Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A High-Dimensional Choice Model for Online Retailing

Zhaohui (Zoey) Jiang; , Jun Li; , Dennis Zhang

To cite this article:
Zhaohui (Zoey) Jiang; , Jun Li; , Dennis Zhang (2025) A High-Dimensional Choice Model for Online Retailing. Management Science 71(4):3320-3339. https://doi.org/10.1287/mnsc.2020.02715

https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Vol. 71, No. 4, April 2025, pp. 3320–3339
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

A High-Dimensional Choice Model for Online Retailing

Zhaohui (Zoey) Jiang,[a,*] Jun Li,[b] Dennis Zhang[c]

[a] Tepper School of Business, Carnegie Mellon University, Pittsburgh, Pennsylvania 15213; [b] Ross School of Business, University of Michigan, Ann Arbor, Michigan 48109; [c] John M. Olin Business School, Washington University in St. Louis, St. Louis, Missouri 63130

*Corresponding author

Contact: zhaohuij@andrew.cmu.edu, https://orcid.org/0000-0003-3354-5851 (Z(Z)J); junwli@umich.edu, https://orcid.org/0000-0002-9237-9147 (JL); denniszhang@wustl.edu, https://orcid.org/0000-0002-4544-775X (DZ)

Received: September 4, 2020
Revised: June 26, 2022; October 21, 2023
Accepted: December 20, 2023
Published Online in Articles in Advance: July 19, 2024

https://doi.org/10.1287/mnsc.2020.02715

Abstract. Online retailers are facing an increasing variety of product choices and diversified consumer decision journeys. To improve many operations decisions for online retailers, such as demand forecasting and inventory management and pricing, an important first step is to obtain an accurate estimate of the substitution patterns among a large number of products offered in the complex online environment. Classic choice models either do not account for these substitution patterns beyond what is reflected through observed product features or do so in a simplified way by making a priori assumptions. These shortcomings become particularly restrictive when the underlying substitution patterns get complex as the number of options increases. We provide a solution by developing a high-dimensional choice model that allows for flexible substitution patterns and easily scales up. We leverage consumer clickstream data and combine econometric and machine learning (graphical lasso, in particular) methods to learn the substitution patterns among a large number of products. We show our method offers more accurate demand forecasts in a wide range of synthetic scenarios when compared with classical models (e.g., the independent and identically distributed Probit model), reducing out-of-sample mean absolute percentage error by 10%–30%. Such performance improvement is further supported by observations from a real-world empirical setting. More importantly, our method excels in precisely recovering substitution patterns across products. Compared with benchmark models, it reduces the percentage deviation from the underlying elasticity matrix by approximately half. This precision serves as a critical input for enhancing business decisions such as assortment planning, inventory management, and pricing strategies.

History: Accepted by Vishal Gaur, operations management.
Supplemental Material: The online appendix and data files are available at https://doi.org/10.1287/mnsc.2020.02715.

Keywords: choice model • machine learning • retail management • high-dimensional • clickstream data

1. Introduction

Digital advancements have transformed the retail space. While online shopping offers greater convenience to consumers, it becomes more complex for retailers to manage. Significant efforts are required for effective pricing, inventory, and assortment decisions, as online retailers often offer a variety of choices that outstrip those of their brick-and-mortar counterparts. To improve operational decisions for these online retailers, an important first step is to achieve an accurate understanding of the substitution patterns among a large number of products offered in this complex online environment. Such substitution patterns across products inform retailers how changes in one product impact other products' sales, which is a crucial input for many decisions (e.g., pricing, recommendation systems, inventory management, and assortment planning). Although classic choice models offer an elegant framework for estimating substitution patterns among competing options, they have limited applicability and poor performance in settings with a large number of products stemming from their inherent restrictions in modeling substitutions based on observed product features, detailed in Section 2.

Estimating flexible substitution patterns among large choice sets presents significant challenges. First, the space of all relevant product features can be very large, and many are difficult to quantify (e.g., the quality of advertising images/videos on the product page) and observe (e.g., consumers' exposure to product advertisements on other related websites). As a result, researchers often fail to control for all relevant factors affecting consumers' choices and the corresponding products' substitution patterns. These factors, when left uncontrolled, introduce important correlations among product utility

shocks, challenging the independent and identically distributed (IID) assumptions common in many choice models. At the same time, being exposed to various advertisement and distribution channels, consumers exhibit intricate patterns of preferences across products, posting challenges for researchers attempting to formulate a priori structural assumptions about unobserved utility shocks.

To tackle these challenges, we propose a choice model that allows for flexible high-dimensional substitution patterns and carries few assumptions. First, we leverage a unique data source in online retailing—consumers' clickstream data—and combine econometric and machine learning methods to estimate the conditional dependencies among products' underlying utility shocks at the clicking phase. Essentially, these conditional dependencies are captured by nonzero entries in the precision matrix (inverse of the variance-covariance matrix for utility shocks). We then use these identified dependencies to guide the estimation of a multinomial probit model at the purchasing phase.

We provide theoretical properties of our proposed method and demonstrate its superior performance with simulated and empirical data when compared with classic models. Specifically, we evaluate our method's performance on synthetic data, considering variations in variance-covariance matrix density, signal-to-noise ratio, click intensity, number of products, as well as the degree to which clickstream data informs purchasing decisions (including the impact of both overlooked and redundant information). Compared with a classical multinomial probit model with IID error terms (IID Probit model) and a probit model with relaxed diagonal elements in its variance-covariance matrix (Probit w/Diag model), our method consistently offers better in- and out-of-sample performances, yielding more precise demand forecasts. In the primary synthetic setup, when benchmarked against the performance of the true underlying model (which illustrates the lowest attainable error), the additional mean absolute percentage error of our model's demand forecast is only 1.26%, which is significantly lower than both the IID Probit model's 21.12% and the Probit w/Diag model's 21.09%. Even when compared against a probit model with all elements fully relaxed in the variance-covariance matrix, our model still offers a better out-of-sample performance, primarily because of the overfitting issues associated with estimating a fully relaxed variance-covariance matrix. We also apply our method to a real-world data set from a large online retailer with a complex product assortment. We show that, in this empirical setting, our method continues to offer better out-of-sample demand forecast performances.

More importantly, with enhanced demand forecasts, our model is able to accurately describe high-dimensional substitution patterns in the data, thereby offering refined insights into both own-product and cross-product price elasticities. Specifically, we evaluate the percentage deviation of our estimated elasticity matrix from the true matrix under the data-generating model. Remarkably, this deviation is only about half of the deviations seen in IID Probit and Probit w/Diag models. This substantial improvement forms a crucial foundation for enhancing a variety of operational decisions, including inventory management, assortment planning, and pricing management.

Our research makes several contributions. First, we offer an approach to estimate flexible high-dimensional substitution patterns for retailers. Such estimation is practically important because many retail decisions, such as pricing, inventory, and assortment, depend critically on knowledge of demand substitution patterns. We demonstrate that our method, compared with traditional choice models, offers more accurate estimates of substitution patterns; helps retailers make better demand forecasting, inventory management, and pricing decisions; and could potentially be further generalized to many other retailing applications, including promotion and assortment decisions. Second, our paper serves as another example of combining state-of-the-art machine learning methods with traditional econometrics to improve daily operations. We contribute to a growing literature attempting to combine machine learning methods with econometric methods to improve business decisions. Finally, we demonstrate with a large-scale, real-world data set that consumer clickstream data can indeed be utilized to better estimate demand substitution patterns, hence contributing to the emerging literature that leverages clickstream data in operational decisions.

2. Literature Review

Overall, we contribute to the following four streams of literature: (1) classic choice models, (2) large-scale choice models, (3) the application of machine learning methods in choice models, and (4) more broadly, retail operations management.

2.1. Classic Choice Models

Choice models with logit random utility shocks (McFadden and Train 2000), that is, multinomial logit models, or MNLs, are widely adopted in demand estimation because their closed-form specification makes estimation and corresponding optimization easier. However, this advantage comes at a cost for demand estimation—the MNL model exhibits the well-known independence of irrelevant alternatives (IIA) property (Train 2009), which can impose unrealistic substitution patterns in demand estimation. For example, the IIA property implies that a focal product with a lower price gains share from all other products in proportion

to their original shares, regardless of how similar (or dissimilar) they are to the focal product (see Train 2009 for more examples). This is problematic because misspecified substitution patterns will result in suboptimal pricing and assortment recommendations.

Two widely used solutions to achieve a non-IIA specification are (1) varying the error components in the latent utility specification (e.g., the nested logit or the generalized extreme value distribution models), or (2) shifting to a random coefficient specification (e.g., mixed logit/probit models). However, both approaches have limitations. The nested logit model requires preexisting knowledge about the product nest structure. The random coefficient model, though it theoretically can approximate any random-utility model to any desired degree of accuracy (McFadden and Train 2000), often fails to achieve ideal performance in reality because of the imperfect choice of explanatory variables and the distributions of the random parameters.

To achieve a fully flexible and realistic substitution pattern, a preferred solution is to use a multinomial probit (MNP) model with the error term following a normal distribution with a flexible covariance matrix (Train 2009, p. 103). However, this classic model is rarely applied because the number of parameters in the model is quadratic in the number of products, making it infeasible to estimate in most practical settings. Moreover, unless we have a large amount of consumer purchasing observations, we cannot achieve efficient estimates of these extremely flexible models.

We contribute to the choice model literature by offering a solution to this computational challenge in an online retail setting. Specifically, we leverage consumer clickstream data with a graphical lasso to build a structural understanding of the variance-covariance matrix among products' utility shocks. We then estimate a flexible MNP choice model leveraging such understanding. This two-step estimation greatly reduces the data as well as computational requirements in estimating the MNP choice model.

2.2. Large-Scale Choice Model

There is another set of literature within choice modeling that specifically focuses on choice models with large-scale data. Studies here have several different emphases, including (1) estimating a *basket demand* as the number of potential product baskets increases drastically with more products in the choice set (Ruiz et al. 2020), (2) estimating models with an improved *computational efficiency* and potentially trading off some accuracy in estimating substitution patterns among choices (Fox 2007, Amano et al. 2019, Chiong and Shum 2019), and (3) recovering a *realistic and flexible substitution pattern* in presence of the large choice sets while maintaining computational feasibility. Our focus is (3), which we discuss in more detail.

The closest studies to our method in this stream of literature are those of Yai et al. (1997) and Dotson et al. (2018), both of which estimate a structured covariance matrix. As mentioned before, estimating a full covariance matrix in probit models allows for flexible substitution patterns but generates a substantial computational burden. This computational burden can be alleviated with an imposed structure on the covariance matrix (instead of estimating a full matrix). Yai et al. (1997) estimate a probit model of route choices where the covariance between any two routes depends on the length of shared route segments. Dotson et al. (2018) estimate a probit model of product choice where covariance between any two products depends on the perceptual distance between choice alternatives and the distance is parameterized by the observed product attributes. Our paper shares the same spirit—enjoying the flexibility of a probit model while also reducing the number of parameters to be estimated. But instead of making assumption a priori (i.e., parameterizing the covariance with the observed product attributes), we let the data tell us about the structure of the variance-covariance matrix by analyzing which products consumers click on simultaneously.

There are other papers approaching the high-dimensionality problem through aggregate demand models. Smith and Allenby (2019) and Smith et al. (2019) propose random partition in estimating a log-linear demand system, and they use a Bayesian approach to flexibly estimate these partitions using supermarket scanner data at the store level. Their work considers single-layer, nonoverlapping partitions, which is suitable for identifying products *across* categories/groups that are separable (i.e., consumer preferences within each separable group are independent of consumption levels in the other group), whereas we focus on the challenges of high dimensionality *within* a large product category/group where all products are substitutable.

In terms of research context and data used, our paper is perhaps closest to Amano et al. (2019), who model a two-stage decision process—forming a consideration set and then finalizing the purchase within that set—using purchasing and clickstream data. The authors estimate aggregate consideration-set probabilities and approximate the across-set substitution using a parsimonious form; the percentage of consumers choosing a consideration set decreases with a higher *average* price among products in the set. These modeling decisions are effective in saving computation time while trading off precision in estimating substitution patterns because the probability of choosing a consideration set depends only on the average price rather than the price of each product in the set. Our model, however, focuses on estimating an almost fully flexible substitution pattern by imposing minimum assumptions. In sum, as we introduced at the beginning of this

subsection, Amano et al. (2019) focus more on (2), whereas our paper focuses more on (3).

