Causal Product Networks: A Data-driven Methodology for Modeling Basket-Shopping Consumer Behavior

Vishal Gaur*, Nur Kaynar *, Ziwei Zhu *

Samuel Curtis Johnson Graduate School of Management, Cornell University, Ithaca, NY, USA. vg77@cornell.edu, nur.kaynar@cornell.edu, zz575@cornell.edu

The purchase of one product affects the purchases of others in a shopping basket. Discovering and measuring these effects across an exponentially large number of possible interactions among numerous products in different categories is extremely challenging. We propose a new causal discovery approach to learn the underlying causal structure among product purchases from observational shopping basket data, filter out non-causal correlations, and construct a causal product network (CPN) to describe these latent interactions. We validate this approach and demonstrate its value by utilizing a large-scale basket-shopping dataset. Our main results are as follows. (1) We show that CPNs represent interactions among products in a shopping basket much more accurately than other candidate network structures of hypothetical consumer behavior, such as a complete network or a correlations-based network. (2) Our unrestricted CPN model, which allows product-level relationships, more accurately fits basket-shopping data than restricted specifications of product interactions used in the previous literature, such as category-level interactions, without requiring significantly more parameters. (3) Comparing brick-and-mortar and online channels, we discover that the brick-and-mortar channel exhibits denser causal connections among product purchases in a basket. (4) Finally, using the constructed CPNs, we demonstrate their application in a multi-category assortment optimization problem and find that our model outperforms a benchmark multinomial logit model that treats each category independently by 20%-42% in total sales. Our results make a significant advancement towards using large-scale transaction data for modeling shopping behavior, generating new managerial insights, and optimizing decisions.

Key words: basket shopping, product network, assortment optimization, causal discovery, empirical research

1. Introduction

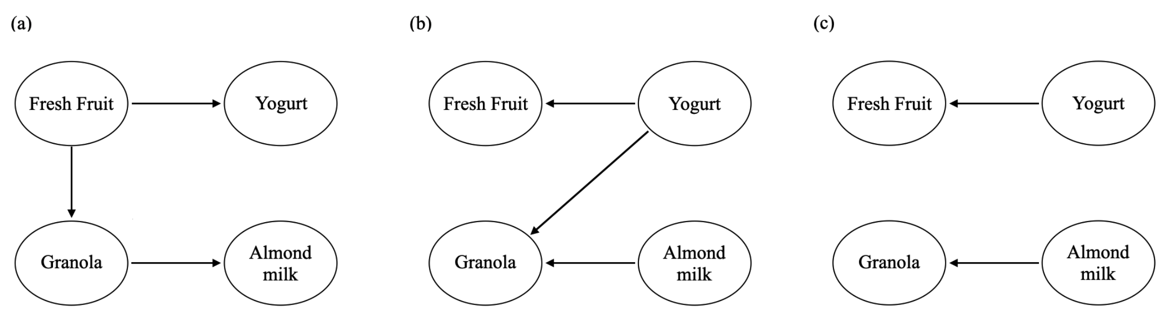
Consider a customer purchasing multiple products across categories in a single shopping visit to a supermarket store. The study of such basket-shopping behavior is important for retailers to address different kinds of questions, such as: (i) utilizing the vast amount of historical basket purchase data, can we identify how buying one product *causally* affects the purchases of additional products, and (ii) using the causal interactions among product purchases, can we optimize retailing decisions, such

* Authors are listed in alphabetical order

as which products to include in the assortment strategy or what prices and promotion strategies to employ to maximize the total expected profit or sales.

We illustrate these questions in Figure 1 by showing different scenarios of interactions among four products: yogurt, fresh fruit, granola, and almond milk; the interactions are depicted as *causal product networks* (CPNs) where nodes symbolize the products and directed edges denote causal relationships between the purchases of these products. In Figure 1(a), fresh fruit purchase influences yogurt and granola purchases, with granola additionally inducing almond milk purchases. In contrast, Figure 1(b) suggests yogurt as the driver that influences fresh fruit and granola purchases, while almond milk directly impacts granola sales. Figure 1(c) features a sparser configuration, highlighting only the direct causal relationships between yogurt and fresh fruit, as well as between almond milk and granola. Given these differences, the assortment strategy must be tailored for each network: offering a wide variety of fresh fruit may be crucial in network (a) to initiate the purchasing sequence, while a focus on yogurt and almond milk variety can be essential in network (b) to exploit their role as purchase drivers, and network (c) suggests that jointly determining the assortments of yogurt with fresh fruit independently of the other products and likewise for almond milk with granola could be an effective strategy. Thus, understanding the specific causal dynamics between these products can provide significant benefits for optimizing the assortment to meet consumer buying habits.

Figure 1 **An illustration of alternative causal networks for four products from different categories**



While considerable advancement has been made in consumer choice for purchases within a single category, it is only recently that new methods are emerging to characterize basket-shopping behavior of consumers. In particular, the Multivariate MNL (MVMNL) and the Bundle Multivariate Logit Model (BundleMVL-K) models have been developed to accommodate a complementary effect across categories and a substitution effect within categories in modeling multiple-purchase decisions (Jasin et al. 2023, Tulabandhula et al. 2023). These models generalize the prior literature on Multinomial Logit Models (MNLs), Nested Multinomial Logit (NMNL), and Mixed MNL (MMNL) models that are used to formulate a single-choice problem by capturing substitution patterns within a single category (Börsch-Supan 1990, McFadden and Train 2000). However, there hasn't yet been an empirical study characterizing the behavior of consumers when they make multiple purchases. As a result, the basket-shopping models proposed in the literature typically assume specific structures of basket-shopping behavior for tractability without the benefit of empirical evidence.

Moreover, modeling causal relationships for basket-shopping consumers is a hard problem. For instance, for the above example with four products, there are 12 possible causal relationships and 2^{12} different causal graphs, which will likely yield different optimal product assortments. As the size of the product network of a retailer increases, it would be astronomically difficult to conduct controlled experimental studies to identify which causal relationships are supported by data and evaluate their relative strength. In this paper, we propose a novel causal structure learning technique to uncover causal relationships among product purchases and construct directed acyclic graphs (DAGs) that can fully reveal the dynamics among many products by utilizing the transaction data collected by retailers. For instance, in Figure 1, basket-shopping data would be used to relate the instances of purchasing fresh fruit, yogurt, granola, and almond milk with each other as well as with other products and iteratively eliminate possible conditional relationships until a DAG is obtained that is supported by data. We use the PC algorithm (Spirtes et al. 2000b), the most popular causal discovery algorithm, which leverages probabilistic independence tests to infer these structures from observational data.

It is important to note a caveat of our research. While we use the term *causal* due to the causal discovery method that we employ, our study is not a causal study in the traditional economics sense of the term because we do not conduct controlled field experiments to measure each relationship. Instead, the causal discovery approach is a data-driven approach that seeks to systematically eliminate relationships that do not have evidence in observational data. The network estimated from this approach will then have to be further tested by retailers and will also have to be examined for omitted variables, which can be done through selected field experiment studies. Thus, we do not claim that the networks discovered in our estimation are truly causal. Our goal is to validate the applicability of this technique for analyzing retail transaction data and extracting important insights.

With the above motivation, we study the following main research questions in this paper: (1) How do causal product networks perform in describing the relationships among product purchases when compared to various other network representations of consumer behavior? In other words, are the assumptions in various theoretical models consistent with the characteristics of basket-shopping consumer behavior discovered from the data by applying causal discovery? (2) Given the empirical evidence for network structures, how can we construct the optimal assortment strategy based on the causal relationships among product purchases, and to what extent does it outperform a benchmark within-category assortment strategy based on Multinomial Logit (MNL) models? (3) Since the technique of causal discovery has not been applied to this problem before, how well-suited is this technique to the challenges of estimating causal graphs for basket-shopping data?

We analyze the first research question in three steps. First, we formalize the representation of interactions among product purchases through a causal network, where the relationships between product purchases are captured using causal links. We learn the existence of these relationships in a data-driven way using the PC algorithm in the context of basket shopping. Note that the causal product networks are learned from the final basket data and they are independent of the sequence in which the items were added to the basket. We then evaluate the goodness of fit of these causal graphs with respect to complete networks and correlation-based networks in representing the behavior of basket-shopping consumers. Complete networks are defined as product networks in which all products are interconnected by undirected edges; correlation-based networks connect statistically correlated products with undirected edges; and causal networks are characterized by edges that signify causal links among products. We hypothesize that causal networks provide a more accurate description of the relationships among products than complete and correlation-based networks (Hypothesis 1).

Second, motivated by the structure of product interactions assumed in the previous literature, including product-level and category-level interactions, we then investigate the validity of various alternative specifications of causal networks. More specifically, we consider three alternative restrictions on the causal network: (1) causal effects that are from category to category, (2) causal effects from products to categories, and (3) causal effects from categories to products. The study of these specifications is not only useful in understanding consumer shopping behavior but also computationally relevant because the number of parameters to be estimated varies across these specifications. The causal network with product-level causal effects requires more parameters, but allows them to be tailored to individual products, which can be beneficial by providing greater flexibility in representing customer purchase behavior. Therefore, we hypothesize that these restrictions do not hold and the product-level causal effects model most effectively describes the causal relationships in basket shopping compared to the three alternative specifications (Hypothesis 2).

Third, our model can be estimated specifically for different types of retailing contexts to characterize differences and similarities in consumer behavior across them. For instance, in the previous literature, Chintala et al. (2023) empirically find that online purchases show lower shopping basket variety compared to brick-and-mortar (B&M) purchases, but the causal relationships in consumer purchase behavior in these two channels have not been investigated. To investigate such an application of the model, we compare the causal relationships in consumer shopping behavior between

B&M and online channels and hypothesize that the online channel exhibits fewer causal relationships among product purchases within shopping baskets compared to the B&M channel (Hypothesis 3), which can be used to explain the lower variety in online shopping baskets.

We conduct this analysis by collaborating with Numerator, a prominent market research company known for its first-party, consumer-sourced data. Numerator gathers purchase data for a large panel of consumers across various B&M and online channels; transaction sources include retailer loyalty data, mobile app purchases, email as well as paper receipts. Their extensive database comprises approximately 8.21 million consumer shopping baskets from more than 17,000 stores and around 3.90 million products in 36 categories, encompassing major retail chains such as Walmart, Costco, and Target, as well as small supermarkets and local grocery stores across the United States. We utilize the rich Numerator data by subdividing it into themes, including pasta, quick service restaurant (QSR), bakery, prepared food, and unprepared food. We extract data for these five themes, each comprising different product categories, and conduct our analysis on each theme to thoroughly study customers' basket-shopping behavior for different types of products. The empirical results presented in this paper are shown separately for the five themes, covering 250 products in 25 categories across B&M and online channels, with the number of baskets ranging from 4,907 to 14,472 for each theme.[1]

Our empirical evidence reveals significantly important results and confirms all three hypotheses. First, we demonstrate that causal product networks provide a more accurate representation of product relationships in shopping baskets compared to complete and correlation-based networks. Moreover, causal product networks require fewer parameters, thereby reducing complexity; e.g., for the pasta theme, the complete network has 2,750 parameters to be estimated, the correlation-based network has 693, and the causal product network has 56. Second, we find that product-level causal effects most effectively describe the causal relationships in basket shopping, outperforming the alternative category-level restrictions. This result shows that consumers' basket-shopping behavior is driven by product-level causality, not category-level interactions as assumed in the prior literature. For example, we find that the purchases of condiments and beverages are statistically independent, but purchasing fruit juice, a product in the beverages category, has a causal effect on the purchase of salad dressing, a product in the condiments category. Thus, product-level effects can provide a more accurate description of basket-shopping consumer behavior. Moreover, we find that the number of edges in the product-level causal model is surprisingly only marginally larger than that in the alternative restriction-based specifications. Third, our analysis reveals that the online shopping channel exhibits fewer causal relationships among product purchases within shopping baskets compared to the B&M channel. Altogether, these results demonstrate the value of causal structure

[1] An exception is that the number of baskets for the Quick Service Restaurant (QSR) theme in the online channel is smaller than 500 due to its limited online availability.

learning to characterize basket-shopping consumer behavior and provide qualitative insights into the relationships between product categories.

To answer the second research question based on the causal network, we propose a mixed-integer program that uses the constructed causal product networks to optimize product assortments for basket shopping. Arguably, if causal modeling more accurately captures consumer behavior, it should result in better assortment decisions than models that ignore causality. Thus, to examine the value of modeling basket-shopping behavior for assortment optimization, we compare the performance of our causal product networks with the classic MNL model; the former incorporates across-category basket-shopping behavior, while the latter considers only within-category choice behavior. Comparing these two models in both B&M and online channels for one theme of product categories (prepared foods), we find that our causal model outperforms the MNL model by approximately 20%-42% in total sales across assortment sizes. Moreover, we observe distinct assortment strategies between causal product network models and MNL models, which reveals insights into the assortment implications of consumer behavior.

Finally, with respect to the third research question, we develop a method to combine causal discovery with simultaneous equations modeling in order to construct the specification of consumer basket-shopping behavior. In the first step of this method, causal discovery is used to construct a DAG of causal relationships across products in our dataset. A key challenge that we face in this analysis is that the amount of basket-shopping data is too large and too sparse to allow efficient application of the PC algorithm. We address this challenge by constructing many smaller subsamples of the data, estimating the DAG for each sample, and statistically determining the most salient relationships. The output of this procedure is an identification of the edges in the graph and their orientation, but not the strength of each edge. Thus, in the second step, we convert the DAG into a simultaneous equations model and use this model to estimate the demand rate for each product and the coefficients along the DAG. This two-step process gives us the final demand rates for products as a function of a given assortment of products and causal relationships. Our method can be scaled to many different retailing settings and can be expanded to incorporate additional types of retail data such as prices, promotions, store layout, etc. to obtain further insights.

The remainder of this paper is organized as follows: §2 summarizes the related literature; §3 explains how to construct causal product networks and quantify the strengths of these relationships using historical basket-shopping data; §4 validates the accuracy of the proposed framework using synthetic data; §5 develops the hypotheses regarding basket-shopping consumer purchase decisions; §6 presents the empirical results; §7 focuses on the assortment optimization problem; and §8 concludes the paper.

2. Literature Review

Our paper is related to the existing literature on choice modeling for basket-shopping behavior, use of graph models in retail analytics, use of causal structure learning, and shopping behavior of customers in multi-channel retailing. In this section, we review the relevant literature on these topics and describe the contributions of our work. Choice modeling for basket-shopping customers broadly relies on representing the complementary and substitution effects among products during the consumer buying process. Multinomial Logit Models (MNL), which are widely used in discrete choice modeling, capture substitution behavior within a category for either a single utility-maximizing choice (Börsch-Supan 1990, McFadden and Train 2000) or multiple purchase decisions (Bai et al. 2023), and can also be used to estimate substitution effects in a shopping basket (Mani et al. 2022). In the recent years, the traditional MNL framework has been extended to model multiple-purchase decisions across categories in basket shopping. The nested-MNL model has been extended to describe the behavior of customers who choose one product from each of two categories and to demonstrate its performance in an assortment optimization problem (Cachon and Kök 2007). The Multivariate Multinomial Logit (MVMNL) model and the Bundle Multinomial Logit (BundleMVL-K) model have been developed to account for both complementary and substitution effects across categories (Jasin et al. 2023, Tulabandhula et al. 2023). These models have made considerable advances in developing the idea that purchasing one product can either increase or decrease the utility of purchasing another product, which can capture complementary or substitution effects between products. Typically, these models make further assumptions regarding the structure of the between-product effects for tractability; for example, assuming that the effects are at the category-level rather than the product-level or assuming symmetry can both significantly reduce the number of parameters.

However, these assumptions have not been tested in real-life data, and the empirical analysis of basket-shopping behavior remains scarce. One recent exception is Ruiz et al. (2020) who propose a random utility model of sequential discrete choice, where customers make purchase decisions one after another and the utility of subsequent purchases depends not only on the attributes of each product but also on all of the previous purchases. This method is based on a complete product network and requires all pairwise product relationships to be estimated, which is a challenging problem. Our research contributes to this literature by constructing causal product networks using basket data, which identify the existence of specific directional pairwise relationships represented as directed acyclic graphs, and by showing that CPNs provide a significantly better fit to real data than complete networks. Our paper also provides insights into basket-shopping behavior and a method to estimate the parameters of choice modeling that can be utilized in optimization models.

