Demand estimation under multi-store multi-product substitution in high density traditional retail

Citation for published version (APA):

DOI:
10.1016/j.ejor.2017.09.014

Document status and date:
Published: 01/04/2018

Document Version:
Accepted manuscript including changes made at the peer-review stage

Please check the document version of this publication:
- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl
providing details and we will investigate your claim.

Download date: 15. Sep. 2023
Demand Estimation Under Multi-Store Multi-Product Substitution in High Density Traditional Retail

Mingchao Wan\textsuperscript{a}, Yihui Huang\textsuperscript{a}, Lei Zhao\textsuperscript{a}, Tianhu Deng\textsuperscript{a}, Jan C. Fransoo\textsuperscript{b}

\textsuperscript{a}Department of Industrial Engineering, Tsinghua University, Beijing, China
\textsuperscript{b}School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, Netherlands

Abstract

In large cities in emerging economies, traditional retail is present in a very high density, with multiple independently owned small stores in each city block. Consequently, when faced with a stockout, consumers may not only substitute with a different product in the same store, but also switch to a neighboring store. Suppliers may take advantage of this behavior by strategically supplying these stores in a coherent manner. We study this problem using consumer choice models. We build two consumer choice models for this consumer behavior. First, we build a Nested Logit model for the consumer choice process, where the consumer chooses the store at the first level and selects the product at the second level. Then, we consider an Exogenous Substitution model. In both models, a consumer may substitute at either the store level or the product level. Furthermore, we estimate the parameters of the two models using a Markov chain Monte Carlo algorithm in a Bayesian manner. We numerically find that the Nested Logit model outperforms the Exogenous Substitution model in estimating substitution probabilities. Further, the information on consumers’ purchase records helps improve the estimation accuracies of both the first-choice probabilities and the substitution probabilities when the beginning inventory level is low. Finally, we show that explicitly including such substitution behavior in the inventory optimization process can significantly increase the expected profit.

Keywords: Retailing; Traditional retail; Demand estimation; Demand substitution; Consumer choice model

1. Introduction

Compared with modern organized retail such as supermarkets and hypermarkets, traditional retail has received little attention from the research community. In developed countries, the market share of traditional retail is small, with less than 10\% in the grocery sector of Europe and North America (Nielsen 2015). However, this is very different in emerging markets and other developing countries, with market share of around 50\% on average across Latin America and Asia, and well above 90\% in some large countries such as Nigeria and India (Nielsen 2015). This market share has remained remarkably stable over the past decade, despite considerable economic growth. Blanco and Fransoo (2013) argue that these mom-and-pop operated and independently owned small retail stores (“nanostores”) are likely...
to remain a large factor in grocery retailing due to the socio-economic and demographic structure, especially in very large cities. For instance, in a city as developed as Beijing that has become over the past decades, small retail stores are still surviving in large numbers. Based on our estimates with a number of suppliers, Beijing still has about 30,000 small retail stores in the core area of the city (within the 5th ring), and about double this number in the entire city jurisdiction. Beijing residents can easily find small retail stores in their neighborhoods. In fact, nanostores are so dense in Beijing that it is not uncommon to have multiple nanostores around one residential neighborhood. The consequence of this convenience is that consumers often not only substitute between products, but also substitute between stores. Then, a consumer who finds her favorite product unavailable may switch to a different product at the same store or simply visit another nearby small retail store. Intuitively, consumers exhibit two layers of demand substitution: multi-store substitution and multi-product substitution. This type of two-layer demand substitution is yet understudied in the literature.

For many products, suppliers such as Coca Cola, Danone, or Unilever, deliver directly to the small retail stores. Investigating the aforementioned multi-store multi-product substitution behavior can help these suppliers better make assortment and inventory decisions and improve their sales performance in this important retail channel. We aim to build a consumer choice model that accounts for this behavior and propose a statistical estimation procedure. Intuitively, the Nested Logit model (NL) is an ideal choice for the demand model as it can naturally model the two-layer substitution: substitution between stores and substitution between products. Therefore, we build a general Nested Logit model where a consumer first selects which store to visit and then decides which product to purchase. The substitution behavior is endogenously modeled. In addition to the Nested Logit model, the two-layer substitution behavior can also be captured by an exogenous substitution matrix. Therefore, we also build an Exogenous Substitution model (ES) to describe the two-layer substitution behavior.

The main contribution of our paper is embodied in three aspects. First, we investigate the understudied consumer behavior of multi-store multi-product substitution, which exists in traditional small retail stores. We apply two suitable consumer choice models: the Nested Logit model and the Exogenous Substitution model. Second, we estimate the model parameters using a Markov chain Monte Carlo algorithm in a Bayesian manner. Finally, we numerically compare the two models and find that the Nested Logit model outperforms the Exogenous Substitution model in estimating substitution probabilities when we only have aggregate sales data. Further, we study the impact of estimation accuracy on the optimal beginning inventory levels and expected profits and find that including substitution behavior significantly increases the expected profit when the profit margin is low.

The rest of this paper is structured as follows: we briefly review the related literature in Section 2. In Section 3, we introduce the two consumer choice models for multi-store and multi-product substitution. Section 4 formulates the likelihood function based on sales data. Then, Sections 5 and 6 present the estimation algorithm and numerical results, respectively. We offer some concluding remarks in Section 7.
2. Literature Review

The body of research related to this work can be divided into three broad sets. The first set includes papers that consider substitution effects in assortment planning problems. The second set predicts a consumer’s behavior when facing a stockout using some predictors, e.g., store loyalty and product loyalty. The third set studies the demand and substitution estimation problem based on different types of sales data and estimation methods.

There is extensive literature on assortment planning problems that consider product substitution. Kök et al. (2009) and Shin et al. (2015) give excellent reviews of the recent papers on assortment planning and its applications. The models of consumer choice and substitution behavior can be classified into three main types:

(1) The utility-based models (e.g., the Multinomial Logit (MNL) model and the Nested Logit (NL) model). Consumers assign a certain utility to each product and choose products with probabilities computed from those utilities. Utility-based models allow consumers to substitute from the unavailable products to all available products. van Ryzin and Mahajan (1999) study the assortment planning and inventory optimization problems under MNL. They show that the optimal assortment always consists of a certain number of the most popular products. Recently, Kök and Xu (2011) study joint assortment optimization and pricing problems under two-level NL (product level and brand level). They show that the optimal assortment of product types within each brand is utility-ordered.

(2) The locational choice model. The locational choice model is commonly used to model consumers’ choice behavior. Each product is viewed as a set of attributes (e.g., price, flavor, size). Consumers choose the product that is nearest to their most preferred point in the attribute space. If the first-choice product is out-of-stock, consumers substitute to other products in an increasing order of their distances from the first-choice product. Gaur and Honhon (2006) consider the assortment planning and inventory optimization problem under the locational choice model and present a heuristic and an upper bound based on a polynomial-time approximation.

(3) The Exogenous Substitution model. Consumers in this model substitute to product $j$ from $i$ with exogenous substitution probability $\alpha_{ij}$ if product $i$ is unavailable. Kök and Fisher (2007) study the assortment optimization problem with the exogenous substitution model and propose an iterative optimization heuristic. Hübner et al. (2016) develop this problem by proposing a new efficient algorithm. Fisher and Vaidyanathan (2014) also study the assortment optimization problem under the exogenous substitution model. But in their model, the choice and substitution probabilities are all at attribute-level.