2.3. Combining Machine Learning and Choice Models

Our paper also contributes to an emerging literature on combining machine learning methods with choice modeling methods. For example, Wan et al. (2017) use a latent factorization approach that incorporates price variation. Jagabathula et al. (2018) propose a model-based embedding technique to segment a large population of consumers into nonoverlapping groups with similar preferences. Chiong and Shum (2019) use random projection to compress a high-dimensional choice set into a lower-dimensional Euclidean space when estimating the choice model. Our paper introduces the idea of using graphical lasso, a tool popular in the machine learning field, to estimate the structure of the covariance matrix of errors in probit models. Here, we briefly summarize this approach.

To uncover the important connections among a large number of products, we utilize machine learning methods, specifically Gaussian graphical models, on consumer clickstream data to uncover a product network that describes the general structure of a substitution matrix among products. The edges in our Gaussian graphical models can be interpreted as the direct influence between two nodes. The graphical lasso method eliminates spurious or misleading relationships by removing nonexistent direct links among a large number of nodes. Graphical lasso was first proposed by Friedman et al. (2008) to estimate the precision matrix and obtain the graphical structure through a penalized maximum-likelihood approach. It was soon applied to different fields, including information networks (Gomez-Rodriguez et al. 2012), biostatistics (Cai et al. 2013), and neuroscience (Allen et al. 2014). Similar to Rothman et al. (2010), we aim to estimate a sparse precision matrix after taking into account the effects of the covariates on the mean utilities.

To the best of our knowledge, graphical lasso has not been utilized in operations models or choice models specifically. Therefore, we contribute to this literature by proposing a new way to incorporate a new machine learning technique with classic econometrics and operations management to improve operations efficiency in online retailing.

2.4. Retail Operations Management

Last but not least, our research contributes to the growing empirical retailing operations literature. With the growing amount of data from online and offline retailers, operations researchers have empirically studying various operations decisions in retailing (e.g., Gallino and Moreno 2014, Kesavan et al. 2016, Ertekin and Agrawal 2020). We contribute to this literature in three aspects. First, we contribute to a growing stream of literature utilizing choice models to understand consumer/firm decisions in retail settings (Lee et al. 2016, Fisher et al. 2018, Bray and Stamatopoulos 2022, Feldman et al. 2022). Second, we contribute to an increasing stream of literature using econometric or machine learning methods to handle high-dimensional challenges in retail settings (Ferreira et al. 2016, Glaeser et al. 2019, Mankad et al. 2019, Cohen et al. 2022). Last, in terms of the data employed, our work relates to a stream of literature in retailing operations that leverages online data, including data from online discussion forums (Netzer et al. 2012), consumer review data (Lee and Bradlow 2011), website performance data (Gallino et al. 2023), and, similar to our approach, online searching/browsing data (Kim et al. 2010, Ringel and Skiera 2016, Ngwe et al. 2019).

We contribute by proposing a new computationally and data-efficient method to estimate choice models with highly flexible substitution patterns. We present the theoretical properties of our proposed method and validate the model with synthetically generated and real-world data sets.

3. A Model with Flexible Substitution Patterns

In this section, we first present the choice model, motivate the importance of estimating a flexible variance-covariance matrix among products' utility shocks, and discuss the two challenges associated with estimating it: high dimensionality and limited variations in data. We then propose our solutions. Specifically, we estimate a variance-covariance matrix from a sparse precision matrix leveraging denser substitution information provided by clickstream data.

3.1. The Model

Consider a product category with J products. Let u_{ij} denote the utility that consumer i obtains when purchasing product j $(=1,2,\ldots,J)$. Consumer i can also choose not to purchase any product; that is, the consumer chooses the outside option that offers utility u_{i0}. Let y_{ij} be a binary variable where $y_{ij}=1$ denotes consumer i purchases product j, and $y_{ij}=0$ otherwise. A consumer will choose the option that maximizes utility. We let

$$u_{ij} = v_{ij} + \epsilon_{ij} = \alpha + X_{ij}\boldsymbol{\beta} + \epsilon_{ij}, \quad \forall j = 1,2,\ldots,J; \text{ and } u_{i0} = \epsilon_{i0}. \tag{1}$$

The vector X_{ij} captures product features that include time-variant, individual-specific characteristics (for example, product reviews, recommendations, prices, and promotions), as well as characteristics that are constant over time and across individuals (for example, product specifications).[1] The coefficient $\boldsymbol{\beta}$ captures

sensitivities to observed product characteristics. Because only differences in utilities are identifiable, we normalize the mean utility of the outside option to be zero. Error term ϵ_{ij} represents consumer i's random utility shock of purchasing product j. Specifically, we define a vector $\boldsymbol{\epsilon}_i := \{\epsilon_{i0}, \epsilon_{i1}, \epsilon_{i2}, \ldots, \epsilon_{iJ}\}$. Vector $\boldsymbol{\epsilon}_i$ is a normally distributed $(J+1)\times 1$ vector, IID across individual consumers, with zero mean and variance-covariance matrix:

$$\begin{aligned}\Sigma_{+0} &= \begin{pmatrix} \sigma_{0,0} & \sigma_{0,1} & \sigma_{0,2} & \cdots & \sigma_{0,J} \\ & \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,J} \\ & & \sigma_{2,2} & \cdots & \sigma_{2,J} \\ & & & \ddots & \vdots \\ & & & & \sigma_{J,J} \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{0,0} & \sigma_{0,1} & \sigma_{0,2} & \cdots & \sigma_{0,J} \\ & \begin{pmatrix} \sigma_{1,1} & \sigma_{1,2} & \cdots & \sigma_{1,J} \\ & \sigma_{2,2} & \cdots & \sigma_{2,J} \\ & & \ddots & \vdots \\ & & & \sigma_{J,J} \end{pmatrix} \end{pmatrix} \\ &= \begin{pmatrix} \sigma_{0,0} & 0 & 0 & \cdots & 0 \\ & \begin{pmatrix} & & \Sigma & & \end{pmatrix} \end{pmatrix}, \end{aligned} \tag{2}$$

where $\sigma_{jk} = \mathrm{cov}(\epsilon_{ij}, \epsilon_{ik})$. For simplicity of our later notations, we use Σ to denote the lower right block in Σ_{+0} after excluding the outside option. Note that we only specify the upper part of Σ_{+0} and Σ because they are symmetric matrices. The model is very similar to the standard probit model in the literature except that the latter assumes the variance-covariance matrix to be the identity matrix (i.e., error terms are IID across options). In other words, the standard probit model is a special case of our more generalized model. One might notice that Σ_{+0} is overidentified (Train 2009) because only differences in utilities are identifiable. In Appendix EC.1, we elaborate on the matrix transformation undertaken to ensure identification.

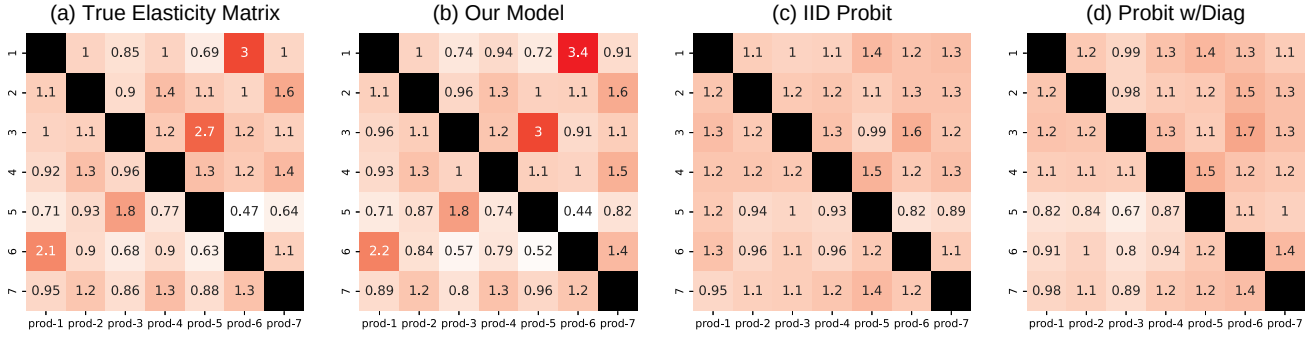
We assume that the covariance between ϵ_{i0} and ϵ_{ij} is zero; that is, $\sigma_{0,j} = 0$ when $j \neq 0$. We make this assumption because (1) the outside option is vastly different from any actual product option—it includes many consumer decisions, for example, not purchasing at all, purchasing from a different e-retailer, or purchasing from a different channel. As a result, the outside option is unlikely to have a strong comovement with any specific product; and (2) consumer purchase or click data are relatively sparse, so we choose to utilize the limited variations from the data to facilitate the estimation and identification of the covariance among products' utility shocks, which is the focus of our study.

Hereafter, we use a flexible Σ (including $\sigma_{0,0}$) to characterize the variance-covariance matrix. The advantage of allowing a flexible Σ matrix is that it imposes minimal assumptions on substitution patterns (Train 2009). We illustrate its importance in Section 3.2.

3.2. The Importance of Estimating a Flexible Variance-Covariance Matrix

We provide a concrete numerical example to motivate the importance of allowing for a flexible Σ. Consider a product category consisting of seven distinct products. In Figure 1(a), we plot the true elasticity matrix among products implied by the underlying true demand model. The off-diagonal elements in the elasticity matrix help illustrate when a product's price changes, what the impacts would be on other products' sales. For instance, based on the matrix's column 3, we find that when the price of product 3 increases, there will be larger impacts on product 5, that is, 2.12 (= 1.8/0.85) times larger compared with the average impact on other products. This elasticity matrix is an important input for the subsequent operational decisions (e.g., inventory and pricing) and, hence, needs to be accurately recovered.

Figure 1. (Color online) Elasticity Matrices (Standardized) Recovered Under Different Models



Note. To standardize each entry in all four elasticity matrices, we divide the corresponding entry in the original elasticity matrices by the geometric mean of the averages of its corresponding column and row.

We compare several choice models' performance in recovering the true elasticity matrix. These models are the same except for Σ. We estimate these models and present their resulting elasticity matrices in Figure 1. Specifically, we consider (1) our model: a probit model with a flexible Σ informed by clickstream information (in Figure 1(b)); (2) IID Probit: the standard IID probit model, where Σ equals the identify matrix (in Figure 1(c)); (3) Probit w/Diag: a probit model with flexible diagonal elements in Σ, but the off-diagonal elements are set to be zeros (in Figure 1(d)).

Comparing results in Figure 1, we can see that our model better recovers the relative magnitudes of the cross-product elasticities. Take product 3 as an example: both the true model and our model suggest that when the price of product 3 increases, there will be larger impacts on product 5 in comparison with the impacts on other products. Yet both IID Probit and Probit w/Diag models fail to reflect this underlying relationship between products—both models wrongly suggest consumers of product 3 would substitute for other products rather evenly. Note that, though Probit w/Diag model is more flexible than IID Probit as it relaxes the diagonal elements, their performances are not significantly different. These observations suggest that estimating a flexible Σ, particularly estimating the off-diagonal elements, is important for recovering the underlying elasticities.