The methodology that we utilize in this paper falls under product networks and directed acyclic graphs (DAGs), which have attracted growing attention in modeling retail operations settings in the recent years. In e-commerce and information systems literature, Huang et al. (2007) represent the purchase behavior of customers in online marketplaces as bipartite consumer-product graphs and study the characteristics of these empirical graphs against behavioral predictions from random graphs. Oestreicher-Singer and Sundararajan (2012) construct co-purchase networks for books sold on Amazon.com, where vertices represent books and edges connect books purchased by the same customer, and study the effect of the network relationships on product demand. Dhar et al. (2014) also use the product network of books sold on Amazon.com and utilize it to predict future product demand as a function of the historical demand of a given product and its linkages with other products. While these papers use graphs to model co-purchase behavior, in the operations management literature, DAGs have been used to represent preference ordering of products in non-parametric consumer choice models. For example, Honhon et al. (2010) study the optimization of assortment and inventory quantities in a product category when stockout-based substitution is specified by a family of DAGs. Jagabathula and Vulcano (2018) study the estimation of partial ordering of preferences, represented as DAGs, from historical purchasing data. Jagabathula et al. (2022) utilize this type of non-parametric choice model for personalized promotions and combine the estimation of DAGs from historical data with an MNL model.

Note that DAGs are used in different ways in the above literature: to represent complementary effects between products in the first group of papers and substitution effects in the second group. Our paper differs from this literature by relying on causal directed acyclic graphs (DAGs) in product networks to describe the causation between product purchases in shopping baskets. In our model, we allow DAGs to represent both complementary and substitution effects; the edges between products in the causal product networks are estimated from data and a positive weight on an edge represents complementarity whereas a negative weight represents substitution. To discover these causal relationships among product purchases, we introduce a technique known as causal discovery, which can infer causal relationships from observational data (Meek 2013, Glymour et al. 2019). This technique employs algorithms to analyze patterns of conditional independence relations among variables to infer the direction and presence of causality (Eberhardt 2017). Based on the causal discovery approach, we propose that our causal product networks can be widely applied to solve operational problems such as choice modeling, assortment optimization, and revenue management, and can also be generalized to incorporate different types of variables.

More generally, the causal discovery approach has been increasingly utilized in the fields of economics and policy-making in the recent years. Hall-Hoffarth (2022) adopts a causal discovery approach to uncover latent macroeconomic Dynamic Stochastic General Equilibrium (DSGE) models, which can be represented by directed acyclic graphs (DAGs). Martinoli et al. (2023) introduce a general procedure for the calibration and validation of macroeconomic simulation models based on the causal search approach and apply it to study the relationship between climate change and economic growth. Eberhardt et al. (2024) propose an optimization-based method for inferring causal structures from observational data, using it to assess the validity of instrumental variables. This approach is demonstrated with the well-known quarter-of-birth and proximity-to-college instruments to estimate the returns to education. Kaynar and Mitrofanov (2024) develop a causal discovery algorithm that combines short-term experimental data with long-term observational data to identify short-term variables influenced by treatment, which in turn affect the long-term outcome. This approach is applied to analyze the long-term effects of subsidies on healthy food products. In addition to these studies in economics, the causal discovery approach has been widely applied across a broad range of areas, including bioinformatics (e.g., Foraita et al. 2020, Triantafillou et al. 2017), healthcare (e.g., Tu et al. 2019, Shen et al. 2021, Hasan and Gani 2022), and environmental science (e.g., Ebert-Uphoff and Deng 2014, Runge et al. 2019, Nowack et al. 2020). Our study introduces this approach to the field of retail operations, and addresses challenges related to the unique aspects of very large scale retailing data, such as sparsity and heterogeneity.

Finally, our paper is related to the existing literature on multi-channel retailing. Research in this literature has modeled and analyzed the shopping behavior of customers across different channels, especially the brick-and-mortar and online channels. For example, Degeratu et al. (2000) compare the price sensitivities of customers between online and brick-and-mortar channels; Venkatesan et al. (2007) analyze the driving factors of multichannel shopping behavior; Chu et al. (2010) investigate brand loyalty and price sensitivities for households in the online and brick-and-mortar channels; Gallino and Moreno (2014) demonstrate that integrating online and offline channels, specifically buying items online and picking them up in a physical store, can lower customer purchases in the online channel while increasing offline purchases and customer traffic; Wang and Goldfarb (2017) study the complementarity and substitution effects in product demand between brick-and-mortar and online channels; Gallino et al. (2023) examine how in-process delays affect customer purchase behavior in the online channel and how customer sensitivity to these delays changes throughout different stages of a shopping trip; and Chintala et al. (2023) show that there are significant differences in shopping basket variety, basket similarity, and other purchase patterns for both brick-and-mortar and online channels. Since our dataset includes both brick-and-mortar and online channels, it enables us to apply the causal discovery approach separately to each of these channels and examine differences between the causal graphs associated with these channels. We conduct this analysis and show that the causal graphs are consistent with the empirical findings in Chintala et al. (2023).

Thus, our paper contributes to the literature on retail operations by developing new insights into basket-shopping behavior that can be utilized in optimization models and can accommodate both complementary and substitution effects. Moreover, our paper provides a comprehensive analysis of the effectiveness of the novel technique of causal discovery in retail operations by testing it across different product groups and shopping channels, and by evaluating its impact on optimal assortments.

3. Causal Product Network Model and Estimation

In this section, we introduce our proposed methodology for learning the interactions among product purchases by constructing a causal product network. Rather than assuming predefined relationships, our approach combines causal structure learning with simultaneous equations modeling to derive these relationships from data, allowing us to analyze the behavior of basket-shopping consumers in a data-driven manner. We first outline the standard assumptions for causal structure learning and explain how to construct causal product networks from historical shopping data using a well-known causal structure learning algorithm in §3.1. This algorithm returns a DAG in which product purchases are represented as nodes, and the causal relations between product purchases are represented using directed edges. It is important to note that causal structure learning alone does not quantify the strength of discovered edges, i.e., the magnitude of the effect on each edge. Thus, in the second step, we convert the DAG into a simultaneous equations model (SEM) and use this model to estimate the demand rate for each product and the coefficients along the DAG in §3.2. This two-step process gives us the final demand rates for products as a function of a given assortment of products and causal relationships.

We begin by introducing the notation we will use throughout the paper. Let $J = \{1, 2, \ldots, m\}$ be the set of categories where m is the total number of categories. Let $I_j = \{1, 2, \ldots, n_j\}$ be the set of products in category $j \in J$ where n_j is the total number of products in category j. Let $N = \sum_{j \in J} n_j$ be the total number of products across categories. Lastly, we define x_{ij} to represent the purchase of product i from category j.

3.1. Causal Structure Learning

Causal structure learning aims to identify causal relationships among variables from observational data without conducting experiments. Causal modeling involves associating a probability distribution $P_{\mathcal{G}}(V)$ with a graph $\mathcal{G} = (V, E)$, where V and E represent the set of variables and the edges included in the graph, respectively. In our setting, V is defined as the set of all product purchases, i.e., $V = \{x_{ij} | i \in I_j, j \in J\}$. Let $par(x_{ij})$ denote the set of nodes that are the parents of product x_{ij}, i.e., $par_{ij} = \{x_{i'j'} | (x_{i'j'} \to x_{ij}) \in E\}$. The underlying assumption is that the distribution $P_{\mathcal{G}}(V)$ is generated by the graph structure in a way that allows factorization: $P_{\mathcal{G}}(V) = \prod_{x_{ij} \in V} P_{\mathcal{G}}(x_{ij} | par_{ij})$.

The following three key assumptions enable bridging the observed data with the underlying causal structure (Eberhardt 2017, Glymour et al. 2019).

Assumption 1 (Causal Markov Condition) *Each variable $x_{ij} \in V$ is conditionally independent of its non-descendants given its parents in the causal graph $\mathcal{G} = (V, E)$.*

Assumption 2 (Faithfulness) *The only independencies present in the probability distribution are those implied by the graph structure through the causal Markov condition.*

Assumption 3 (Causal Sufficiency) *All common causes of any pair of variables in V are included within the set V.*

The first assumption, the causal Markov condition follows from how we have defined the probability distribution in terms of the causal structure (Pearl 2009) and permits us to transition from the causal graph to the observed probabilistic independencies. In contrast, the faithfulness condition represents an additional assumption that ensures an independence in the data is due to the underlying graph, rather than, for example, two causal pathways canceling each other out (Spirtes et al. 2000a). Lastly, the causal sufficiency assumption ensures that all common causes of any pair of variables in the set V are also contained within V, thereby excluding the existence of any unobserved confounders. These assumptions establish a correspondence between the observed data and the underlying graph, enabling causal structure learning algorithms to leverage probabilistic independence relations implied by the data to learn the underlying causal relations. The following theorem formalizes this correspondence:

Theorem 1 ***(Spirtes et al. 2000b)***. *Let graph $\mathcal{G}^*$ be the true data generating graph. Under the causal Markov condition, causal faithfulness and causal sufficiency assumptions, we have:*

(i) for all $x_{ij}, x_{i'j'} \in V$, x_{ij} and $x_{i'j'}$ are adjacent in true graph $\mathcal{G}^*$ if and only if x_{ij} and $x_{i'j'}$ are dependent conditional on any $C \in C_{(ij)(i'j')}$ and

(ii) for all $x_{ij}, x_{i'j'}, x_{i''j''} \in V$ such that x_{ij} is adjacent to $x_{i'j'}$, $x_{i'j'}$ is adjacent to $x_{i''j''}$, and x_{ij} and $x_{i''j''}$ are not adjacent in the true graph $\mathcal{G}^*$, $x_{ij} \rightarrow x_{i'j'} \leftarrow x_{i''j''}$ is in the true graph $\mathcal{G}^*$ if and only if x_{ij} and $x_{i''j''}$ are dependent conditional on every set $C \in C_{(ij)(i''j'')}$ such that $x_{i'j'} \in C$,

where $C_{(ij)(i'j')} = \{C \mid C \subseteq V \setminus \{x_{ij}, x_{i'j'}\}\}$ stores all possible conditioning sets for $x_{ij}, x_{i'j'} \in V$.

Theorem 1 establishes that both the adjacency relations and the orientation of edges in a graph can be inferred from conditional independence relations. The PC algorithm, which is a popular and widely used method in causal structure learning, utilizes this correspondence between (conditional) independence relations and the underlying graph to iteratively learn the underlying causal relations from observational data (Spirtes et al. 2000b). Specifically, the PC algorithm uses the principles outlined in Theorem 1 and operates in two main phases: the first phase discovers the skeleton of the causal structure without specifying the directions of discovered edges and the second phase handles orienting the edges of the skeleton. The pseudo-codes for these two phases of the PC algorithm are provided in Algorithms 1 and 2, respectively.

In Algorithm 1, we start with a complete undirected graph in which each pair x_{ij} and $x_{i'j'}$ in V is connected by an undirected edge. We then evaluate each pair of variables, x_{ij} and $x_{i'j'}$, for probabilistic independence by considering subsets C in $C_{(ij)(i'j')} = \{C \mid C \subseteq V \setminus \{x_{ij}, x_{i'j'}\}\}$, i.e., all subsets of V excluding the focal pair of variables. The size of each subset, denoted by $|C|$, incrementally increases throughout the testing process. If x_{ij} and $x_{i'j'}$ are found to be independent given a subset C, we remove the edge between them, following part (i) of Theorem 1. The output of Algorithm 1 is the skeleton of the causal structure, which represents the potential causal relations without directionality among the variables and it serves as the input for the next phase of the algorithm. Algorithm 2 focuses on orienting the edges within the graph following part (ii) of Theorem 1. We proceed by examining conditional independence relations between variable triplets $x_{ij}, x_{i'j'}, x_{i''j''} \in V$ such that x_{ij} is adjacent to $x_{i'j'}$ and $x_{i'j'}$ is adjacent to $x_{i''j''}$ and x_{ij} and $x_{i''j''}$ are not adjacent in the skeleton returned by Algorithm 1. If x_{ij} and $x_{i''j''}$ are probabilistically dependent with respect to conditioning set $C = \{x_{i'j'}\}$ then we orient $x_{ij}—x_{i'j'}—x_{i''j''}$ as $x_{ij} \rightarrow x_{i'j'} \leftarrow x_{i''j''}$. The algorithm then employs Meek's rules for further edge orientation, ensuring that all the orientations are consistent with Theorem 1 and the resulting graph does not include cycles (Meek 2013; see Appendix A for illustration). The PC algorithm offers flexibility in selecting independence tests appropriate for the specific domain and enables the choice of preferred methods for correcting multiple hypothesis testing. These choices depend on factors such as sample size, the number of variables, whether the variables are categorical or continuous, and the assumptions made about the parametric nature of the causal relationships.

In general, the independence structure seen in observational data is not guaranteed to uniquely identify the underlying causal graph. Two graphs with different structures are said to be *Markov equivalent* if they have the same independence structure (Verma and Pearl 1990). In the context of causal discovery, the Markov equivalence class of the true, data-generating graph is the limit of what can be learned about the causal structure from the independence structure in the data. The PC algorithm has been shown to be asymptotically correct, meaning in the large-sample limit it discovers the true data-generating graph up to an equivalence class, given Assumptions 1-3 (Spirtes et al. 2000b). This implies that the edge set E returned by Algorithm 2 represents a Markov equivalence class of graphs and it might include undirected edges. In the context of our study, having $(x_{i'j'} \rightarrow x_{ij}) \in E$ implies purchasing product i' from category j' has a *direct* causal impact on

Algorithm 1: PC algorithm - Skeleton Discovery (Spirtes et al. 2000b)

Input: I_j, J and $V = \{x_{ij}$ for $i \in I_j, j \in J\}$.
Output: $\mathcal{G} = (V, E)$.
Initialization: $\mathcal{G} = (V, E)$ where $E = \big\{(x_{ij} — x_{i'j'}),\ \ \forall x_{ij}, x_{i'j'} \in V$ such that $x_{ij} \neq x_{i'j'}\big\}$,
$\mathcal{N}_{\mathcal{G}}(x_{ij}) = \{x_{i'j'} \in V \setminus x_{ij}\}$.
1. Initialize $\ell = 0$.
2. **for** an ordered pair $x_{ij}, x_{i'j'} \in V$ where $x_{i'j'} \in \mathcal{N}_{\mathcal{G}}(x_{ij})$ and $|\mathcal{N}_{\mathcal{G}}(x_{ij}) \setminus x_{i'j'}| \geq l$:
3.   **for** $C \in C_{(ij)(i'j')}$ with $|C| = \ell$:
4.   **if** $x_{ij} \perp\!\!\!\perp x_{i'j'} \mid C$:
5.   Remove the edge $(x_{ij} — x_{i'j'})$ from E.
6.   Update $\mathcal{N}_{\mathcal{G}}(x_{ij}) = \mathcal{N}_{\mathcal{G}}(x_{ij}) \setminus x_{i'j'}$ and $\mathcal{N}_{\mathcal{G}}(x_{i'j'}) = \mathcal{N}_{\mathcal{G}}(x_{i'j'}) \setminus x_{ij}$.
7.   **Break**.
8. **if** $\ell < |V| - 2$:
9.   Update $\ell = \ell + 1$. Go to Step 2.
10. **else** $\ell = |V| - 2$:
11.   **Break**.
12. Return $\mathcal{G} = (V, E)$.

Algorithm 2: PC algorithm - Orienting Edges (Spirtes et al. 2000b)

Input: $\mathcal{G} = (V, E)$.
Output: $\mathcal{G} = (V, E)$.
1. **for** $x_{ij}, x_{i'j'}, x_{i''j''} \in V$ where $\{(x_{ij} — x_{i'j'}), (x_{i'j'} — x_{i''j''})\} \subseteq E$ & $\{(x_{ij} — x_{i''j''})\} \nsubseteq E$:
2.   **if** $x_{ij} \not\perp\!\!\!\perp x_{i''j''} \mid x_{i'j'}$:
3.   Orient $x_{ij} — x_{i'j'} — x_{i''j''}$ as $x_{ij} \rightarrow x_{i'j'} \leftarrow x_{i''j''}$.
4.   Update $E = \big(E \setminus \{(x_{ij} — x_{i'j'}), (x_{i'j'} — x_{i''j''})\}\big) \cup \{(x_{ij} \rightarrow x_{i'j'}), (x_{i'j'} \leftarrow x_{i''j''})\}$.

