



# Estimating Primary Demand for a Heterogeneous-Groups Product Category under Hierarchical Consumer Choice Model

Haengju Lee & Yongsoon Eun

To cite this article: Haengju Lee & Yongsoon Eun (2016) Estimating Primary Demand for a Heterogeneous-Groups Product Category under Hierarchical Consumer Choice Model, IIE Transactions, 48:6, 541-554, DOI: [10.1080/0740817X.2015.1078524](https://doi.org/10.1080/0740817X.2015.1078524)

To link to this article: <http://dx.doi.org/10.1080/0740817X.2015.1078524>

 View supplementary material 

 Accepted author version posted online: 22 Sep 2015.  
Published online: 22 Sep 2015.

 Submit your article to this journal 

 Article views: 158

 View related articles 

 View Crossmark data 

# Estimating primary demand for a heterogeneous-groups product category under hierarchical consumer choice model

Haengju Lee and Yongsoon Eun

Department of Information and Communication Engineering, DGIST 333 Techno Jungang-daero, Hyeonpung-myeon, Dalseong-gun, Daegu, 711-873, Republic of Korea

## ABSTRACT

This article discusses the estimation of primary demand (i.e., the true demand before the stockout-based substitution effect occurs) for a heterogeneous-groups product category that is sold in department store settings, based on historical sales data, product availability, and market share information. For such products, a hierarchical consumer choice model can better represent purchasing behavior. This means that choice occurs on multiple levels: a consumer might choose a particular product group on the first level and purchase a product within that chosen group on the second level. Hence, in the present study, we used the nested multinomial logit choice model for the hierarchical choice and combined it with non-homogeneous Poisson arrivals over multiple periods. The expectation-maximization algorithm was applied to estimate the primary demand while treating the observed sales data as an incomplete observation of that demand. We considered the estimation problem as an optimization problem in terms of the inter-product-group heterogeneity, and this approach relieves the revenue management system of the computational burden of using a nonlinear optimization package. We subsequently tested the procedure with simulated data sets. The results confirmed that our algorithm estimates the demand parameters effectively for data sets with a high level of inter-product-group heterogeneity. Supplementary materials are available for this article. Go to the publisher's online edition of *IIE Transaction* for further discussions and detailed proofs.

## ARTICLE HISTORY

Received 1 January 2014  
Accepted 1 June 2015

## KEYWORDS

Demand untruncation;  
consumer choice analysis;  
nested multinomial logit  
model; EM algorithm

## 1. Introduction

Demand estimation is critical for retailers, as it enables optimal assortment decisions and optimal inventory management. Most retail stores save their sales transaction history and, based on thus data, attempt demand estimation in various ways (e.g., using time-series models). One of the challenges of demand estimation arises from temporary stockouts of products. When facing a stockout of a product that they had intended to purchase, consumers either leave without making any purchase or purchase, as a substitute, a related product that is available in the store. The first scenario is referred to as a lost sale and the second as substitute demand. Existing inventory systems, which store only sales transactions and product inventories, are not capable of logging lost sales and substitute demand. Thus, the saved transaction data are referred to as truncated or censored. Therefore, estimating demand based on censored sales transaction data, without considering the effect of stockouts, would yield a result different from the true demand. We refer to this true demand, which is to say the first-choice demand when all products are in stock, as the primary demand.

Another central demand characteristic is the hierarchical structure of consumers' purchasing behavior. When a product category with *heterogeneous* groups is sold in a traditional department store setting (i.e., where a variety of products with different brands and different features are sold in the same store), consumer choice seems to occur hierarchically. This means that

choice occurs on multiple levels: a consumer might choose a particular product group on the first level and then purchase a product within that group on the second level. Indeed, Kök and Xu (2011) introduced two hierarchical purchasing behaviors, brand-primary process and type-primary process. (See Section 2 for a detailed literature review.) If hierarchical choice behavior exists, it affects the product substitution pattern and, in turn, the substitute demand and lost sales. Thus, in primary demand estimation, it is important to take into account hierarchical consumer choice behavior.

A widely employed demand-estimation approach is to use discrete choice models such as Multinomial Logit (MNL) or Nested Multinomial Logit (NMNL) (see Ben-Akiva and Lerman (1985) and Train (2009)). Vulcano *et al.* (2012) integrated the MNL model with the Expectation-Maximization (EM) algorithm. This strategy, however, is limited when a product category with heterogeneous groups is considered, due to the MNL model's Independence from Irrelevant Alternatives (IIA) property (Ben-Akiva and Lerman, 1985; Train, 2009)). As Kannan and Wright (1991) demonstrated, it is the NMNL discrete-choice model that is suitable for hierarchical consumer choice behavior. Accordingly, in the present study, we integrated the NMNL model with the EM algorithm. Our work can be viewed as having extended the work of Vulcano *et al.* (2012) to cases wherein hierarchical consumer choice behavior exists.

We followed the procedure of Vulcano *et al.* (2012) wherein the estimation problem is viewed in terms of primary demand

and the observed sales data are treated as incomplete observations of that demand. The demand parameters to be estimated are (i) the inter-product-group similarity; (ii) the preference vector of products; and (iii) the arrival rates. The first two originate from the NMNL choice model. To estimate all three of them, the EM algorithm was applied. Accordingly, first, the hierarchical consumer choice behavior that might underlie the demand structure was identified. To that end, the demand parameters were estimated under two different hierarchical structures: the brand-primary model and the type-primary model. The one with the greater incomplete data log-likelihood was chosen as the demand structure. The substitute demand and lost sales were then estimated accordingly. The estimation required only censored sales transaction data, product availability, and market share information.

Unlike the case in Vulcano *et al.* (2012), direct application of the EM algorithm does not yield a closed-form solution in its M-step, due to the newly introduced inter-product-group similarity parameter. This is a complex multivariate maximum likelihood estimation problem. Thus, rather than resorting to nonlinear optimization packages in each M-step, we restructured the estimation problem as an optimization problem with respect to the inter-product-group similarity parameter.

We tested the proposed procedure on simulated data sets and compared the estimation results with those of Vulcano *et al.* (2012). The proposed method effectively reveals the hidden hierarchical demand structure and demand parameters for data sets with a high level of inter-product-group heterogeneity. The following are the most interesting numerical results.

1. Our algorithm predicted the choice hierarchy based on the substitution patterns within the product groups in the data set. This was numerically demonstrated by considering the two hierarchical structures: the brand-primary model and the type-primary model.
2. We showed numerically why more accurate demand estimation is important for optimal assortment decision-making and inventory management.
3. We showed how to estimate market share, one of the input variables in our estimation problem, in cases where it is unknown.

In a revenue management system that deals with hundreds of thousands of products, both speed and performance are important. Computationally intensive demand-estimation methods for solving assortment problems and inventory problems within a small processing time window, unfortunately, are not practically viable. Our contribution to the literature is the development of a simple optimization problem for estimation of the primary demand for a product category with heterogeneous groups. The optimization is effectively achieved without any need to resort to a nonlinear optimization package, thus relieving a revenue management system of what otherwise would be a heavy computational burden.

## 2. Literature review

A brief review is given on the existing literature that addresses the problem of estimating stockout-based substitution rates and lost sales.

Anupindi *et al.* (1998) presented a method to estimate demand parameters, arrival rates, and substitution probabilities. They used the EM method to obtain the maximum likelihood estimation while treating the time of stockout as missing data under a periodic-review system. To manage the rapidly growing number of parameters, they imposed a constraint that at most two products are out of stock at any given period.

Campo *et al.* (2003) studied the impact of stockouts on purchase quantity and substitution in different consumer segments. They found that stockouts may lead to the purchase of smaller quantities, and asymmetric (non-IIA) choice shifts. They proposed a modification of the MNL model for more general stockout-based substitution. Consumers who typically purchase multiple units per purchase might decide to buy fewer packs in case of stockouts to limit the risk of buying a less familiar alternative. Consumers are exposed to the loss of intrinsic utility following the choice of a less preferred item. They allowed for heterogeneous consumer reactions to stockouts using latent class estimation.