We now discuss the implications for subsequent pricing decisions, emphasizing the importance of estimating a flexible Σ in enhancing operational decision making. To illustrate, we experiment with an exogenous change to this product category and see how prices should be adjusted accordingly. In particular, suppose product 3's price is reduced by \$30 (around 20%). Let us see how prices for other products should be adjusted accordingly. As a benchmark, we calculate the "true" optimal product price adjustments (in percentage) under the underlying demand model and report them in column (1) in Table 1. In this calculation, the optimal prices are the ones maximizing the overall profit of the product category. We also calculate the optimal product price adjustments under our model, IID Probit, and Probit w/Diag, and report them in columns (2), (3), and (4), respectively. As can be seen from the table, the price adjustments proposed by our model are close to the ones suggested by the true model, yet both IID Probit and Probit w/Diag models fail to propose the desirable price adjustments. In particular, as shown in Figure 1(a), the cross-price elasticities associated with product 3 are higher for product 5. This observation is reflected in the pricing decisions under the true model and also our model—product 5 has a larger price adjustment as a result of product 3's price changes. In comparison, both IID Probit and Probit w/Diag models fail to recognize the close substitutability between products 3 and 5 and suggested price changes of similar scales for all products. These observations again suggest that estimating a flexible Σ matrix, particularly estimating the off-diagonal elements, is important for more accurate pricing decisions.

Table 1. Changes in Price Recommendations Under Different Models in A Numeric Example

	Optimal price changes (%) under:			
	True model	Our model	IID Probit	Probit w/Diag
Product 1	−2.82	−2.86	−2.63	−2.84
Product 2	−4.00	−3.95	−2.61	−3.41
Product 3	–	–	–	–
Product 4	−2.56	−2.48	−1.60	−2.18
Product 5	−9.05	−9.30	−3.29	−3.93
Product 6	−2.77	−2.58	−2.95	−3.95
Product 7	−4.06	−4.16	−2.57	−3.91

Note. The table reports price adjustments recommended by various models in response to a \$30 price reduction for product 3.

3.3. Overview of Our Solutions

So far, we have illustrated the importance of estimating a flexible variance-covariance matrix (Σ). Yet estimating Σ is computationally challenging. It contains $[(J+1)J/2-1]$ free parameters to be estimated, which increases quadratically with the number of products. For example, if the product category includes 50 products, there will be $(50+1)\times 50/2-1=1{,}274$ parameters to be estimated. In other words, the estimation problem becomes increasingly complex as the number of products rises, leading to potential risks of spurious estimation and overfitting. On the other hand, in contrast with the scale of the problem, consumer purchase data usually exhibit limited variation, as a consumer only selects one (or none) out of the many options. As a result, most of the outcome variables take zero as values. The limited variation among the outcome variables further adds to the difficulty of estimating a flexible Σ.

We offer two solutions to tackle these challenges. To mitigate the risks of spurious estimations and overfitting associated with high dimensionality challenges, we estimate a sparse precision matrix, defined as the inverse of the variance-covariance matrix. To overcome the limited variation in purchase data, we leverage clickstream data, which provide denser information on what consumers consider as substitutable products. This approach also allows us to transform a quadratic estimation problem into a linear problem under minimal assumptions. We first summarize the insights and then discuss the details of these solutions in Section 3.4.

To address the challenges of spurious estimations and overfitting that arise with high dimensionality, we leverage one critical property of Gaussian variables: the

precision matrix of Gaussian variables indicates conditional dependence. In particular, an element of the precision matrix is zero if and only if the two corresponding variables are independent conditional on all other variables. Theoretical details of this property are formally introduced in Section 3.4.1. In our setting, this property intuitively suggests a nonzero element in the precision matrix indicates that the utility shocks of the two associated products are correlated *conditional on* utility shocks of all other products. In other words, the two products experience common utility shocks unique to themselves but not to others. For instance, two sweaters both have unique lace trims, whereas others do not; two cosmetic products are both promoted by a YouTube influencer, but other products are not. Whereas there are many potential product pairs, most product pairs likely do not share sufficiently unique utility shocks. Consequently, the precision matrix is likely sparse. Moreover, consumers typically consider a handful of options in their decision processes even though many more options are available, which also likely leads to a sparse precision matrix. Later in Section 5.1, we provide empirical evidence to support our conjecture of sparse precision matrices in several representative real-world retail contexts. In scenarios where the precision matrix is sparse, our method not only yields more reliable and stable results but is also practically more appealing. By understanding which products are connected (or more connected than others), retail managers can focus their limited attention on the most related products within their portfolio and improve the accuracy of their decisions.

Next, to overcome the limited variation in purchase data, we leverage clickstream data that provide denser information on what consumers consider as substitutable products. We leverage consumer clickstream information to estimate nonzero elements in the precision matrix, which can inform the estimation of variance-covariance matrix during the subsequent purchasing stage. Clickstream data are helpful for two reasons. First, clickstream data are typically denser than purchasing data. Many consumers may click through a website but not end up purchasing anything. According to recent surveys, the average conversion rate in online shopping is often less than 4% across many product categories and shopping channels.[2] Moreover, the consumers who decide to purchase click on many more options than they purchase. Second and more important, we often cannot rely on purchasing data to estimate the nonzero entries in the precision matrix because most consumers choose at most one option, which only provides information on how one option is preferred over others, but not how options' utilities are correlated. Hence, the identification based on purchasing data needs to rely on aggregate-level variations across time/markets. In practice, we may not have such sufficient variations to estimate the precision matrix among utility shocks. On the other hand, if we use clickstream data, we can exploit individual-level variations to estimate which products are close substitutes—in reality, many consumers click more than one product, which provides valuable information on the options a consumer values at the same time.[3] Whereas clickstream data offers these benefits, it is crucial to recognize that the clicking utility, which drives consumers' clicking behaviors, may not perfectly align with the corresponding purchasing utility, which ultimately governs consumers' purchasing decisions. To effectively harness clickstream data, we introduce two mild assumptions in Section 3.4.2. Under these assumptions, we are able to significantly reduce the complexity of the estimation from quadratic to linear in the number of products. We also conduct sensitivity analyses to evaluate the impact of these assumptions and demonstrate the robustness of our approach. Next, we delve into a formal theoretical discussion of the aforementioned intuitions and details of our estimation.

3.4. Estimation

We begin by introducing the theoretical properties of the precision matrix and then describe our two-stage estimation process. In the first stage, we leverage clickstream data to estimate a sparse precision matrix of utility shocks at the clicking phase. In the second stage, we solve the original purchasing-phase problem, estimating the probit model with a flexible variance-covariance matrix.

3.4.1. Theoretical Property of the Precision Matrix.

The precision matrix associated with a multivariate Gaussian distribution has a property based on the Hammersley and Clifford (1971) theorem:

Property. *If ϵ_{ijt} follows multivariate Gaussian distribution with the variance-covariance matrix being Σ and the corresponding precision matrix being Φ $(=\Sigma^{-1})$, the conditional independence between ϵ_j and ϵ_k given other variables $(\epsilon_{-\{j,k\}})$ is equivalent to $\Phi_{jk}=0$. Mathematically, $\Phi_{jk}=0 \iff \epsilon_j \perp \epsilon_k | \epsilon_{-\{j,k\}}$.*

We define the *support* of a matrix M as the (0,1)-matrix with jk^{th} entry equal to one if the jk^{th} entry of M is nonzero, and equal to zero otherwise. The property given suggests that the support of the precision matrix for Gaussian variable ϵ_{ijt} represents the conditional dependence among products' utility shocks. In the language of graph theory, by estimating the precision matrix, we are looking for a *product network* with edges representing the direct links between any pair of products.

Note that we choose to assume a sparse precision matrix as opposed to a sparse variance-covariance matrix. This choice is motivated precisely by the property

shown—the covariance matrix reflects the *marginal* dependence among products ϵ's, whereas the precision matrix reflects the *conditional* dependence among products ϵ's. As a result, the precision matrix is more likely to be sparse because the common shocks have been partialled out. Nevertheless, the efficacy of our method is not reliant on the precision matrix being sparse; however, in instances where it is sparse, our methodology is more effective in mitigating the risks of spurious estimations and overfitting compared with conventional approaches.

3.4.2. Stage 1: Gain Structural Understanding of Σ Based on Clickstream Data. We utilize clickstream data to gain insights into the structure of the variance-covariance matrix among products' utility shocks at the purchasing phase. Note that our objective is not to explain clicking behavior per se but, rather, to harness clickstream data for an improved stage 2 estimation. Hence, we intentionally construct a parsimonious empirical model to capture consumer clicking behaviors.

A Clicking Model. Consumer i decides whether to click on product j to access its details based on the product's perceived utility u^c_{ij}, referred to as "clicking utility":

$$u^c_{ij} = X^c_{ij}\beta^c + \epsilon^c_{ij}, \tag{3}$$

where the vector X^c_{ij} captures observed product features, and ϵ^c_{ij} captures the unobserved utility shocks. Note the features entering clicking and purchasing utilities (i.e., X^c and X) and their impacts (i.e., β^c and β) can be different. Let us denote the variance-covariance matrix of ϵ^c as Σ^c (and the corresponding precision matrix as Φ^c) with the jk^{th} entry being σ^c_{jk}.

Linking Clicking with Purchasing. We specify a parsimonious model to link consumer clicking and purchasing decisions. Whereas the impact from observed product features is explicitly accounted for, we focus on how clicking and purchasing utilities are connected via unobserved utility shocks, specifically, ϵ^c_{ij} and ϵ_{ij}. The two terms intuitively capture features unobservable to researchers but that influence consumers' clicking and purchasing decisions. Here is how ϵ^c_{ij} and ϵ_{ij} are connected. Upon clicking, consumers access a detailed product page, with the newfound information postclick denoted as v_{ij}. We model how consumers combine information already known from the clicking phase (ϵ^c_{ij}) with the newly obtained v_{ij} as $\epsilon_{ij} = \delta_j\epsilon^c_{ij} + v_{ij}$, where $\delta_j(>0)$ represents the level of attention consumers allocate to information available prior to viewing the detailed product page (e.g., the main product image), and it is allowed to be product specific.

The newfound information postclick (v_{ij}) may exhibit correlation with preclick information (ϵ^c_{ij}). For example, detailed positive reviews observed after clicking could be positively correlated with an overall good rating observed before clicking. Specifically, v_{ij} can be decomposed into a component that is proportional to ϵ^c_{ij} and an orthogonal residual term e such that $v_{ij} = \rho_j\epsilon^c_{ij} + e_{ij}$, where $\text{cov}(e_{\cdot}, \epsilon^c) = 0$. The ρ_j, which governs the extent to which the newfound information correlates with preclick information, is also allowed to be product specific. Combining these, we have

$$\epsilon_{ij} = w_j\epsilon^c_{ij} + e_{ij}, \text{ where } w_j = \rho_j + \delta_j. \tag{4}$$

The parameter w summarizes the direct impact of preclick information and its potential correlation with the newly discovered information from the detailed product page. Intuitively, it captures the extent to which information obtained prior to clicking is *predictive* of purchasing utility shocks. For simplicity of notation, we denote the vector of weights for all products as $\boldsymbol{w} := \{w_1, w_2, \ldots w_J\}$.

We make the following two assumptions:

Assumption 1. *The vector $\boldsymbol{w}$ is a nonzero vector.*

Assumption 2. *The residual term e's represents idiosyncratic shocks across products; that is, e_{ij} is independent of e_{ik} $(\forall k \neq j)$.*

Assumption 1 is a mild assumption. Intuitively, Assumption 1 implies that consumer clicking behavior offers directional insights into predicting their subsequent purchasing decisions. This is supported by empirical evidence in Appendix EC.2. Whereas it is not strictly a theoretical necessity, if many w's approach zero, the advantage of our approach diminishes (as will be shown in a set of synthetic analyses in Section 4). Note that w_j can vary by product, allowing the amount of directional information obtained from clickstream data to differ across products.

Assumption 2 requires that, in the newfound information postclick, the component (e) orthogonal to the prior knowledge behaves as an idiosyncratic shock. In other words, e's do not introduce significant new inter-product correlations. Note that this assumption permits the postclick acquired information to exhibit correlations across products; it requires only that e, which is entirely unrelated to the information available preclick, remains uncorrelated across products. This assumption is arguably mild given the rich information encapsulated in clickstream data. Consumers arrive at a click through a variety of channels—filters, sorting, keyword searches, recommendations, etc. Each channel contains specific aspects of product information. Collectively, they capture comprehensive information regarding a product and how its utility could be correlated with

similar products. Furthermore, because of the sequential consumer journey from clicking to purchasing, additional interproduct connections may remain undiscovered and prove irrelevant to the purchasing decision if two products lack sufficient similarities to be clicked on simultaneously in the first place. Nonetheless, it remains possible that some interproduct correlations are not captured entirely through ϵ^c. We simulate these scenarios in Section 4 and find that our approach consistently outperforms traditional methods even when correlations are not fully captured by clickstream data.