Most of the papers on assortment planning do not show how to estimate the purchase and substitution parameters in their model except for Kök and Fisher (2007) and Fisher and Vaidyanathan (2014). Kök and Fisher (2007) estimate substitution probabilities with inventory-transactions data (data recorded at every sale and replenishment). Fisher and Vaidyanathan (2014) use a Maximum Likelihood Estimation (MLE) method to estimate demands and substitution probabilities with aggregate sales data.
The second stream of literature predicts a consumer’s response to stockouts, including in-store substitution, cross-store substitution, deferment, and cancellation. Campo et al. (2000) build a MNL model with three types of predictors: product characteristics (e.g., product importance, pack size), consumer characteristics (e.g., store loyalty, shopping frequency), and situation characteristics (e.g., time constraint, required product quantity). Then they fit the model with data collected by questionnaires and estimate the probability of each response. Zinn and Liu (2001) extend their work by using different predictors such as perceived store characteristics and demographics.

The third stream of literature studies the problem of estimating demand and substitution behavior when considering stockouts, which is the most related to our work. Table 1 categorizes the related literature according to five dimensions: available data, consumer arrival, first-choice, substitution, and methodology.

The second column in Table 1 indicates the available data for demand estimation. There are two types of available data: store-level data, including inventory levels at the beginning and end of each period; consumer-level data, including the purchase records of all consumers and the product availability throughout all the periods. The following three columns identify how each paper models a consumer’s behavior by three aspects: arrival, first-choice, and substitution. Most papers use one model to describe both first-choice and substitution behavior (e.g., Conlon and Mortimer 2013, Musalem et al. 2010, Talluri and van Ryzin 2004, van Ryzin and Vulcano 2014, Vulcano et al. 2010, 2012). However, Kök and Fisher (2007) and Fisher and Vaidyanathan (2014) use exogenous substitution probabilities to allow a more flexible substitution structure. The last column states the methodology of estimation.
<table>
<thead>
<tr>
<th>Research Source</th>
<th>Available Data Type</th>
<th>Consumer</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anupindi et al. (1998)</td>
<td>Store-level data</td>
<td>Poisson</td>
<td>Exogenous prob.</td>
</tr>
<tr>
<td>Conlon and Mortimer (2013)</td>
<td>Store-level data</td>
<td>None</td>
<td>Random-coefficient</td>
</tr>
<tr>
<td>Fisher and Vaidyanathan (2014)</td>
<td>Store-level data</td>
<td>None</td>
<td>Exogenous prob.</td>
</tr>
<tr>
<td>Jain et al. (2015)</td>
<td>Both</td>
<td>Stochastic Demand</td>
<td>One Product (No Choice and Sub.)</td>
</tr>
<tr>
<td>Musalem et al. (2010)</td>
<td>Store-level data</td>
<td>None</td>
<td>Random-coefficient</td>
</tr>
<tr>
<td>van Ryzin and Vulcano (2014)</td>
<td>Consumer-level data</td>
<td>Bernoulli</td>
<td>Rank-based model</td>
</tr>
<tr>
<td>Vulcano et al. (2010)</td>
<td>Consumer-level data</td>
<td>Bernoulli</td>
<td>MNL</td>
</tr>
<tr>
<td>Vulcano et al. (2012)</td>
<td>Store-level data</td>
<td>Poisson</td>
<td>MNL</td>
</tr>
</tbody>
</table>

\(^1\)E-M is short for Expectation-Maximization algorithm.
\(^2\)MLE is short for Maximum Likelihood Estimation.
\(^3\)MCMC is short for Markov chain Monte Carlo algorithm.
Anupindi et al. (1998) is the first work to estimate primary demand considering stockouts and consumer substitution behavior. They assume that consumer arrivals follow a Poisson distribution and that choice probabilities are exogenous. They develop the estimation model with Expectation-Maximization (E-M) algorithm under one stockout and two stockouts in each period when only store-level data are available. However, they need to estimate the choice probabilities for every possible set of available alternatives. Consequently, the main limitation of their method is that the number of parameters to be estimated grows rapidly with the number of stockouts in each period.

Musalem et al. (2010) is the closest to our paper. They study the estimation method for the effect of stock-out based substitution with store-level data, which can be accessed by periodic inventory review systems. Their model combines Markov chain Monte Carlo (MCMC) with sampling using a Bayesian method to simulate the transition of the inventory and then estimate the effect of substitution and lost sales.

Our paper differs from Musalem et al. (2010) in several aspects. First, we capture consumers’ multi-store multi-product substitution behavior using the Nested Logit model and the Exogenous Substitution model, while Musalem et al. (2010) use a Random-coefficient model. Second, in addition to demand estimation, we also investigate the value of information on purchase records and the impact of estimation accuracy on inventory policy and expected profit.

3. Model

In this section, we first describe the basic model setting, which includes multiple stores and multiple products over a finite horizon. Then, we build two consumer choice models, respectively. Finally, we formulate the likelihood function based on store-level data. Tables 2 and 3 summarize the major notation.

3.1. Basic Model Setting

We consider a market with $I$ small retail stores. Without loss of generality, we assume all stores offer $J$ substitutable products. Let $I = \{1, 2, \ldots, I\}$ represent the set of all stores and $J = \{1, 2, \ldots, J\}$ represent the set of all possible products. We define $C = I \times J$ as the set of all possible store-product combinations. For example, $C = \{(1, 1), (1, 2), (2, 1), (2, 2)\}$ represents the case where there are two stores both offering two products. We also define set $C' = C \cup \{(0, 0)\}$ to include the no-purchase option. Throughout this paper, we use subscript $i$ to denote a “store”, and subscript $j$ to denote a “product”.

3.2. Consumer Choice: Nested Logit Model

We assume a pool of finite number of consumers. Consumers have independent choices. Each consumer’s choice follows the Nested Logit model (see Figure 1 for the two-store two-product case). They first choose at store level and then at product level. In Appendix A, we also consider the case where a consumer first chooses at product level and then at store level. Let $\nu = (v_{1,1}, v_{1,2}, \ldots, v_{1,J}, v_{2,1}, \ldots, v_{I,J})$ be the preference weight vector for all the store and product combinations. Furthermore, we assume that all the stores have the same dissimilarity parameter $\gamma$, which is
Table 2: Notation (available data and missing data)

**Available data**

- **I** = \{1, 2, \ldots, I\}, the set of all stores.
- **J** = \{1, 2, \ldots, J\}, the set of all products.
- **C** = \{(i, j) | i \in I, j \in J, product \ j \ is \ offered \ at \ store \ i\}, the set of all possible combinations of stores and products.
- **C'** = \(C \cup \{(0, 0)\}\), the set of all possible alternatives, including no-purchase option (0, 0).
- **T** = Total number of periods.
- **N_t** = Total number of consumers in period \(t\).
- **b_{tij}** = The beginning inventory of product \(j\) at store \(i\) in period \(t\).
- **e_{tij}** = The ending inventory of product \(j\) at store \(i\) in period \(t\).
- **s_{tij}** = \(b_{tij} - e_{tij}\), the total sales of product \(j\) at store \(i\) in period \(t\).
- **B** = \(\{b_{tij}, t \in \{1, \ldots, T\}, i \in I, j \in J\}\), the set of all beginning inventory data.
- **S** = \(\{s_{tij}, t \in \{1, \ldots, T\}, i \in I, j \in J\}\), the set of all sales data.