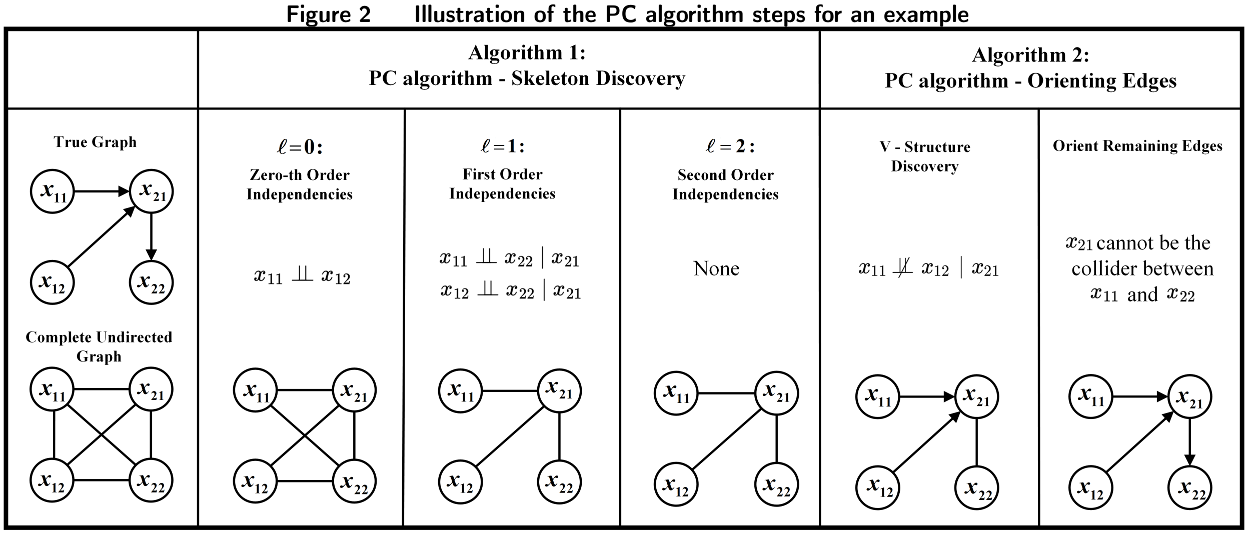
Meek rules (See Appendix A for illustration):
4. **(R1) for** $x_{ij}, x_{i'j'}, x_{i''j''} \in V$ where $x_{ij} \notin \mathcal{N}_{\mathcal{G}}(x_{i''j''})$ and $\{(x_{ij} \rightarrow x_{i'j'}), (x_{i'j'} — x_{i''j''})\} \subseteq E$:
5.   Update $E \leftarrow \big(E \cup \{(x_{i'j'} \rightarrow x_{i''j''})\}\big) \setminus \{(x_{i'j'} — x_{i''j''})\}$.
6. **(R2) for** $x_{ij}, x_{i'j'}, x_{i''j''} \in V$ where $\{(x_{ij} \rightarrow x_{i'j'}), (x_{i'j'} \rightarrow x_{i''j''}), (x_{ij} — x_{i''j''})\} \subseteq E$:
7.   Update $E \leftarrow \big(E \cup \{(x_{ij} \rightarrow x_{i''j''})\}\big) \setminus \{(x_{ij} — x_{i''j''})\}$.
8. **(R3) for** $x_{ij}, x_{i'j'}, x_{i''j''}, x_{i'''j'''} \in V$ where $x_{i'j'} \notin \mathcal{N}_{\mathcal{G}}(x_{i'''j'''})$ &
9. $\{(x_{ij} — x_{i'j'}), (x_{i'j'} \rightarrow x_{i''j''}), (x_{ij} — x_{i'''j'''}), (x_{i'''j'''} \rightarrow x_{i''j''}), (x_{ij} — x_{i''j''})\} \subseteq E$:
10.   Update $E \leftarrow \big(E \cup \{(x_{ij} \rightarrow x_{i''j''})\}\big) \setminus \{(x_{ij} — x_{i''j''})\}$.
11. **(R4) for** $x_{ij}, x_{i'j'}, x_{i''j''}, x_{i'''j'''} \in V$ where $x_{ij} \notin \mathcal{N}_{\mathcal{G}}(x_{i''j''})$ &
12. $\{(x_{ij} \rightarrow x_{i'j'}), (x_{i'j'} \rightarrow x_{i''j''}), (x_{ij} — x_{i'''j'''}), (x_{i'''j'''} — x_{i''j''}), (x_{i'j'} — x_{i'''j'''})\} \subseteq E$:
13.   Update $E \leftarrow \big(E \cup \{(x_{i'''j'''} \rightarrow x_{i''j''})\}\big) \setminus \{(x_{i'''j'''} — x_{i''j''})\}$.
14. **Return** $\mathcal{G} = (V, E)$ where $E = E$.

purchasing product i from category j. Whereas having $(x_{i'j'} — x_{ij}) \in E$ implies an unspecified causal relationship between the products. In other words, $(x_{i'j'} — x_{ij})$ suggests that there is an edge between $x_{i'j'}$ and x_{ij}, but it could point in either direction.

Figure 2 illustrates the steps of the PC algorithm for an example. For a given true data-generating graph over product purchases x_{11} and x_{21} in category 1, and x_{12} and x_{22} in category 2, the PC algorithm conducts a series of conditional independence tests using the data generated accordingly with this graph. Algorithm 1 starts with testing for marginal independencies ($\ell = 0$). Since x_{11} and

x_{12} are marginally independent, denoted as $x_{11} \perp\!\!\!\perp x_{12}$, the edge between x_{11} and x_{12} is removed. Then, we start testing for conditional independencies. At the first order ($\ell = 1$), since $x_{11} \perp\!\!\!\perp x_{22} \,|\, x_{21}$ and $x_{12} \perp\!\!\!\perp x_{22} \,|\, x_{21}$, the edges between x_{11} and x_{22} and x_{12} and x_{22} are removed. Since no further independencies are found at the second order ($\ell = 2$), we obtain the skeleton of the graph. Using this skeleton, Algorithm 2 orients the edges using the dependency relations. Since $x_{11} \not\perp\!\!\!\perp x_{12} \,|\, x_{21}$, we must have $x_{11} \rightarrow x_{21} \leftarrow x_{12}$ following part (ii) of Theorem 1. Finally, we orient the remaining edges using Meek's rules, resulting in $x_{21} \rightarrow x_{22}$, forming the directed acyclic graph that captures the causal relationships. In this example, we are able to identify a unique graph, indicating that there is only one graph in the Markov equivalence class of the underlying graph. [2]

Figure 2 Illustration of the PC algorithm steps for an example



3.2. Simultaneous Equations Model

While the PC algorithm discovers the *existence* of causal relationships among product purchases, it doesn't quantify the strength of their influences. To estimate these causal effects, we convert the DAG into a SEM and use this model to estimate the demand rate for each product and the coefficients along the DAG. Note that the previous section establishes that the distribution $P_{\mathcal{G}}(V)$ is generated by the graph structure in a way that allows factorization: $P_{\mathcal{G}}(V) = \prod_{x_{ij} \in V} P_{\mathcal{G}}(x_{ij} | par(x_{ij}))$. In the SEM, we build on this principle by considering only the parents of each product as the explanatory variables in the equation corresponding to that product. When the CPN is a DAG, then the SEM can be transformed into a triangular matrix and solved to obtain unbiased estimates of all coefficients. However, as discussed in Section 3.1, the edge set E may include undirected edges due to Markov

[2] As previously mentioned, unique identification of the causal graphs is not always guaranteed.

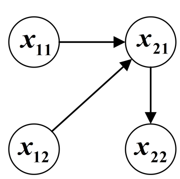
equivalence, making it challenging to identify the exact set of parents of variables. To account for this, we define the parent set par_{ij} for each variable x_{ij} as $par_{ij} = \{x_{i'j'} | (x_{i'j'} \rightarrow x_{ij}) \in E\} \cup \{x_{i'j'} | (x_{i'j'} — x_{ij}) \in E\}$, i.e., undirected edges are included in the equations of both products. Then, the corresponding simultaneous equations can be written as:[3]

$$x_{ij} = \alpha_{ij} + \sum_{x_{i'j'} \in par_{ij}} \beta^{ij}_{i'j'} x_{i'j'} + \epsilon_{ij}, \quad \forall i \in I_j, j \in J. \tag{1}$$

In this SEM, we represent the purchase probability of each product x_{ij} using an intercept α_{ij} that represents the baseline purchase rate, a linear sum of the effects from parent products within par_{ij}, and an error term ϵ_{ij}. The coefficient $\beta^{ij}_{i'j'}$ indicates the estimated strength of the causal effect of the purchase of product $x_{i'j'}$ on the purchase of product x_{ij}; $\beta^{ij}_{i'j'}$ can take either positive or negative values representing situations where the demand for one product may cause an increase or a decrease in the demand for another product. This model is estimated using basket-shopping transaction data, where each basket represents a row in the estimation dataset. We suppress the index for baskets for ease of notation.

Let $\boldsymbol{\beta}$ be the $(N \times N)$-matrix of coefficients $\beta^{ij}_{i'j'}$. Figure 3 illustrates how the graph $\mathcal{G}$ informs construction of the coefficient matrix $\boldsymbol{\beta}$. In this example, each non-zero element $\beta^{ij}_{i'j'}$ within the $(N \times N)$ matrix corresponds to a directed edge from $x_{i'j'}$ to x_{ij} in $\mathcal{G}$, representing the magnitude of that effect. Zero entries in the matrix reflect the absence of a direct causal edge between the respective product purchases in the graph.

Figure 3 **An example of a causal graph and its corresponding coefficient matrix β**



$$\boldsymbol{\beta} = \begin{array}{c} \\ x_{11} \\ x_{12} \\ x_{21} \\ x_{22} \end{array} \begin{array}{c} \begin{array}{cccc} x_{11} & x_{12} & x_{21} & x_{22} \end{array} \\ \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ \beta^{21}_{11} & \beta^{21}_{12} & 0 & 0 \\ 0 & 0 & \beta^{22}_{21} & 0 \end{bmatrix} \end{array}$$

[3] The coefficients of this model can be biased due to endogeneity caused by undirected edges in the CPN. We find that the proportion of undirected edges in our CPNs is small and allows us to obtain good fit results. We also utilize an alternative approach of forcing edges to be oriented by comparing the statistical significance of each directed edge and selecting the direction with the higher significance. We adopt this approach in §7 of the paper for assortment optimization.

4. Numerical Validation with Synthetic Data

In real-life data, the counterfactual is not observed and the true causal graph is unknown. Therefore, in this section, we present a simulation study to validate the ability of our proposed method consisting of the PC algorithm and the corresponding SEM for recovering the true model of causal relationships and their strengths from product purchases.

Our analysis proceeds in two steps. First, we synthetically generate basket data for a large set of causal graphs with varying numbers of products and categories. Recall that m represents the total number of categories and n_j denotes the number of products in category j, where $j \in \{1, \dots, m\}$. We randomly construct 100 causal product networks with 15 to 50 products in each graph as follows: m is uniformly distributed between 3 and 5, n_j is uniformly distributed between 5 and 10 for each category j, and we vary graph densities by using edge inclusion probability that is uniformly distributed between 0.1 and 0.5. For each given causal product network, we assume that the purchases within a basket are driven by (*i*) base purchase rates and (*ii*) the underlying causal product network and the corresponding edge coefficients that represent the strength of these relations. Specifically, we assume that the decision to purchase a product follows a Bernoulli distribution, where the purchase probability is the sum of its base purchase rate and the effects from its purchased parent products as shown in (1). We simulate synthetic basket-shopping data for each graph with the number of baskets varying uniformly between 1000 and 10,000. In the second step, we utilize the simulated data to learn the causal network using the PC algorithm and estimate parameters using the SEM, allowing us to compare the estimated causal graph and parameters directly with the true values. These steps are summarized in Algorithm 3.

Table 1 assesses the performance of the PC algorithm across these 100 randomly generated causal product networks by comparing its output to the true causal graphs. In skeleton discovery, we assess the algorithm's ability to accurately identify the *existence* and *non-existence* of edges between variables. In edge orientation, we evaluate how precisely the algorithm determines the directions of the edges.

The results in Table 1 indicate that the true positive rate and true negative rate for identifying the skeleton of causal structure among product purchases are 97.37% and 98.80%, respectively, demonstrating the algorithm's accuracy in discovering the skeleton. The percentage of edges that were not discovered is 2.63%, and the percentage of non-existing edges that were incorrectly identified is 1.20%. Turning to edge orientation, the last two columns of Table 1 show the final result after completing this phase. We find that 82.35% of the true edges are oriented correctly, 7.38% are oriented incorrectly, and the rest 7.65% are underdetermined, i.e., these edges were discovered during the skeleton discovery phase but their orientation could not be determined in the edge orientation phase. The decline in true positive rate from 97.37% to 82.35% after edge orientation is due to the

Algorithm 3: Simulation of Basket Shopping

Input: Number of simulations $S = 100$.
Output: $\mathcal{G}_s$, $\bar{\mathcal{G}}_s$ for $s \in \{1, \ldots, S\}$, err_α, and err_β.
for $s \in \{1, \ldots, S\}$:
  Constructing a random causal product network $\mathcal{G}_s$
  Sample total number of categories m_s from $U(3, 5)$.
  Sample total number of products n_{js} for each category $j \in \{1, \ldots, m_s\}$ from $U(5, 10)$.
  Define the total number of products N as $N = \sum_{j \in \{1,\ldots,m_s\}} n_{js}$.
  Sample edge inclusion probability p_s from $U(0.1, 0.5)$.
  Construct a random DAG $\mathcal{G}_s$ with N nodes using the edge inclusion probability p_s.
  Construct the parent set par_{ijs} for each product purchase using the graph $\mathcal{G}_s$.
  Generating model parameters over $\mathcal{G}_s$:
  for $i \in \{1, \ldots, n_{js}\}, j \in \{1, \ldots, m_s\}$:
    Sample α_{ijs} from $U(0, 0.5)$.
    Sample $\beta^{ij}_{i'j's} \sim U(0, 1)$ for $x_{i'j'} \in par_{ijs}$ such that $\sum_{x_{i'j'} \in par_{ijs}} \beta^{ij}_{i'j's} \le 1 - \alpha_{ijs}$.
    Set $\beta^{ij}_{i'j's} = 0$ for $x_{i'j's} \notin par_{ijs}$.
  Generating basket-shopping data:
  Sample number of baskets T from $U(10^3, 10^4)$.
  Generate basket-shopping data $x^t_{ijs} \sim Bernoulli(a_{ijs} + \sum_{x_{i'j'} \in par_{ijs}} \beta^{ij}_{i'j's} x^t_{i'j's})$
  for $i \in \{1, \ldots, n_{js}\}, j \in \{1, \ldots, m_s\}, t \in T,$.
  Graph learning and estimation of model parameters:
  Run the PC algorithm over the generated data. Let $\bar{\mathcal{G}}_s$ denote the learned graph.
  Estimate α_{ijs} and $\beta^{ij}_{i'j's}$ for $i \in \{1, \ldots, n_{js}\}, j \in \{1, \ldots, m_s\}$ using Equation (1) over
  the generated data and graph $\bar{\mathcal{G}}_s$. Let $\bar{\alpha}_{ijs}$ and $\bar{\beta}^{ij}_{i'j's}$ denote these estimates.
Compute estimation errors:
Compute $err_\alpha = \left\{(\alpha_{ijs} - \bar{\alpha}_{ijs})/\alpha_{ijs}, \text{ for } i \in \{1, \ldots, n_{js}\}, j \in \{1, \ldots, m_s\}, s \in \{1, \ldots, S\}\right\}$.
Compute $err_\beta = \left\{(\beta^{ij}_{i'j's} - \bar{\beta}^{ij}_{i'j's})/\beta^{ij}_{i'j's}, \text{ for } i, i' \in \{1, \ldots, n_{js}\}, j, j' \in \{1, \ldots, m_s\}, s \in \{1, \ldots, S\}\right\}$.
Return $\mathcal{G}_s$, $\bar{\mathcal{G}}_s$ for $s \in \{1, \ldots, S\}$, err_α, and err_β.

challenge in causal discovery in converting undirected edges from the skeleton into directed edges, as not all causal relationships can be unambiguously determined from observational data due to Markov equivalence classes as discussed in §3.1 and the limitations of data such as sample size.

Table 1 PC algorithm performance in skeleton discovery and edge orientation

True Edge	Discovered Edge	Skeleton Discovery Result (%)	Edge Orientation	Edge Orientation Result (%)
Exists	Discovered	97.37	Oriented Correctly	82.35
			Oriented Incorrectly	7.38
			Underdetermined	7.65
	Not Discovered	2.63		
Does not exist	Discovered	1.20	Oriented	1.18
			Underdetermined	0.03
	Not Discovered	98.80		

Note: The percentages represent the proportion of edges in each case, given whether the true edge exists or does not exist.

After evaluating the PC algorithm's accuracy for causal structure identification within basket-shopping data, we shift our focus to assessing the performance of parameter estimation using the SEM constructed from the graphs discovered by the PC algorithm. Let $\bar{\mathcal{G}}_s$ denote the graph discovered in simulation scenario s and $\overline{par}_{ijs}$ denote the parents of product purchase x_{ij} in graph $\bar{\mathcal{G}}_s$. We use $\bar{\cdot}$ notation to distinguish estimates from the true values. The SEM corresponding to the graph $\bar{\mathcal{G}}_s$ is formalized using (1) as follows:

$$x_{ijs} = \bar{\alpha}_{ijs} + \sum_{x_{i'j'} \in \overline{par}_{ijs}} \bar{\beta}^{ij}_{i'j's} x_{i'j's} + \epsilon_{ijs}, \quad \forall i \in \{1, \ldots, n_{js}\}, j \in \{1, \ldots, m_s\}. \tag{2}$$

Note that the structure of the discovered graph is used to directly determine the form of the simultaneous equations. Upon estimating this model, we compare the parameters' estimates against the true values to determine their accuracy. For this purpose, we define err_α to store the normalized errors between the estimated and true values of base purchase rates for all products across the 100 scenarios corresponding to the 100 random graphs, defined as $err_\alpha = \{(\alpha_{ijs} - \bar{\alpha}_{ijs})/\alpha_{ijs}, \text{ for } i \in \{1, \ldots, n_{js}\}, j \in \{1, \ldots, m_s\}, s \in \{1, \ldots, S\}\}$. Similarly, err_β, defined as $err_\beta = \{(\beta^{ij}_{i'j's} - \bar{\beta}^{ij}_{i'j's})/\beta^{ij}_{i'j's}, \text{ for } i, i' \in \{1, \ldots, n_{js}\}, j, j' \in \{1, \ldots, m_s\}, s \in \{1, \ldots, S\}\}$, stores the normalized errors for causal effects of product purchases across all product pairs over the 100 scenarios.