Kalyanam *et al.* (2007) provided an empirical study on the impact of the individual item's presence on the sales of other items. The mere presence of certain items increases the sales volume of the whole assortment. They studied the impact of the absence of each individual item on the category sales. Their model allowed for a more general substitution pattern than the MNL model. Using a hierarchical Bayes framework, they estimated parameters for a joint model of sales and stockouts. They developed a system of equations that accounted for the inter-relationship between sales, demand, and stockouts. The presence of individual items could increase the overall category sales by drawing attention of consumers who seek variety. However, some stockouts increased category demand because of the elimination of clutter by removing non-popular items.

Kök and Fisher (2007) developed a method to estimate the substitution and demand for products and proposed an EM method to solve the assortment planning problem. The M-step resorted to the nonlinear optimization technique due to the complexity of the likelihood function.

Bruno and Vilcassim (2008) proposed a model to account for varying levels of observed product availability. They pointed out that the set of products available in a particular store is not observable even though the aggregate level of distribution can be observable. They simulated potential product assortment vectors that each consumer faces in the actual store visit by drawing multivariate Bernoulli vectors based on the observed aggregate level of availability. The parameters were estimated by using the simulation.

Conlon and Mortimer (2009) considered an estimation of demand for periodic inventory system where the choice set of an individual consumer is latent. They developed a method incorporating the observation by using the EM algorithm to estimate primary demand across the unobserved choice sets. Just as in the model of Anupindi *et al.* (1998), the number of parameters to be estimated in the E-step grows exponentially when multiple products are simultaneously out of stock.

Musalem *et al.* (2010) considered substitution effects due to stockouts under the periodic-review inventory system (incomplete information about product availability). Instead of the maximum likelihood estimation method, they used the Markov

chain Monte-Carlo simulation method to estimate model parameters.

Vulcano *et al.* (2012) combined the MNL choice model with the EM algorithm to estimate primary demand. They applied the EM algorithm to estimate demand parameters for the incomplete sales transaction data with hidden variables of lost sales and recaptured demand. They developed a simple procedure consisting of closed-form expressions in each EM iteration. This method is efficient for a product category with homogeneous groups; however, it is limited for a product category with heterogeneous groups where hierarchical consumer choice behavior exists.

Fisher and Vaidyanathan (2014) considered a demand estimation procedure and retail assortment optimization. They allowed assortments to vary by stores, subject to a maximum number of different assortments. They viewed a Stock-Keeping Unit (SKU) as a set of attribute levels. By applying maximum likelihood estimation to sales history, they estimated the demand for attribute levels and substitution probabilities. This estimation enabled the retailer to estimate the demand for any potential SKU, including those not currently carried.

We also provide a brief review on the literature related to hierarchical consumer choice behavior.

Kannan and Wright (1991) applied the NMNL model to the coffee market and showed that the NMNL model performs better than the MNL model for the market where hierarchical consumer choice behavior exists.

Kök and Xu (2011) introduced two hierarchical purchase behaviors, namely, the brand-primary process and the type-primary process, and studied optimal product assortment planning and pricing under the two hierarchical consumer choice models. In the brand-primary process, consumers choose a brand first and then choose a product type within that brand. This process primarily models consumer choice behavior when different product types satisfy functional needs of consumers with strong brand loyalty (e.g., ice creams with different flavors). The type-primary process, on the other hand, models consumer behavior of selecting a product type first and then choosing a brand (e.g., regular coffee versus decaffeinated coffee). In this case, product types are functionally differentiated.

Gallego and Topaloglu (2014) studied constrained assortment optimization problems under the nested logit model. They showed that the optimal assortment under cardinality constraints (i.e., limiting the number of products) can be obtained by solving a linear program. They also showed how to obtain an assortment with a performance guarantee of two for the NP-hard assortment optimization problem under space constraints (i.e., limiting the total space consumption). The joint assortment optimization and pricing problem can be effectively solved by using the developed method for constrained assortment optimization.

### 3. Model description

We consider  $T$  period sales transaction data of a product category with heterogeneous groups (e.g., coffee, womens' dresses) that is sold in the same store. Arriving consumers follow the two-stage hierarchical choice process. Depending on how the choice process occurs, products are partitioned into  $K$  groups.

In the brand-primary process, consumers choose a brand first and then a product type within that chosen brand. In the type-primary process, consumers choose a product type first and then a brand within that chosen product type. The product group corresponds to the product brand in the brand-primary model and the product type in the type-primary model. Hence, the product group represents the higher or first level of consumers' choice.

Time periods are indexed by  $t = 1, \dots, T$ , and product groups are indexed by  $k = 1, \dots, K$ . The set of all the products in group  $k$  is denoted by  $B_k$ , and the set of available products in period  $t$  is denoted by  $S_{kt}$ , satisfying  $S_{kt} \subset B_k$ . Time period  $t$  is decided such that any product in  $S_{kt}$  is available and any unavailable product is unavailable throughout period  $t$ . This assumption is reasonable in the perpetual inventory system, in which actual stockout events can be observed. It should also be noted that perpetual inventory systems are becoming more widely used in retail environments. Even in periodic inventory systems, retailers tend to collect inventory information more frequently to mitigate the limitation of missing exact stockout events. The observed number of purchases of product  $j$  from group  $k$  in period  $t$  is denoted by  $z_{kjt}$ . If product  $j$  from group  $k$  is not available in period  $t$ , we define  $z_{kjt} = 0$ . We do not consider product returns, which implies  $z_{kjt} \geq 0$ . We assume that for all product  $j$  from group  $k$ , there is at least one period  $t$  where  $z_{kjt} > 0$ . Otherwise, we do not consider the product in our analysis.

The total number of purchases in period  $t$  is  $m_t = \sum_{k=1}^K \sum_{j \in B_k} z_{kjt}$ . For the sake of notational convenience, we define  $\mathbf{B} = (B_1, B_2, \dots, B_K)$ ,  $\mathbf{S}_t = (S_{1t}, S_{2t}, \dots, S_{Kt})$ , and  $\mathbf{z}_t = (z_{1t}, \dots, z_{Kt})$  where  $z_{kt} = (z_{kjt}, \text{ for all } j \in B_k)$ . We assume that the number of consumer arrivals is Poisson-distributed with mean  $\lambda_t$  in period  $t$ . The vector of arrival rates is denoted by  $\lambda = (\lambda_1, \dots, \lambda_T)$ .

#### 3.1. Hierarchical choice process

Arriving consumers in period  $t$  follow the two-stage hierarchical choice process as shown in Fig. 1. From available products  $S_t$ , they first select a product group to purchase or no-purchase. The no-purchase decision is viewed by a fictitious product group indexed by a zero. If they choose group  $k$  ( $k \neq 0$ ), then they next select a product to purchase from  $S_{kt}$ . If they choose the no-purchase option 0 on the first stage, they leave without any purchase.