Under these assumptions, clickstream data provide information about the variance-covariance matrix Σ in latent purchasing utilities. Lemma 1 states this formally.

Lemma 1. *Under Assumptions 1 and 2, the variance-covariance matrix Σ among purchasing utilities can be expressed as a function of the variance-covariance matrix Σ^c among clicking utilities. Specifically, let $\sigma_{jk} = \mathrm{cov}(\epsilon_j, \epsilon_k)$, $\sigma^c_{jk} = \mathrm{cov}(\epsilon^c_j, \epsilon^c_k)$, $\sigma_{jj} = \mathrm{var}(\epsilon_j)$, $\sigma^c_{jj} = \mathrm{var}(\epsilon^c_j)$, and we have:*

$$\textit{Off-diagonal elements}: \ \sigma_{jk} = w_j w_k \cdot \sigma^c_{jk}$$

$$\textit{On-diagonal elements}: \ \sigma_{jj} = w_j^2 \cdot \sigma^c_{jj} + (1-\delta_j)^2 \cdot \mathrm{var}(e_j). \tag{5}$$

Proofs for the lemma are provided in Appendix EC.3.

Lemma 1 shows how we could leverage Σ^c—to be estimated from the clickstream data in stage 1—to more effectively estimate Σ. In particular, with knowledge of Σ^c, our focus in stage 2 narrows down to estimating the $\boldsymbol{w}$ vector, which captures the extent to which information obtained preclick is predictive of purchasing utility shocks in the given empirical context. Regarding the on-diagonal elements σ_{jj}, we make no assumptions regarding the relative magnitude of σ^c_{jj} and $\mathrm{var}(e_j)$. Rather, $\{\sigma_{jj}\}_{j=1,..,J}$ will be directly estimated from data in stage 2 and is allowed to vary by product.

Next, we describe how to estimate Σ^c from a sparse precision matrix. As mentioned in Section 3.3, the precision matrix among product utilities is likely to be sparse. We apply the graphical lasso algorithm developed by Friedman et al. (2008) to estimate a sparse precision matrix Φ^c. Specifically, $\hat{\Phi}^c$ is the solution to the following optimization problem over nonnegative definite matrices Φ^c:

$$\hat{\Phi}^c = \arg\min_{\Phi^c} (\mathrm{tr} S^c \Phi^c - \mathrm{logdet}\Phi^c + \lambda \|\Phi^c - \mathrm{diag}(\Phi^c)\|_1), \tag{6}$$

where S^c is the empirical covariance matrix of vector $\boldsymbol{u}^c$ conditional on product features X^c. Note that u^c is not directly observed in most empirical settings; we discuss our solution to this challenge in estimating S^c later in this subsection. The term $\mathrm{tr} S^c \Phi^c - \mathrm{logdet}\Phi^c$ is the negative log likelihood, where *tr* denotes the trace, and *det* denotes the determinant. The term $\|\Phi^c - \mathrm{diag}(\Phi^c)\|_1$ is the sum of the absolute values of off-diagonal coefficients of Φ^c. See Appendix EC.4 for the proof.

The graphical lasso estimator uses an L_1 penalty to enforce sparsity on the precision matrix Φ^c through the tuning parameter λ: the larger the tuning parameter λ is, the sparser the precision matrix. There are several methods to set the tuning parameter λ: cross-validation (CV), Bayesian information criterion (BIC), or adaptive methods. We discuss these methods in detail in Appendix EC.5. The optimization problem in Equation (6) can be solved using the block-wise coordinate descent approach in Banerjee et al. (2008). See Rothman et al. (2008) for the guaranteed convergence of this estimator.

We now delve into the estimation of S^c. Note that we typically do not directly observe the clicking utility u^c_{ij}, nor the underlying latent error term ϵ^c_{ij}. We only observe z_{ij}, which is a binary variable indicating whether a consumer i clicks a product j. Under standard and mild assumptions, we can model $z_{ij} = \mathbf{1}(u^c_{ij} > c)$, where c is the cost of clicking (c can be further refined to c_i to accommodate individual-specific clicking costs). We show through extensive simulation studies in Section 4.2 and Appendix EC.7 that $\mathrm{cov}(u^c_{ij}, u^c_{ik} | X^c_{ij}, X^c_{ik})$ can be linearly approximated by $\mathrm{cov}(z_{ij}, z_{ik} | X^c_{ij}, X^c_{ik})$. Instead of solving Equation (6), we solve for $\hat{\Phi}^{c,z}$ as the solution to the following optimization problem in Equation (7) over nonnegative definite matrices Φ^c:

$$\hat{\Phi}^{c,z} = \arg\min_{\Phi^c} (\mathrm{tr} S_{zz} \Phi^c - \mathrm{logdet}\Phi^c + \lambda \|\Phi^c - \mathrm{diag}(\Phi^c)\|_1), \tag{7}$$

where S_{zz} is the empirical covariance matrix of vector $\boldsymbol{z}$ conditional on product features X^c. See Appendix EC.6 for the detailed algorithms.

Upon obtaining the estimated $\hat{\Phi}^{c,z}$, we can compute the corresponding covariance matrix associated with the sparse precision matrix ($\hat{\Sigma}^{c,z} = (\hat{\Phi}^{c,z})^{-1}$). This becomes the crucial input to stage 2 of our estimation approach. In Section 4.2, we demonstrate with synthetic data that estimating Equation (7) provides a close estimate $\hat{\Sigma}^{c,z}$ to the true variance-covariance matrix Σ^c. For simplicity of notation, in the remainder of the paper, we drop z from the superscript and use $\hat{\Sigma}^c$ and $\hat{\Phi}^c$ to denote $\hat{\Sigma}^{c,z}$ and $\hat{\Phi}^{c,z}$, respectively.

3.4.3. Stage 2: Estimate the Probit Model for Consumer Purchase. Knowing the variance-covariance matrix associated with consumers' clicking utilities, we now return to the original problem: estimating the choice model using purchase data. We focus on estimating the

following parameters: (1) the weights $\boldsymbol{w}$, (2) the diagonal elements in Σ (denoted as diag(Σ),) and (3) coefficients of product characteristics, that is, α and $\boldsymbol{\beta}$. Specifically, we summarize $\{\alpha,\boldsymbol{\beta},\boldsymbol{w},\text{diag}(\Sigma)\}$ into a parameter vector, denoted as $\boldsymbol{\theta}$, and estimate $\boldsymbol{\theta}$ by maximizing the likelihood function

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{N}\prod_{j=1}^{J} \Pr(y_{ij} = 1 \mid \boldsymbol{\theta})^{d_{ij}}, \tag{8}$$

where d_{ij} indicates whether consumer i purchases product j or not based on the purchasing data, $\prod_{j=1}^{J} \Pr(y_{ij} = 1 \mid \boldsymbol{\theta})^{d_{ij}}$ is the probability of observing the actual choice d_{ij} made by consumer i, and $\Pr(y_{ij} = 1 \mid \boldsymbol{\theta})$ is the conditional choice probability (CCP) of observing consumer i purchase product j given a set of parameters summarized in $\boldsymbol{\theta}$. Specifically, CCP is calculated as

$$\begin{aligned}\Pr(y_{ij} = 1 \mid \boldsymbol{\theta}) &= \Pr(u_{ij} > u_{ik},\ \forall k \neq j)\\ &= \Pr(v_{ij} + \epsilon_{ij} > v_{ik} + \epsilon_{ik},\ \forall k \neq j)\\ &= \int_{\epsilon_{i1}=-\infty}^{\epsilon_{ij}+v_{ij}-v_{i1}} \cdots \int_{\epsilon_{ij}=\epsilon_{i0}-v_{ij}}^{\infty} \cdots \\ &\quad \int_{\epsilon_{ij}=-\infty}^{\epsilon_{ij}+v_{ij}-v_{iJ}} F(\epsilon_i)\ d\epsilon_{i0}\, d\epsilon_{i1}\, d\epsilon_{i2} \ldots d\epsilon_{iJ},\end{aligned} \tag{9}$$

where $F(\cdot)$ is the cumulative distribution function for $\boldsymbol{\epsilon}_i$.

We estimate the parameters in Equation (8) using a simulated maximum-likelihood (SML) estimator because the conditional choice probability based on normally distributed error terms does not have a closed-form expression. In particular, the simulation of $F(\cdot)$ is performed using the logit-smoothed accept-reject simulator.[4]

4. Model Performance on Synthetic Data

In this section, we demonstrate the performance of the proposed method and compare our method to classic choice models under a wide range of synthetic settings. We first describe how we construct these synthetic settings. We then discuss the key points in the estimation process and the estimation performance of the two stages.

4.1. Synthetic Data Generation

We simulate consumer clicking and purchasing data based on the model introduced in the previous section. Specifically, consumers' purchasing decisions are driven by purchasing utilities $u(= X\boldsymbol{\beta} + \epsilon)$, and their clicking decisions are driven by clicking utilities $u^c(= X^c\boldsymbol{\beta^c} + \epsilon^c)$. Note that consumers might access different information or exhibit different sensitivities to the same product features when clicking versus purchasing. This is reflected in the synthetic experiment design as follows.

4.1.1. Baseline Simulation Setup.

For the baseline case, we consider a choice setting with 50 periods, 200 consumers per period, 30 products, and 3 observed product features: x_1, x_2, and x_3, where x_3 denotes price.[5] The relative scale of the underlying parameters and the relative distribution of product features are set based on real-world empirical settings, which we introduce in Section 5. In particular, we set the coefficients of the observed product features in u_{ij} as follows: $\alpha = 0$, $\beta_1 = 0.1$, $\beta_2 = 0.167$, $\beta_3 = -0.017$. We assume the coefficients for the observed product features in clicking utility u^c_{ij} to be $\beta^c_1 = 0.033$, $\beta^c_2 = 0.033$, $\beta^c_3 = -0.007$, which are different from the coefficients in u_{ij}. This difference captures that, in reality, consumers might consider product features differently when clicking versus when purchasing the product. For example, we assume the price coefficients are -0.007 and -0.017 for clicking and purchasing, respectively, which intuitively reflects that customers may click expensive products but may be more sensitive to prices when purchasing.

The purchasing utility shocks $\boldsymbol{\epsilon}$ are simulated based on Σ, which is set by taking an inverse of a sparse precision matrix (with a density around 10%, reflecting the real-world example in Section 5.2) with its nonzero off-diagonal entry positions randomly determined. The resulting baseline Σ has 25% nonzero elements. We later perform sensitivity analyses to study the impact of Σ's density. We simulate the unobserved terms in clicking utilities $\boldsymbol{\epsilon}^c$ from $\boldsymbol{\epsilon}$ as guided by Equation (4), where their linkage weights ($\boldsymbol{w}$) are randomly determined for each product with an average of 0.6 in the baseline setup. Intuitively, the level of $\boldsymbol{w}$ governs the extent to which preclick information is predictive of purchasing utility shocks. The baseline level corresponds to a "click-to-purchase R^2" of 75%, representing the proportion of variations in ϵ attributed to ϵ^c. This setup is derived from empirical analyses conducted based on data from a major online e-commerce platform.[6] We adjust $\boldsymbol{w}$ (by setting the w values for some products toward zero) and the corresponding R^2 in later sensitivity analyses to evaluate their effects.

The resulting signal-to-noise ratio (SNR) in consumer purchasing utility for the baseline setup is around two. In generating the clicking incidences, we assume that consumer i clicks on product j if that consumer's clicking utility exceeds a clicking cost c. This cost c, which moderates the overall consumer clicking intensity, is calibrated such that, on average, a consumer clicks around two products in the baseline setup.[7] We later conduct sensitivity analyses regarding this clicking intensity.