To fully evaluate the value of causal graph learning, we introduce a benchmarking method in which the SEM is constructed using a different criterion: including all products whose sales are significantly correlated, regardless of the causal graph. Specifically, for each product i from category j, we define the set $corr_{ijs}$ consisting of all significantly correlated purchases of products at a significance level of 0.05 in scenario s. Then, the corresponding simultaneous equations model for this correlation-based structure is defined as:

$$x_{ijs} = \alpha^{corr}_{ijs} + \sum_{x_{i'j'} \in corr_{ijs}} \beta^{ijcorr}_{i'j's} x_{i'j's} + \epsilon_{ijs}, \quad \forall i \in \{1, \ldots, n_{js}\}, j \in \{1, \ldots, m_s\}. \tag{3}$$

Similar to err_α and err_β, we define err^{corr}_α and err^{corr}_β to represent the errors between the correlation-based estimates α^{corr}_{ijs} and $\beta^{ijcorr}_{i'j's}$ and the actual values α_{ijs} and $\beta^{ij}_{i'j's}$ in scenario s, respectively. Then, by comparing the results of the model (2) with respect to the model (3), we can evaluate the performance of causal graph learning for accurately discovering the true model.

We conduct the estimation and compute the statistics of err_α, err_β, err^{corr}_α, and err^{corr}_β across all the scenarios as follows: (i) when the true estimate $\beta^{ij}_{i'j's}$ equals zero because the true graph does not include the corresponding edge, the measure of $\beta^{ij}_{i'j's}$ is undefined due to the denominator being zero; (ii) additionally, when the estimated value $\bar{\beta}^{ij}_{i'j's}$ equals zero because the edge does not exist in our discovered graph while the true value is non-zero, the error term err_β equals -1, which corresponds to the 2.63% as shown in Table 1, (iii) in all other cases, the error measures can be calculated because both the true values and the estimated values are non-zero. In the following analysis, we will focus on the error measures for case (iii) only, as these cases provide more meaningful statistics for comparing approaches. The proportions of the other two cases are presented in Table 1.

Table 2 presents the statistical characteristics of the error terms across different estimators. The mean values of err_α and err_β are 1.70% and −0.03%, respectively, which are closer to zero compared to the corrected estimates err_α^{cor} and err_β^{cor} with mean values of −88.99% and −10.75%. This suggests that the estimates from the PC algorithm (err_α and err_β) are less biased than those from the correlation-based structure. The root mean squared errors (RMSE) values further support this finding, with err_α and err_β showing lower variability (322.60% and 13.98%) compared to err_α^{cor} and err_β^{cor} (815.99% and 21.10%). This indicates that the estimates from the PC algorithm are more consistent.

To assess the symmetry and tailedness of the distributions, we compute the skewness and kurtosis of these error terms. The skewness of err_β is 0.17, which is closer to zero than the skewness of err_β^{cor} (−1.79), suggesting that err_β is more symmetrically distributed. The kurtosis of err_β (18.87) and err_β^{cor} (6.71) are both higher than 3, indicating that both have heavier tails than the normal distribution. In contrast, the error measures associated with the base purchase rate (α), specifically err_α and err_α^{cor}, exhibit more significant skewness and heavier tails. The skewness for err_α and err_α^{cor} is 12.29 and −26.73, respectively, with kurtosis values of 982.36 and 763.69, indicating distributions with substantial skewness and extreme tails.

Recognizing the influence of outliers on these statistics, we identified and removed outliers that lie beyond three standard deviations from the mean. This process led to the detection of 6, 6, 76, and 90 outliers for err_α, err_α^{cor}, err_β, and err_β^{cor}, respectively, which represent approximately 0.02%, 0.02%, 1.99% and 2.13% for the corresponding α and β terms. After removing these outliers, the skewness of err_α and err_β significantly improved, moving closer to zero (3.92 and −0.08), compared to err_α^{cor} and err_β^{cor} (−8.43 and −0.96). Similarly, the kurtosis of err_α and err_β reduced to 51.10 and 2.76. These values are closer to 3 compared with err_α^{cor} and err_β^{cor} (125.29 and 1.69), indicating that err_α and err_β now exhibit distributions closer to normality than err_α^{cor} and err_β^{cor} respectively.

Apart from the statistics of the error measures, we also visualize their distributions. Figure 4 depicts the distributions of err_α and err_α^{corr}, and Figure 5 depicts the distributions of err_β and err_β^{corr}. Figure 4 shows that the values of err_α are centered close to zero, while the values of err_α^{corr} exhibit more bias as they have heavier tails and significant skew away from zero. Similarly, Figure 5 shows that the distribution of err_β closely resembles a normal distribution, whereas the values of err_β^{corr} are shifted from zero.

Table 2 Statistics of the error terms for different estimators

	All Estimates				Without Outliers			
Measures	err_α	err_α^{cor}	err_β	err_β^{cor}	err_α	err_α^{cor}	err_β	err_β^{cor}
Mean (%)	1.70	-88.99	-0.03	-10.75	-3.27	-56.58	-0.29	-9.15
RMSE (%)	322.60	815.99	13.98	21.10	50.86	121.22	10.35	17.03
Skewness	12.29	-26.73	0.17	-1.79	3.92	-8.43	-0.08	-0.96
Kurtosis	982.36	763.69	18.87	6.71	51.10	125.29	2.76	1.69
N	2822	2822	3821	4225	2816	2816	3745	4135

Note: (1) RMSE refers to the root mean squared error of the error terms; (2) N refers to the observation number; (3) Outliers are observations that lie beyond three standard deviations from the mean of the error measure.

Figure 4 Frequency distributions of err_α and err_α^{cor}.

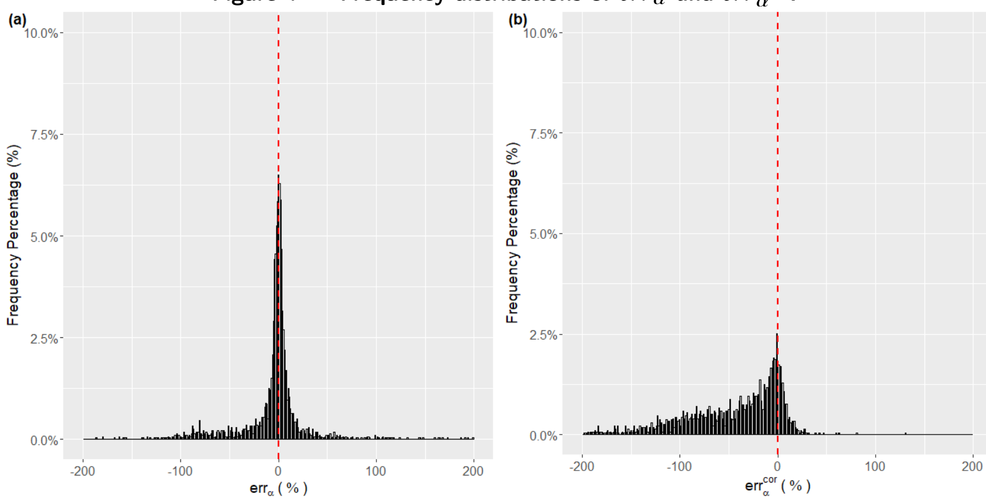
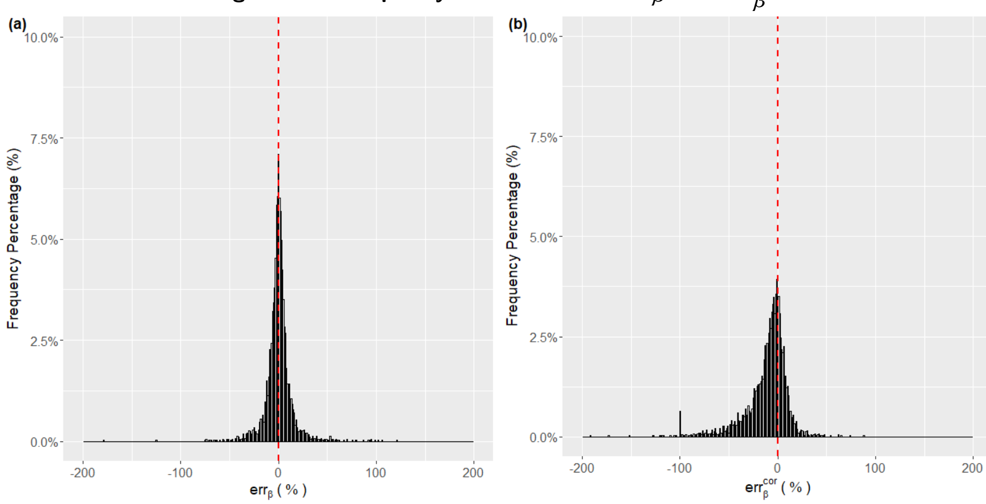


Figure 5 Frequency distributions of err_β and err_β^{cor}.



We conclude this analysis with two findings. First, the causal graph learning method is significantly superior to the correlations-based benchmark for recovering the true graph. Second, the PC algorithm and SEM are able to correctly identify the causal structure for 82.35% of the cases and accurately estimate their true parameters. Having demonstrated the PC algorithm's ability to accurately learn product networks and the importance of selecting the right specifications in the SEM for parameter estimation with synthetic data, we now turn our attention to evaluating our framework for real-world datasets and using it to generate insights.

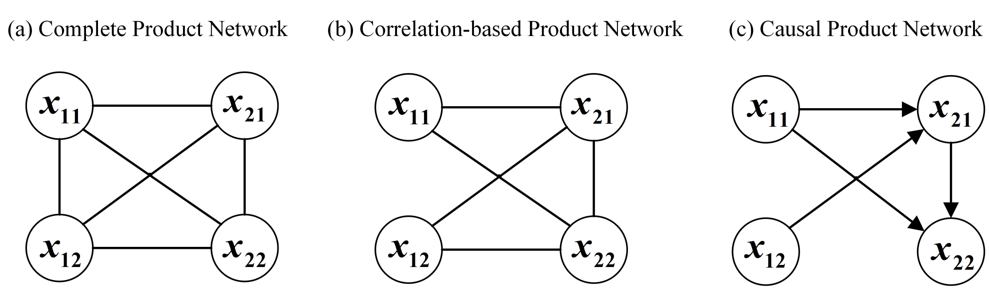
5. Hypothesis Development

In this section, we develop three hypotheses to characterize purchase decisions of basket-shopping consumers. The first hypothesis examines various product networks to model interactions among products in a shopping basket, with the goal of finding the most accurate representation of these interactions. The second hypothesis investigates the validity of various alternative specifications of causal relationships in basket shopping, including causal effects between individual products, from categories to products, from products to categories, and between categories. The third hypothesis examines the differences in consumer behavior between brick-and-mortar and online channels and focuses on characterizing the sparsity of causal connections among product purchases across these channels.

5.1. Product Interaction Networks

There are three alternative ways in which the interactions between products in a shopping basket can be represented. First, inspired by Tulabandhula et al. (2023), we model each product as interconnected with every other. We refer to this structure as the *complete network*. Next, we explore an interaction model in which each product is interconnected only with products that show significant correlation (i.e., p-value ≤ 0.05) with the same idea as the benchmarking model in §4. We refer to this as the *correlation network*. The last network we consider is a *causal* product network obtained through the PC algorithm.

Figure 6 **Different types of product interaction networks that can arise from the same purchase data by making different modeling assumptions**



Note: The products x_{11}, x_{12} and x_{21}, x_{22} belong to categories 1 and 2, respectively.

Figure 6 illustrates examples of these networks showing interactions among four products across two categories. Figure 6(a) represents the complete network; Figure 6(b) represents the correlation network omitting the edge between x_{11} and x_{22} due to the lack of significant correlation in the basket data; lastly, Figure 6(c) represents the causal network, removing the edge between x_{12} and x_{22} because even though they are correlated there is no direct causal link—this relationship is mediated by the path $x_{12} \rightarrow x_{21} \rightarrow x_{22}$. Unlike the first two, the causal network orients the edges to indicate the direction of relations, whereas the relationships in the other networks are symmetric.

Note that all three networks in Figure 6 can be obtained from the same given basket-shopping dataset in real-life. For instance, using only correlations may lead researchers to incorrectly infer a direct connection between purchases of x_{12} and x_{22}, when instead, this relation might be mediated by purchases of x_{21}, suggesting no direct link. The proposed causal discovery methodology can differentiate these types of indirect relationships. In other words, causal networks, unlike complete or correlation-based networks, can remove spurious connections between product purchases and provide directional relations. This results in a more accurate and parsimonious representation of purchase relationships. Thus, as the first application of our methodology, we compare these networks on our dataset and test the hypothesis that a causal network offers a more precise description of purchase relationships compared to complete and correlation-based networks (Hypothesis 1).

Hypothesis 1 (Network Structures) *A causal product network more accurately represents product purchase relationships in basket shopping compared to complete and correlation-based networks.*

To evaluate this hypothesis, we estimate all three models on our dataset as follows. The SEM for the complete network representation is given by:

$$x_{ij} = \alpha_{ij}^{comp} + \sum_{i' \in I_{j'}, j' \in J} \beta_{i'j'}^{ijcomp} x_{i'j'} + \epsilon_{ij}, \quad \forall i \in I_j, j \in J. \tag{4}$$

Let $corr_{ij}$ store product purchases significantly correlated with the purchase of product i from category j (i.e., p-value ≤ 0.05). The corresponding SEM for the correlation network representation is defined as:

$$x_{ij} = \alpha_{ij}^{corr} + \sum_{x_{i'j'} \in corr_{ij}} \beta_{i'j'}^{ijcorr} x_{i'j'} + \epsilon_{ij}, \quad \forall i \in I_j, j \in J. \tag{5}$$

Let $\mathcal{G} = (V, E)$ be the causal graph derived from the PC algorithm and par_{ij} store the parents of the purchase of product i from category j in graph $\mathcal{G}$. The SEM for the causal network representation is defined as:

$$x_{ij} = \alpha_{ij} + \sum_{x_{i'j'} \in par_{ij}} \beta_{i'j'}^{ij} x_{i'j'} + \epsilon_{ij}, \quad \forall i \in I_j, j \in J. \tag{6}$$

These models differ from each other only in the number of parameters. Thus, we assess which model provides a better fit to test Hypothesis 1.

5.2. Specifications for Causal Representations in Basket Shopping

Building on the hypothesis that a causal network more accurately represents basket-shopping behavior, we now investigate various specifications within causal structures. The prior literature suggests that there can be different kinds of assumptions regarding consumers' basket-shopping behavior. For example, Song and Chintagunta (2006) and Jasin et al. (2023) assume that interactions between two products from distinct categories depend solely on the categories to which they belong. Alternatively, we can generalize the interactions by considering those effects from products to categories and from categories to products. The former suggests that the effects from any product to different products within the same category are identical. The latter implies that the effects from all the products within a category to another product in a different category are identical. We can apply these restrictions to the simultaneous equations models and can empirically investigate which of these alternative specifications most effectively represents the data.

Figure 7 **Alternative causal product network specifications to Figure 6(c) with different types of imposed restrictions**

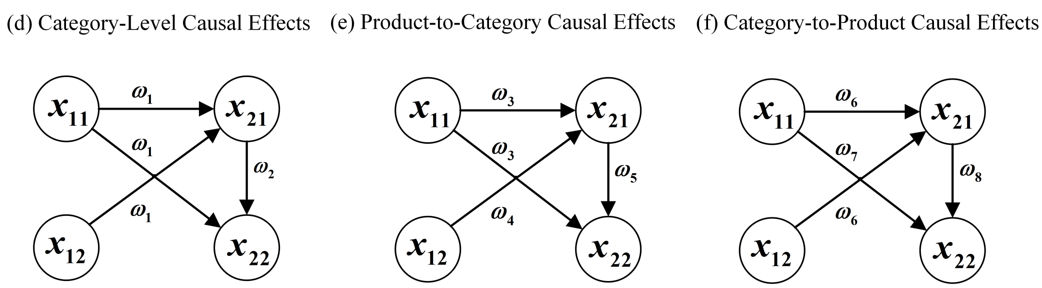


Figure 7 demonstrates these specifications using the causal network shown in Figure 6(c). The products x_{11},x_{12} and x_{21},x_{22} belong to categories 1 and 2, respectively. The parameters ω_p for $p = 1, 2, \ldots, 8$ correspond to different magnitudes of causal effects. Figures 7(d)-(f) show different conceptual representations of consumer behavior. Specifically, Figure 7(d) considers category-level specification, ensuring that the magnitude of causal effects from x_{11} to x_{21}, x_{11} to x_{22}, and x_{12} to x_{21} are identical. Figure 7(e) considers product-to-category specification, ensuring the magnitude of causal effects from x_{11} to x_{21} and x_{11} to x_{22} are identical. The last specification is depicted in Figure 7(f) where we restrict the magnitude of causal effects from x_{11} to x_{21} and x_{12} to x_{21} to be identical.