We use the NMNL model (Ben-Akiva and Lerman, 1985) in order to model the hierarchical consumer choice process. The utility of a consumer associated with purchasing product  $j$  from

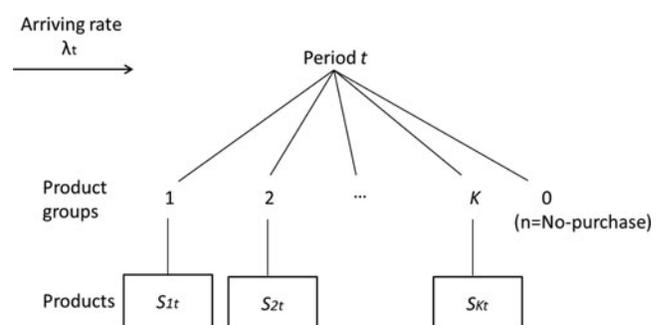


Figure 1. Hierarchical consumer choice model.

group  $k$  is given by

$$U_{kj} = u_{kj} + \epsilon_k + \epsilon_{kj},$$

where  $u_{kj}$  represents the expected utility of the product, which is identical across all consumers;  $\epsilon_k$  is the random shared unobserved utility component attributable to group  $k$ ; and  $\epsilon_{kj}$  is the remaining random total unobserved utility component. The utility of the no-purchase option is  $U_0 = u_0 + \epsilon_0$ . Note that consumers solely compare differences of utilities, so the absolute level of utility is irrelevant to consumers' choice (Train, 2009). Hence, the constant parts of the utilities are normalized to be  $u_0 = 0$ . This can be done by subtracting  $u_0$  from all the utilities. In the NMNL model,  $\epsilon_k$  and  $\epsilon_{kj}$  are assumed to be independent for all  $j \in B_k, k = 1, \dots, K$ . This implies that products in the same group are correlated as the covariance of two utilities in group  $k$  is  $\text{var}(\epsilon_k) > 0$ . After a further normalization of utilities,  $\epsilon_{kj}$  are independent and identically distributed zero-mean Gumbel random variables with scale parameter 1. The variable  $\epsilon_k$  are distributed so that  $\max_{j \in B_k} U_{kj}$  are Gumbel distributed with scale parameter  $\mu \in [0, 1]$ , and  $\epsilon_0$  is a zero-mean Gumbel random variable with scale parameter  $\mu$ . The zero-mean Gumbel random variable with scale parameter  $\mu$  has a distribution function of  $F(x) = \exp[-\exp[-(x \times \mu + \gamma)]]$ , where  $\gamma$  is Euler's constant and has an appropriate value of 0.5722. The preference (or attractiveness) of product  $j \in B_k$  is defined by  $v_{kj} = e^{u_{kj}}$ . Note that the preference of no-purchase is  $v_0 = 1$ . The vector of the preference weight is denoted by  $\mathbf{v} = (v_{kj}, \forall j \in B_k, k = 1, \dots, K)$ . Under the above conditions, the probability of no-purchase, denoted by  $P_0(\mathbf{S}_t; \mathbf{v}, \mu)$ , and the probability of purchasing product  $j$  from group  $k$  ( $j \in S_{kt}$ ), denoted by  $P_{kj}(\mathbf{S}_t; \mathbf{v}, \mu)$ , in period  $t$  are calculated in the following lemma.

**Lemma 1.** (Ben-Akiva and Lerman, 1985) *The probability of no-purchase in period  $t$  is*

$$P_0(\mathbf{S}_t; \mathbf{v}, \mu) = \frac{1}{\sum_{m=1}^K (\sum_{l \in S_{mt}} v_{ml})^\mu + 1}. \tag{1}$$

*The probability of purchasing product  $j$  from group  $k$  ( $j \in S_{kt}$ ) in period  $t$  is given by*

$$P_{kj}(\mathbf{S}_t; \mathbf{v}, \mu) = \frac{v_{kj} \times (\sum_{l \in S_{kt}} v_{kl})^{\mu-1}}{\sum_{m=1}^K (\sum_{l \in S_{mt}} v_{ml})^\mu + 1}. \tag{2}$$

We call scale parameter  $\mu$  the inter-product-group similarity because the greatest product similarity is achieved when  $\mu = 1$  (note that  $\mu \in [0, 1]$ ). When  $\mu = 1$ , the NMNL model reduces to the standard MNL model where all products are horizontally heterogeneous. In this case, the probability of purchasing product  $j$  from group  $k$  is  $v_{kj} / \sum_{m=1}^K \sum_{l \in S_k} v_{ml} + 1$ . As  $\mu$  goes to zero, the inter-product-group heterogeneity increases.

Now we consider the substitution pattern in the NMNL model. When a product is out of stock, the demand may be substituted by one of the available products. Due to the substitution effect, purchase probabilities of other available products increase. To know the substitution effect in the NMNL model, we introduce the Independence from Irrelevant Nests (IIN) property (Train, 2009). The ratio of the purchase probability of product  $j$  from group  $k$  and that of product  $l$  from group

$m$  is

$$\frac{P_{kj}(\mathbf{S}_t; \mathbf{v}, \mu)}{P_{ml}(\mathbf{S}_t; \mathbf{v}, \mu)} = \frac{v_{kj} \times (\sum_{i \in S_{kt}} v_{ki})^{\mu-1}}{v_{ml} \times (\sum_{i \in S_{mt}} v_{mi})^{\mu-1}}.$$

The MNL model has a restrictive choice behavior of the IIA property where the ratio of purchase probabilities of two alternatives are the same under two different choice sets containing the two alternatives (Ben-Akiva and Lerman, 1985; Train, 2009). In the NMNL model, the IIA property holds for products within a group because the above ratio is  $v_{kj}/v_{ml}$  when  $k = m$ . However, it does not hold for products in different groups. The ratio depends on preferences of products in group  $k$  and group  $m$  but not on preferences of products in a group different from  $k$  and  $m$ . This property is called the IIN because it is similar to the IIA on a group level (Train, 2009). Due to the IIN property, the NMNL model has a special substitution pattern. The substitution effect occurs in the product group level. A product's stockout creates a greater effect on products in the same group than products in different groups. Specifically, the percentage increase of purchase probability of products in the same group is greater than that of products in different groups. However, if the products are horizontally heterogeneous as in the MNL model (i.e.,  $\mu = 1$ ), the percentage increase of purchase probabilities are all the same. We state this result in the following lemma.

**Lemma 2.** *In the NMNL model with  $\mu < 1$ , a product's stockout leads to a higher percentage increase of purchase probability in the same group than those in other groups. However, when  $\mu = 1$  (i.e., the MNL model), the percentage increase of purchase probabilities are same for all the products.*

The following example shows how the substitution effect occurs in the level of product group for the NMNL model.

**Example 1:** *Assume that there are four products to sell in a store: product 1, product 2, product 3, and product 4. Products 1 and 2 belong to group 1, and products 3 and 4 belong to group 2. The inter-product-group similarity is given by  $\mu = 0.5$ . The preference vector is  $(v_1, v_2, v_3, v_4) = (1.5, 0.8, 1, 0.4)$ . Note that the preference of no-purchase is normalized to be one. Using the NMNL model, we see that the purchase probability changes when product 1 is out of stock. In Table 1, we summarize the purchase probabilities and the no-purchase probability. The probabilities are in the third column when all of the products are available. The probabilities after the stockout of product 1 are in the fourth column. The ratio of probabilities after to before the stockout is in the fifth column. All of the probabilities are increased after the stockout. However, the ratio of product 2 is greater than others. Therefore, product 2, which is in the same group (with the product 1), has been greatly influenced by the stockout event.*

**Table 1.** Purchase probabilities after and before the stockout of product 1 in the NMNL model

Group	Product	Before stockout	After stockout	Ratio of after to before
1	1	0.2673	NA	NA
	2	0.1426	0.2906	2.0383
2	3	0.2284	0.2746	1.2022
	4	0.0914	0.1098	1.2022
No-purchase		0.2703	0.3249	1.2022