4.1.2. Alternative Simulation Setups. Apart from the baseline setup, we also consider extensive variations of it in alternative simulation settings. These variations help us understand how our model performs under different contexts. In particular, we experiment with different (1) signal-to-noise ratios (increase or decrease by 50%), (2) sample sizes (4,000, 6,000, and 15,000), (3) density of the variance-covariance matrix (10% and 50%), (4) numbers of products (50, 80 and 100), (5) clicking intensities (increase or decrease the average number of clicks to 1.1 and 2.5), (6) click-to-purchase R^2 governed by w (set w values toward zero for an increasing number of products so that the corresponding R^2 values are 60%, 45%, and 15%), and (7) we further explore setups in which there exists a "nonidiosyncratic information gap" between the clicking and purchasing phases, elaborated further later.

Recall that our model captures the correlation between the unobserved utility terms ϵ and ϵ^c in the clicking and purchasing phases. In the main model, we assume the orthogonal deviation, e, between ϵ and ϵ^c (recall from Equation (4)) can be attributed to idiosyncratic shocks that are independent across products. In alternative simulation scenarios, we further explore the relationship between ϵ and ϵ^c and consider a nonidiosyncratic information gap from clicking to purchasing. In this context, e is no longer assumed to be independent across products and instead contains correlational patterns not present in Σ^c. Intuitively, we simulate scenarios where consumers might miss crucial interproduct connections during the clicking phase but learn them in the purchasing phase. We examine three particular scenarios characterized by nonidiosyncratic information gaps of 25%, 50%, and 75%. These scenarios are visualized in Appendix EC.8 Figure EC.2.

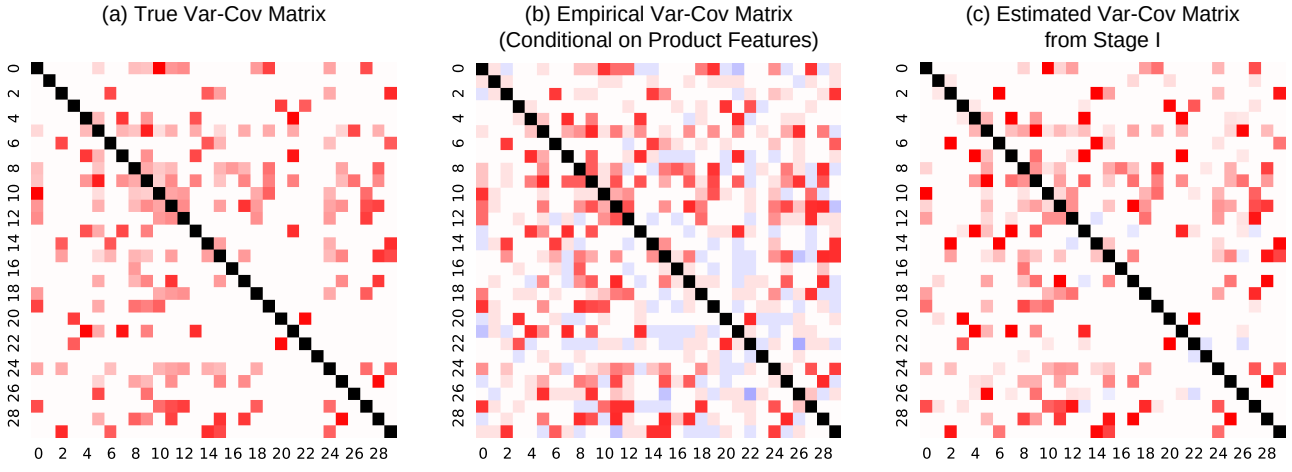
Lastly, contrasting scenarios in (6) and (7) is informative. Whereas (7) represents scenarios where key product connections are overlooked in clickstream data, (6) represents scenarios where information learned from clickstream becomes redundant. Specifically, in (6), we lower click-to-purchase R^2 by adjusting the levels of w for many products toward zero, resulting in many key connections in the clicking phase becoming inconsequential during the purchasing phase. For visualizations for (6), refer to Appendix EC.8 Figure EC.3.

4.2. Stage 1 Estimation Performance

We now discuss the stage 1 performance. Recall that in stage 1, we leverage clickstream data and estimate a sparse precision matrix $\hat{\Phi}^c$ among clicking utility shocks. To demonstrate its performance and contribution to stage 2, we compare the estimated click-phase variance-covariance matrix $\hat{\Sigma}^c$ $(= (\hat{\Phi}^c)^{-1})$ against the true click-phase Σ^c in underlying data-generation process.

Figure 2 illustrates the stage 1 performance visually. From left to right, we present (1) the underlying click-phase variance-covariance matrix (Σ^c), (2) the variance-covariance matrix of consumer clicking dummies conditional on observed product features (S_{zz}, which is an intermediate step in stage 1 as recalled from Equation (7)), and (3) the estimated click-phase variance-covariance matrix ($\hat{\Sigma}^c$) using our approach from stage 1. As shown in the figure, the variance-covariance matrices $\hat{\Sigma}^c$ and Σ share a similar structure, indicating our approach successfully recovers the underlying interproduct connections through utility shocks. See Appendix EC.9 for further discussions of stage 1 performance across various scenarios.

Figure 2. (Color online) Figure Illustration of Stage 1 Performance



Notes. This figure demonstrates the model's performance in estimating the variance-covariance matrix. The tuning parameter is selected using adaptive CV with inverse square weights. The zero elements in the matrix are colored white. Red colors represent positive elements. Blue colors represent negative elements. The darker the color, the larger the absolute value of the corresponding element. In true Σ^c, we specify only positive covariance values among the utility shocks of products, in line with the empirical observations presented in Figure EC.8. Nevertheless, from a theoretical standpoint, the performance insights would remain applicable even if the covariance entries were negative.

Table 2. Stage 1 Performance in Recovering the Click-Phase Variance-Covariance Matrix Σ^c

	Recovering nonzero entries		Recovering entry magnitudes	
Simulation setting	False nonzero (%) (1)	False zero (%) (2)	Correlation (3)	R^2 (%) (4)
Main setup	3.8	1.8	0.965	93.1
Signal-to-noise ratio				
SNR − 50%	2.9	1.8	0.964	92.9
SNR + 50%	5.9	0.9	0.964	93.0
Density of Σ				
=50%	6.8	5.7	0.959	92.0
=10%	4.9	0.0	0.979	95.8
Sample size				
=4,000 (20 days)	11.2	4.5	0.948	89.8
=6,000 (30 days)	5.3	2.7	0.961	92.4
=15,000 (75 days)	0.9	0.9	0.969	93.9
Clicking intensity (number of clicks)				
=1.1	2.4	4.5	0.956	91.5
=2.5	3.5	1.8	0.969	93.8
Number of products				
=50	1.3	2.9	0.970	94.0
=80	0.5	6.5	0.972	94.4
=100	0.0	5.6	0.971	94.3

Notes. The table reports the correlation between estimated and true values within Σ^c (in Column-(1)), the corresponding *R*-square (in Column (2)). Columns (1) and (2) report false nonzero rate and false zero rate. False nonzero (false zero) measures the percentage of zero (nonzero) entries in the Σ^c matrix labeled as nonzero (zero) by our method.

Table 2 details our stage 1 performance on two fronts. Firstly, it shows our approach's ability to accurately recover key interproduct connections by identifying which entries in Σ^c are nonzero (in columns (1) and (2)). Secondly, it demonstrates our approach's ability to capture the magnitude of each entry in Σ^c by calculating the correlation between estimated and true values within Σ^c (in column (3)) and the R^2 measuring the extent to which our estimated values explain variations of the true values in Σ^c (in Column (4)). Overall, our approach performs well on both fronts, demonstrated by the low false nonzero and false zero rates, along with a correlation nearing one and an R^2 approaching 100%.[8] This also confirms that we can use click dummies to estimate Σ^c in stage 1. Note that these results are calculated based on the adaptive CV tuning parameter approach, which tends to be the most effective tuning parameter selection method in our setting. Details of alternative approaches are discussed in Appendix EC.5. To demonstrate the stage 1 convergence, we illustrate how performance measures change under different sample sizes in Appendix EC.10.

Next, we discuss how the stage 1 performance varies in different simulation settings. For *SNR*, our approach performs well with higher and lower SNRs. We notice that the false nonzero rate is slightly higher with a higher SNR. This is expected because in cases where the noise (unobserved shocks) is relatively low compared with the signal (observed features), it is morechallenging to detect and estimate the variance-covariance matrix of the unobserved shocks. For *Density of* Σ, our approach maintains good performance with fewer (10%) and more (50%) nonzero entries in the variance-covariance matrix. For *Sample Size*, our approach performs well even when the sample size is reduced by 40% (to 6,000 observations). As anticipated, when the sample size is further reduced, the error rates begin to increase more noticeably. On the other hand, as the sample size increases, the model's performance further enhances. For *Clicking Intensity (governed by c)*, our approach performs relatively well under different clicking intensities. We tend to have higher false-zero rates in cases with lower clicking intensities. This is intuitive: when fewer products are clicked on, there are fewer clicking observations, thus leading to less accurate estimates. For *Number of Products*, our approach is able to scale up to handle a larger choice set.

In this stage, we do not analyze the sensitivity around the click-to-purchase R^2, nor the implication of the nonidiosyncratic information gap. This is because the aforementioned sensitivities affect only stage 2 performance (to be discussed later), but not that of stage 1.

4.3. Stage 2 Estimation Performance

Given the estimated click-phase $\hat{\Sigma}^c$, we proceed to stage 2. Here, we use consumer purchasing data to estimate the choice model, including the sensitivities

($\boldsymbol{\beta}$) to observed product characteristics as well as the weights ($\boldsymbol{w}$) linking unobserved utility shocks between clicking and purchasing phases (and the diagonal elements in Σ). We compare the in- and out-of-sample performance of our model with alternative, commonly used choice models: (1) the standard probit model with IID utility shocks (i.e., IID Probit), (2) a probit model with relaxed diagonal entries in its variance-covariance matrix Σ (i.e., Probit w/Diag), (3) a probit model with a fully relaxed Σ without utilizing the stage 1 inputs (i.e., Probit w/Full), and (4) the standard logit model (i.e., IID Logit). In alternative Models (1)–(3), we keep the probit specification to ensure comparability with our model. Nevertheless, we include alternative Model (4) to ensure the comparison will not be affected if we use logit models. We begin with results from the baseline scenario and will turn to alternative scenarios later in this subsection.

We estimate all five models and compare the estimation results. The sample used for estimation is described in Section 4.1. We generate another random sample of the same size and using the same model parameters for out-of-sample tests. We report the estimation results in Table 3. As shown in the table, most models produce accurate estimates for $\boldsymbol{\beta}$, yielding values close to the true $\boldsymbol{\beta}^{\text{true}}$ from the underlying data generation process, with Probit w/Full and IID Logit models as notable exceptions. The $\boldsymbol{\beta}$ estimated from IID Logit appears significantly different, yet upon closer examination, it essentially scales the true value by a nearly consistent factor. This discrepancy is anticipated, given that probit and logit models operate on different scales, making only the normalized $\boldsymbol{\beta}$ relevant. For the Probit w/Full model, the relaxation of numerous elements in the variance-covariance matrix (with 30 products, there are $30 + (30 \times 29)/2 - 1 = 464$ parameters in Σ) seems to compromise its precision in determining the true scale for $\boldsymbol{\beta}$.

Regarding these models' performance in fitting the data as reflected in the log likelihood, our method consistently outperforms most alternative models, achieving a higher log-likelihood value. This suggests that our method better explains in-sample individual choices. Whereas the Probit w/Full outperforms us in in-sample fitting, largely because of its substantially larger number of parameters, its out-of-sample performance is significantly weaker, evidenced by an out-of-sample log likelihood even lower than that of the IID Probit.