The study of these specifications is useful in understanding consumer shopping behavior because the identification of the correct model has implications for decision-making tasks such as assortment planning, pricing, and promotions. The four models, including Figure 6(c) and Figures 7(d)-(f), exhibit varying levels of generalization: considering causal effects between products represents the most general structure while considering causal effects between categories is the most restricted. These restrictions, if proven accurate, can have critical managerial implications. Additionally, they are also computationally relevant because the number of parameters to be estimated decreases with the restrictions. On the one hand, the more general models allow parameters to be tailored to individual products, but on the other hand, having more parameters might be computationally burdensome and might not provide the best model fit. Thus, balancing this tradeoff, we hypothesize that models capturing product-level interactions provide a more accurate representation of shopping data even when penalizing for model complexity (Hypothesis 2).

Hypothesis 2 (Specification Comparison) *Across the four causal specifications, the product-level causal effects model illustrated in Figure 6(c) most effectively describes the causal relationships in basket shopping (in terms of model fit) compared to alternatives illustrated in Figures 7(d)-(f). In other words, the restrictions imposed by the alternative models do not hold.*

We now formally outline the restrictions applied to the SEM for these specifications. Let $\mathcal{G} = (V, E)$ be the causal graph derived by the PC algorithm. First, we introduce the category-level specification. Among the existing causal edges in E, let $E^{j}_{j'} \subseteq E$ store the edges from products in category j' to products in category j, i.e., $E^{j}_{j'} = \{(x_{i'j'} \to x_{ij}) \in E,\ \forall i' \in I_{j'}, i \in I_j\}$. We restrict the coefficients of all the edges in $E^{j}_{j'}$ to be identical for any $j, j' \in J$. The corresponding SEM for this structure is defined as:

$$x_{ij} = \alpha_{ij} + \sum_{x_{i'j'} \in par_{ij}} \beta^{ij}_{i'j'} x_{i'j'} + \epsilon_{ij}, \quad \forall i \in I_j, j \in J \tag{7a}$$

$$\text{subject to} \quad \beta^{i_1 j}_{i'_1 j'} = \beta^{i_2 j}_{i'_2 j'}, \quad i_i \in I_j, i'_1 \in I_{j'}, j, j' \in J : (x_{i'_1 j'} \to x_{i_1 j}), (x_{i'_2 j'} \to x_{i_2 j}) \in E^{j}_{j'}. \tag{7b}$$

Next, we introduce the product-to-category specification. Let $E^{j}_{i'j'} \subseteq E$ store the edges from product i' in category j' to any product in category j, i.e.,$E^{j}_{i'j'} = \{(x_{i'j'} \to x_{ij}) \in E,\ \forall i \in I_j\}$. We restrict the coefficients of all the edges in $E^{j}_{i'j'}$ to be identical for any $i' \in I_{j'}, j, j' \in J$. The corresponding SEM for this structure is defined as:

$$x_{ij} = \alpha_{ij} + \sum_{x_{i'j'} \in par_{ij}} \beta^{ij}_{i'j'} x_{i'j'} + \epsilon_{ij}, \quad \forall i \in I_j, j \in J \tag{8a}$$

$$\text{subject to} \quad \beta^{i_1 j}_{i'j'} = \beta^{i_2 j}_{i'j'} \quad i_1, i_2 \in I_j, i' \in I_{j'}, j, j' \in J : (x_{i'j'} \to x_{i_1 j}), (x_{i'j'} \to x_{i_2 j}) \in E^{j}_{i'j'}. \tag{8b}$$

Lastly, we introduce the category-to-product specification. Let $E^{ij}_{j'} \subseteq E$ store the edges from any product in category j' to product i from category j, i.e., $E^{ij}_{j'} = \{(x_{i'j'} \to x_{ij}) \in E,\ \forall i' \in I_{j'}\}$. We restrict the coefficients of all the edges in $E^{ij}_{j'}$ to be identical for any $i \in I_j, j, j' \in J$. The corresponding SEM for this structure is defined as:

$$x_{ij} = \alpha_{ij} + \sum_{x_{i'j'} \in par_{ij}} \beta^{ij}_{i'j'} x_{i'j'} + \epsilon_{ij}, \quad \forall i \in I_j, j \in J \tag{9a}$$

$$\text{subject to} \quad \beta^{ij}_{i'_1 j'} = \beta^{ij}_{i'_2 j'}, \quad i \in I_j, i'_1, i'_2 \in I_{j'}, j, j' \in J : (x_{i'_1 j'} \to x_{ij}), (x_{i'_2 j'} \to x_{ij}) \in E^{ij}_{j'}. \tag{9b}$$

These SEMs correspond to different restrictions on the causal effects in Figure 6(c) as follows: Figure 7(d) visualizes the specification in Equation (7) where the causal effects between categories are identical, i.e., $\beta^{22}_{11} = \beta^{21}_{12} = \omega_1$; Figure 7(e) corresponds to Equation (8) where the causal effects from any product to a category are identical, i.e., $\beta^{21}_{11} = \beta^{22}_{11} = \omega_3$; Figure 7(f) corresponds to Equation (9) where the causal effects from any category to a product are identical, i.e., $\beta^{21}_{11} = \beta^{21}_{12} = \omega_6$. Thus, the three specifications are applied as restrictions to the product-level model presented in Equation (6). This enables us to test Hypothesis 2 using a likelihood ratio test or AIC criterion.

5.3. Sparsity of Causal Product Networks Across Channels

The power of our causal network learning approach for archival data is that it can be not only used to construct evidence regarding different model specifications but also applied separately to different retailing contexts to analyze differences in basket-shopping behavior of customers. We now turn our attention to investigating differences in these networks across brick-and-mortar and online channels. Indeed, the prior literature has examined assortment optimization problems across these two channels. For example, Lo and Topaloglu (2022) examines an assortment problem for a physical store aimed at maximizing the total expected revenue of an omnichannel retailer. Gopalakrishnan et al. (2023) compare the impacts of assortment width, defined as the number of unique categories in which a retailer offers products, and assortment depth, defined as the average number of products offered within each category, on order delivery timeliness. Sapra and Kumar (2023) examine the joint assortment optimization strategy for omnichannel retailers, taking into account three classes of customers who use only the brick-and-mortar channel, only the online channel, or both channels.

The models in the above literature are based on differences in both the cost economics and consumer behavior across the two channels. With regard to empirical evidence for the latter, Chintala et al. (2023) show that there are systematic differences between online and offline grocery shopping. Specifically, online purchases have significantly lower shopping basket variety than brick-and-mortar. However, they do not study causal relationships in consumer purchase behavior and highlight the need for considering causal relationships. Building on this, we hypothesize that the online channel exhibits fewer causal relationships among product purchases within shopping baskets compared to the brick-and-mortar channel (Hypothesis 3). We believe that this difference in sparsity can provide insights for omnichannel assortment optimization by revealing how consumer behavior varies across online and offline environments.

Hypothesis 3 (Sparsity) *Compared to the brick-and-mortar channel, the online channel exhibits fewer causal relationships among product purchases within shopping baskets.*

Our method to test Hypothesis 3 differs slightly from the previous hypotheses. We first construct the causal graphs separately for online and brick-and-mortar baskets; then instead of estimating the SEM model, we compare the resulting graphs with respect to the distribution of edges between pairs of nodes.

6. Empirical Evidence

In this section, we introduce our dataset, describe the sample construction process, and provide statistical summaries of the samples in §6.1. We then empirically test the proposed hypotheses and present the results of our analysis in §6.2.

6.1. Data

We collaborate with Numerator, a prominent market research company known for its first-party, consumer-sourced data. Numerator gathers purchase data from both brick-and-mortar and online channels for a large panel of consumers through multiple sources: retailer loyalty data, mobile app purchases, email receipts, as well as uploaded paper receipts. Their extensive database comprises customer purchase records from more than 17,000 stores, encompassing major retail chains such as Walmart, Costco, and Target, as well as small supermarkets and local grocery stores across the United States. We use their dataset for the year 2021.

The first question we face is how to utilize all available data to test our hypotheses. Learning causal graphs is known to be NP-hard (Chickering et al. 2004). A key challenge that we face in this analysis is that the amount of basket-shopping data is too large and too sparse to allow efficient application of the PC algorithm. To address this problem, we construct many smaller subsamples of the data, estimate the DAG for each sample, and statistically determine the most salient relationships as described below. Another question is how to maintain consistency with the causal sufficiency assumption (Assumption 3). Theoretically, all the product information in the real world should be included in the estimation model, but including millions of products increases the computational cost exponentially. To address this issue, we select five different themes for our analysis: each theme is a collection of multiple representative departments and each department includes several products. For each theme, we consider the most frequently purchased products in each selected department to represent shopping baskets as these products are representative of shopping baskets. Further, to include information about other products within and outside these departments in the product networks, we create aggregated control variables for the total amount of purchases of other products within each selected department and of products outside the selected departments. These control variables help account for information about other product purchases in shopping baskets, thereby supporting the causal sufficiency assumption in our empirical study. This design of our study provides us with sufficient information to construct and analyze product networks for the five themes and demonstrate the usefulness of this approach.

Specifically, we focus on the following five themes: pasta, quick service restaurant (QSR), bakery, prepared food, and unprepared food, labeled sequentially as Themes 1 through 5. We chose these themes based on two criteria: (1) items within each theme are available in both brick and mortar and online shopping channels, and (2) the themes are distinct enough to test the hypotheses across a range of scenarios. Theme 1 comprises five departments: pasta & noodle, meat, produce, dairy, and condiments. Theme 2 includes ten departments: QSR beverages, QSR breakfast, QSR sandwiches & wraps, QSR sauces & condiments, QSR Mexican, QSR snack & Sides, QSR desserts, QSR entrees, QSR Italian, and QSR salads. Theme 3 comprises five departments: bakery sweet goods, in-store bakery, packaged bakery, baking & cooking and dairy. Theme 4 includes five departments: frozen foods, canned, deli & prepared foods, beverages, and condiments. Theme 5 comprises five departments: produce, shelf stable meals, meat, herbs & spices, and condiments. Note that product category and department are used synonymously in our data set. The lists of the most frequently purchased product sets for the five themes across both brick-and-mortar and online channels, in the brick-and-mortar channel only, and in the online channel only are shown in Appendix B.

To test Hypotheses 1 and 2, we extract and analyze basket data for each theme separately, including purchases from both brick-and-mortar and online channels, as follows. For each theme, we identify the most frequently purchased products within each department in 2021 combined across both channels and choose a total of 50 products. In Themes 1, 3, 4, and 5, which consist of five departments each, we select the top 10 products from each department. In Theme 2, which includes ten departments, we select the top 5 products from each department. After establishing the product set for each theme using the entire dataset, we retrieve all the shopping basket data, including both online and brick-and-mortar channels, for 500 randomly chosen customers who had purchased products from at least three departments within the theme in a single basket. We refer to this collection of datasets as Data I, which comprises five datasets, each corresponding to a different theme.

To test Hypothesis 3, since the most frequently purchased products differ between the two channels, we consider two product sets for each channel: the most frequently purchased products in brick-and-mortar stores and the most frequently purchased products online. We refer to these as the top brick-and-mortar products and top online products, respectively. For each product set, we retrieve baskets for each theme, separating brick-and-mortar and online purchases, termed brick-and-mortar baskets and online baskets. Thus, we create different datasets for four scenarios for each theme: (1) baskets constructed from top brick-and-mortar product sets using brick-and-mortar purchases, (2) baskets constructed from top brick-and-mortar product sets using online purchases, (3) baskets constructed from top online product sets using online purchases, and (4) baskets constructed from top online product sets using brick-and-mortar purchases. We refer to this collection of datasets as Data II, which comprises twenty datasets (= 4 scenarios × 5 themes). By considering the most popular products from both brick-and-mortar stores and online platforms and analyzing purchases made through both these channels, we aim to show that the differences between brick-and-mortar purchases and online purchases are consistent across the datasets regardless of the type of products being considered.

Tables 3(a) and 3(b) present the statistical summaries for the total number of products purchased per basket for Data I and Data II, respectively. As shown in Table 3(a), the median number of purchased items is the same across the five themes, while the mean and the maximum quantity varies, ranging from 29 to 236 items. In Table 3(b), the median quantity of purchased items in B&M baskets remains at 2, whereas the median quantity in online baskets is higher. On average, online baskets also contain more items compared to B&M baskets. Additionally, the number of baskets for each theme varies between 5,000 and 15,000 across Data I and Data II, with the exception of online Theme 2 (Quick service restaurants theme) sample. Note that the sample size for Theme 2 in the online channel is significantly smaller, with only 189 baskets for the top brick-and-mortar channel products and 372 baskets for the top online channel products. We believe this is due to the nature of QSRs, which make it attractive for customers to make purchases in person rather than waiting for online orders and delivery.

6.2. Testing Hypotheses 1 and 2

For testing Hypothesis 1, we need to generate the causal product network, the complete graph, and the correlation-based graph for each theme. Using Data I, we first construct causal product networks for each theme using the PC algorithm as detailed in Algorithms 1 and 2, and using a significance level of 0.05 for partial correlations to test for conditional independence. To improve the reliability of the learned causal graphs, we resample the data with replacement and construct the final graph based on edge frequencies derived from these samples as proposed in the literature (Friedman et al. 1999, Imoto et al. 2002, Mooij et al. 2020). For each theme, we create 50 subsamples, each consisting of all the orders from 100 randomly selected customers, and apply the PC algorithm to discover a causal graph for each sample.[4] The final causal graph is constructed by

[4] We start with our customer pool of 500 randomly selected customers. However, constructing causal graphs for all their baskets was computationally intensive, taking several hours, and had to be repeated for 5 themes, each 50 times. To decrease runtime, we switched to subsampling and focused on orders from 100 randomly selected customers from the pool.

Table 3 Summary statistics of shopping baskets in Data I and Data II

(a) Data I, used for testing Hypotheses 1 and 2

	Total Quantity of Items in Each Basket							
	Min	25%	50%	75%	Max	Mean	SD	N
Theme 1 (Pasta)	1	1	2	4	29	3.02	2.96	14,689
Theme 2 (QSR)	1	1	2	3	105	2.74	2.84	8,858
Theme 3 (Bakery)	1	1	2	3	28	2.51	2.16	7,332
Theme 4 (Prepared Food)	1	1	2	4	61	2.78	2.70	8,345
Theme 5 (Unprepared Food)	1	1	2	4	236	3.17	4.34	13,586

(b) Data II, used for testing Hypothesis 3

Top Brick-and-Mortar Products																
	Brick-and-Mortar Channel								Online Channel							
	Min	25%	50%	75%	Max	Mean	SD	N	Min	25%	50%	75%	Max	Mean	SD	N
Theme 1	1	1	2	4	29	2.92	2.61	14,472	1	2	4	6	27	4.80	3.67	6,664
Theme 2	1	1	2	3	47	2.75	2.57	9,247	1	1	2	3	48	3.78	5.45	189
Theme 3	1	1	2	3	41	2.53	2.12	7,252	1	2	3	4	25	3.39	2.39	6,125
Theme 4	1	1	2	3	58	2.67	2.75	6,684	1	1	3	4	60	3.50	3.30	5,563
Theme 5	1	1	2	4	170	3.16	3.77	12,901	1	2	4	6	46	4.81	4.07	5,914
Top Online Products																
	Brick-and-Mortar Channel								Online Channel							
	Min	25%	50%	75%	Max	Mean	SD	N	Min	25%	50%	75%	Max	Mean	SD	N
Theme 1	1	1	2	4	258	2.89	4.10	13,023	1	2	5	8	62	5.78	4.54	6,613
Theme 2	1	1	2	3	44	2.52	2.29	4,907	1	1	2	3	50	2.90	3.98	372
Theme 3	1	1	2	3	72	2.73	2.53	9,039	1	2	3	5	24	3.69	2.66	6,502
Theme 4	1	1	2	4	39	2.79	2.77	5,871	1	2	3	5	53	4.02	3.64	6,240
Theme 5	1	1	2	4	165	2.93	3.09	13,056	1	2	4	7	35	5.39	4.42	6,340

Note: (1) Themes 1-5 correspond to Pasta, Quick Service Restaurant (QSR), Bakery, Prepared Food Items, and Unprepared Food Items, respectively, where each theme consists of several product categories; (2) N refers to the total number of baskets; (3) SD refers to the standard deviation of product quantity in shopping baskets; (4) The size of product sets for all the themes is 50; (5) Quantity of items refers to the total number of products included within our product set.

including edges that appear in more than 30% of the networks. Both the implementation of the PC algorithm and the testing for independence were conducted using the *pcalg* package (Markus Kalisch et al. 2012). We then generate the complete product network for each theme in Data I by simply connecting all products by undirected edges. In the correlation-based product network, we determine edges between nodes through pairwise Pearson correlation tests at the 0.05 significance level. Having thus obtained all three graphs, we build simultaneous equations models for the complete product network, the correlation-based product network, and the causal product network.

