Jasin, S., Najafi, S., and Zhang, H. (2022). *Assortment Optimization with Multi-Item Basket Purchase under the Multivariate MNL Model.* Available at SSRN: https://ssrn.com/abstract=3818886.

Maddah, B. and Bish, E. K. (2007). "Joint pricing, assortment, and inventory decisions for a retailer's product line". *Naval Research Logistics (NRL)*, 54(3), pp. 315–330.

Mahajan, S. and Ryzin, G. van (2001). "Stocking Retail Assortments Under Dynamic Consumer Substitution". *Operations Research*, 49(3), pp. 334–351.

Manchanda, P., Ansari, A., and Gupta, S. (1999). "The "Shopping Basket": A Model for Multicategory Purchase Incidence Decisions". *Marketing Science*, 18(2), pp. 95–114.

McFadden, D. (1973). "Conditional Logit Analysis of Qualitative Choice Behaviour". In: *Frontiers in Econometrics*. Ed. by P. Zarembka. New York, NY, USA: Academic Press New York, pp. 105–142.

Moon, S. and Russell, G. J. (2008). "Predicting Product Purchase from Inferred Customer Similarity: An Autologistic Model Approach". *Management Science*, 54(1), pp. 71–82.

Murphy, K. P. (2012). "Machine Learning: a Probabilistic Perspective". In: MIT press.

Netessine, S. and Rudi, N. (2003). "Centralized and Competitive Inventory Models with Demand Substitution". *Operations Research*, 51(2), pp. 329–335.

Pearson, B. (2015). *Walmart inventory cuts: 5 ways to make room for the best shoppers.* Forbes. Accessed on 05/05/2023.

Petro, G. (2020). *Walmart Challenges Amazon With 'Less Is More' Strategy.* Forbes. Accessed on 05/05/2023.

Richards, T. J., Hamilton, S. F., and Yonezawa, K. (2018). "Retail Market Power in a Shopping Basket Model of Supermarket Competition". *Journal of Retailing*, 94(3), pp. 328–342.

Rooderkerk, R. P., DeHoratius, N., and Musalem, A. (2022). "The past, present, and future of retail analytics: Insights from a survey of academic research and interviews with practitioners". *Production and Operations Management*, 31(10), pp. 3727–3748.

Rooderkerk, R. P. and Kök, A. G. (2019). "Omnichannel Assortment Planning". In: *Operations in an Omnichannel World*. Ed. by S. Gallino and A. Moreno. Springer International Publishing, pp. 51–86.

Rusmevichientong, P., Shmoys, D., Tong, C., and Topaloglu, H. (2014). "Assortment Optimization under the Multinomial Logit Model with Random Choice Parameters". *Production and Operations Management*, 23(11), pp. 2023–2039.

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

Ryzin, G. van and Mahajan, S. (1999). "On the Relationship Between Inventory Costs and Variety Benefits in Retail Assortments". *Management Science*, 45(11), pp. 1496–1509.

Schäfer, F., Hense, J., and Hübner, A. (2023). "An analytical assessment of demand effects in omni-channel assortment planning". *Omega*, 115, p. 102749.

Schlapp, J. and Fleischmann, M. (2018). "Technical Note—Multiproduct Inventory Management Under Customer Substitution and Capacity Restrictions". *Operations Research*, 66(3), pp. 740–747.

Schneider, F. and Klabjan, D. (2013). "Inventory control in multi-channel retail". *European Journal of Operational Research*, 227(1), pp. 101–111.

Schraudolph, N. and Kamenetsky, D. (2008). "Efficient Exact Inference in Planar Ising Models". In: *Advances in Neural Information Processing Systems*. Ed. by D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou. Vol. 21. Curran Associates, Inc.

Segev, D. (2019). "Assortment Planning with Nested Preferences: Dynamic Programming with Distributions as States?" *Algorithmica*, 81, pp. 393–417.

Seifert, R. W., Thonemann, U. W., and Sieke, M. A. (2006). "Integrating direct and indirect sales channels under decentralized decision-making". *International Journal of Production Economics*, 103(1), pp. 209–229.