### 3.2. Incomplete data likelihood function and market share

The sales data are incomplete observations of primary demands. Using the preference vector  $\mathbf{v} = (v_{kj}, \forall j \in B_k, k = 1, \dots, K)$ , inter-product-group similarity  $\mu$ , and arrival rates  $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_T)$ , we write the incomplete data likelihood function of observing the sales data:

$$\mathcal{L}(\mathbf{v}, \mu, \boldsymbol{\lambda}) = \prod_{t=1}^T \left( \Pr(m_t | \mathbf{v}, \mu, \boldsymbol{\lambda}) \frac{m_t!}{\prod_{k=1}^K \prod_{j \in S_{kt}} z_{kjt}!} \times \prod_{k=1}^K \prod_{j \in S_{kt}} \left[ \frac{P_{kj}(\mathbf{S}_t; \mathbf{v}, \mu)}{1 - P_0(\mathbf{S}_t; \mathbf{v}, \mu)} \right]^{z_{kjt}} \right), \quad (3)$$

where

$$\begin{aligned} \Pr(m_t | \mathbf{v}, \mu, \boldsymbol{\lambda}) &= \Pr(m_t \text{ customers purchase in period } t | \mathbf{v}, \mu, \boldsymbol{\lambda}) \\ &= \frac{(\lambda_t (1 - P_0(\mathbf{S}_t; \mathbf{v}, \mu)))^{m_t} e^{-\lambda_t (1 - P_0(\mathbf{S}_t; \mathbf{v}, \mu))}}{m_t!}. \end{aligned}$$

The number of consumers purchasing in period  $t$ , which is  $m_t$ , follows a Poisson distribution with mean  $\lambda_t (1 - P_0(\mathbf{S}_t))$  and  $P_{kj}(\mathbf{S}_t; \mathbf{v}, \mu) / (1 - P_0(\mathbf{S}_t; \mathbf{v}, \mu))$  is the conditional probability of purchasing product  $j$  from group  $k$  given that a consumer makes a purchase. The direct maximization of the incomplete data likelihood function with respect to  $\mathbf{v}$ ,  $\mu$ , and  $\boldsymbol{\lambda}$  is not promising due to the likelihood function (and the log of it) having a complex form without much structure even for  $\mu = 1$  as reported in Vulcano *et al.* (2012).

The continuum of maxima for the MNL model (Vulcano *et al.*, 2012) still holds for the NMNL model. Let  $(\mathbf{v}^*, \mu^*, \boldsymbol{\lambda}^*)$  be a maximizer of the incomplete data likelihood function in Equation (3). For any  $\alpha > 0$ , we define a new preference vector  $\mathbf{v}^0 = \alpha \mathbf{v}^*$  and a new arrival rate:

$$\lambda_t^0 = \frac{\sum_{k=1}^K (\sum_{l \in S_{kt}} \alpha v_{kl}^*)^{\mu^*} + 1}{\sum_{k=1}^K (\sum_{l \in S_{kt}} \alpha v_{kl}^*)^{\mu^*} + \alpha \mu^*} \lambda_t^*.$$

We have  $\lambda_t^* (1 - P_0(\mathbf{S}_t; \mathbf{v}^*, \mu^*)) = \lambda_t^0 (1 - P_0(\mathbf{S}_t; \mathbf{v}^0, \mu^*))$ . Therefore,  $(\mathbf{v}^0, \mu^*, \boldsymbol{\lambda}^0)$  has the same likelihood value and is also a maximizer for any  $\alpha > 0$ . Note that this is the extension of the solution presented by Vulcano *et al.* (2012), in which the new preference vector is  $\mathbf{v}^0 = \alpha \mathbf{v}^*$  and the new arrival rate is  $\lambda_t^0 = (\alpha \sum_{l \in S_t} v_l^* + 1) / (\sum_{l \in S_t} v_l^* + 1) \lambda_t^*$ . Note that the new arrival rate in the MNL model can be obtained by applying  $\mu = 1$  to the new arrival rate in the NMNL model. To resolve the multiple optimal solutions, Vulcano *et al.* (2012) assumed that there is an exogenous estimate of the the store's market share (or market potential). We denote the market share as  $s = 1 - P_0(\mathbf{B}; \mathbf{v}, \mu) = \sum_{k=1}^K \sum_{j \in B_k} P_{kj}(\mathbf{B}; \mathbf{v}, \mu)$ , where  $P_0(\mathbf{B}; \mathbf{v}, \mu)$  is the probability of no-purchase and  $P_{kj}(\mathbf{B}; \mathbf{v}, \mu)$  is the purchase probability of product  $j$  from group  $k$  under the complete product assortment (i.e., no stockouts). This article also assumes that market share  $s$  is exogenously given.

### 3.3. Likelihood function based on primary demand

Instead of solving the Maximum Likelihood Estimation (MLE) of the incomplete data likelihood function in Equation (3), we view the estimation problem with respect to the primary

demands as described by Vulcano *et al.* (2012). The primary demand of a product is the number of consumers whose first choice is that product.

Let  $Z_{kjt}$  be the random number of purchases of product  $j$  from group  $k$  in period  $t$ . The number can be broken down into two terms: the primary demand of that product and the substitute demand from other unavailable products. Therefore,

$$Z_{kjt} = X_{kjt} + Y_{kjt},$$

where  $X_{kjt}$  is the primary demand of product  $j$  from group  $k$  and  $Y_{kjt}$  is the substitute demand to product  $j$  from other unavailable products. We also let  $X_{0t}$  be the primary demand of no-purchase and  $Y_{0t}$  be the lost sales in period  $t$ .

If the primary demands (i.e.,  $X_{kjt}$  and  $X_{0t}$ ) are available, the likelihood function can be written as

$$L(\mathbf{v}, \mu) = \prod_{k=1}^K \prod_{j \in B_k} P_{kj}(\mathbf{B}; \mathbf{v}, \mu)^{\sum_{t=1}^T X_{kjt}} \times P_0(\mathbf{B}; \mathbf{v}, \mu)^{\sum_{t=1}^T X_{0t}}.$$

We let  $N_{kj} = \sum_{t=1}^T X_{kjt}$  and  $N_0 = \sum_{t=1}^T X_{0t}$ , which are the total primary demands over all the periods. The log-likelihood function is

$$LL(\mathbf{v}, \mu) = \sum_{k=1}^K \sum_{j \in B_k} N_{kj} \ln P_{kj}(\mathbf{B}; \mathbf{v}, \mu) + N_0 \ln P_0(\mathbf{B}; \mathbf{v}, \mu). \quad (4)$$

## 4. EM algorithm under the NMNL model

In this section, we describe the E-step of the EM algorithm to calculate the conditional expected log-likelihood function when purchase data  $\mathbf{z}_t$  and parameter estimators  $(\hat{\mathbf{v}}, \hat{\mu})$  are given. We then describe the M-step to calculate the MLEs of the conditional expected log-likelihood function.

### 4.1. E-Step

In the E-step, we attempt to obtain the conditional expected values of primary demands that are denoted by  $\hat{X}_{kjt} = E[X_{kjt} | \mathbf{z}_t, \hat{\mathbf{v}}, \hat{\mu}]$  and  $\hat{X}_{0t} = E[X_{0t} | \mathbf{z}_t, \hat{\mathbf{v}}, \hat{\mu}]$ . We also let  $\hat{\lambda}_t = E[\lambda_t | \mathbf{z}_t, \hat{\mathbf{v}}, \hat{\mu}]$ . First, if product  $j$  from group  $k$  is unavailable in period  $t$  (i.e.,  $j \in B_k \setminus S_{kt}$ ), we have  $\hat{X}_{kjt} = \hat{\lambda}_t \times P_{kj}(\mathbf{B}; \hat{\mathbf{v}}, \hat{\mu})$  and  $\sum_{m=1}^K \sum_{l \in S_{mt}} z_{mlt} = \hat{\lambda}_t \times (1 - P_0(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu}))$ . Combining the two equations, we have

$$\hat{X}_{kjt} = \frac{P_{kj}(\mathbf{B}; \hat{\mathbf{v}}, \hat{\mu})}{1 - P_0(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu})} \sum_{m=1}^K \sum_{l \in S_{mt}} z_{mlt}.$$