To test Hypothesis 2, we begin our analysis with the final causal graph obtained above. To this graph, we impose the category-to-category, product-to-category, and category-to-product specifications in the corresponding simultaneous equations models as discussed in §5.2.

Table 4 presents the model fit results with labels (a)-(f) denoting the six configurations in Hypotheses 1 and 2, as depicted in Figures 6 and 7. The top half of the table shows comparative analysis using Akaike Information Criterion (AIC) scores, which balances the number of parameters and model fit, and the bottom half provides insights into the number of parameters in each model. Note that each estimate is based on an SEM consisting of 50 equations for 50 products. The number of independent variables varies across the specifications and is also different in Theme 2 since it consists of 10 categories.

Comparing the results for (a), (b), and (c) across all themes in Table 4, we find that the causal product network consistently shows the lowest AIC scores. This indicates that the causal networks provide a more accurate representation of product relationships in shopping baskets compared to complete and correlation-based networks, confirming Hypothesis 1. Additionally, comparing the AIC score of (c) with the three alternative causal specifications (d), (e), and (f), we find that the product-level causal model consistently shows the best scores. This shows that it most effectively captures the causal relationships in basket shopping and the restrictions imposed by the other models do not hold, substantiating Hypothesis 2. This result provides us with an important finding that consumers' basket-shopping behavior is driven by product-level causality, not category-level interactions, which is relevant for the design of choice models for basket-shopping behavior. Moreover, note that there is a very highly significant difference in the AIC scores of model (c) versus all the other models, demonstrating the quality of the fit.

Turning to the bottom half of the table, causal product networks require fewer parameters, thereby reducing estimation complexity as well as the potential for multicollinearity; e.g., for the pasta theme, the complete network has 2,750 parameters to be estimated, the correlation-based network has 693, and the causal product network has 56. Further, we find that the number of edges in the product-level causal model is surprisingly only marginally larger than that in the alternative restriction-based specifications, e.g., for the pasta theme, these numbers are 56 parameters for the causal product network, 14 for the category-level specification, 42 for the product-to-category specification, and 39 for the category-to-product specification. We make two inferences from these results: (i) the number of causal relationships in basket-shopping behavior is not too large, making it computationally tractable and managerially feasible to analyze and manage these relationships, and (ii) the correct specification of these relationships is critical for model fit, which should materially affect decision-making tasks based on choice modeling.

6.3. Testing Hypothesis 3

To test Hypothesis 3, we create causal product networks separately for each of the 20 datasets in Dataset II representing both brick-and-mortar and online channels. Following the same approach as described in the previous subsection, for each theme and product set, we extract 50 subsamples from the corresponding dataset, each consisting of orders from 100 randomly selected customers. We then use the PC algorithm to discover the causal graph for each sample.

Table 4 Model fit results for the six network structures in Hypotheses 1 and 2

AIC Scores						
Network Structure	(a)	(b)	(c)	(d)	(e)	(f)
Theme 1 (Pasta)	669,880	668,745	315,548	699,523	698,639	690,543
Theme 2 (QSR)	449,416	447,761	364,352	480,381	475,279	470,493
Theme 3 (Bakery)	178,631	177,365	147,093	197,965	196,526	189,143
Theme 4 (Prepared Food)	403,995	409,765	272,623	445,457	444,797	443,339
Theme 5 (Unprepared Food)	1,054,567	1,052,783	531,967	1,135,675	1,132,544	1,117,816
Number of Parameters						
Network Structure	(a)	(b)	(c)	(d)	(e)	(f)
Theme 1 (Pasta)	2,750	693	56	14	42	39
Theme 2 (QSR)	3,000	842	48	22	42	39
Theme 3 (Bakery)	2,750	496	53	14	42	42
Theme 4 (Prepared Food)	2,750	520	63	15	52	49
Theme 5 (Unprepared Food)	2,750	493	36	13	29	32

Note: (1) The size of product sets for all the themes is 50; (2) The numbers of baskets are 14689, 8858, 7332, 8345, and 13586 for themes 1-5, respectively.

To compare the sparsity of the brick-and-mortar and online causal product networks statistically, we used a paired t-test. Each pair in the test represents the frequency of occurrence of a specific directed edge in the brick-and-mortar causal network and the frequency of the same directed edge in the online causal network across the 50 samples. We exclude edges that were never found to be present in either the brick-and-mortar or online settings in any of the samples. This exclusion ensures that the analysis focuses only on relevant edges that were observed at least once.

Table 5 Average frequency of occurrence of directed edges across 50 simulations: Test for Hypothesis 3

	Top B&M Products				**Top Online Products**			
	B&M	Online	Difference	Num. of Pairs	B&M	Online	Difference	Num. of Pairs
Theme 1 (Pasta)	4.40	3.71	0.87*** (8.20)	1322	5.75	5.34	0.41** (7.70)	1379
Theme 2 (QSR)	5.80	0.80	6.99*** (14.91)	618	10.87	2.93	7.94*** (15.63)	432
Theme 3 (Bakery)	3.16	2.80	0.69*** (7.03)	982	4.67	4.71	-0.05 (6.65)	1201
Theme 4 (Prepared Food)	3.52	2.96	0.46** (8.65)	1072	5.04	4.73	0.31* (8.04)	1193
Theme 5 (Unprepared Food)	3.90	3.08	0.90*** (6.95)	1308	4.95	4.46	0.48*** (6.99)	1392

Note: (1) *p<0.1; **p<0.05; ***p<0.01; (2) The values for B&M, Online, and Difference correspond to the average occurrence frequencies of a directed edge across 50 simulations for B&M, Online, and the B&M-Online comparison, respectively; (3) The number of observations represents the number of directed edges that appear at least once in B&M or Online causal structures.

The results of the paired t-test are presented in Table 5. We show the average frequency of directed edge occurrences across 50 simulations for both brick-and-mortar and online channels, along with the differences between these frequencies and the total number of edge pairs (i.e., the number of directed edges that appear at least once in either the brick-and-mortar or online causal structures across 50 subsamples). Our results show that the online channel has a smaller number of edges than the brick-and-mortar channel in 9/10 cases; six of these are statistically significant at a 0.01 significance level, two at 0.05, and one at 0.1. For Theme 3, there is a significant difference when top brick-and-mortar products are considered, but not when the top online products are considered. Overall, these results support our hypothesis that online causal product networks are sparser than their brick-and-mortar counterparts. This implies that purchases in a basket in the online channel are more likely to be independent of each other, with limited complementary and substitution effects, than those in the brick-and-mortar channel. Thus, assortments and inventories for online channels can be planned independently for products more easily than for the brick-and-mortar channel. Moreover, this also suggests that there is a greater need for recommendation systems and cross-promotions to increase basket size in online channel than in the brick-and-mortar channel.

Robustness. We conduct robustness tests for Hypotheses 1 and 2 by separately analyzing the B&M and online channels for all five themes to verify the consistency of results for these hypotheses across both channels. Our results are presented in Appendix C and support both hypotheses, except for the online channel in Theme 2, where insufficient data was available.

7. Assortment Optimization

In this section, we assess the value of modeling basket-shopping behavior for assortment optimization. For this, we construct and compare optimal assortments across two scenarios: when demand is given by our causal product networks method, and when it is assumed to follow a traditional multinomial logit model that considers choice within categories but treats each category as independent. We train and evaluate both choice models on our Numerator dataset. Arguably, if causal modeling more accurately captures consumer behavior, it should result in better assortment decisions than a model that only considers choice behavior within a category and ignores downstream effects across categories. Thus, we seek to estimate the difference in performance between these two solution approaches to quantify the value of modeling basket-shopping behavior.

We conduct this study using Data II for the prepared food theme (Theme 4; see § 6.1 for details) considering the B&M and online channels separately. For the B&M channel, the product set consists of the top B&M products, while for the online channel, the product set comprises the top online products. Thus, our approach consists of the following steps. First, using the method presented in §6, we construct causal product networks using the PC algorithm across 50 subsamples and include a causal edge if it appears in more than 30% of these subsamples. To ensure that the final causal product network is free of cycles, we prioritize the inclusion of causal edges based on their frequency of appearance across the subsamples, starting with the most frequent. Second, using simultaneous equations modeling, we then estimate the base purchase probability for each product and the strengths of causal effects within the constructed product network. After this step, we have a complete specification of the CPN-based purchase model, which serves as input for assortment optimization. Third, we propose a mixed-integer program to optimize product assortments for basket shopping considering the discovered causal product networks. Finally, using the same dataset, we estimate a traditional MNL model for each product category, use it to optimize the assortment under MNL and compare the optimization results under MNL with those under causal graphs.

Our mixed-integer programming model aims to select the optimal product assortment that maximizes total demand by considering the individual demand for each product and the effect of each product's purchase on others within the discovered causal network. We use total demand as the objective function instead of sales revenue or profit since our dataset does not include cost and price values. Nevertheless, our method can be easily adapted to include these parameters when they are available.

Let y_{ij} indicate the inclusion of a product in the assortment, where $y_{ij} = 1$ if and only if product i from category j is included in the assortment. Let d_{ij} represent the total demand rate for product i from category j. Additionally, we define z_{ij} as a dummy binary variable and M as a large positive constant to make sure demand rate d_{ij} remains non-negative. Lastly, we define K as the parameter representing the maximum number of products included in the assortment. We denote the problem of finding optimal assortment of size K given a causal product network $\mathcal{G} = (V, E)$ and edge coefficients $\beta^{ij}_{i'j'}$ for $(x_{i'j'} \to x_{ij}) \in E$ as CPN-MIP$(\mathcal{G}, \boldsymbol{\beta})$ and formulate it as follows:

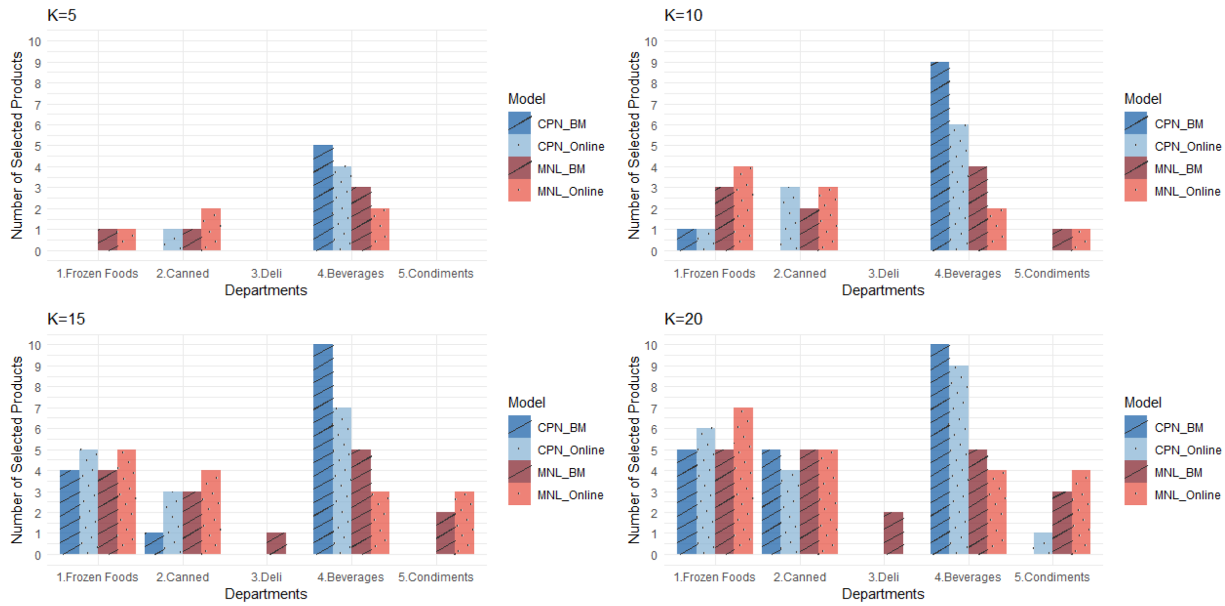
$$\text{CPN-MIP}(\mathcal{G},\boldsymbol{\beta}): \quad \begin{aligned}
\underset{\mathbf{y},\mathbf{d},\mathbf{z}}{\text{maximize}} \quad & \sum_{i\in I_j, j\in J} d_{ij} && \text{(10a)}\\
\text{subject to} \quad & d_{ij} \le a_{ij}\cdot y_{ij} + \sum_{(i',j')\in par_{ij}} \beta^{ij}_{i'j'} \cdot d_{i'j'} + M\cdot(1-z_{ij}), \quad \forall i\in I_j, j\in J, && \text{(10b)}\\
& a_{ij}\cdot y_{ij} + \sum_{(i',j')\in par_{ij}} \beta^{ij}_{i'j'} \cdot d_{i'j'} \le M\cdot z_{ij}, \quad \forall i\in I_j, j\in J, && \text{(10c)}\\
& a_{ij}\cdot y_{ij} + \sum_{(i',j')\in par_{ij}} \beta^{ij}_{i'j'} \cdot d_{i'j'} \ge -M\cdot(1-z_{ij}), \quad \forall i\in I_j, j\in J, && \text{(10d)}\\
& y_{ij} \le z_{ij}, \quad \forall i\in I_j, j\in J, && \text{(10e)}\\
& d_{ij} \le M\cdot y_{ij}, \quad \forall i\in I_j, j\in J, && \text{(10f)}\\
& \sum_{i\in I_j, j\in J} y_{ij} \le K, && \text{(10g)}\\
& y_{ij} \in \{0,1\}, \quad \forall i\in I_j, j\in J, && \text{(10h)}\\
& z_{ij} \in \{0,1\}, \quad \forall i\in I_j, j\in J. && \text{(10i)}
\end{aligned}$$

The objective function maximizes the total demand across all the products in the assortment. Constraint (10b) ensures that the demand d_{ij} for product i from category j does not exceed the sum of its base demand rate, i.e., $a_{ij} \cdot y_{ij}$ and the spillover demand from its parent products, i.e., $\sum_{(i',j') \in par_{ij}} \beta^{ij}_{i'j'} \cdot d_{i'j'}$. Notice that $a_{ij} \cdot y_{ij} + \sum_{(i',j') \in par_{ij}} \beta^{ij}_{i'j'} \cdot d_{i'j'}$ can be negative if $\beta^{ij}_{i'j'} \leq 0$ for some $(i', j') \in par_{ij}$. The term $M \cdot (1 - z_{ij})$ in constraint (10b) ensures that the demand rate d_{ij} remains non-negative, and enables us to capture substitution and complementarity effects in the same model. Constraint (10c) ensures that z_{ij} is 1 when $a_{ij} \cdot y_{ij} + \sum_{(i',j') \in par_{ij}} \beta^{ij}_{i'j'} \cdot d_{i'j'}$ is positive and constraint (10d) ensures that z_{ij} is 0 when $a_{ij} \cdot y_{ij} + \sum_{(i',j') \in par_{ij}} \beta^{ij}_{i'j'} \cdot d_{i'j'}$ is negative. Constraint (10e) ensures that if $z_{ij} = 0$, which occurs when $a_{ij} \cdot y_{ij} + \sum_{(i',j') \in par_{ij}} \beta^{ij}_{i'j'} \cdot d_{i'j'}$ is negative, then product i from category j is not included in the assortment, i.e., $y_{ij} = 0$. Constraint (10f) ensures that if product i in category j is not selected to be in the assortment, i.e., $y_{ij} = 0$, the demand rate d_{ij} cannot be positive. Lastly, constraint (10g) ensures the total number of products selected in the assortment does not exceed K.

To examine the value of considering the causal relations among product purchases in assortment decisions, we compared the performance of CPN-MIP$(\mathcal{G}, \boldsymbol{\beta})$, which considers both within-category and across-category relations, with that of the traditional MNL model, which focuses solely on within-category relations. Based on van Ryzin and Mahajan (1999), we first estimate a choice model separately for each product category in Data II. Then, using this choice model as input, we maximize the objective function of total expected sales across all categories in the theme subject to the linking constraint that at most K products are stocked. The estimation dataset and the objective function of the optimization problem are identical across the CPN and MNL models, with the only difference being in the assumed choice model. To solve the constrained MNL assortment optimization problem, we utilize the majorization result in van Ryzin and Mahajan (1999) for within category optimization to generate candidate solutions and apply the capacity constraint to obtain the best solution across categories. We report the pseudocode for this optimization in Appendix D.