Smith, S. A. and Agrawal, N. (2000). "Management of Multi-Item Retail Inventory Systems with Demand Substitution". *Operations Research*, 48(1), pp. 50–64.

Bibliography

Song, I. and Chintagunta, P. K. (2006). "Measuring Cross-Category Price Effects with Aggregate Store Data". *Management Science*, 52(10), pp. 1594–1609.

Sopadjieva, E., Dholakia, U. M., and Benjamin, B. (2017). *A Study of 46,000 Shoppers Shows That Omnichannel Retailing Works*. Harvard Business Review. Accessed on 05/05/2023.

Spitzer, F. (1971). "Markov Random Fields and Gibbs Ensembles". *American Mathematical Monthly*, 78(2), pp. 142–154.

Timonina-Farkas, A., Katsifou, A., and Seifert, R. W. (2020). "Product assortment and space allocation strategies to attract loyal and non-loyal customers". *European Journal of Operational Research*, 285(3), pp. 1058–1076.

Toh, K. C., Todd, M. J., and Tutuncu, R. H. (1999). "SDPT3 — A Matlab software package for semidefinite programming, Version 1.3". *Optimization Methods and Software*, 11(1-4), pp. 545–581.

Topaloglu, H. (2013). "Joint Stocking and Product Offer Decisions Under the Multinomial Logit Model". *Production and Operations Management*, 22(5), pp. 1182–1199.

Train, K. (2002). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Transchel, S., Buisman, M. E., and Haijema, R. (2022). "Joint assortment and inventory optimization for vertically differentiated products under consumer-driven substitution". *European Journal of Operational Research*, 301(1), pp. 163–179.

Tulabandhula, T., Sinha, D., and Patidar, P. (2020). *Multi-Purchase Behavior: Modeling and Optimization*. Available at SSRN: https://ssrn.com/abstract=3626788.

Tutuncu, R., Toh, K., and Todd, M. (2003). "Solving semidefinite-quadratic-linear programs using SDPT3". *Mathematical Programming Ser. B*, 95, pp. 189–217.

Vasilyev, A., Maier, S., and Seifert, R. W. (2023). "Assortment optimization using an attraction model in an omnichannel environment". *European Journal of Operational Research*, 306(1), pp. 207–226.

Vulcano, G., Ryzin, G. van, and Ratliff, R. (2012). "Estimating Primary Demand for Substitutable Products from Sales Transaction Data". *Operations Research*, 60(2), pp. 313–334.

Wächter, A. and Biegler, L. (2006). "On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming". *Mathematical Programming*, 106, pp. 25–57.

Wainwright, M. J. and Jordan, M. I. (2006). "Log-determinant relaxation for approximate inference in discrete Markov random fields". *IEEE Transactions on Signal Processing*, 54(6), pp. 2099–2109.

Andrey VASILYEV

PHONE: +41789571800
EMAIL: vasiljev.andr@gmail.com

EDUCATION

Sep. 2018 - *Jul. 2023* *(expected)*	PhD in Operations Research **École Polytechnique Fédérale de Lausanne (EPFL)**, Lausanne, Switzerland Chair of Technology and Operations Management (TOM)
Sep. 2017 - *Aug. 2018*	MSc in Operational Research with Data Science (with distinction, GPA 85/100) **University of Edinburgh**, Edinburgh, UK
Sep. 2013 - *Jul. 2017*	BSc in Mathematics (GPA 9.05/10) **National Research University "Higher School of Economics" (HSE)**, Moscow, Russia