Next, if product  $j$  from group  $k$  is available (i.e.,  $j \in S_{kt}$ ), we have  $\hat{X}_{kjt} = \hat{\lambda}_t \times P_{kj}(\mathbf{B}; \hat{\mathbf{v}}, \hat{\mu})$  and  $z_{kjt} = \hat{\lambda}_t \times P_{kj}(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu})$ . From the two equations, we have

$$\hat{X}_{kjt} = \frac{P_{kj}(\mathbf{B}; \hat{\mathbf{v}}, \hat{\mu})}{P_{kj}(\mathbf{S}_t; \hat{\mathbf{v}}, \hat{\mu})} \times z_{kjt}.$$

Finally, the conditional expected value of the primary demand of no-purchase can be derived from  $\hat{X}_{0t} = \hat{\lambda}_t \times P_0(\mathbf{B}; \hat{\mathbf{v}}, \hat{\mu})$ ,  $\hat{X}_{0t} + \sum_{k=1}^K \sum_{l \in B_k} \hat{X}_{kl} = \hat{\lambda}_t$ , and  $s = 1 - P_0(\mathbf{B}; \hat{\mathbf{v}}, \hat{\mu})$ , where  $s$  is the

market share. We have

$$\hat{X}_{0t} = \frac{1-s}{s} \sum_{k=1}^K \sum_{j \in B_k} \hat{X}_{kjt}.$$

We summarize the results in the following lemma.

**Lemma 3.** *The conditional expected value of the primary demand of product  $j$  from group  $k$  in period  $t$  is*

$$\hat{X}_{kjt} = \begin{cases} \frac{P_{kj}(B; \hat{v}, \hat{\mu})}{1 - P_0(S; \hat{v}, \hat{\mu})} \sum_{m=1}^K \sum_{l \in S_{mt}} z_{mlt} & \text{for all } j \in B_k \setminus S_{kt} \\ \frac{P_{kj}(B; \hat{v}, \hat{\mu})}{P_{kj}(S; \hat{v}, \hat{\mu})} \times z_{kjt} & \text{for all } j \in S_{kt} \end{cases}$$

and the conditional expected primary demand of no-purchase is given by

$$\hat{X}_{0t} = \frac{1-s}{s} \sum_{k=1}^K \sum_{j \in B_k} \hat{X}_{kjt},$$

where  $s$  is the exogenously given market share of the store.

By using the estimated primary demand of  $\hat{X}_{kjt}$ , we estimate the substitute demand for product  $j$  from group  $k$  as follows:

$$\hat{Y}_{kjt} = z_{kjt} - \hat{X}_{kjt}. \tag{5}$$

Note that the substitute demand is negative when the product is not available. The number of lost sales is estimated as

$$\hat{Y}_{0t} = - \sum_{k=1}^K \sum_{j \in B_k} \hat{Y}_{kjt}. \tag{6}$$

The total primary demand of product  $j$  in group  $k$  over all periods is estimated as  $\hat{N}_{kj} = \sum_{t=1}^T \hat{X}_{kjt}$  and  $\hat{N}_0 = 1 - s/s \sum_{k=1}^K \sum_{j \in B_k} \hat{N}_{kj}$  (because  $\hat{X}_{0t} = 1 - s/s \sum_{k=1}^K \sum_{j \in B_k} \hat{X}_{kjt}$  in Lemma 3). Note that  $\hat{N}_{kj} > 0$  and  $\hat{N}_0 > 0$  because  $\sum_{t=1}^T z_{kjt} > 0$  (i.e., it is assumed that  $z_{kjt} > 0$  in at least one period  $t$ ). The conditional expected log-likelihood in Equation (4) for a given parameter set  $(\hat{v}, \hat{\mu})$ , defined by  $\hat{LL}(v, \mu) = E[LL(v, \mu) | z_t, \hat{v}, \hat{\mu}]$ , is

$$\begin{aligned} \hat{LL}(v, \mu) &= \sum_{k=1}^K \sum_{j \in B_k} \hat{N}_{kj} \times \ln P_{kj}(B; v, \mu) + \hat{N}_0 \times \ln P_0(B; v, \mu) \\ &= \sum_{k=1}^K \sum_{j \in B_k} \hat{N}_{kj} \times \ln v_{kj} + (\mu - 1) \\ &\quad \times \sum_{k=1}^K \left( \sum_{j \in B_k} \hat{N}_{kj} \right) \times \ln \sum_{l \in B_k} v_{kl} \\ &\quad - \frac{\hat{N}_0}{1-s} \times \ln \left( \sum_{m=1}^K \left( \sum_{l \in B_m} v_{ml} \right) + 1 \right). \end{aligned} \tag{7}$$

### 4.2. M-Step

In the M-step, we find estimators to maximize the conditional expected log-likelihood given in Equation (7). In this section, we find the critical point and study its properties. All of the proofs are in Online Supplement D.

The conditional expected log-likelihood in Equation (7) is a multivariate function that depends on preference vector  $v$  and inter-product-group similarity  $\mu$ . Unlike the MNL model by Vulcano *et al.* (2012), the closed-form solution of the parameters is difficult to obtain due to the additional parameter  $\mu$ . The key idea behind our approach is to calculate the critical point of the preference vector while keeping  $\mu$  constant. Then we can obtain the closed-form solution for the preference vector as a function of  $\mu$ . Furthermore, the critical point is unique. We state the result in the following proposition.

**Proposition 1.** *For a given value of  $\mu$  (i.e., constant  $\mu$ ), the conditional expected log-likelihood function in the E-step is unimodal, and its unique optimal solution is*

$$v_{kj}^*(\mu) = \frac{\hat{N}_{kj}}{\sum_{l \in B_k} \hat{N}_{kl}} \times \left( \frac{\sum_{l \in B_k} \hat{N}_{kl}}{\hat{N}_0} \right)^{1/\mu}, \forall j \in B_k, \tag{8}$$

$$k = 1, \dots, K.$$

The log-likelihood function in Equation (7) is not jointly concave (see Vulcano *et al.* (2012) for the case where  $\mu = 1$ ); however, it is unimodal and its unique optimal solution has a closed form. Note that  $v_{kj}^*(\mu) > 0$  because  $\hat{N}_{kj}$  and  $\hat{N}_0$  are strictly positive. We first consider the special case where  $\mu = 1$  (i.e., the MNL model). In this case, the optimal solution becomes  $v_{kj}^*(\mu) = \hat{N}_{kj}/\hat{N}_0$ , which is the ratio of the total primary demand of product  $j$  to the total primary demand of no-purchase. Note that the preference of no-purchase is normalized to be one. Hence, if the total primary demand of a product is greater than the total primary demand of no-purchase, the preference weight of the product is greater than one (and *vice versa*).

We now consider that  $\mu$  is strictly less than one. The  $v_{kj}^*(\mu)$  in Equation (8) is the product of two terms. The first term is the ratio of the total primary demand of product  $j$  to the total primary demand of its group  $k$ , which is between zero and one. This term represents the popularity of that product within its group (i.e., it is more popular in its group as this term approaches to one). Note that this term does not depend on  $\mu$ . The equation in the parenthesis of the second term is the ratio of the total primary demand of its group  $k$  to the total primary demand of no-purchase. The value of  $\mu$  gives a different effect on the second term depending on whether or not the ratio is greater than one. We call group  $k$  a *popular group* if its total primary demand is greater than the total primary demand of no-purchase (i.e.,  $\sum_{l \in B_k} \hat{N}_{kl} > \hat{N}_0$ ) and call group  $k$  an *unpopular group* if  $\sum_{l \in B_k} \hat{N}_{kl} < \hat{N}_0$ . We also call group  $k$  a *neutral group* if  $\sum_{l \in B_k} \hat{N}_{kl} = \hat{N}_0$ . As the level of inter-product-group heterogeneity increases (i.e., smaller value of  $\mu$ ), the product in a popular group has an increased preference weight, but the product in an unpopular group has a decreased preference weight. The preference weight of a product in a neutral group does not depend on the heterogeneity level. Therefore, the NMNL model assigns more preference weights to products in popular groups as the level of inter-product-group heterogeneity becomes higher.