Figure 8 presents a visual comparison of assortment strategies for the prepared food theme deployed by the CPN and MNL models for different values of K, where K takes on the values 5, 10, 15, and 20. The bar graphs highlight the number of selected products from five distinct departments within the prepared food theme: frozen foods, canned food, deli, beverages, and condiments. For each value of K, the comparison between the CPN and MNL models is illustrated separately for both brick-and-mortar and online channels, represented by different colored bars within each category. We observe a consistent trend where the CPN model tends to allocate more selections towards the beverages category across all K values. This emphasis is due to the fact that the CPN model generally does not find a substitution effect among products in the beverages category, with the only exceptions being the negative effects from ‘Cola - Coca Cola’ to ‘Citrus Berry Soda’ and

Figure 8 Comparison of assortment strategies for the CPN and MNL models (prepared food theme)



from 'Cola - Pepsi' to 'Still Water' in the brick-and-mortar channel. Except for these instances, all other beverage relationships are found to be positive, implying that a diverse assortment of beverages enhances total demand. In contrast, the MNL model naturally assumes substitution effects within each category, leading to a smaller assortment. We also find that the CPN model discovers that there is no interaction between sugary beverages and diet beverages. The MNL model would automatically force substitution unless these products are treated as separate categories; this type of manual classification is difficult to do in very large datasets with many latent attributes.

As the value of K increases, the CPN model strategically includes more products from the frozen and canned foods categories. This strategy illustrates how the CPN model leverages the underlying relationships among products from different categories to optimize the assortment. In contrast, the MNL model focuses on optimizing selections within individual categories based on their internal demand dynamics, rather than on their interactions with other categories. Consequently, the MNL model may not fully capture the additional consumer demand that could be generated from strategically placed cross-category products.

Table 6 demonstrates that the assortment strategies derived from the CPN model outperform those from the MNL model by 20.00% to 41.72% in total expected sales across different assortment sizes and channels. These differences highlight the advantages of employing a causal modeling approach in assortment planning, enabling retailers to develop more sophisticated strategies that account for the complex nature of cross-category purchase behaviors.

Table 6 Percentage difference in expected sales between the CPN- and MNL-based assortment strategies

Basket Sizes	$K = 5$	$K = 10$	$K = 15$	$K = 20$
Brick-and-Mortar (B&M) Channel	20.00	27.35	25.87	25.16
Online Channel	23.00	41.72	41.24	25.53

Note: The percentage difference between the CPN and MNL models is calculated as (Expected Sales from CPN/Expected Sales from $MNL - 1) \times 100\%$.

8. Conclusion

Our paper demonstrates the usefulness of causal product networks as a methodology for studying large-scale basket-shopping data. We show that causal product networks can effectively and comprehensively capture both complementarity and substitution effects within the same model, they can be estimated from existing archival data without the need to conduct costly field experiments, and they reveal practically useful insights regarding customers' shopping behavior. Using extensive data from Numerator, a market research company, we show that this method more accurately represents product relationships in shopping baskets and requires fewer parameters than either complete or correlation-based networks. Further, it also describes causal relationships in shopping baskets more effectively than category-level specifications. Finally, our analysis demonstrates that online shopping channels exhibit fewer causal relationships among product purchases compared to brick-and-mortar channels, highlighting differences in consumer behavior across retail contexts. To examine the value of causal product networks in multiple category assortment optimization, we propose a mixed-integer program that uses the constructed networks. We find that our causal model outperforms the MNL model by 20.00% to 41.72% in total sales across various assortment sizes and both brick-and-mortar and online channels. Moreover, we observe distinct assortment strategies between the causal product network and MNL models, which reveals insights into the assortment implications of consumer behavior. These results demonstrate the value of causal structure learning to characterize consumer basket-shopping behavior and provide insights into the relationships between product categories.

While our paper takes the first step towards using causal discovery to model basket-shopping behavior, there are several important limitations to note. First, our model can be enriched by incorporating additional retail data such as prices, promotions, and store layouts. Expanding the model to include these factors would allow for a more nuanced understanding of the various factors that influence consumer behavior, potentially leading to even more effective assortment strategies and more accurate identification of causal relationships. Second, while our approach leverages observational data to infer causal relationships, it lacks the experimental validation typically required in traditional economic studies. Future research can focus on integrating controlled field experiments to validate the causal networks identified through our model. Finally, our study does not account for customer heterogeneity. Future research could explore incorporating differences in consumer preferences and behavior patterns by utilizing panel data. This approach would allow the model to capture individual-level variations, enabling more personalized and accurate predictions of basket-shopping.

References

Bai Y, Feldman J, Segev D, Topaloglu H, Wagner L (2023) Assortment optimization under the multi-purchase multinomial logit choice model. *Operations Research* ISSN 0030-364X.

Börsch-Supan A (1990) On the compatibility of nested logit models with utility maximization. *Journal of Econometrics* 43(3):373–388, ISSN 0304-4076.

Cachon GP, Kök AG (2007) Category management and coordination in retail assortment planning in the presence of basket shopping consumers. *Management Science* 53(6):934–951, ISSN 0025-1909.

Chickering M, Heckerman D, Meek C (2004) Large-sample learning of bayesian networks is np-hard. *Journal of Machine Learning Research* 5:1287–1330.

Chintala SC, Liaukonytė J, Yang N (2023) Browsing the aisles or browsing the app? how online grocery shopping is changing what we buy. *Marketing Science* ISSN 0732-2399.

Chu J, Arce-Urriza M, Cebollada-Calvo JJ, Chintagunta PK (2010) An empirical analysis of shopping behavior across online and offline channels for grocery products: the moderating effects of household and product characteristics. *Journal of Interactive Marketing* 24(4):251–268, ISSN 1094-9968.

Degeratu AM, Rangaswamy A, Wu J (2000) Consumer choice behavior in online and traditional supermarkets: The effects of brand name, price, and other search attributes. *International Journal of research in Marketing* 17(1):55–78, ISSN 0167-8116.

Dhar V, Geva T, Oestreicher-Singer G, Sundararajan A (2014) Prediction in economic networks. *Information Systems Research* 25(2):264–284, ISSN 1047-7047.

Eberhardt F (2017) Introduction to the foundations of causal discovery. *International Journal of Data Science and Analytics* 3:81–91, ISSN 2364-415X.

Eberhardt F, Kaynar N, Siddiq A (2024) Discovering causal models with optimization: Confounders, cycles, and instrument validity. *Management Science* .

Ebert-Uphoff I, Deng Y (2014) Causal discovery from spatio-temporal data with applications to climate science. *2014 13th International Conference on Machine Learning and Applications*, 606–613 (IEEE).

Foraita R, Friemel J, Günther K, Behrens T, Bullerdiek J, Nimzyk R, Ahrens W, Didelez V (2020) Causal discovery of gene regulation with incomplete data. *Journal of the Royal Statistical Society Series A: Statistics in Society* 183(4):1747–1775.

Friedman N, Goldszmidt M, Wyner A (1999) Data analysis with bayesian networks: a bootstrap approach. *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*, 196–205.

Gallino S, Karacaoglu N, Moreno A (2023) Need for speed: The impact of in-process delays on customer behavior in online retail. *Operations Research* 71(3):876–894.

Gallino S, Moreno A (2014) Integration of online and offline channels in retail: The impact of sharing reliable inventory availability information. *Management Science* 60(6):1434–1451.

Glymour C, Zhang K, Spirtes P (2019) Review of causal discovery methods based on graphical models. *Frontiers in genetics* 10:524, ISSN 1664-8021.

Gopalakrishnan S, Matta M, Yourdshahy MI, Choudhary V (2023) Go wide or go deep? assortment strategy and order fulfillment in online retail. *Manufacturing & Service Operations Management* 25(3):846–861, ISSN 1523-4614.

Hall-Hoffarth EW (2022) Causal discovery of macroeconomic state-space models. *Available at SSRN 4212251* .

Hasan U, Gani MO (2022) Kcrl: A prior knowledge based causal discovery framework with reinforcement learning. *Machine Learning for Healthcare Conference*, 691–714 (PMLR).

Honhon D, Gaur V, Seshadri S (2010) Assortment planning and inventory decisions under stockout-based substitution. *Operations research* 58(5):1364–1379.

Huang Z, Zeng DD, Chen H (2007) Analyzing consumer-product graphs: Empirical findings and applications in recommender systems. *Management science* 53(7):1146–1164, ISSN 0025-1909.

Imoto S, Kim SY, Shimodaira H, Aburatani S, Tashiro K, Kuhara S, Miyano S (2002) Bootstrap analysis of gene networks based on bayesian networks and nonparametric regression. *Genome Informatics* 13:369–370.

Jagabathula S, Mitrofanov D, Vulcano G (2022) Personalized retail promotions through a directed acyclic graph–based representation of customer preferences. *Operations Research* 70(2):641–665, ISSN 0030-364X.

Jagabathula S, Vulcano G (2018) A partial-order-based model to estimate individual preferences using panel data. *Management Science* 64(4):1609–1628, ISSN 0025-1909.

Jasin S, Lyu C, Najafi S, Zhang H (2023) Assortment optimization with multi-item basket purchase under multivariate mnl model. *Manufacturing & Service Operations Management* ISSN 1523-4614.

Kaynar N, Mitrofanov D (2024) Long-term policy impact estimation with causal structure learning. *Available at SSRN 4913070* .

Lo V, Topaloglu H (2022) Omnichannel assortment optimization under the multinomial logit model with a features tree. *Manufacturing & Service Operations Management* 24(2):1220–1240, ISSN 1523-4614.

Mani V, Thomas DJ, Bansal S (2022) Estimating substitution and basket effects in retail stores: Implications for assortment planning. *Management Science* 68(7):5002–5024.

Markus Kalisch, Martin Mächler, Diego Colombo, Marloes H Maathuis, Peter Bühlmann (2012) Causal inference using graphical models with the R package pcalg. *Journal of Statistical Software* 47(11):1–26, URL `http://dx.doi.org/10.18637/jss.v047.i11`.

Martinoli M, Moneta A, Pallante G (2023) Calibration and validation of macroeconomic simulation models: A general protocol by causal search. *Available at SSRN 4325509* .

McFadden D, Train K (2000) Mixed mnl models for discrete response. *Journal of applied Econometrics* 15(5):447–470, ISSN 0883-7252.

Meek C (2013) Causal inference and causal explanation with background knowledge. *arXiv preprint arXiv:1302.4972* .

Mooij JM, Magliacane S, Claassen T (2020) Joint causal inference from multiple contexts. *Journal of Machine Learning Research* 21(99):1–108.

Nowack P, Runge J, Eyring V, Haigh JD (2020) Causal networks for climate model evaluation and constrained projections. *Nature communications* 11(1):1415.

Oestreicher-Singer G, Sundararajan A (2012) The visible hand? demand effects of recommendation networks in electronic markets. *Management science* 58(11):1963–1981, ISSN 0025-1909.

Pearl J (2009) *Causality* (Cambridge university press).

Ruiz FJ, Athey S, Blei DM (2020) Shopper. *The Annals of Applied Statistics* 14(1):1–27.

Runge J, Bathiany S, Bollt E, Camps-Valls G, Coumou D, Deyle E, Glymour C, Kretschmer M, Mahecha MD, Muñoz-Marí J, et al. (2019) Inferring causation from time series in earth system sciences. *Nature communications* 10(1):2553.

Sapra A, Kumar S (2023) Optimal joint assortment for an omni-channel retailer. *Information Systems Research* ISSN 1047-7047.

Shen X, Ma S, Vemuri P, Castro MR, Caraballo PJ, Simon GJ (2021) A novel method for causal structure discovery from ehr data and its application to type-2 diabetes mellitus. *Scientific reports* 11(1):21025.

Song I, Chintagunta PK (2006) Measuring cross-category price effects with aggregate store data. *Management Science* 52(10):1594–1609, ISSN 0025-1909.

Spirtes P, Glymour C, Scheines R, Kauffman S, Aimale V, Wimberly F (2000a) Constructing bayesian network models of gene expression networks from microarray data .

Spirtes P, Glymour CN, Scheines R (2000b) *Causation, prediction, and search* (MIT press), ISBN 0262194406.

Triantafillou S, Lagani V, Heinze-Deml C, Schmidt A, Tegner J, Tsamardinos I (2017) Predicting causal relationships from biological data: Applying automated causal discovery on mass cytometry data of human immune cells. *Scientific reports* 7(1):12724.

Tu R, Zhang K, Bertilson B, Kjellstrom H, Zhang C (2019) Neuropathic pain diagnosis simulator for causal discovery algorithm evaluation. *Advances in Neural Information Processing Systems* 32.

Tulabandhula T, Sinha D, Karra SR, Patidar P (2023) Multi-purchase behavior: Modeling, estimation, and optimization. *Manufacturing & Service Operations Management* 25(6):2298–2313, ISSN 1523-4614.

van Ryzin G, Mahajan S (1999) On the relationship between inventory costs and variety benefits in retail assortments. *Management Science* 45(11):1496–1509.

Venkatesan R, Kumar V, Ravishanker N (2007) Multichannel shopping: causes and consequences. *Journal of Marketing* 71(2):114–132, ISSN 0022-2429.

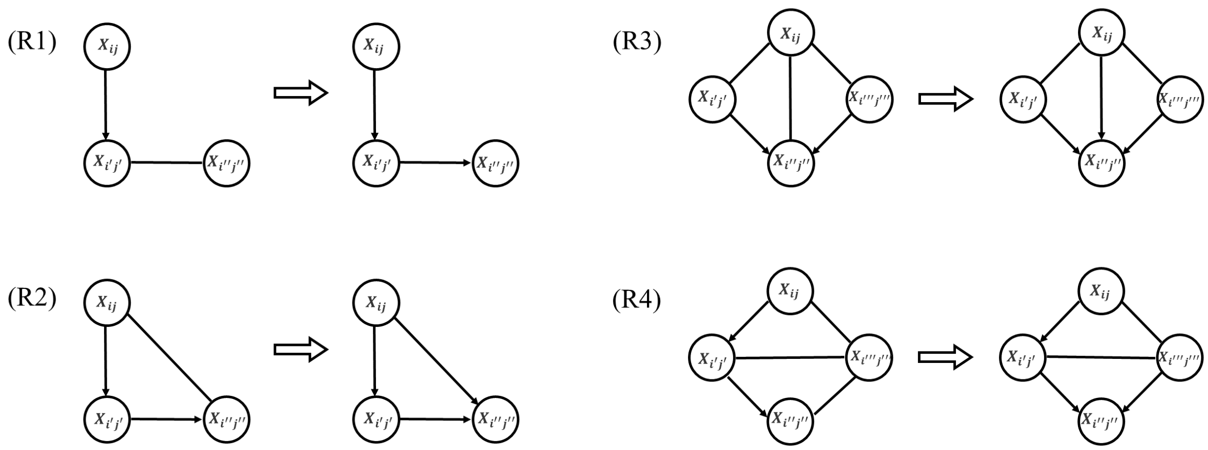
Verma T, Pearl J (1990) Equivalence and synthesis of causal models. *Proceedings of the Sixth Annual Conference on Uncertainty in Artificial Intelligence*, 255–270.

Wang K, Goldfarb A (2017) Can offline stores drive online sales? *Journal of Marketing Research* 54(5):706–719, ISSN 0022-2437.

Appendix A: Graphical illustration of four Meek rules

Figure 9 graphically illustrates the four Meek rules that correspond to (R1), (R2), (R3), and (R4) in Algorithm 2.

Figure 9 Meek rules



Appendix B: Product Lists of Five Themes

Tables 7, 8, and 9 display the most frequently purchased products across both the brick-and-mortar and online channels, the brick-and-mortar channel only, and the online channel only, respectively. Due to Numerator's data disclosure policy, we have concealed the brand information and are only displaying the product lists without brand details.

(See next page for the tables.)