**Missing data**

- **a_{tn}** = \(a_{tn_11}, \ldots, a_{tn_{IJ}}\), where \(a_{tnj}\) is an indicator variable taking 1 only when consumer \(n\) in period \(t\) finds product \(j\) at store \(i\) available.
- **A** = \(\{a_{tn}, t \in \{1, \ldots, T\}, n \in \{1, \ldots, N_t\}\}\), the set of all \(a_{tn}\).
- **w_{tn}** = \(w_{tn_00}, w_{tn_11}, \ldots, w_{tn_{IJ}}\), where \(w_{tnj}\) is an indicator variable taking 1 only when consumer \(n\) in period \(t\) buying product \(j\) at store \(i\).
- **W** = \(\{w_{tn}, t \in \{1, \ldots, T\}, n \in \{1, \ldots, N_t\}\}\), the set of all purchase record.

Table 3: Notation (parameters to be estimated)

**Parameters to be estimated**

- **\(\nu\)** = \((\nu_1, \ldots, \nu_J)\), the preference vector.
- **\(\gamma\)** = the store dissimilarity parameter.
- **\(p_{tij}\)** = The probability of a consumer preferring product \(j\) at store \(i\) when all the products are available.
- **\(P_{tnj}\)** = The probability that consumer \(n\) in period \(t\) purchases product \(j\) at store \(i\).
- **\(\alpha_{tij\rightarrow i'j'}\)** = The substitution probability from product \(j\) at store \(i\) to product \(j'\) at store \(i'\).
similar to the model of Berry (1994), Cardell (1997) and Conlon and Mortimer (2013). Following the Nested Logit model, the probability of a consumer choosing product \( j \) at store \( i \) as her first-choice is given by

\[
p_{ij}(v, \gamma) = \frac{\left( \sum_{j' \in J} \exp(v_{ij'}) \right)^{\gamma} \times \exp(v_{ij})}{1 + \sum_{i' \in I} \left( \sum_{j' \in J} \exp(v_{i'j'}) \right)^{\gamma - 1}}
\]

(1)

However, the products may not always be available. We consider a total of \( T \) periods. Let \( N_t \) be the number of consumers arrived in period \( t \). Consumers arriving in period \( t \) are labelled from 1 to \( N_t \) by their arrival order. We characterize the product availability by an indicator vector \( a_{tn} = (a_{tn1}, \ldots, a_{tnIJ}) \), where \( a_{tni j} = 1 \) if consumer \( n \) in period \( t \) finds that product \( j \) is available at store \( i \) and \( a_{tni j} = 0 \) otherwise. When considering the availability, the probability that consumer \( n \) in period \( t \) purchases product \( j \) at store \( i \) becomes

\[
P_{tnij}(v, \gamma | a_{tn}) = \frac{\left( a_{tni j} \exp(v_{ij}) \right) \left( \sum_{j' \in J} a_{tni j'} \exp(v_{i'j'}) \right)^{\gamma - 1}}{1 + \sum_{i' \in I} \left( \sum_{j' \in J} a_{tni j'} \exp(v_{i'j'}) \right)^{\gamma - 1}}.
\]

(2)

In the Nested Logit model, choice probability varies with the set of available products. In other words, substitution behavior is considered intrinsically in the Nested Logit model. When only one alternative (say product \( j \) at store \( i \)) is unavailable, we can compute the “equivalent” substitution probability from product \( j \) at store \( i \) to product \( j' \) at store \( i' \) by:

\[
\alpha_{ij \rightarrow i' j'} = \frac{P_{ti' j'}(v, \gamma | a_{tn}) - p_{ij}}{p_{ij}},
\]

(3)

where \( P_{i' j'}(v, \gamma | a_{tn}) \) is the choice probability of product \( j' \) at store \( i' \) when only product \( j \) at store \( i \) is out-of-stock. When multiple stockouts occur simultaneously, we cannot distinguish the unsatisfied demands and thus cannot get the corresponding substitution probabilities.
3.3. Consumer Choice: Exogenous Substitution Model

In another model, we model the consumer shopping behavior as a two-step choice process (see Figure 2). In the first step, a consumer decides her first-choice shopping store and product without knowing the availability. Let the choice probability of choosing product $j$ at store $i$ be $p_{ij}$. The same as in the Nested Logit model, we also include no-purchase option in the first step of the exogenous substitution model. In the second step, the consumer immediately makes purchase if her first-choice is available. Otherwise, she may choose to buy product $j'$ at store $i'$ with exogenous substitution probability $\alpha_{ij \rightarrow i'j'}$, or just walk away without any purchase. We assume that if the substitute product $(i', j')$ is also out-of-stock, then the sale is lost permanently, i.e. there is no second substitution attempt. Many studies in this field make the same single-attempt assumption, such as Smith and Agrawal (2000), Netessine and Rudi (2003), Kök (2003), and Kök and Fisher (2007). Kök (2003) proves that a multi-attempt substitution model can be approximated with a single-attempt model by a higher substitution probability.

The probability that consumer $n$ in period $t$ purchases product $j$ at store $i$ now consists of two parts: the first-choice part and the substitution part from other unavailable products, and is specified as follows:

$$P_{ni}(p, \alpha|a_n) = \begin{cases} p_{ij} + \sum_{(i', j') \in C} (1 - a_{ni}) \alpha_{i'j' \rightarrow ij}, & a_{ni} = 1; \\ 0, & a_{ni} = 0. \end{cases} \label{eq:exogenous-substitution}$$

The probability that consumer $n$ in period $t$ does not purchase anything is $P_{n00}(p, \alpha|a_n) = 1 - \sum_{(i,j) \in C} P_{ij}(p, \alpha|a_n)$.

Compared with traditional utility-based models (e.g., Musalem et al. 2010, Conlon and Mortimer 2013), the Exogenous Substitution model provides a more flexible substitution structure by introducing an exogenous substitution matrix $\alpha$, which is similar to Kök and Fisher (2007). The key difference between our model and the model in Kök and Fisher (2007) is the modeling of the first-choice probabilities. In our model, the first-choice probabilities are exogenous. But Kök and Fisher (2007) use a Multinomial logit framework to model the first-choice probabilities.
3.4. The Difference Between Nested Logit Model and Exogenous Substitution Model

The key difference between the Nested Logit model and the Exogenous Substitution model is the substitution structure. In the Nested Logit model, the substitution structure is determined by the utilities and dissimilarity parameters. Consequently, the substitution structure of the Nested Logit model has the following two main features. First, the substitution probabilities are correlated with the first-choice probabilities. Second, all substitution probabilities should be nonzero. Next, we explain these two features in detail.

In the Nested Logit model, any change of model parameters would lead to the simultaneous change of first-choice probabilities and substitution probabilities. Figure 3 illustrates the simultaneous change under the two-store two-product setting. The change of utility of product 1 at store 1 (Figure 3a) and the change of the store dissimilarity parameter (Figure 3b) both lead to the simultaneous change of the first-choice probability of no-purchase option \(p_{00}\), the in-store substitution probability \(\alpha_{11\rightarrow12}\), and the cross-store substitution probability \(\alpha_{11\rightarrow21}\). We observe that the first-choice probabilities and substitution probabilities are correlated in the Nested Logit model. On the contrary, the first-choice probabilities \(p\) and substitution probabilities \(\alpha\) can be determined separately (or more flexibly) in the Exogenous Substitution model.