It is difficult to maximize the conditional expected log-likelihood with respect to both the preference vector and the inter-product-group similarity at the same time. To resolve the

difficulty, our proposed method is to make the EM algorithm a sub-module of an optimization problem that is called repeatedly with a different value of  $\mu$ . By Proposition 1, the M-step has a simple and closed-form calculation. The optimization problem is defined in terms of  $\mu$  to maximize the incomplete data likelihood in Equation (3), calling the EM algorithm as its sub-module. Specifically, let  $v_{kj}^{EM}(\mu)$  be the outcome of the EM algorithm for a given  $\mu$ . The value of  $\mu$  that maximizes the incomplete data likelihood function  $\mathcal{L}(v^{EM}(\mu), \mu, \hat{\lambda})$  defined in Equation (3) is the optimal value of  $\mu^*$  and the corresponding optimal preference weight for product  $j$  from group  $k$  is obtained by  $v_{kj}^{EM}(\mu^*)$ . The number of arrivals in period  $t$  is estimated by

$$\hat{\lambda}_t = \hat{X}_{0t} + \sum_{k=1}^K \sum_{j \in B_k} \hat{X}_{kj}.$$

As the conditional expected log-likelihood function  $\hat{L}L(v, \mu)$  in Equation (7) is continuous in both  $v$  and  $\hat{v}$ , the converged value of  $v_{kj}^{EM}(\mu)$  is the stationary point of the incomplete data likelihood function for a given  $\mu$  (Wu, 1983). Therefore, the outcome of the optimization problem is the stationary point of the incomplete data likelihood function with the optimal level of the inter-product-group similarity.

The next proposition shows that market share constraint  $s = 1 - P_0(\mathbf{B}; v, \mu)$  is kept through the iterations of the EM algorithm. Therefore, if the initial values of the preference vector and the inter-product-group similarity satisfy the market share constraint, all of the updated solutions in the iterations of the EM-algorithm also satisfy the market share constraint.

**Proposition 2.** For a given  $\mu$ , the optimal preference vector satisfies the market share constraint. In other words,

$$s = 1 - P_0(\mathbf{B}; v^*(\mu), \mu).$$

### 4.3. Overall estimation procedure

We numerically search for the optimal value of  $\mu^*$  that maximizes the incomplete data log-likelihood function in Equation (3) from the converged preference vector from the EM algorithm for each  $\mu$ . We start from  $\mu = 1$  and decrease  $\mu$  by a small amount  $\delta$ . The optimal value of  $\mu$  is achieved when the value of the incomplete data log-likelihood does not improve by decreasing the value of  $\mu$ . The summarized procedure is as follows.

*Step 1.* Initialize  $\mu = 1$  and set  $\delta = 0.05$  and  $\epsilon = 0.0001$ .

1.1 (Initialization.) Set  $X_{kjt} = z_{kjt}$  if  $j \in S_{kt}$  and  $X_{kjt} = 0$  if  $j \in B_k \setminus S_{kt}$ ,  $\forall j \in B_k$ ,  $k = 1, \dots, K$ , and  $t = 1, \dots, T$ . Let  $X_{0t} = 1 - s/s \sum_{k=1}^K \sum_{j \in B_k} X_{kj}$  for  $t = 1, \dots, T$ . Then we have  $N_{kj} = \sum_{t=1}^T X_{kjt}$ ,  $\forall j \in B_k$ ,  $k = 1, \dots, K$  and  $N_0 = \sum_{t=1}^T X_{0t}$ . The preference vector is

$$v_{kj} = \frac{N_{kj}}{\sum_{l \in B_k} N_{kl}} \times \left( \frac{\sum_{l \in B_k} N_{kl}}{N_0} \right)^{1/\mu}, \forall j \in B_k, \\ k = 1, \dots, K.$$

1.2 (E-Step.) Update  $X_{kjt}$  and  $X_{0t}$   $\forall j \in B_k$ ,  $k = 1, \dots, K$ ,  $t = 1, \dots, T$  by Lemma 3. Let  $N_{kj} = \sum_{t=1}^T X_{kjt}$   $\forall j \in B_k$ ,  $k = 1, \dots, K$  and  $N_0 = \sum_{t=1}^T X_{0t}$ .

1.3 (M-step.) Set

$$v_{kj}^{new} = \frac{N_{kj}}{\sum_{l \in B_k} N_{kl}} \times \left( \frac{\sum_{l \in B_k} N_{kl}}{N_0} \right)^{1/\mu}, \forall j \in B_k, \\ k = 1, \dots, K.$$

1.4 (EM-stopping criterion.) Check  $\sum_{k=1}^K \sum_{j \in B_k} |v_{kj}^{new} - v_{kj}| \leq \epsilon$ , or the number of iterations is greater than the maximum number of iterations,

1.4.1 If true, then go to Step 2.

1.4.2 If false, set  $v_{kj} = v_{kj}^{new} \forall j \in B_k$ ,  $k = 1, \dots, K$  and go to 1.2 of the E-Step.

*Step 2.* Set  $\lambda_t = X_{0t} + \sum_{k=1}^K \sum_{j \in B_k} X_{kj}$  and compute  $\log \mathcal{L}^{new} = \log \mathcal{L}(v, \mu, \lambda)$  using Equation (3).

*Step 3.* (Stopping criterion.) Check  $\log \mathcal{L}^{new} \leq \log \mathcal{L}$ , or  $\mu - \delta < 0$ .

3.1 If true, the procedure stops with the estimation result.

3.2 If false, set  $\mu = \mu - \delta$  and go to step 1.1 of the initialization of EM algorithm.

### 4.4. Brand-primary choice model and type-primary choice model

If the choice hierarchy of purchasing products is known (either the brand primary or the type-primary), then the proposed method is applied to estimate the demand parameters. Now, assume that there is no knowledge about the choice hierarchy. The hidden choice structure then can be revealed by applying the proposed method to both the brand-primary choice model and the type-primary choice model. The one with the higher likelihood value is chosen for the choice hierarchy model. We show this by using simulated data in Section 5.

## 5. Numerical study

We generated simulation data from a known demand structure where consumers follow the NMNL choice process. The proposed EM algorithm based on the NMNL choice model was applied to the simulated data to show how the algorithm estimates the demand parameters. We compared the results with the original demand parameters along with those by Vulcano *et al.* (2012).

We simulated the data with a homogeneous arrival rate of  $\lambda = 50$  over a selling horizon of  $T = 15$  periods. We assumed that consumers followed the brand-primary choice process. There were two product brands, brand A and brand B, and three different product types in each brand, 1, 2, 3. The preference vector was given by  $\mathbf{v} = ((v_{A1}, v_{A2}, v_{A3}), (v_{B1}, v_{B2}, v_{B3})) = ((1, 0.5, 0.1), (0.9, 0.4, 0.05))$ . The preference weight of no-purchase was normalized to be one,  $v_0 = 1$ . We used an inter-product-group similarity of  $\mu = 0.3$ . In this case, the exogenously given market share is  $s = 0.6919$ . The randomly simulated sales data are summarized in Table 2. The product unavailability was exogenously set before the simulation and is represented as "NA." Note that arrivals are unobservable in the sales data because no-purchases are unobservable (the number of arrivals is the sum of purchases and no-purchases).