EXPERIENCE

Sep. 2018 - *Jul. 2023* *(expected)*	Doctoral Assistant **École Polytechnique Fédérale de Lausanne (EPFL)**, Lausanne, Switzerland PhD Thesis: "Quantitative Methods for Omnichannel Decision-Making" Supervisor: Prof. Ralf W. Seifert Methodology highlights: discrete choice modeling, mixed-integer programming, expectation-maximization algorithm, newsvendor model, simulated annealing, Markov random fields, Ising model, sparse maximum likelihood estimation, Gibbs sampling Other responsibilities: co-supervising 4 MSc theses, teaching (see the Teaching Assistantship section)
Jul. 2022 - *Sep. 2022*	Visiting Doctoral Researcher **University College London (UCL)**, London, UK Two-month research visit at UCL Dept. of Statistical Science, hosted by Dr. Sebastian Maier
Mar. 2021 - *Aug. 2021*	Junior Applied Scientist / Research Fellow **Zalando SE**, Berlin, Germany (working remotely at Zalando's Swiss subsidiary Fision AG) Zalando AI Fellowship Program Project: "Replenishment Recommendations for Zalando Fulfillment Solutions Partners" • Designed an algorithm based on an optimization problem under chance constraints • Implemented the algorithm and a baseline method as modules of Python package • Developed a simulation framework for performance evaluation of the algorithms • Simulations showed that on average, following the recommendations leads to a 70% reduction in missed sales while ensuring a low risk of overstocking • Implemented the algorithm and unit tests in Kotlin for production • The algorithm went live a couple of weeks after the end of the internship

SKILLS AND INTERESTS

Areas of expertise:	choice modeling, machine learning, integer programming, heuristic methods, discrete event simulation, time series forecasting, asymptotic methods
Coding skills:	Python (including PyTorch), R, basics of MATLAB, basics of Kotlin
Optimization tools:	PuLP, YALMIP (complemented by Gurobi, CPLEX, MOSEK, CBC, IPOPT), XPress
Other computer skills:	Microsoft Office (Excel, Word, PowerPoint), LaTeX
Hobbies:	reading, history, hiking, chess, table tennis, traveling

Teaching Assistantship

- Value Chain Management in Practice at EPFL (Sep. 2018 - Dec. 2018, Sep. 2019 - Dec. 2019, Sep. 2020 - Dec. 2020, Sep. 2021 - Dec. 2021)
- Network Analytics at EPFL (Feb. 2019 - Jun. 2019)
- Calculus at HSE Moscow Institute of Electronics and Mathematics (Feb. 2016 - Jun. 2016, Sep. 2016 - Dec. 2016, Feb. 2016 - May 2017)
- Algebra at HSE Department of Computer Science (Apr. 2015 - Jun. 2015)

Online Courses

Deep Neural Networks with PyTorch, Bayesian Statistics: From Concept to Data Analysis, Introduction to Machine Learning (in Russian), Analyzing and Visualizing Data with Excel, Statistical Inference and Modeling for High-throughput Experiments

Summer and Winter Schools

Jan. 19-24 2020	Winter School on Data Science, Optimization and Operations Research, Zinal, Switzerland
Jan. 13-18 2019	Winter School on Optimization, Zinal, Switzerland
Jul. 11-29 2016	CROC IT Summer School (Business Analytics), Moscow, Russia

Conferences and Talks

Oct. 16-19 2022	2022 INFORMS Annual Meeting Indianapolis, IN, USA
Sep. 17 2020	Richemont-EPFL Research Day: Digital Technologies & Data Sciences for New Retail Lausanne, VD, Swizerland
Jun. 15-16 2017	Differential Equations and Related Problems of Mathematics IX Priokskaya Scientific Conference, Zaraisk, Moscow region, Russia
Apr. 10 2017	Dynamical Systems and Differential Equations Lomonosov 2017, Moscow, Russia
Jun. 10-11 2016	Differential Equations and Related Problems of Mathematics VIII Priokskaya Scientific Conference, Konstantinovo, Ryazan region, Russia

Publications

Vasilyev, A., Maier, S., and Seifert, R. W. (2023). Assortment optimization using an attraction model in an omnichannel environment. *European Journal of Operational Research*, 306(1):207–226.

Parusnikova, A. and Vasilyev., A. (2019) On the exact Gevrey order of formal Puiseux series solutions to the third Painlevé equation. *Journal of Dynamical and Control Systems*, 25(4):681–690.

Working Papers

Vasilyev, A., Maier, S., and Seifert, R. W. Omnichannel assortment optimization given basket shopping behavior.

Vasilyev, A., Maier, S., and Seifert, R. W. Optimizing omnichannel assortments and inventory provisions using an attraction model.