More importantly, we are interested in how well our model predicts the demand for each product. This could be a crucial foundation for important decisions such as assortment and pricing. We measure demand estimation error under model m using mean absolute percentage error (MAPE) and mean absolute error (MAE):

$$\text{MAPE}^m = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{J}\sum_{j=1}^{J}\frac{|D^m_{j,t} - D^a_{j,t}|}{D^a_{j,t}}, \text{ and}$$

$$\text{MAE}^m = \frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{J}|D^m_{j,t} - D^a_{j,t}|, \tag{10}$$

where $D^a_{j,t}$ summarizes the total actual demand of product j in period t, and $D^m_{j,t}$ is its counterpart predicted by

Table 3. Estimation Results from Our Model and Benchmark Models (Under the Main Synthetic Setup)

Panel A: Parameter estimates

	True	Our model	Probit w/Diag	IID Probit	Probit w/Full	IID Logit
β_1	0.100	0.101 (0.002)	0.100 (0.001)	0.107 (0.002)	0.135 (0.021)	0.255 (0.004)
β_2	0.167	0.169 (0.006)	0.166 (0.002)	0.175 (0.002)	0.227 (0.018)	0.418 (0.005)
β_3	−0.017	−0.017 (0.001)	−0.017 (0.000)	−0.018 (0.000)	−0.022 (0.002)	−0.041 (0.000)
Log likelihood	—	−13,605	−13,738	−13,756	−13,426	−13,901
Time (per search, in minutes)	—	5.1	2.6	0.1	40	0.1
Number of parameters	—	61	32	3	467	3

Panel B: Out-of-sample model performance

	True	Our model	Probit w/Diag	IID Probit	Probit w/Full	IID Logit
Log likelihood	−13,668	−13,722	−13,794	−13,778	−13,874	−13,935
MAPE (%)	15.52	15.72	18.80	18.80	17.46	26.71
Additional error	—	(1.26%)	(21.09%)	(21.12%)	(12.48%)	(72.09%)
MAE	12.43	13.34	14.24	13.82	15.00	17.11
Additional error	—	(7.32%)	(14.59%)	(11.18%)	(20.75%)	(37.70%)

Notes. Numbers in the parentheses in panel A are standard errors of the estimates. Percentages in the parentheses in panel B are increases of a model's MAPE relative to the true model's MAPE. For example, the MAPE of our model is 1.26%(= (15.72% − 15.52%)/15.52%) higher than the MAPE of the true model.

model m. In particular, with N consumers arriving on each day,

$$D^a_{j,t} = \sum_{i=1}^{N} d_{ij,t}\ , \text{ and } \quad D^m_{j,t} = \sum_{i=1}^{N} \Pr(y^m_{ij,t} = 1), \tag{11}$$

where $d_{ij,t}$ indicates whether consumer i arriving at time t purchases product j or not in the synthetic data, and $\Pr(y^m_{ij,t} = 1)$ is the conditional choice probability predicted by model m.

The out-of-sample demand prediction errors are reported in panel B of Table 3. Column (1) suggests the demand prediction accuracy under the true model, which helps us understand the *irreducible error*, that is, the inherent variability of the setting that no model can perfectly capture. Even with knowledge of the data-generation process, achieving perfect day-to-day demand prediction is unattainable, as the specific realizations of error terms in individual consumers' utilities are unobservable. In sum, our model demonstrates an out-of-sample demand prediction performance that is very close to the true model and significantly better than the alternative models. In particular, our model has an out-of-sample MAPE of 15.72%, which is only slightly larger than the true model's 15.52%. To emphasize this achievement, we compute the additional error relative to the irreducible error inherent in the true model. In terms of MAPE, our model incurs an extra error of merely 1.26% $\left(= (15.72\% - 15.52\%)/15.52\%\right)$, significantly lower than the IID Probit model (21.12%), the Probit w/Diag (21.09%), the Probit w/Full (12.48%), and the IID Logit (72.09%).

More crucial than mere aggregate demand prediction metrics is a model's ability to accurately represent the underlying product substitution patterns and guide subsequent operational decisions. As outlined in Section 3.2, this capability hinges on the estimation of a flexible Σ. Hence, even when overall prediction improvements sometimes appear modest, the importance of precisely estimating Σ cannot be overstated. A comprehensive discussion on this topic is in Section 6.

Before delving into that, we discuss below how the stage 2 performance varies in different simulation scenarios. Table 4 compares the out-of-sample demand prediction performance (in MAPE and MAE) under various models across scenarios. In this and subsequent discussions, we primarily compare our model with the IID Probit and Probit w/Diag. We omit comparisons with IID Logit because of its lack of direct comparability, and Probit w/Full is excluded because of the intensive computational resources it demands, coupled with its limited potential for out-of-sample performance.

Signal-to-Noise Ratio. Our model performs well with higher and lower SNRs, consistently outperforming benchmark models. With a lower SNR, the performance gap between our model and benchmarks slightly broadens—with an increased impact from the unobserved utility terms, capturing the intricate connections through these terms becomes increasingly crucial.

Density of Σ. With a denser Σ, representing greater interconnectivity among products through ϵ, the advantages of our approach become more pronounced. In contrast, when Σ is sparse, our advantage diminishes.

Nonidiosyncratic Information Gap. As the information gap expands (i.e., more key connections are missed from clicks), the demand prediction error of our model grows. When viewed through the lens of the additional error beyond the true model's MAPE and MAE, in the primary setting without any information gap, the error is only 1%; in an alternative setting with a 75% information gap, this error climbs to 15%. In sum, our approach's advantage diminishes as the nonidiosyncratic information gap widens. Yet even when only 25% of crucial information is learned from clicks (corresponding to a 75% information gap), our method still outperforms benchmark models.

Click-to-Purchase R^2. Our advantage diminishes as we lower the click-to-purchase R^2 (governed by w). With a 15% click-to-purchase R^2, our performance advantage over benchmark models is fairly narrow. This decline is expected, as a diminishing click-to-purchase R^2 indicates a rise in redundant information from clicks. The model comparisons at $R^2 = 15\%$ appear analogous to situations with a low Σ density, as the connections identified from clicks eventually prove to be inconsequential in the purchase phase. Recall from Section 4.1 that our main synthetic setup aligns more closely with empirical evidence, suggesting that the performance gain in some real-world settings is likely to resemble results observed under a higher R^2 (e.g., $> 60\%$).

Other Alternative Settings. Our model consistently performs well and outperforms benchmark models under varying clicking intensities. Furthermore, it effectively scales to manage larger choice sets with an increased number of products.

In sum, we demonstrate that our model outperforms several classic models in terms of the out-of-sample demand prediction. It is worthwhile to note that the underlying data-generation process used in this section is generic and flexible. Alternative models (such as IID Probit and Probit w/Diag) are special cases of our more generalized data-generating process. In fact, when these alternative models are the underlying truth and we simulate data accordingly, our method offers comparable performance to the underlying true models. For instance, when data are simulated based on the IID Probit model, our method can recognize that

Table 4. Stage 2 Performance in Demand Prediction Accuracy

	MAPE (%)				MAE			
Simulation setting	True (1)	Our (2)	With Diag (3)	IID Prob. (4)	True (5)	Our (6)	With Diag (7)	IID Prob. (8)
Main setup	15.52	15.72 (1%)	18.80 (21%)	18.80 (21%)	12.43	13.34 (7%)	14.24 (15%)	13.82 (11%)
Signal-to-noise ratio								
SNR − 50%	21.11	22.48 (6%)	26.76 (27%)	27.86 (32%)	15.92	17.14 (8%)	18.73 (18%)	18.66 (17%)
SNR + 50%	12.23	12.37 (1%)	14.29 (17%)	14.13 (16%)	10.56	11.30 (7%)	11.77 (11%)	11.48 (9%)
Density of Σ								
=50%	13.56	14.04 (4%)	17.54 (29%)	19.44 (43%)	11.64	12.79 (10%)	13.79 (18%)	14.12 (21%)
=10%	17.02	17.65 (4%)	18.86 (11%)	19.47 (14%)	13.37	13.53 (1%)	14.11 (6%)	14.14 (6%)
Clicking intensity (number of clicks)								
=1.1	15.52	15.81 (2%)	18.98 (22%)	19.06 (23%)	12.43	13.48 (8%)	14.29 (15%)	13.84 (11%)
=2.5	15.52	15.51 (0%)	18.80 (21%)	18.80 (21%)	12.43	13.23 (6%)	14.24 (15%)	13.82 (11%)
Number of products								
=50	10.47	10.72 (2%)	12.07 (15%)	13.41 (28%)	12.41	12.96 (4%)	14.07 (13%)	14.88 (20%)
=80	7.47	7.61 (2%)	8.34 (12%)	9.98 (34%)	12.89	13.81 (7%)	15.01 (16%)	16.43 (27%)
=100	6.14	6.50 (6%)	7.04 (15%)	8.40 (37%)	11.62	12.67 (9%)	13.86 (19%)	15.46 (33%)
Click-to-purchase R^2								
=60%	15.34	15.59 (2%)	19.12 (25%)	18.91 (23%)	12.39	13.44 (8%)	14.30 (15%)	14.82 (20%)
=45%	15.34	15.65 (2%)	16.76 (9%)	17.97 (17%)	13.74	13.76 (0%)	14.23 (4%)	14.52 (6%)
=15%	17.14	17.13 (0%)	18.25 (6%)	18.55 (8%)	14.01	14.43 (3%)	14.60 (4%)	14.42 (3%)
Nonidiosyncratic information gap								
=25%	15.52	16.52 (6%)	18.80 (21%)	18.80% (21%)	12.43	13.62 (10%)	14.24 (15%)	13.82 (11%)
=50%	15.52	17.29% (11%)	18.80 (21%)	18.80% (21%)	12.43	13.91 (12%)	14.24 (15%)	13.82 (11%)
=75%	15.52	17.88 (15%)	18.80 (21%)	18.80 (21%)	12.43	14.01 (13%)	14.24 (15%)	13.82 (11%)

Notes. The table reports MAPE (%) and MAE as defined in Equation (10) for various models. The percentages in parentheses represent the increase of a model's MAPE (MAE) relative to the true model's MAPE (MAE).

the variance-covariance matrix is an identity matrix and has a similar performance as the IID Probit model.

5. Empirical Evidence and Application

In this section, we provide empirical evidence of sparse precision matrices from several real-world settings. We then demonstrate how to apply our method to a data set from a leading international online retailer and discuss the estimation results.

5.1. Empirical Evidence for a Sparse Precision Matrix

An assumption of our approach is the precision matrix among products' utility shocks is sparse. We believe this is a plausible assumption in many online retail settings and provide empirical supports for it in three representative contexts. Because we cannot directly observe the products' utility shocks, we provide empirical evidence using consumer clickstream data. In particular, we look at consumer *coclick* behavior measured by the *click correlation matrix*, which reflects the degree to which two products are clicked together by the same consumer. In Figure EC.7 in Appendix EC.12, we plot the click correlation matrices calculated based on three real-world empirical settings: (1) a leading international online retailer (to be introduced in Section 5.2), (2) a major U.S. online furniture retailer, and (3) a major U.S. online hotel booking company. As can be seen from Figure EC.7, the click correlation matrices in all

three settings are fairly sparse, which supports our assumption that the underlying precision matrix is likely sparse. The reason is that if many different pairs of products all had significant nonzero correlations among their unobserved shocks, we would observe more uniform coclicking behavior across product pairs. That is not what we observe from Figure EC.7. This makes intuitive sense: the coclicks are likely generated from some shared underlying properties. Only an extremely large number of different shared underlying characteristics can support such a large number of different coclick combinations, which is unlikely.

5.2. An Empirical Application

5.2.1. Empirical Setting and Data. We now illustrate how we can apply our method to a real empirical data set. We obtain detailed information on product prices, product features, and customer clickstream and purchase decisions from a large international retailer. The data are generated on the mobile app of the sponsor's e-commerce platform. Specifically, the data set contains the clicking and purchasing history of all users searching kitchen appliances through the retailer's mobile app platform during an observation period from April 1–14, 2018.[9] There are 46 products with at least one impression during the study period and 32,217 consumers. For each consumer impression, the data set includes the product's features and its price at the time of the impression. For this empirical analysis, we define a session as a day.