Figure 3: The effect of the Nested Logit model parameters on the first-choice and substitution probabilities

(a) The effect of utility \(v_{11}\)  
(b) The effect of store dissimilarity parameter \(\gamma\)

In addition, the substitution probabilities in the Nested Logit model, defined in (3), are always positive. It can be straightforwardly proved that any stockout would increase the choice probabilities of all the other alternatives in the Nested Logit model. In other words, \(p^{(ij)}_{rj}\) must be larger than \(p_{rj}\) in (3). Thus all the substitution probabilities in the Nested Logit model cannot be zero. But this may not be true in the real world.

The Exogenous Substitution model’s substitution structure has more freedom than that of the Nested Logit model. Yet the Exogenous Substitution model has more parameters to estimate. In our setting with \(I\) stores and \(J\) products, the Nested Logit model has \(I \times J\) utilities and one store dissimilarity parameter to estimate, while the Exogenous
Substitution model has $I \times J$ first-choice probabilities and $(I \times J)(I \times J - 1)$ substitution probabilities to estimate.

4. Likelihood Function Formulation

Since most traditional small retail stores do not have a formal management information system to record every transaction, the information we can obtain to formulate likelihood function is very limited. What we can observe in practice is only the inventory level of each product at the beginning and end of each period. The sales of product $j$ at store $i$ in period $t$ can be directly obtained by taking the difference of the beginning and end inventory. Let $b_{tij}$, $e_{tij}$ and $s_{tij}$ be the beginning inventory level, ending inventory level, and sales of product $j$ at store $i$ in period $t$, respectively.

In addition to the inventory data, we also use the total number of consumers in each period who decide to make purchase decisions, $N_t$, which includes the consumers choosing no-purchase option. Information about $N_t$ can be obtained from the following methods: (1) recording the total number of transactions in each period, see Kök and Fisher (2007); (2) using demographic information to estimate the number of consumers that would make purchase decisions, see Berry et al. (1995); (3) using market share to estimate the market size, see Vulcano et al. (2012). In the following, we proceed to use the store-level aggregate data to formulate the likelihood function and then estimate our model parameters $\theta$ ($\theta = (v, \gamma)$ in the Nested Logit model and $\theta = (p, \alpha)$ in the Exogenous Substitution model).

There have been two approaches in the literature to formulate the likelihood function using store-level data. The first one uses only sales information without concerning each consumer’s purchase record (Anupindi et al. 1998, Conlon and Mortimer 2013). They treat each stockout time as the missing data and then use the Expectation-Maximization algorithm to estimate the model parameters and missing data iteratively. The main drawback is that the likelihood function becomes complicated and hard to solve when there is more than one stockout in each period. The other approach formulates the likelihood function by consumer-level transaction data, and then sums up the likelihood of all possible transaction data that satisfy store-level aggregate data (Chen and Yang 2007, Musalem et al. 2009, 2010).

We formulate the likelihood function following the second approach. Let $w_{tn} = (w_{tn00}, w_{tn11}, \ldots, w_{tnIJ})$ be the choice indicator vector of consumer $n$ in period $t$, where $w_{tnij}$ takes 1 only when consumer $n$ in period $t$ purchases product $j$ at store $i$. Recall that indicator variable $a_{tnij}$ represents the availability of product $j$ at store $i$ when consumer $n$ arrives in period $t$. Therefore, $(a_{tn}, w_{tn})$ can capture the available set and final purchase decision of consumer $n$ in period $t$. The likelihood function of consumer $n$ in period $t$ can be written as:

$$L_{tn}(\theta|a_{tn}, w_{tn}) = \prod_{(i,j) \in C} \left[ P_{tnij}(\theta|a_{tn}) \right]^{w_{tnij}}. \quad (5)$$

The overall likelihood function is the product of (5) over all periods and all consumers:

$$L(\theta|A, W) = \prod_{t=1}^{T} \prod_{n=1}^{N_t} L_{tn}(\theta|a_{tn}, w_{tn}), \quad (6)$$

where $A = \{a_{tn}\}$ and $W = \{w_{tn}\}$. 

11
The overall likelihood function (6) is based on individual’s available set and purchase information, which are expressed by \((A, W)\). However, we cannot observe such detailed information directly in practice. Instead, what we obtain is store-level aggregate data: sales and beginning inventory of each product in each store. The set of all possible \((A, W)\) that are consistent with store-level data is

\[
\Omega(S,B) = \left\{(A, W) \mid \sum_{n=1}^{N_t} w_{tni,j} < b_{ti,j}, w_{tni,j} \leq a_{tni,j}\right\}, \tag{7}
\]

where \(S = \{s_{ti,j}\}\) and \(B = \{b_{ti,j}\}\). The first constraint in (7) is used to describe that product \(j\) at store \(i\) is available to consumer \(n\) only when its sales before \(n\) does not exceed the beginning inventory. The second constraint ensures that the choice indicator \(w\) is consistent with sales. The last constraint ensures that consumers can purchase product \(j\) at store \(i\) only when it is available. Summing over all possible \((A, W)\), the likelihood function conditioning on store-level aggregate data \((S, B)\) can be expressed as:

\[
L(\theta | S, B) = \sum_{(A, W) \in \Omega(S,B)} L(\theta | A, W) P(A, W | S, B), \tag{8}
\]

where \(P(A, W | S, B)\) is the probability of observing \((A, W)\) conditioning on \((S, B)\).

The traditional way to estimate model parameters \(\theta\) is directly seeking \(\theta^*\) that maximizes the likelihood function (8). However, since there are too many possible \((A, W)\) for a given \((S, B)\), it is unrealistic to calculate the summation over all possible \((A, W)\). Instead, we treat \((A, W)\) as latent parameters and then use a MCMC-based Bayesian to estimate the model parameter \(\theta\).

5. Estimation

In this section, we describe the MCMC estimation method (Algorithm 1). There are two types of parameters: the augmented data \((A, W)\) as latent data/parameters and the parameters of our consumer choice model \(\theta\) (i.e., \((v, \gamma)\) in the NL model, or \((p, \alpha)\) in the ES model).

At the beginning of the algorithm (Step 0), we initialize the values of the purchase record \((A, W)\) and the model parameters \(\theta\). We also initialize the iteration counter \(r\).

In the \(r^{th}\) iteration, we first update the purchase record \((A, W)\) given \(\theta^{(r-1)}\) (Step 1). The Gibbs sampling algorithm is an MCMC sampling algorithm that allows us to sample variables sequentially based on their conditional distributions when direct sampling is difficult. In this paper, instead of updating the entire purchase record simultaneously, we apply the Gibbs sampling algorithm to update a segment of the purchase record by swapping the choices of two consumers at a time (while maintaining those of other consumers unchanged) and repeat until the entire purchase record is updated (see Chen and Yang 2007).