**Table 2.** Simulated purchases and no-purchases over 15 periods for a product category with two brands and three types in the brand-primary choice process (NA denotes non-availability)

Brand	Type	Periods															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Observable data (purchases)	A	1	11	11	12	13	4	NA									
		2	4	4	6	4	5	9	9	12	11	11	19	16	NA	NA	NA
		3	2	0	0	1	0	4	2	3	0	2	4	4	17	8	6
B	1	10	7	9	8	13	12	7	10	NA							
		2	7	3	5	5	10	5	5	6	11	9	19	11	14	13	13
		3	0	1	0	0	0	0	2	0	1	0	3	0	3	3	1
Non-observable data	No-purchases	14	15	11	17	16	18	16	15	15	17	22	23	20	21	19	
	$\lambda$	48	41	43	48	48	48	41	46	38	39	67	54	54	45	39	

To estimate the demand parameters, we applied the EM algorithm based on the NMNL choice model to the observable sales data in Table 2. We assumed that the hierarchical type is unknown. Thus, we set up two different hierarchical models: the brand-primary model and the type-primary model. The value of incomplete data log-likelihood under the brand-primary model was  $-130.5036$  with an estimated inter-product-group similarity of 0.25, and that under the type-primary model is  $-140.5106$  with an estimated inter-group similarity of one (i.e., the type-primary model was reduced to the MNL model). This means that the MNL model fit the data sets better than the NMNL model if a wrong hierarchy was imposed. The value of the incomplete data log-likelihood was greater under the brand-primary model; thus, we selected the brand-primary model for the data. The estimations of  $\mu$  and  $\nu$  are given in Table 3. The second column shows the true values of the parameters; the third column shows the estimation results obtained using the EM algorithm based on the MNL, denoted by EM\_MNL; and the fourth column shows the estimation results obtained using the EM algorithm based on the NMNL, denoted by EM\_NMNL. Note that the parameter estimations by EM\_MNL can also be obtained from the initial solution of EM\_NMNL (initially, we set  $\mu = 1$ ). The percentage in the parentheses denotes the bias from the true value. The parameter bias of EM\_NMNL is smaller than that of EM\_MNL except for  $\hat{\nu}_{B2}$ .

The MLEs are consistent, asymptotically efficient, and normal. Thus, the asymptotic variance-covariance matrix of the estimators is  $-\text{Hess}[\log \mathcal{L}]^{-1}$ , where  $\text{Hess}[\log \mathcal{L}]$  is the matrix of the second derivative of the incomplete data log-likelihood with respect to the parameters evaluated at the estimated values of the parameters (Ben-Akiva and Lerman, 1985). In Online Supplement C, we present the second derivatives of the incomplete

data log-likelihood. The Asymptotic Standard Errors (ASEs) of all of the parameters are the square root of the diagonal elements of  $-\text{Hess}[\log \mathcal{L}]^{-1}$ . The ASE for the demand estimators of EM\_NMNL is in the fifth column of Table 3. For example, the ASE of  $\mu$  is 0.0142 and the ASE of  $\nu_{A1}$  is 0.5080. Due to the asymptotic property for the large sample data, the ASE can be reduced as the number of data increases (i.e., more period data).

The estimated substitute demands by EM\_MNL and by EM\_NMNL are given in Table 4. Two interesting points are observed by comparing the estimations obtained from the two algorithms. First, stockouts of product A1 in periods 6, 7, and 8 tend to be substituted more often by products in brand A than products in brand B under EM\_NMNL than under EM\_MNL. This is expected because all of the products' purchase probabilities increase of the same percentage under the MNL model, whereas the purchase probabilities of products in the same product group increase with a higher percentage under the NMNL model after the stockout (see Lemma 2). Second, the estimated value of lost sales is much smaller under the NMNL model than under the MNL model. The estimated  $\hat{\mu} = 0.25$ , which is far lower than one. This means a higher substitution within brands after stockouts (less lost sales). The primary demand estimations obtained using EM\_MNL and EM\_NMNL are given in Table 5. Note that the primary demands of A2 and A3 are smaller in periods 6, 7, and 8 under the NMNL model than under the MNL model. It is also noteworthy that the primary demands of brand B are larger in periods 6, 7, and 8 under the NMNL model than under the MNL model. The MNL model does not consider group-wise substitution. It thus overestimates the primary demand if its brand is the same as that of the unavailable product and underestimates the primary demand otherwise.

The average of the arrival rate estimations obtained using EM\_MNL is 57.6 with 15.2% bias from the true value of 50, and that by EM\_NMNL is 45.05 with a bias of  $-9.91\%$ . The estimation bias is also reduced in this case. The percentage of lost sales was estimated as

$$\text{Pr}(\text{lostsales}) = \frac{\sum_{t=1}^T Y_{0t}}{\sum_{k=1}^K \sum_{j \in B_k} N_{kj}} = \frac{157.9}{597.9} = 26.4\%$$

under the MNL model, and it was estimated as

$$\text{Pr}(\text{lostsales}) = \frac{27.7}{467.7} = 5.92\%$$

**Table 3.** Estimated demand parameters of EM\_MNL and EM\_NMNL for data in Table 2

Parameters	True value	EM_MNL (bias in percent)	EM_NMNL (bias in percent)	Asymptotic standard errors for EM_NMNL
$\hat{\mu}$	0.3	1 (70.00)	0.25 (-16.67)	0.0142
$\hat{\nu}_{A1}$	1	0.7388 (-26.12)	1.1317 (13.17)	0.5080
$\hat{\nu}_{A2}$	0.5	0.4134 (-17.32)	0.5301 (6.02)	0.1217
$\hat{\nu}_{A3}$	0.1	0.1124 (12.40)	0.0982 (-1.80)	0.0040
$\hat{\nu}_{B1}$	0.9	0.6136 (-31.82)	0.8868 (-1.47)	0.3687
$\hat{\nu}_{B2}$	0.4	0.3372 (-15.70)	0.5006 (25.15)	0.1112
$\hat{\nu}_{B3}$	0.05	0.0303 (-39.40)	0.0440 (-12.00)	0.0008

**Table 4.** Substitute demand estimations,  $\hat{Y}_{kj}$ , for simulated data in Table 2

	Brand	Type	Periods														
			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
EM_MNL	A	1	0	0	0	0	0	-11.4	-9.5	-11.7	-11.1	-10.6	-21.7	-15.0	-23.9	-16.8	-14.0
		2	0	0	0	0	0	2.0	2.0	2.7	4.6	4.6	7.9	6.7	-13.4	-9.4	-7.9
		3	0	0	0	0	0	0.9	0.5	0.7	0	0.8	1.7	1.7	9.2	4.4	3.3
	B	1	0	0	0	0	0	2.7	1.6	2.3	-9.2	-8.8	-18.0	-12.4	-19.8	-14.0	-11.7
		2	0	0	0	0	0	1.1	1.1	1.4	4.6	3.8	7.9	4.6	7.6	7.1	7.1
		3	0	0	0	0	0	0	0.5	0	0.4	0	1.3	0	1.6	1.6	0.5
	Lost sales	0	0	0	0	0	4.5	3.8	4.7	10.7	10.3	21.0	14.5	38.5	27.2	22.7	
EM_MNL	A	1	0	0	0	0	0	-10.3	-8.6	-10.6	-8.3	-7.9	-16.1	-11.1	-13.2	-9.3	-7.8
		2	0	0	0	0	0	5.2	5.2	6.9	6.7	6.7	11.6	9.7	-6.2	-4.4	-3.6
		3	0	0	0	0	0	2.3	1.2	1.7	0	1.2	2.4	2.4	15.5	7.3	5.5
	B	1	0	0	0	0	0	1.0	0.6	0.8	-7.5	-7.2	-14.8	-10.2	-12.1	-8.5	-7.1
		2	0	0	0	0	0	0.4	0.4	0.5	6.5	5.3	11.2	6.5	8.9	8.3	8.3
		3	0	0	0	0	0	0	1.2	0	0.6	0	1.8	0	1.9	1.9	0.6
	Lost sales	0	0	0	0	0	1.4	1.1	0.7	2.0	1.9	3.9	2.6	5.1	4.7	4.1	

under the NMNL model. The total substitute rate is estimated by

$$\frac{\sum_{t=1}^T \sum_{k=1}^K \sum_{j \in S_{kt}} Y_{kjt}}{\sum_{k=1}^K \sum_{j \in B_k} N_{kj}} = \frac{112.4}{597.9} = 18.80\%$$

under the MNL model, and it was estimated as 157.2/467.7 = 33.62% under the NMNL model. As mentioned earlier, a higher substitution within a product group results in the reduction of lost sales and a higher substitution rate.