Similar to other retail settings, we observe a large portion of consumers make no purchase: around 96.11% of the consumers leave without buying anything. It is likely that some consumers who did not purchase anything are just casual "browsers." With most of the consumers being browsers, it is difficult to illustrate our model's performance—there is little room left for any demand forecasting improvement, as one can just forecast that the consumers are not going to purchase anything and achieve a fairly high accuracy. To alleviate this issue, we balance our sample, allowing us to focus on studying how well our model recovers the substitution patterns among all the offered products. After balancing, on average, a consumer clicks on 0.595 products after seeing the initial product display. The percentiles of the number of clicks per consumer are reported in Appendix EC.12.[10] Among all consumers in our final sample, 1,254 purchase an item.

In our analysis, we focus on the following product features as observed Xs: (1) consumer preference score ($score_{ij,t}$), which is the individual-specific consumer preference score calculated based on the *i*2*i* algorithm, which intuitively represents the likelihood that a consumer will purchase this product based on that consumer's previous activities on the platform calculated by a machine learning algorithm; (2) display record ($display_{j,t}$), which is the number of times the product is displayed historically; (3) product price ($price_{j,t}$); (4) the number of good ratings ($\#good\_rating_{j,t}$); (5) the percentage of good ratings ($\%good\_rating_{j,t}$); and (6) the popularity score ($popular_{j,t}$). The summary statistics for these product features are reported in Table EC.4 in Appendix EC.12.

5.2.2. Estimation Results. We apply our method to this data set. We first estimate the Φ^c (and its corresponding Σ^c) using clickstream data; details are reported in Appendix EC.12. There are around 10% nonzero elements in the estimated precision matrix. Next, we estimate the parameters in the choice model as well as the weights linking the clicking and purchasing phases. Results from our method are presented in the first column of Table 5. We also estimate the parameters using alternative models: the IID Probit model and the Probit w/Diag model. The results are shown in the last two columns in Table 5. We also compare different models' out-of-sample demand prediction performance using MAPE and MAE and report them in panel B of Table 5. As can be seen, our model offers lower out-of-sample demand prediction errors. For example, the out-of-sample MAE of our model is 32.2, which is significantly lower than that under the IID Probit model (49.3) and the Probit w/Diag (49.4).

6. Managerial Implications

We have so far illustrated the efficacy of our approach in both simulated and real-world empirical contexts. This methodology has the potential to better guide firms in various business decisions, from assortment planning to inventory management and pricing. In this section, we discuss the primary advantage of our model in enhancing operational decisions: its ability to more accurately capture product substitution patterns. Such accuracy enables businesses to adapt effectively to changes in related products, thereby enhancing operational decisions. Specifically, we demonstrate that our approach recovers more precise product substitution patterns in comparison with benchmark models. The substitution patterns can be represented by the own- and cross-price elasticities. We begin by outlining the method used to compute these elasticities.

An estimated model m predicts demand for product j as D_j^m, as recalled from Equation (11). (In this section, we omit t in the notation for simplicity.) Based on the predicted demand, we can derive the model-implied substitution patterns. Specifically, we calculate the own-price and cross-price elasticities, reflecting how a change in one product's price influences its demand and the demand of other products. Let e_{kj} denote the sensitivity of product j's demand in response to changes in product k's price. For

Table 5. Parameter Estimates and Model Performance on the Empirical Example

Panel A: Parameter estimates

	Full model		IID Probit		Probit w/Diag	
β_{price}	−0.059	(0.003)	−0.041	(0.001)	−0.055	(0.002)
$\beta_{display}$	2.733	(0.946)	4.683	(0.705)	3.489	(0.899)
β_{score}	8.757	(0.471)	6.699	(0.229)	7.307	(0.267)
$\beta_{\#_{good\_rating}}$	6.251	(1.339)	6.432	(1.034)	6.158	(1.404)
$\beta_{\%_{good\_rating}}$	−11.550	(0.877)	−11.619	(0.648)	−11.613	(0.969)
$\beta_{popular}$	4.421	(0.701)	1.513	(0.500)	4.409	(0.609)
Log likelihood	−1,652.6		−1,859.9		−1,659.7	

Panel B: Out-of-sample model performance

	Full model	IID Probit	Probit w/Diag
MAE	32.185	49.347	49.364
MAPE	13.41%	27.28%	16.07%

Notes. The β estimates make intuitive sense: consumer utilities decrease with product prices and increase with the website recommendation intensity (*display*), the match between individual consumer and the product (*score*), the number of good ratings #*good_rating*. The coefficient $\beta_{\%_{good\_rating}}$ is negative, which is counterintuitive at the first sight, yet this actually makes intuitive sense: when controlling for #*good_rating*, a higher %*good_rating* means a lower total number of ratings, which might reduce consumer purchasing utilities. We calculate the out-of-sample MAE based on Equation (10), with N (the number of consumers per day) set to 200.

any given demand model m, we can write

$$e_{kj}^m = \frac{\%\Delta D_j^m}{\%\Delta p_k} = \frac{\Delta D_j^m}{\Delta p_k} \cdot \frac{p_k}{D_j^m}. \tag{12}$$

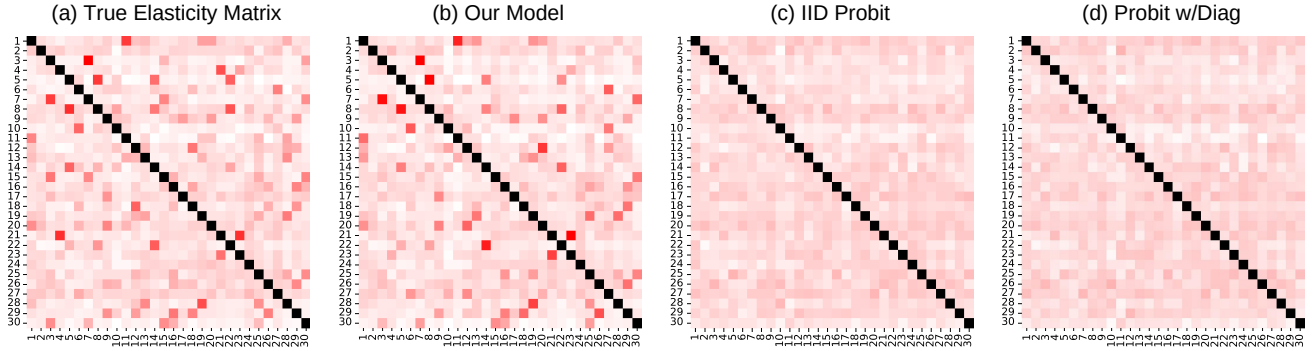
Let E^m summarize the price elasticity matrix under model m, with the kj^{th} entry being e_{kj}^m.

Figure 3 provides a visual comparison of the elasticity matrices. Figure 3(a) represents the elasticity matrix under the true model (E^{True}), Figure 3(b) represents our model, and Figure 3, (c) and (d) represents the benchmark models. Comparing Figure 3, (a) and (b) reveals that our model effectively captures the fundamental patterns present in the true elasticity matrix. In contrast, Figure 3, (c) and (d) deviate significantly from the true model. Notably, the entries in Figure 3, (c) and (d) display a striking uniformity, a consequence of the inherent constraints in Σ under these alternative models.

To quantitatively evaluate the accuracy of our model against benchmark models across various simulation scenarios, we compute the deviations in estimated elasticity matrices from the "true" elasticity matrix (E^{True}). For this analysis, we use an absolute percentage distance metric to quantify the relative deviations: $\%Dist^m = (1/J^2) \cdot \sum_{k=1}^{J} \sum_{j=1}^{J} (|e_{kj}^m - e_{kj}^{True}| / |e_{kj}^{True}|)$. Our conclusions remain robust when employing alternative measures, such as Euclidean distance and non-percentage-based absolute distance. We calculate the overall $\%Dist^m$ across all product pairs and further dissect this measurement separately for only own-price elasticity and cross-price elasticity.

The results are presented in Table 6. As shown, elasticity matrices recovered by our model align more

Figure 3. (Color online) Elasticity Matrices (Standardized) Recovered Under Different Models



Note. To standardize each entry in all four elasticity matrices, we divide the corresponding entry in the original elasticity matrices by the geometric mean of the averages of its corresponding column and row.

Table 6. The Distance from the Model Recovered Elasticity Matrix to the True Matrix

	Distance between all pairs (%)			Own price distance (%)			Cross price distance (%)		
Simulation setting	Our	With Diag	IID Prob	Our	With Diag	IID Prob	Our	With Diag	IID Prob
Main	17.8	33.5	34.0	3.3	4.1	4.9	18.3	34.5	35.1
Signal-to-noise ratio									
SNR − 50%	17.6	39.7	40.2	2.7	4.0	4.4	18.1	41.0	41.5
SNR + 50%	19.3	29.2	28.3	3.7	3.9	5.0	19.8	30.1	29.1
Density of Σ									
=50%	26.2	38.5	41.4	4.4	4.5	6.6	26.9	39.7	42.7
=10%	14.8	18.2	17.8	2.8	2.9	2.8	15.2	18.7	18.3
Clicking intensity (number of clicks)									
=1.1	20.4	33.7	34.3	4.0	4.1	5.0	21.0	34.8	35.4
=2.5	16.7	33.5	34.0	3.4	4.1	4.9	17.2	34.5	35.1
Number of products									
=50	25.2	36.9	37.5	5.0	5.7	7.8	25.6	37.6	38.1
=80	25.5	34.5	38.7	5.0	6.1	10.4	25.7	34.9	39.0
=100	25.1	34.0	38.2	5.7	6.4	10.9	25.3	34.3	38.5
Click-to-purchase R^2									
=60%	16.8	33.6	34.4	2.2	4.1	4.9	17.4	34.6	35.4
=45%	19.7	26.7	27.7	3.0	3.2	3.8	20.3	27.6	28.5
=15%	15.7	16.3	16.8	3.2	2.0	2.5	16.2	16.9	17.3
Nonidiosyncratic information gap									
=25%	21.5	33.5	34.0	3.8	4.1	4.9	22.1	34.5	35.1
=50%	25.0	33.5	34.0	3.4	4.1	4.9	25.8	34.5	35.1
=75%	28.6	33.5	34.0	3.5	4.1	4.9	29.5	34.5	35.1

Notes. The table reports the absolute percentage distance between a model-implied elasticity matrix and the true elasticity matrix. "Between all pairs" considers all elasticities among J products; "own price" considers only own-price elasticity, namely, e^m_{kj} when $k = j$; and "cross price" considers only cross-price elasticity, namely, e^m_{kj} when $k \neq j$. All distance values are expressed as percentages in this table.

closely with the true matrix. Notably, the absolute percentage distance in our models is roughly half of those in the IID Probit and Probit w/Diag models. The most pronounced enhancement can be seen in the accurate recovery of cross-price elasticity.

Upon examining the comparison across various simulation settings, we observe that our model accurately recovers substitution patterns in most situations. As anticipated, the exception occurs when clicks are less predictive of purchases or when there is a substantial nonidiosyncratic information gap. As the click-to-purchase predictability decreases, there is increasing redundant information from clicks. As nonidiosyncratic information gap widens, there are increasing interproduct connections missing from clicks. In both cases, the advantages of our model naturally diminish. Nonetheless, even in these circumstances, as long as some relevant information can be learned from clickstream data, our model can still manage to uncover some key substitution patterns, as shown in Appendix EC.11.

In sum, the significant enhancement achieved by our model in accurately recovering the true substitution pattern serves as a critical foundation for optimizing business strategies across various areas such as inventory management, assortment planning, and pricing management. Examples illustrating how a better substitution pattern translates into better operational decisions, including inventory management and pricing strategies, are provided in Appendix EC.13.

7. Conclusion

To improve demand estimation and subsequent operational decisions, an important first step is to achieve a realistic understanding of the substitution patterns among products. Yet it is challenging to do so among a large number of products and in the complex online environment. We propose a methodology that combines a novel machine learning method (namely, graphical lasso) with classic choice models to tackle this challenge. There are three major innovations in our method. First, we leverage consumer clickstream data to learn products' connections through utility shocks. Second, we introduce the graphical lasso method to the choice modeling framework to help identify accurate substitution patterns. Third, we propose a framework that links the click and purchase utility shocks. This framework allows us to transform a quadratic estimation problem into a linear one under mild assumptions, and it is generalizable to various real-world contexts.