More specifically, for each period \(t\), we first partition the total \(N_t\) consumers into \([N_t/2]\) pairs of consumers (Step 1.1). We then (in Step 1.2) swap the consumer choices in each pair of consumers to obtain a new purchase record (i.e., the available set \(A'\) and consumer choices \(W'\)). We use the Gibbs sampling method to sample between the two
purchase records (namely, \((A, W)\) and \((A', W')\)), and accept the proposed \((A', W')\) using the acceptance/rejection rule with the acceptance ratio as

\[
\Pr[(A', W')|\theta^{(r-1)}] = \frac{\Pr[(A', W')|\theta^{(r-1)}]}{\Pr[(A, W)|\theta^{(r-1)}] + \Pr[(A, W)|\theta^{(r-1)}]}
\]

In Appendix B, we provide an illustrative example on the partition of consumers and the generation of the new purchase record \((A', W')\). We repeat Steps 1.1 and 1.2 for all the periods \(t, t = 1, \ldots, T\) to update the entire purchase record \((A, W)\). Note that we drop the period index \(t\) for notation simplicity. We refer interested readers to Musalem et al. (2009, 2010) for more details.

In Step 2, given the purchase record \((A, W)\) updated in Step 1, we use a hybrid Gibbs sampling method (embedding the Metropolis-Hastings algorithm within Gibbs sampling) to sample the model parameters \(\theta\).

The Gibbs sampling method allow us to sample \(\theta\) (i.e., \(v\) and \(\gamma\) in the NL model, \(p\) and \(\alpha\) in the ES model) sequentially in each iteration. However, sampling \(\theta\) from their conditional posterior distributions is challenging. We use the independent Metropolis-Hastings algorithm to sample \(\theta\), in which the proposal distribution is set to be \(\pi(\theta)\). More specifically, we sample a proposal \(\theta'\) from \(\pi(\theta)\) and accept \(\theta'\) as \(\theta^{(r)}\) using the acceptance/rejection rule with the acceptance ratio as

\[
\Pr[\text{accept}] = \min \left\{ \frac{L(\theta'|A, W)}{L(\theta^{(r-1)}|A, W)}, 1 \right\}.
\]

Otherwise, we reject \(\theta'\) and set \(\theta^{(r)}\) to be \(\theta^{(r-1)}\). The derivation of Equation (10) is in Appendix C.

As mentioned above, to avoid the dimensionality issue, we update \(\theta\) sequentially in each iteration. Next, we use the estimation of the NL model parameters as an example, where we estimate \(v\) and \(\gamma\) sequentially.

Given the purchase record \((A, W)\) and store dissimilarity parameter \(\gamma^{(r-1)}\), the p.d.f. of the conditional posterior distribution of \(v\) is proportional to

\[
\pi(v) L(v, \gamma = \gamma^{(r-1)}|A, W).
\]

We set the proposal distribution to be \(\pi(v)\) and use the independent Metropolis-Hastings algorithm to sample a proposal \(v'\). The acceptance probability of \(v'\) is

\[
\Pr[\text{accept}] = \min \left\{ \frac{L(v', \gamma^{(r-1)}|A, W)}{L(v^{(r-1)}, \gamma^{(r-1)}|A, W)}, 1 \right\}.
\]

We set \(v^{(r)}\) to be \(v'\) if we accept it. Otherwise, we set \(v^{(r)}\) to be \(v^{(r-1)}\).

Next, given \((A, W)\) and \(v^{(r)}\), similarly we sample \(\gamma'\) from its prior distribution \(\pi(\gamma)\) and accept it with probability

\[
\Pr[\text{accept}] = \min \left\{ \frac{L(\gamma', v^{(r)}|A, W)}{L(\gamma^{(r-1)}, v^{(r-1)}|A, W)}, 1 \right\}.
\]

The sequential update of the model parameters \(\theta\) of the ES model is similar, where we first update the first-choice probabilities \(p\) and then the substitution probabilities from \(\alpha_{11}\) (short notation for \([\alpha_{11}, \alpha_{12}, \ldots, \alpha_{11-iJ}]\)), the substitution probability vector from product 1 at store 1 to other store-product combinations.
Algorithm 1 The simultaneous estimation algorithm

**Input:** \(S, B, N, \pi(\theta), R\) (Maximum number of iterations).

**Output:** \(\hat{\theta}\)

**Step 0.** Initialization.
- Set the prior distribution of \(\theta, \pi(\theta)\), to include researchers’ belief.
- Sample \(\theta^{(0)}\) according to \(\pi(\theta)\).
- Sample \((A, W) \in \Omega(S, B)\).
- Set the iteration counter \(r = 1\).

**Step 1.** Update the purchase record \((A, W)\) given \(\theta^{(r-1)}\).

For \(t = 1, \ldots, T\), do

**Step 1.1 Partition.**
Randomly partition the total \(N_t\) consumers into \(\lfloor N_t/2 \rfloor\) pairs.

**Step 1.2 Updating the purchase record \((A, W)\) given \(\theta^{(r-1)}\).**

For each pair of consumers in the partition, do
- Swap their purchase choices and obtain a new purchase record \((A', W')\).
- Sample RAND from the uniform distribution \(U(0, 1)\).
- If \(\text{RAND} < \text{Pr[accept]} = \frac{L(\theta^{(r-1)}|A', W')}{L(\theta^{(r-1)}|A', W') + L(\theta^{(r-1)}|A, W)}\)
  - Set \(A = A', W = W'\).

**Step 2.** Update \(\theta^{(r)}\) given \((A, W)\).

- Sample \(\theta'\) from the prior (proposal) distribution \(\pi(\theta)\).
- Sample RAND from the uniform distribution \(U(0, 1)\).
- If \(\text{RAND} < \text{Pr[accept]} = \min \left\{ \frac{L(\theta'|A, W)}{L(\theta^{(r-1)}|A, W)} + 1 \right\}\)
  - Set \(\theta^{(r)} = \theta'\).
- Else
  - Set \(\theta^{(r)} = \theta^{(r-1)}\).

**Step 3** Stopping rule.
- If \(r \leq R\)
  - Set \(r = r + 1\) and go to **Step 1**.
- Else (Termination)
  - Take the average of second half of \(\theta^{(r)}, r = R/2 + 1, \ldots, R\), as the estimation \(\hat{\theta}\),
    \[\hat{\theta} \leftarrow (\theta^{(R/2+1)} + \ldots + \theta^{(R)})/(R/2).\]
  - Calculate the empirical cumulative distribution of second half of \(\theta^{(r)}\), and then obtain the double-sided 95% confidence interval.
The estimation process terminates at the preset maximum number of iterations $R$. Similar to Chen and Yang (2007) and Musalem et al. (2009, 2010), we discard the first half of estimation results and take the average of the second half of $\theta^{(r)}$, $r = R/2 + 1, \ldots, R$ as $\hat{\theta}$.

6. Numerical Experiment

In this section, we perform numerical experiments to study the two demand estimation (consumer choice) models, namely, the Nested Logit model and the Exogenous Substitution model. First, we examine the estimation error of the proposed estimation procedure. Second, we study the value of information on consumer purchase records. Finally, we compare the two demand estimation models in terms of the optimal beginning inventory levels and expected profit. In the following, we start with the description of the experimental design.

6.1. Design of Experiment

The framework of the numerical experiment is shown in Figure 4. Given a specific setting on the consumer arrival rates and beginning inventory levels, the simulation module simulates consumers’ first-choice and substitution behavior. The output of the simulation model is each product’s sales at each store. The estimation module then estimates the parameters of a chosen demand estimation model (“belief”), based on which the optimization module solves for the optimal beginning inventory levels. Finally, the evaluation module evaluates the optimal beginning inventory levels under the true model (“reality”).