Table 7 Top B&M and online products

Theme 1 (Pasta)	Pasta & Noodle	Spaghetti, Dry, Macaroni, Penne, Rotini
	Meat	Steaks, Ground Beef, Chicken, Lunch Packs, Whole Cuts & Roasts, Pork Chop, Sausage-Pork
	Produce	Bananas, Onions, Tomatoes, Apples, Grapes, Avocado, Cucumber, Potatoes, Lettuce, Strawberries
	Dairy	Milk, Coffee Creamers, Shredded Cheese, Cream Cheese, Sour Cream, Greek Yogurt
	Condiments	Nut Butters, Salad Dressings, Salsa, Hummus, Mayonnaise, Jam, Jelly & Marmalades
Theme 2 (QSR)	QSR Beverages	Cola, Pepper-Style Soda, Hot Coffee, Citrus & Berry Soda
	QSR Breakfast	Hash Browns, Donuts, Sausage Breakfast Sandwich, Ham Breakfast Sandwich
	QSR Sandwiches & Wraps	Beef Burger, Crispy Chicken Sandwich
	QSR Sauces & Condiments	Ketchup, BBQ Sauce, Mild Sauce, Ranch Dip, Hot Sauce
	QSR Mexican	Chicken Quesadilla, Steak Taco, Veggie Burrito, Steak Nachos, Steak Burrito
	QSR Snack & Sides	French Fries, Potato Chips
	QSR Desserts	Milk Shake, Pie, Cookies, Ice Cream Cone
	QSR Entrees	Chicken Nuggets, Bone-In Chicken
	QSR Italian	Hand Tossed Pizza, Pan Pizza, Thin Crust Pizza, Breadsticks
	QSR Salads	Chicken Cobb Salad, Chicken Garden Salad, Southwest Chicken Salad
Theme 3 (Bakery)	Bakery Sweet Goods	Donuts, Muffins, Cakes, Cookies, Snack Pies, Pastries, Pies
	In-Store Bakery	Bread & Breadsticks, Rolls, Bagels, Croissants, Tortillas, Italian Bread, Buns
	Packaged Bakery	Tortillas, Bagels, Buns, English Muffins, Rolls, Wheat Bread, White Bread
	Baking & Cooking	Pasta & Pizza Sauces, Sugar, Barbecue Sauce, Baking Chips, Pudding, Custard & Mousse Mix, Marshmallows, Tomato Sauce, Paste & Puree, Gelatin & Jello Mix, Cake Mixes, Dessert Syrups
	Dairy	Milk, Coffee Creamers, Shredded Cheese, Cream Cheese, Sour Cream, Greek Yogurt
Theme 4 (Prepared Food)	Frozen Foods	Frozen Breakfast, Frozen Chicken, French Fries, Single Serve Meals, Packaged Ice Cream, Pizza Bites/Rolls
	Canned	Prepared Beans, Canned Tuna, Canned Tomatoes, Variety Beans, Applesauce, Fruit Cups, Canned Green Beans
	Deli & Prepared Foods	Chicken-Prepared, Deli Salad, Pork-Deli, Ready-Made Sandwiches & Wraps-Prepared, Turkey-Deli, Sushi-Prepared
	Beverages	Cola, Sports Drinks, Citrus & Berry Soda, Still Water, Seltzers & Sparkling Water
	Condiments	Nut Butters, Salad Dressings, Salsa, Hummus, Mayonnaise, Jam, Jelly & Marmalades
Theme 5 (Unprepared Food)	Produce	Bananas, Onions, Tomatoes, Apples, Grapes, Avocado, Cucumber, Potatoes, Lettuce, Strawberries
	Shelf Stable Meals	Canned Soups, Mac & Cheese, Ramen & Noodle Soups, Potato Mixes, Pasta Dishes, Rice Dishes
	Meat	Steaks, Ground Beef, Chicken, Lunch Packs, Whole Cuts & Roasts, Pork Chop, Sausage, Pork-Mainstream
	Herbs & Spices	Mexican Seasoning, Grill Seasoning, Garlic Powder & Garlic Salt, Chili Seasoning
	Condiments	Nut Butters, Salad Dressings, Salsa, Hummus, Mayonnaise, Jam, Jelly & Marmalades

Note: (1) This table displays product names without brand information, and duplicate names across different brands have been removed; (2) The top products are those that appear most frequently in both online and brick-and-mortar baskets.

Table 8 Top B&M products

Theme 1 (Pasta)	Pasta & Noodle	Spaghetti, Penne, Dry, Rotini, Macaroni
	Meat	Chicken, Lunch Packs, Pork-Mainstream, Steaks, Ground Beef, Pork Chop, Sausage-Pork, Whole Cuts & Roasts
	Produce	Cucumber, Apples, Onions, Tomatoes, Potatoes, Bananas, Strawberries, Grapes, Avocado, Lettuce
	Dairy	Greek Yogurt, Shredded Cheese, Coffee Creamers, Milk, Sour Cream, Cream Cheese
	Condiments	Salsa, Nut Butters, Jam, Jelly & Marmalades, Hummus, Mayonnaise, Salad Dressings
Theme 2 (QSR)	QSR Beverages	Citrus & Berry Soda, Pepper-Style Soda, Frappuccino, Cola
	QSR Breakfast	Sausage Breakfast Sandwich, Hash Browns, Donuts, Ham Breakfast Sandwich
	QSR Sandwiches & Wraps	Beef Burger, Crispy Chicken Sandwich
	QSR Sauces & Condiments	Ranch Dip, Mild Sauce, BBQ Sauce, Ketchup, Hot Sauce
	QSR Mexican	Steak Taco, Veggie Burrito, Chicken Quesadilla, Steak Nachos, Chicken Taco
	QSR Snack & Sides	French Fries
	QSR Desserts	Pie, Milk Shake, Cookies
	QSR Entrees	Bone-In Chicken, Chicken Nuggets
	QSR Italian	Breadsticks, Pan Pizza, Thin Crust Pizza, Hand Tossed Pizza
	QSR Salads	Chicken Garden Salad, Chicken Cobb Salad, Southwest Chicken Salad
Theme 3 (Bakery)	Bakery Sweet Goods	Cookies, Packaged Muffins, Snack Pies, Pastries, Pies, Muffins, Donuts, Cakes
	In-Store Bakery	Tortillas, Italian Bread, Buns, Croissants, Bread & Breadsticks, Rolls, Bagels
	Packaged Bakery	Bagels, Buns, Wheat Bread, Sandwich Bread, White Bread, Tortillas, English Muffins, Rolls
	Baking & Cooking	Sugar, Cake Mixes, Gelatin & Jello Mix, Barbecue Sauce, Dessert Syrups, Pudding, Custard & Mousse Mix, Baking Chips, Marshmallows, Pasta & Pizza Sauces, Tomato Sauce, Paste & Puree
	Dairy	Greek Yogurt, Shredded Cheese, Coffee Creamers, Milk, Sour Cream, Cream Cheese
Theme 4 (Prepared Food)	Frozen Foods	French Fries, Syrup Carriers-Frozen Breakfast, All Other Single Serve Meals, Packaged Ice Cream, Nutrition Single Serve Meals, Frozen Chicken, Prepared Entrees-Frozen Breakfast
	Canned	Fruit Cups, Canned Tomatoes, Applesauce, Canned Green Beans, Canned Tuna, Variety Beans, Prepared Beans
	Deli & Prepared Foods	Turkey-Deli, Ready-Made Sandwiches & Wraps-Prepared, Pork-Deli, Chicken-Prepared, Deli Salad, Sushi-Prepared
	Beverages	Still Water, Seltzers & Sparkling Water, Pepper & Skipper Soda, Sports Drinks, Citrus & Berry Soda, Cola
	Condiments	Salsa, Nut Butters, Jam, Jelly & Marmalades, Hummus, Mayonnaise, Salad Dressings
Theme 5 (Unprepared Food)	Produce	Cucumber, Apples, Onions, Tomatoes, Potatoes, Bananas, Strawberries, Grapes, Avocado, Lettuce
	Shelf Stable Meals	Ramen & Noodle Soups, Pasta Dishes, Canned Soups, Rice Dishes, Potato Mixes, Mac & Cheese
	Meat	Chicken, Lunch Packs, Pork-Mainstream, Steaks, Ground Beef, Pork Chop, Sausage-Pork, Whole Cuts & Roasts
	Herbs & Spices	Chili Seasoning, Garlic Powder & Garlic Salt, Mexican Seasoning, Grill Seasoning
	Condiments	Salsa, Nut Butters, Jam, Jelly & Marmalades, Hummus, Mayonnaise, Salad Dressings

Note: (1) This table displays product names without brand information, and duplicate names across different brands have been removed; (2) The top products are those that appear most frequently in brick-and-mortar baskets.

Table 9 Top online products

Theme 1 (Pasta)	Pasta & Noodle	Spaghetti, Penne, Dry, Rotini, Macaroni, Lasagna, Bow-Tie
	Meat	Chicken, Lunch Packs, Pork-Mainstream, Pork Bacon, Steaks, Ground Beef, Sausage-Pork
	Produce	Cucumber, Salad Greens, Apples, Onions, Tomatoes, Bananas, Grapes, Avocado, Lettuce
	Dairy	Cream, Butter, Shredded Cheese, Natural Sliced Cheese, Coffee Creamers, Milk, Cream Cheese
	Condiments	Salsa, Nut Butters, Jam, Jelly & Marmalades, Hummus, Mayonnaise, Salad Dressings
Theme 2 (QSR)	QSR Beverages	Cola, Hot Coffee, Lemonade
	QSR Breakfast	Bagels, Bakery Other, Hash Browns, Muffins
	QSR Sandwiches & Wraps	Grilled Chicken Sandwich, Crispy Chicken Sandwich, Bacon Breakfast Sandwich
	QSR Sauces & Condiments	Hot Sauce, Ketchup, Mild Sauce, Beef Burger, Jam & Jelly
	QSR Mexican	Chicken Quesadilla, Steak Taco
	QSR Snack & Sides	French Fries
	QSR Desserts	Cookies, Milk Shake, Pie
	QSR Entrees	Chicken Nuggets, Bone-In Chicken
	QSR Italian	Hand Tossed Pizza, Breadsticks, Cheese Bread, Pan Pizza
	QSR Salads	Chicken Garden Salad, Chicken Cobb Salad, Southwest Chicken Salad
Theme 3 (Bakery)	Bakery Sweet Goods	Cookies, Packaged Muffins, Snack Pies, Muffins, Donuts, Pies
	In-Store Bakery	Tortillas, Buns, Croissants, Bread & Breadsticks, Rolls, Ciabatta
	Packaged Bakery	Bagels, Buns, Wheat Bread, White Bread, Tortillas, English Muffins, Rolls
	Baking & Cooking	Distilled Vinegar, Sugar, Barbecue Sauce, Pudding, Custard & Mousse Mix, Baking Chips, Marshmallows, Pasta & Pizza Sauces, Tomato Sauce, Paste & Puree
	Dairy	Cream, Butter, Shredded Cheese, Natural Sliced Cheese, Coffee Creamers, Milk, Cream Cheese
Theme 4 (Prepared Food)	Frozen Foods	Frozen Mixed Vegetables, French Fries, Syrup Carriers-Frozen Breakfast, Frozen Broccoli, Packaged Ice Cream, Nutrition Single Serve Meals, Frozen Chicken, Prepared Entrees-Frozen Breakfast
	Canned	Fruit Cups, Canned Tomatoes, Applesauce, Canned Green Beans, Canned Tuna, Canned Corn, Variety Beans, Prepared Beans
	Deli & Prepared Foods	Turkey-Deli, American Cheese (Deli), Ready-Made Sandwiches & Wraps-Prepared, Pork-Deli, Deli Salad
	Beverages	Still Water, Seltzers & Sparkling Water, Fruit Juice, Sports Drinks, Cola
	Condiments	Salsa, Nut Butters, Jam, Jelly & Marmalades, Hummus, Mayonnaise, Salad Dressings
Theme 5 (Unprepared Food)	Produce	Fresh Cucumber, Salad Greens, Apples, Onions, Tomatoes, Bananas, Grapes, Avocado, Lettuce
	Shelf Stable Meals	Ramen & Noodle Soups, Pasta Dishes, Canned Soups, Rice Dishes, Mac & Cheese
	Meat	Chicken, Lunch Packs, Pork-Mainstream, Pork Bacon, Steaks, Ground Beef, Sausage-Pork
	Herbs & Spices	Chili Seasoning, Onion Powder, Grill Seasoning, Cinnamon (Powder & Sticks), Mexican Seasoning, Garlic Powder & Garlic Salt
	Condiments	Salsa, Nut Butters, Jam, Jelly & Marmalades, Hummus, Mayonnaise, Salad Dressings

Note: (1) This table displays product names without brand information, and duplicate names across different brands have been removed; (2) The top products are those that appear most frequently in online baskets.

Appendix C: Robustness Checks (Testing Hypotheses 1 and 2 for B&M and Online Channels Separately)

Having established that brick-and-mortar and online channels can differ in terms of causal connections in Hypothesis 3, we conduct robustness checks to investigate whether Hypotheses 1 and 2 hold for each channel separately using Data II. For causal product network construction, we follow the subsampling steps discussed in §6.2. For each theme and product set, we create 50 distinct sub-samples from the corresponding brick and mortar and online dataset, separately. The final causal networks are constructed by including edges that appear in more than 30% of the corresponding networks. To test Hypothesis 1, we construct the complete and correlation-based product networks for each channel, considering each of the five themes and product sets using Data II. To test Hypothesis 2, we base our analysis on the final brick and mortar and online causal networks and explore the category level, product-to-category, and category-to-product specifications, as discussed in §6.2.

Tables 10 and 11 present the AIC scores for each theme within the B&M and online channels, respectively. For the B&M channel, Table 10 shows that causal product networks consistently have the lowest AIC scores across all five themes, supporting both Hypotheses 1 and 2. For the online channel, Table 11 shows that causal product networks have the lowest AIC scores across all themes except QSR. This exception is likely due to the limited availability of QSR options in the online channel, as our dataset contains only 372 baskets for this theme. Overall, the online channel supports Hypotheses 1 and 2, except for the QSR theme due to insufficient data.

Table 10 AIC scores of six network structures (B&M channel)

AIC Scores						
Network Structure	(a)	(b)	(c)	(d)	(e)	(f)
Theme 1 (Pasta)	434,287	434,459	379,375	472,665	470,332	462,954
Theme 2 (QSR)	400,257	398,786	302,397	426,723	425,004	418,412
Theme 3 (Bakery)	206,276	205,360	164,234	222,585	221,609	216,441
Theme 4 (Prepared Food)	318,785	325,329	204,119	354,248	353,327	351,881
Theme 5 (Unprepared Food)	873,727	872,505	200,014	904,788	904,314	896,668
Number of Parameters						
Network Structure	(a)	(b)	(c)	(d)	(e)	(f)
Theme 1 (Pasta)	2,750	784	57	14	43	43
Theme 2 (QSR)	3,000	826	46	21	41	35
Theme 3 (Bakery)	2,750	465	42	14	34	34
Theme 4 (Prepared Food)	2,750	438	51	15	37	38
Theme 5 (Unprepared Food)	2,750	671	36	12	28	25

Note: (1) The size of product sets for all the themes is 50; (2) The numbers of baskets are 14472, 9247, 7252, 6684, and 12901 for themes 1-5, respectively.

Table 11 AIC scores of six network structures (online channel)

AIC Scores						
Network Structure	(a)	(b)	(c)	(d)	(e)	(f)
Theme 1 (Pasta)	517,146	515,696	393,831	545,159	541,324	533,827
Theme 2 (QSR)	5,320	9,944	11,869	20,413	18,927	19,593
Theme 3 (Bakery)	206,276	205,360	164,234	222,585	221,609	216,441
Theme 4 (Prepared Food)	428,622	430,103	55,925	460,231	460,192	457,196
Theme 5 (Unprepared Food)	524,803	523,068	352,612	546,854	544,392	538,789
Number of Parameters						
Network Structure	(a)	(b)	(c)	(d)	(e)	(f)
Theme 1 (Pasta)	2,750	996	87	15	55	61
Theme 2 (QSR)	3,000	113	34	27	33	32
Theme 3 (Bakery)	2,750	599	52	13	41	39
Theme 4 (Prepared Food)	2,750	612	11	6	10	9
Theme 5 (Unprepared Food)	2,750	660	55	15	43	44

Note: (1) The size of product sets for all the themes is 50; (2) The numbers of baskets are 6613, 372, 6502, 6240, and 6340 for themes 1-5, respectively.

Appendix D: Greedy Algorithm for MNL Model

The greedy algorithm described in Algorithm 3 is used to find the optimal assortment strategy for the Multinomial Logit (MNL) model. It iteratively selects items that maximize incremental utility until the assortment size limit is reached, returning the optimal assortment strategy and the objective value of the MNL model.

Algorithm 4: Greedy Algorithm for MNL Model

Input: $I_j = \{1, ..., n\}$, $J = \{1, ..., m\}$, K and $q_{ij}, i \in I_j, j \in J$ where $q_{1j} \geq \cdots \geq q_{nj}$.
Output: $Y = \{y_{ij} \mid i \in I_j, j \in J\}$, Utility U.
Initialization: $y_{ij} = 0, \forall i \in I_j, j \in J$, $IU_j = 0$ and $k_j = 0, \forall j = 1, \ldots, m$.
1. Initialize $S = 0$ and $U = 0$.
2. **While** $S < K$:
3. **for** each j in $\{1, ..., m\}$:
4. Compute incremental utility $IU_j = \frac{\sum_{w=0}^{k_j+1} q_{wj}}{\sum_{w=0}^{k_j+1} q_{wj} + q_{0j}} - \frac{\sum_{w=0}^{k_j} q_{wj}}{\sum_{w=0}^{k_j} q_{wj} + q_{0j}}$.
5. Select $\max\{IU_1, \ldots, IU_m\}$ and the corresponding index j'.
6. Update $k_{j'} = k_{j'} + 1$, $y_{k_{j'} j'} = 1$ and $S = \sum_{j=1}^{m} k_j$.
7. Compute $U = \sum_{r=1}^{m} \frac{\sum_{w=0}^{k_m} q_{wr}}{\sum_{w=0}^{k_m} q_{wr} + q_{0r}}$.
8. Return Y and U.