Our method performs well in a wide range of synthetic scenarios in providing more accurate demand forecasts as well as own-/cross-product elasticity estimates. Compared with classical choice models, our method consistently offers more accurate out-of-sample demand forecasts. For example, under the primary synthetic setup, our model's out-of-sample MAPE is nearly the same as the true model (which represents the best achievable performance) and is significantly lower than the IID Probit model, the probit with relaxed diagonal

elements, and even the probit model with fully relaxed variance-covariance elements. Applying our method to a real online retail setting, we show that our method continues to offer more accurate demand forecasts. Most importantly, we demonstrate that our method recovers more precise product substitution patterns. Specifically, the absolute percentage deviation of our estimated elasticity matrix from the true model's is about half of those of the IID Probit and the diagonal probit models. With more accurate substitution estimates, our method provides a critical input to enhance operational decision making in a variety of applications, including inventory management and pricing decisions.

Practically, our proposed method can be readily applied by online retailers to various business settings with large choice sets. This includes demand estimation for inventory and transshipment decisions, as well as a variety of other operational decision areas such as promotion and assortment. We find that capturing products' connections via the graphical lasso approach will lead to a more accurate demand estimation and more accurate own-price and cross-price elasticity estimates. This highlights the importance of incorporating information learned from consumers' clicking activity as well as combining state-of-the-art machine learning methods with the choice modeling framework.

Our analysis is, of course, not without limitations. First, this study focuses on demand forecasting at the product level, which assists retailers in making important operational decisions such as demand forecasting and product pricing. It would be interesting to see how clickstream information could help inform demand estimation for each consumer, which could help retailers make personalized promotion and assortment decisions. Second, we learn the substitution pattern by observing the consumer clickstream activities among the offered products, which cannot be applied to new products. An important extension is to learn substitution patterns among new and existing products based on preexisting products. This could help retailers determine what new products to introduce, as well as the optimal ordering quantities and prices for newly introduced products. Last, whereas our method relies on consumers' click data in online retailing, it would be interesting and important to investigate how to modify our methods to suit settings where clicks are completely unobservable, such as offline retailing or in service industries.

Endnotes

[1] We do not explicitly consider the product availability information in our model. If such information is available to researchers, there are ways to incorporate it into the model: we can either (1) change consumers' choice sets over time to remove out-of-stock products, or (2) adjust the prices for out-of-stock products to be infinite to capture the fact that consumers cannot purchase out-of-stock products.

[2] See the Adobe Digital Index 2020 report at https://www.slideshare.net/adobe/adi-consumer-electronics-report-2020, and research from Episerver retail clients at https://www.episerver.com/reports/2019-b2c-ecommerce-benchmark-report/.

[3] Note that the product(s) consumers click/purchase may also depend on the retailer's recommendation system. One can capture the impact of recommendation systems by including related variables in Xs in Equation (1). We show an empirical example in Section 5 where the aggregate effect of recommendation systems is explicitly controlled for. Estimating such a model helps tease out substitutions driven by recommendation systems. However, in cases where there are no obvious ways to explicitly control for the impact of recommendation systems, the results should be taken as a given if the system remains unchanged. One would need to reestimate the model if the system changes.

[4] Note that we do not directly incorporate individual consumers' clicking decisions in the stage 2 choice model for three reasons. First, we would like to retain the generality of our method without making structural assumptions about individual consumers' search processes and how insights from their clicking behavior could be employed to refine their choice sets in the purchasing phase. Second, our method is constructed to facilitate policy simulations for operational decisions such as pricing. If clickstream data are essential for the stage 2 estimation, we would need to simulate all possible subsets of products clicked by a consumer at various price vectors, posting a combinatorial problem. Lastly, using only purchase data and not making additional assumptions would allow us to make fair comparisons between our model and other choice models. It allows us to better assess the value of estimating a flexible Σ rather than ignoring unobserved substitutability or making assumptions about it.

[5] Note that we intentionally choose to use only few observed features in this synthetic experiment. This choice helps reflect that, in reality, practitioners or researchers often do not get to observe many factors impacting consumers' decisions. Without properly accounting for these features explicitly in the mean utility part, these features are left in unobserved terms. This is a key motivation of our approach—with key features left in products' unobserved terms, the variance-covariance matrix among products' unobserved terms would exhibit important substitution patterns.

[6] In Appendix EC.2, we explore the extent to which purchase behavior can be explained by clicking behavior in this major online e-commerce platform. Our analysis indicates that a significant portion of explainable variations in purchase decisions are captured by clicking decisions.

[7] This is motivated by many papers studying consumer behavior in online retailing environment: for example, Amano et al. (2019) find that "on average, a search session contains 2.45 products"; Chen and Yao (2017) find that, in their data, "495 consumers made a total of 1,140 click-throughs, with an average of 2.30 click-throughs per consumer."

[8] This performance is fairly impressive given that a randomly guessed $\hat{\Sigma}^c$ with the same density as ours would have a false nonzero (zero) rate as high as 75% (25%).

[9] There are two reasons why we choose the kitchen appliances category as our example application. First, the application of the discrete choice model typically requires a single-choice setting. The category of kitchen appliances is suitable, as most consumers purchase zero or one product. We observe that fewer than 0.1% of consumers purchase more than one product from this category in a day. We therefore exclude them from the analyses. A second reason the kitchen appliance category is a good fit is that consumers purchase kitchen appliances infrequently. Unlike in other categories

such as consumer packaged goods, consumers usually do not stockpile kitchen appliances. In particular, we observe merely 3.86% of consumers have impressions on multiple days. For the purpose of our analysis, we treat the same consumer on different days as different consumers.

[10] Note that the clicking intensity in this empirical setting is lower than in many other online retail settings. This is not too surprising because these data come from a mobile app setting. A consumer's cost of clicking a product might be higher when browsing happens through a mobile as compared with a desktop—it is more difficult to open multiple tabs and compare products in a mobile app environment. In other online retail settings with higher clicking intensity, we would expect an even stronger performance gain from using our approach.

References

Allen EA, Damaraju E, Plis SM, Erhardt EB, Eichele T, Calhoun VD (2014) Tracking whole-brain connectivity dynamics in the resting state. *Cerebral Cortex* 24(3):663–676.

Amano T, Rhodes A, Seiler S (2019) Large-scale demand estimation with search data. Harvard Business School Working Paper 19-022, Harvard Business School, Boston.

Banerjee O, Ghaoui LE, d'Aspremont A (2008) Model selection through sparse maximum likelihood estimation for multivariate Gaussian or binary data. *J. Machine Learning Res.* 9:485–516.

Bray RL, Stamatopoulos I (2022) Menu costs and the bullwhip effect: Supply chain implications of dynamic pricing. *Oper. Res.* 70(2): 748–765.

Cai TT, Li H, Liu W, Xie J (2013) Covariate-adjusted precision matrix estimation with an application in genetical genomics. *Biometrika* 100(1):139–156.

Chen Y, Yao S (2017) Sequential search with refinement: Model and application with click-stream data. *Management Sci.* 63(12): 4345–4365.

Chiong KX, Shum M (2019) Random projection estimation of discrete-choice models with large choice sets. *Management Sci.* 65(1):256–271.

Cohen MC, Zhang R, Jiao K (2022) Data aggregation and demand prediction. *Oper. Res.* 70(5):2597–2618.

Dotson JP, Howell JR, Brazell JD, Otter T, Lenk PJ, MacEachern S, Allenby GM (2018) A Probit model with structured covariance for similarity effects and source of volume calculations. *J. Marketing Res.* 55(1):35–47.

Ertekin N, Agrawal A (2020) How does a return period policy change affect multichannel retailer profitability? *Manufacturing Service Oper. Management* 23(1):210–229.

Feldman J, Zhang DJ, Liu X, Zhang N (2022) Customer choice models vs. machine learning: Finding optimal product displays on Alibaba. *Oper. Res.* 70(1):309–328.

Ferreira KJ, Lee BHA, Simchi-Levi D (2016) Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing Service Oper. Management* 18(1):69–88.

Fisher M, Gallino S, Li J (2018) Competition-based dynamic pricing in online retailing: A methodology validated with field experiments. *Management Sci.* 64(6):2496–2514.

Fox JT (2007) Semiparametric estimation of multinomial discrete-choice models using a subset of choices. *RAND J. Econom.* 38(4):1002–1019.

Friedman J, Hastie T, Tibshirani R (2008) Sparse inverse covariance estimation with the graphical lasso. *Biostatistics* 9(3):432–441.

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

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

Glaeser CK, Fisher M, Su X (2019) Optimal retail location: Empirical methodology and application to practice: Finalist-2017 *M&SOM* practice-based research competition. *Manufacturing Service Oper. Management* 21(1):86–102.

Gomez-Rodriguez M, Leskovec J, Krause A (2012) Inferring networks of diffusion and influence. *ACM Trans. Knowledge Discovery Data* 5(4):1–37.

Hammersley JM, Clifford P (1971) Markov fields on finite graphs and lattices. Unpublished manuscript, University of Oxford, Oxford, UK.

Jagabathula S, Subramanian L, Venkataraman A (2018) A model-based embedding technique for segmenting customers. *Oper. Res.* 66(5):1247–1267.

Kesavan S, Kushwaha T, Gaur V (2016) Do high and low inventory turnover retailers respond differently to demand shocks? *Manufacturing Service Oper. Management* 18(2):198–215.

Kim JB, Albuquerque P, Bronnenberg BJ (2010) Online demand under limited consumer search. *Marketing Sci.* 29(6):1001–1023.

Lee TY, Bradlow ET (2011) Automated marketing research using online customer reviews. *J. Marketing Res.* 48(5):881–894.

Lee J, Gaur V, Muthulingam S, Swisher GF (2016) Stockout-based substitution and inventory planning in textbook retailing. *Manufacturing Service Oper. Management* 18(1):104–121.

Mankad S, Shunko M, Yu Q (2019) How to find your most valuable service outlets: Measuring influence using network analysis. Preprint, submitted April 4, http://dx.doi.org/10.2139/ssrn.3366127.

McFadden D, Train K (2000) Mixed MNL models for discrete response. *J. Appl. Econometrics* 15(5):447–470.

Netzer O, Feldman R, Goldenberg J, Fresko M (2012) Mine your own business: Market-structure surveillance through text mining. *Marketing Sci.* 31(3):521–543.

Ngwe D, Ferreira KJ, Teixeira T (2019) The impact of increasing search frictions on online shopping behavior: Evidence from a field experiment. *J. Marketing Res.* 56(6):944–959.

Ringel DM, Skiera B (2016) Visualizing asymmetric competition among more than 1,000 products using big search data. *Marketing Sci.* 35(3):511–534.

Rothman AJ, Levina E, Zhu J (2010) A new approach to Cholesky-based covariance regularization in high dimensions. *Biometrika* 97(3):539–550.

Rothman AJ, Bickel PJ, Levina E, Zhu J (2008) Sparse permutation invariant covariance estimation. *Electr. J. Statist.* 2:494–515.

Ruiz FJ, Athey S, Blei DM (2020) Shopper: A probabilistic model of consumer choice with substitutes and complements. *Ann. Appl. Statist.* 14(1):1–27.

Smith AN, Allenby GM (2019) Demand models with random partitions. *J. Amer. Statist. Assoc.* 115(529):47–65.

Smith AN, Rossi PE, Allenby GM (2019) Inference for product competition and separable demand. *Marketing Sci.* 38(4):690–710.

Train KE (2009) *Discrete Choice Methods with Simulation* (Cambridge University Press, New York).

Wan M, Wang D, Goldman M, Taddy M, Rao J, Liu J, Lymberopoulos D, McAuley J (2017) Modeling consumer preferences and price sensitivities from large-scale grocery shopping transaction logs. *Proc. 26th Internat. Conf. World Wide Web* (International World Wide Web Conferences Steering Committee, Geneva), 1103–1112.

Yai T, Iwakura S, Morichi S (1997) Multinomial probit with structured covariance for route choice behavior. *Transportation Res. Part B Methodological* 31(3):195–207.