We would like to note that the consumer choice models in the estimation and optimization modules are consistent with each other. One chooses a consumer choice model based on his “belief” and estimates the model parameters and then uses the same model to optimize the beginning inventory levels accordingly. In addition, the consumer choice models in the simulation and evaluation modules are the same, since we want to use the “reality” to evaluate a given set of beginning inventory levels.
Based on what we discussed above, we construct the following four sets of experiments:

1. Simulation: NL $\rightarrow$ Estimation: NL $\rightarrow$ Optimization: NL $\rightarrow$ Evaluation: NL;
2. Simulation: NL $\rightarrow$ Estimation: ES $\rightarrow$ Optimization: ES $\rightarrow$ Evaluation: NL;

We present the estimation results in Section 6.2 and the inventory optimization and evaluation results in Section 6.3.

### 6.1.1. Demand Simulation

In the experiment, we study the situation where two small retail stores sell the same set of two products. We consider a multi-period setting with $T = 20$ periods. The total number of consumer arrivals in each period is drawn from a Poisson distribution with mean $\lambda = 200$. Consumer arrivals in different periods are independent of each other.

In Appendix D, we further consider a setting with two stores and three products.

We use two consumer choice models to describe a consumer’s behavior in the simulation process. In the Nested Logit model, the product utilities are $(v_{11}, v_{12}, v_{21}, v_{22}) = (0.5, 0, -0.5, -1)$ and the store dissimilarity parameter $\gamma$ is 0.4. In the Exogenous Substitution model, the first-choice probabilities are $(p_{11}, p_{12}, p_{21}, p_{22}) = (0.27, 0.16, 0.18, 0.11)$. The exogenous substitution matrix is

$$
\begin{pmatrix}
(1,1) & (1,2) & (2,1) & (2,2) \\
(1,1) & - & 0.66 & 0.11 & 0.06 \\
(1,2) & 0.72 & - & 0.09 & 0.05 \\
(2,1) & 0.15 & 0.09 & - & 0.59 \\
(2,2) & 0.13 & 0.08 & 0.66 & - \\
\end{pmatrix}
$$

To make the Nested Logit model and the Exogenous Substitution model comparable, we set the parameters of the two models to be consistent with each other. That is, the first-choice probabilities in the Nested Logit model and Exogenous Substitution model are exactly the same; the substitution matrix in the Exogenous Substitution model is parallel to the virtual substitution probabilities constructed by (3) in the Nested Logit model. Because of the existence of no-purchase option, the sum of first-choice probabilities and the sum of each row in the substitution matrix of the Exogenous Substitution model are both strictly less than 1. In addition, we also study a symmetric substitution matrix which cannot be well approximately by an NL model and present the experiment in Appendix E.

We summarize the setting of beginning inventory levels in the simulation module in Table 4. We investigate the impact of beginning inventory levels by reducing the beginning inventory level of (1, 1) (denoting product 1 at store 1) from 65 to 45 and fixing the beginning inventory levels of (1, 2), (2, 1) and (2, 2) at high inventory positions. In each simulation setting, the beginning inventory levels remain constant across all 20 periods. (Actually, our estimation procedure can be applied to any arbitrary inventory policies as long as each period’s beginning and ending inventory
levels can be observed. Here we fix the beginning inventory levels for simplicity.) For notation simplification, we omit the period index \( t \) and use \( b_{ij} \) to represent the beginning inventory level of \((i, j)\).

### Table 4: Beginning inventory levels in the simulation model

<table>
<thead>
<tr>
<th>Store-Product</th>
<th>Arrival Rate ((\lambda p_{ij}))</th>
<th>Beginning Inventory ((b_{ij}))</th>
<th>Corresponding Stockout Prob.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1, 1)</td>
<td>53.02</td>
<td>65, 60, 55, 50, 45</td>
<td>5%, 15%, 40%, 63%, 85%</td>
</tr>
<tr>
<td>(1, 2)</td>
<td>32.16</td>
<td>45</td>
<td>1%</td>
</tr>
<tr>
<td>(2, 1)</td>
<td>35.54</td>
<td>50</td>
<td>1%</td>
</tr>
<tr>
<td>(2, 2)</td>
<td>21.56</td>
<td>32</td>
<td>1%</td>
</tr>
</tbody>
</table>

6.1.2. Demand Estimation

In the demand estimation module, we estimate the demand parameters based on the assumed consumer choice models, e.g., either the Nested Logit model or the Exogenous Substitution model, using Algorithm 1.

We use weakly informative prior distributions. In the Nested Logit model, utility parameters \( v_{ij} \) are drawn from uniform distribution \( U(-2, 1) \), and store dissimilarity parameter \( \gamma \) is drawn from \( U(0.1, 0.9) \). In the Exogenous Substitution model, the prior distributions of the first-choice probabilities and substitution probabilities are all flat Dirichlet distributions. The total number of iterations in each estimation process is set to be 20,000 and the last 10,000 iterations are used for estimation.

6.2. Estimation Results

We report the estimation results of the Nested Logit model and the Exogenous Substitution model, under the assumptions that the underlying “reality” (simulation) follows either the Nested Logit model or the Exogenous Substitution model.

For the purpose of explicit comparison of the two estimation models, when the estimation model is the Nested Logit model, we convert the estimation results from the utilities and store dissimilarity to the first-choice probabilities and substitution probabilities, computed by (1) and (3). As \((1, 2), (2, 1)\) and \((2, 2)\) have a low stockout probability, we only present the estimation results of \(\alpha_{11\rightarrow12}, \alpha_{11\rightarrow21} \) and \(\alpha_{11\rightarrow22}\).

6.2.1. When Consumer Choices Follow the Nested Logit Model

When consumer choices follow the Nested Logit model, Figures 5 and 6 show the results of the two estimation models under different beginning inventory levels of \((1, 1)\). In each plot, the dash lines with the corresponding markers
Figure 5: Estimation results when the simulation model and estimation model are both NL

(a) First-choice probability

(b) Substitution probability

Figure 6: Estimation results when the simulation model is NL and estimation model is ES

(a) First-choice probability

(b) Substitution probability
represent the true values of the model parameters. We also report the confidence intervals of the estimated parameters in Appendix F.

From Figures 5a and 6a, we observe that the two estimation models obtain similar estimation results on the first-choice probabilities. More specifically, when the beginning inventory $b_{11}$ is low, we underestimate $p_{11}$ by no more than 17.3% and overestimate $p_{12}$ by no more than 15.5%. As $b_{11}$ increases, the estimation of $p_{11}$ and $p_{12}$ become more accurate. On the other hand, the estimation results of $p_{21}$ and $p_{22}$ are less affected by $b_{11}$ and remain close to their true values.

Intuitively, when $b_{11}$ is low, more unsatisfied demand from (1, 1) can be recaptured by other available alternatives. Under our numerical setting, in-store substitution outweighs cross-store substitution and thus most unsatisfied demand from (1, 1) is recaptured by (1, 2). Thus, the sales of (1, 1) is smaller and the sales of (1, 2) is larger, in contrast to their primary demand. Consequently, $p_{11}$ is underestimated and $p_{12}$ is overestimated. As $b_{11}$ increases, the chance of stockout and the consequent substitution decreases. Therefore, the estimation accuracies of the first-choice probabilities increase.

Figures 5b and 6b show that, when the underlying “reality” is governed by the Nested Logit model, the Nested Logit estimation model significantly outperforms the Exogenous Substitution estimation model in estimating substitution probabilities. While the former is able to estimate the substitution probabilities accurately, the latter can hardly distinguish among the different substitution probabilities, especially when $b_{11}$ is relatively high. The results may indicate the importance of capturing the consumer behaviors and identifying the correct consumer choice model. We will continue this discussion in Section 6.2.2.

6.2.2. When Consumer Choices Follow the Exogenous Substitution Model

When the consumer choices follow the Exogenous Substitution model, Figures 7 and 8 show the results of the two estimation models under different beginning inventory levels of (1, 1). Again, in each plot, the dash lines with the corresponding markers represent the true values of the model parameters. We also report the confidence intervals of the estimated parameters in Appendix F.

We observe similar results as in the previous section. The only difference is that the estimation accuracy of substitution probabilities of the Nested Logit model is relatively lower, comparing Figures 5b and 7b. We may have expected that the Exogenous Substitution model performs better (than the Nested Logit model) when the underlying “reality” is governed by the same model, but the results are counter-intuitive.

As discussed in Section 3.4, the substitution probabilities are jointly determined by utilities and store dissimilarity parameter in the Nested Logit model. Thus, the sales data of all the periods can be used to estimate the substitution probabilities. However, in the Exogenous Substitution model, the substitution probabilities have larger freedom and, more importantly, can only be estimated by the sales data of the stockout periods. These differences explain the results observed above.
Figure 7: Estimation results when the simulation model is ES and estimation model is NL

(a) First-choice probability

(b) Substitution probability

Figure 8: Estimation results when the simulation model and estimation model are both ES

(a) First-choice probability

(b) Substitution probability
In Section 6.2.3, we investigate whether the information on the purchase records can help improve the estimation accuracies of the Exogenous Substitution estimation model (substitution probabilities in particular).

6.2.3. The Value of Information on Consumers’ Purchase Records

In a small retail store, the store owner may be able to obtain the consumers’ purchase records by installing a Point of Sale (PoS) system. We are interested in how much the availability of the consumers’ purchase records may help improve the estimation accuracy of the Exogenous Substitution model.

We use the same data sets as in Section 6.2.2, together with the true purchase records to estimate the model parameters. In the estimation process, we use the true purchase records instead of the simulated purchase records in Step 1 of Algorithm 1. Other steps in Algorithm 1 remain the same. We present the estimation results in Figure 9. For comparison purpose, we also present the estimation results without the purchase records (i.e., the estimation results as in Figure 8b) using dash-dot lines. Again, the dash lines with the corresponding markers represent the true values of the model parameters.

Figure 9: Estimation results when the simulation model and estimation model are both ES and purchase records are known

(a) First-choice probability

(b) Substitution probability

The first main observation is that the information on consumers’ purchase records helps improve the estimation accuracies of first-choice probabilities when the beginning inventory level is low. As observed in Figure 9a, the estimation results of first-choice probabilities are centered around their true values, at all the beginning inventory levels. Recall that without purchase records, the first-choice probabilities are underestimated when the beginning inventory level is low (as shown in Figure 8a). The results suggest that, when the beginning inventory level is low, we can estimate the first-choice probabilities better with consumers’ purchase records.
Another observation is that the information on consumers’ purchase records significantly improves the estimation accuracies of substitution probabilities when the beginning inventory level is low. Figure 9b shows that, when $b_{11}$ is low, the estimation results of substitution probabilities are more accurate with purchase records information (than those without purchase records information). As $b_{11}$ increases, the improvement becomes less significant. This is reasonable. As the beginning inventory level increases, fewer stockouts occur and thus fewer substitution behaviors are observed. Consequently, the estimation accuracies of substitution probabilities are low even with purchase records information.

Combining the above observations, the value of information on consumers’ purchase records is high when the beginning inventory level is low. The information on the consumers’ purchase records can help improve the estimation accuracy of both the first-choice probabilities and the substitution probabilities. As the beginning inventory level increases, the value of information on the consumers’ purchase records decreases. In other words, in order to obtain a good estimation of the Exogenous Substitution parameters (substitution probabilities in particular), we need to set a low beginning inventory level and obtain the purchase records at the same time.

Note that the individual transaction data can help better estimate the number of consumer arrivals. However, even if available in the type of environment that we study, individual transaction data would only record consumers who have made a purchase, but lack information on consumers without purchase. Therefore, accurate estimation of the total number of consumer arrivals (including those without purchasing) is challenging in any retail setting. One may need to resort to exogenous information for the estimation. In Kök and Fisher (2007), the authors use the sales transaction of all products to estimate the total number of consumer arrivals, with an assumption that the number of consumers who visited the store but did not purchase anything is negligible. Musalem et al. (2010) suggest that the total number of consumer arrivals can possibly be obtained using demographic information to estimate the size of the market. Vulcano et al. (2012) also suggest to use a retailer’s market share or potential to estimate the market size.

In our setting, traditional small retail stores adapt well to their residential neighborhood and store owners often possess good knowledge of their consumers. Rather than knowing the exact number of consumer arrivals each day, it is reasonable to assume that a store owner has a good knowledge about the average number of consumer arrivals per day. Therefore, we have experimented on replacing the total number of consumer arrivals in each period with the mean value ($\lambda = 200$) yet keeping the remaining settings the same as in Section 6.1.1. The results of these experiments are very similar to those reported above, so this method appears to be quite robust. Consequently, using the average number of consumer arrivals may be in general a reasonable and robust way to deal with this challenging problem.

6.3. Inventory Optimization Results

In this subsection, we explore the impact of estimation accuracy on the inventory policies and expected profit. First, we introduce a single-period centralized control newsvendor model. Next, we apply optimization via simulation (OvS) to obtain the optimal beginning inventory levels and then compute the corresponding expected profit. Finally, we study the impact of the estimation accuracy on the optimal inventory decisions and the expected profit.
In traditional small retail stores, brand owners control the order quantities by routine store visiting and convincing up or down store owners’ order quantities. Besides, there is no price competition among store owners. Therefore, it is reasonable to construct a centralized control model. We assume that all the products have the same unit price \( m \) and cost \( c \). Summing the expected profit over all store-product combinations, the objective function is given by

\[
\max_{b \geq 0} E[\Pi(b, d^E)] = \sum_{(i, j) \in C} mE[\min(d^E_{ij}, b_{ij})] - cb_{ij},
\]  

where \( d^E_{ij} \) represents the effective demand of \((i, j)\), including both first-choice and substitution demand.

We use OvS to solve (14) and obtain the optimal beginning inventory levels. First, we build the two consumer choice models using AnyLogic University 7.0.2. Second, given the consumer choice model and its parameters, we use the OptQuest optimization engine embedded in the software to solve (14), to find the candidate optimal beginning inventory levels. Third, we use the “cleanup” procedure (Nelson 2010, Hong and Nelson 2009) to select the best inventory policy among the candidates, to address the variability issue in OvS. Finally, once we obtain the optimal beginning inventory levels, we compute the corresponding expected profit by simulation.

6.3.1. Baseline Results-Unit Cost = 9

Traditional small retail store owners operate in a relatively low profit margin environment. In our experiments, we set the unit price \( m = 10 \) and the unit cost \( c = 9 \) (i.e., profit margin of around 10%). Table 5 shows the inventory optimization results under different simulation and estimation models. In Table 5, we also present the optimization results of another two models as benchmarks. The first model “Est: W/o Sub.” directly uses the percentages of sales as the choice probabilities and ignores stockout and substitution effects. The other model is the underlying true model with the exact values of the simulation model parameters (“True Value”).

First, we observe in Table 5 that, the optimal beginning inventory level is lower if we underestimate the first-choice probabilities. In Table 5, under both the Nested Logit estimation model and the Exogenous Substitution estimation model, \( b^*_11 \) increases with \( b_{11} \) in the simulation model. Only slight violation can be found when \( b_{11} \) is high (e.g., 55-65), but this is due to the fact that we cannot avoid the variability issue in OvS.

Furthermore, optimal beginning inventory levels are higher with substitution. As shown in Table 5, under either simulation model, the optimal beginning inventory levels of “Est: NL” and “Est: ES” are both higher than those without any substitution estimation (“Est: W/o Sub.”) for a given setting of \( b_{11} \). The total demand is stochastically larger when consumers are more likely to substitute. Higher substitution indeed leads to lower inventory under a fixed service level. However, higher substitution may not decrease inventory levels when the goal is to maximize profit. A seller may have incentives to keep more inventory in order to realize more sales. Considering the impact of stockout and substitution can bring additional benefits to holding inventory as it can be further used to satisfy spill-over demands from other products. Then, because the marginal benefit of holding inventory is increasing, the optimal inventory level is increasing. This observation suggests that store owners should carry more with substitution.
Table 5: Optimal beginning inventory levels ($b^*_i$) under different simulation and estimation models when unit cost is 9

<table>
<thead>
<tr>
<th>$b_{11}$ in Sim. Model</th>
<th>Sim. Model: NL</th>
<th>Sim. Model: ES</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$b_{11}^<em>$  $b_{12}^</em>$  $b_{21}^<em>$  $b_{22}^</em>$</td>
<td>$b_{11}^<em>$  $b_{12}^</em>$  $b_{21}^<em>$  $b_{22}^</em>$</td>
</tr>
<tr>
<td>Est: W/o Sub.</td>
<td>45</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>42</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>44</td>
</tr>
<tr>
<td>Est: NL</td>
<td>45</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>46</td>
</tr>
<tr>
<td>Est: ES</td>
<td>45</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>55</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>60</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>65</td>
<td>48</td>
</tr>
<tr>
<td>True Value</td>
<td>49</td>
<td>29</td>
</tr>
</tbody>
</table>
It is worth mentioning that, Rajaram and Tang (2001) find that optimal beginning inventory levels under substitution may be higher or lower than those without substitution, depending on the underlying parameters. Our results imply that the optimal beginning inventory levels are higher when considering stockout and substitution. Considering the impact of stockout and substitution can bring additional benefits to holding inventory as it can be further used to satisfy spill-over demands from other products. Then, because the marginal benefit of holding inventory is increasing, the optimal inventory level is increased.

Figure 10: Expected profits (evaluation results) under different simulation and estimation models when unit cost is 9

(a) Simulation model: NL

(b) Simulation model: ES

We now study the evaluation results. In Figure 10, we plot the evaluation results under different simulation and estimation models. Note that the expected profit is based on the optimal inventory level under the estimated parameters $\hat{\theta}$ of the corresponding estimation model (“belief”), evaluated under the true parameters $\theta$ of the corresponding simulation model (“reality”). Figure 10a shows the case when the simulation model is the Nested Logit model and Figure 10b shows the case when the simulation model is the Exogenous Substitution model. Again, in each graph, we add evaluation results of “Est: W/o Sub.” and “True value” as benchmarks.

An important finding is that considering stockout and substitution can significantly increase the expected profit. In Figure 10a, under all the five beginning inventory levels, the expected profits of “Est: NL” and “Est: ES” are much higher than those of “Est: W/o Sub.”. This observation also holds under the Exogenous Substitution simulation model (see Figure 10b). The results imply that it is beneficial to recognize the substitution behavior when facing stockouts and to take it into account in demand estimation and inventory decisions, even if the substitution estimation itself may not be very accurate.
6.3.2. Varying Unit Cost

Next, we further examine our observations in Section 6.3.1 under different profit margins. To do so, we fix the unit price \( m = 10 \) and vary the unit price as \( c = 7 \) and \( c = 5 \). Figures 11 and 12 show the expected profits of these two cases.

Figure 11: Expected profits (evaluation results) under different simulation and estimation models when unit cost is 7

Comparing Figures 10, 11, and 12, we observe that, when the profit margin increases (or the unit cost decreases), the gaps between the expected profits of the three estimation models (i.e., estimating without substitution, the Nested Logit model and the Exogenous Substitution model) become smaller. When the profit margin is high (or the unit cost is low), the store owners tend to increase the beginning inventory levels, which results in less stockouts and also less accurate estimation of the substitution probabilities. Therefore, the benefit of considering the substitution behavior diminishes.

We repeat the simulation and estimation process under the reversed NL sequence, in which consumers first select among different products and then choose stores in Appendix A. We then apply the analysis process to the two-store three-product case in Appendix D and study the case where the simulation model is the ES model and employ a symmetric substitution matrix that cannot be well approximately by an NL model in Appendix E. We find that the estimation results are similar to those of the original experiment.

7. Conclusions

In this paper, we have studied the demand estimation problem in traditional small retail stores when considering consumers’ multi-store multi-product substitution behavior. We describe this consumer behavior by two consumer
choice models, namely the Nested Logit model and the Exogenous Substitution model. Then we apply the Markov chain Monte Carlo algorithm to estimate parameters of the two demand models.

Based on our numerical experiment, our main observations are:

(1) When consumer choices follow the Nested Logit model, the Nested Logit estimation model significantly outperforms the Exogenous Substitution estimation model in estimating substitution probabilities. When consumer choices follow the Exogenous Substitution model and the substitution matrix is constructed following the Nested Logit model, this observation still holds.

(2) Consumers’ purchase records can help improve the estimation accuracy of both the first-choice probabilities and the substitution probabilities of the Exogenous Substitution model when the beginning inventory level is low.

(3) When the profit margin is low (as normally faced by traditional small retailers), it is beneficial to recognize and consider the substitution behavior in demand estimation and inventory decisions. Considering substitution increases the optimal beginning inventory levels as well as the expected profit. However, such benefit diminishes as the profit margin increases.

In this paper, we study the demand estimation and inventory control problem from the perspective of the brand owner. In traditional retail, the decision-making of the small retail stores is boundedly rational and is significantly influenced by the sales agent of the brand owner, as the primary objective of the small retail stores is subsistence rather than profit maximization (Boulaksil et al. 2014). Hence, for the purpose of this study it is reasonable to assume that small retail store owners would effectively accept the brand owner’s recommended inventory levels.
Admittedly, while this paper opens up an interesting yet understudied research field in retail operations management, its results remain exploratory. There are ample research opportunities for future work. From the modeling perspective, besides the two classic consumer choice models studied in this paper, one may consider other consumer choice models (e.g., rank-based model) and study their effects. A more complete choice model would be the mixed MNL. A potential future research direction is to extend our estimation framework to the mixed MNL demand model. From the algorithmic perspective, one may extend or develop more efficient demand estimation algorithms to handle more realistic cases. Finally, we assume stable assortment in our study. It is interesting to further consider the product assortment decisions under multi-store multi-product substitution.

Acknowledgments

This research was jointly supported by The National Natural Science Foundation of China (NSFC) under Project No. 71361130017 and The Netherlands Organization for Scientific Research (NWO) under Project No. 629.001.012.

References