We used the likelihood ratio index to compare the estimation results of EM\_NMNL and that of EM\_MNL. The null hypothesis of the likelihood ratio test was  $H_0: \mu = 1$ , and the alternative hypothesis was  $H_1: \mu \neq 1$ . Let  $\log \mathcal{L}(\text{EM\_MNL})$  be the log-likelihood by EM\_MNL and  $\log \mathcal{L}(\text{EM\_NMNL})$  be the log-likelihood by EM\_NMNL. Note that the MNL model was a NMNL model with a constraint of  $\mu = 1$ . The likelihood ratio index given by  $-2(\log \mathcal{L}(\text{EM\_MNL}) - \log \mathcal{L}(\text{EM\_NMNL}))$  had the chi-square distribution with a degree of freedom equal to the number of constraints, which is one in this case (see Train, 2009). As  $\log \mathcal{L}(\text{EM\_MNL}) = -140.5106$  and  $\log \mathcal{L}(\text{EM\_NMNL}) = -130.5046$ , the likelihood ratio index was 20.0120. This value is greater than the 5%

quantile of the chi-squared value with one degree of freedom, 3.841. Therefore, we rejected the hypothesis that the consumer choice model was MNL (i.e.,  $\mu = 1$ ).

We also considered in and out-of-sample statistics for the model selection criteria. First, we computed the Root Mean Squared Errors (RMSEs) of the two algorithms. We estimated the number of purchases of product  $j$  from group  $k$  as  $\hat{\lambda}_t \times P_{kj}(\mathbf{S}_t; \hat{\nu}, \hat{\mu})$ . The RMSE of EM\_MNL was 20.2714 and that of EM\_NMNL was 14.4218. Next, we computed the Root Mean Forecast Squared Errors (RMFSEs). We simulated additional sales date from period 16 to period 20 and forecasted the number of purchases of each product obtained using the two algorithms. The RMFSE of EM\_MNL was 14.4108 and that of EM\_NMNL was 11.5745. The RMSE and RMFSE of EM\_NMNL were smaller than those of EM\_MNL, respectively. Hence, the parameters of EM\_NMNL fit the data better than those of EM\_MNL.

Lastly, we considered other model selection criteria: the Akaike Information Criterion (AIC) and Bayes Information Criterion (BIC). Note that BIC penalizes extra parameters more heavily than does the AIC. Both the AIC and BIC have two terms: a negative log-likelihood and a penalty term penalizing a complex model (i.e., more parameters), which are

**Table 5.** Primary demand estimations,  $\hat{X}_{kj}$ , for simulated data in Table 2

	Product	Periods															$\hat{N}_{kj}$
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
EM_MNL	A1	11	11	12	13	4	11.4	9.5	11.7	11.1	10.6	21.7	15.0	23.9	16.8	14.0	196.7
	A2	4	4	6	4	5	7.0	7.0	9.3	6.4	6.4	11.1	9.3	13.4	9.4	7.9	110.1
	A3	2	0	0	1	0	3.1	1.5	2.3	0	1.2	2.3	2.3	7.8	3.6	2.7	29.9
	B1	10	7	9	8	13	9.3	5.4	7.7	9.2	8.8	18.0	12.4	19.8	14.0	11.7	163.4
	B2	7	3	5	5	10	3.9	3.9	4.6	6.4	5.2	11.1	6.4	6.4	5.9	5.9	89.8
	B3	0	1	0	0	0	0	1.5	0	0.6	0	1.7	0	1.4	1.4	0.5	8.1
	No-purch.	15.1	11.6	14.2	13.8	14.2	15.4	12.8	15.9	15.0	14.4	29.4	20.2	32.3	22.8	19.0	266.2
	$\hat{\lambda}_t$	49.1	37.6	46.2	44.8	46.2	49.9	41.6	51.6	48.7	46.6	95.4	65.7	104.9	74.0	61.7	864.1
EM_NMNL	A1	11	11	12	13	4	10.3	8.6	10.6	8.3	7.9	16.1	11.1	13.2	9.3	7.8	154.2
	A2	4	4	6	4	5	3.8	3.8	5.1	4.3	4.3	7.4	6.3	6.2	4.4	3.6	72.3
	A3	2	0	0	1	0	1.7	0.8	1.3	0	0.8	1.6	1.6	1.5	0.7	0.5	13.4
	B1	10	7	9	8	13	11	6.4	9.2	7.5	7.2	14.8	10.2	12.1	8.5	7.1	141.1
	B2	7	3	5	5	10	4.6	4.6	5.5	4.5	3.7	7.8	4.5	5.1	4.7	4.7	79.7
	B3	0	1	0	0	0	0	1.8	0	0.4	0	1.2	0	1.1	1.1	0.4	7.0
	No purch.	15.1	11.6	14.2	13.8	14.2	14	11.6	14.1	11.1	10.6	15	17.4	12.8	10.7	208.3	
	$\hat{\lambda}$	49.1	37.6	46.2	44.8	46.2	45.4	37.8	45.8	36.2	34.5	70.7	48.6	56.5	41.5	34.8	676.0

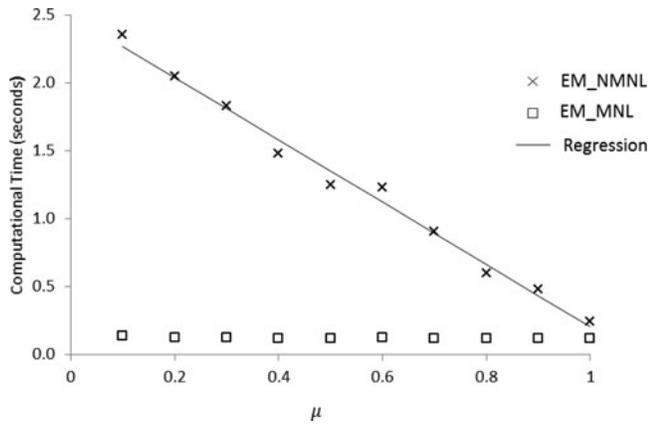


Figure 2. Computational time (in seconds) with respect to  $\mu$ .

defined by

$$\begin{aligned} \text{AIC} &= -2 \log L + 2 \times d, \\ \text{BIC} &= -2 \log L + d \times \log |data|, \end{aligned}$$

where  $d$  is the number of parameters and  $|data|$  is the number of observations. The number of observations was  $\sum_{t=1}^T \sum_{k=1}^K \sum_{j \in B_k} z_{kjt} = 699$ . The number of parameters was 21 for the MNL model and 22 for the NMNL model. The AIC of EM\_MNL was 323.0212 and that of EM\_NMNL was 305.0092. The BIC of EM\_MNL was 418.5639 and that of EM\_NMNL was 405.1015. The values of AIC and BIC for the EM\_NMNL were smaller than those obtained for the EM\_MNL. Accordingly, one parameter increase of the NMNL model (i.e.,  $\mu$ ) is justified.

### 5.1. Computational time and $\mu$

We compared the computational time of EM\_NMNL with that of EM\_MNL. Both algorithms were implemented using Python (version 2.7), and we used Microsoft Windows 7 on a CPU with an Intel Core i5 processor and 4 GB of RAM. The computational time difference as the two algorithms depends on the true value of  $\mu$  because EM\_NMNL uses  $\mu = 1$  as an initial value (see Section 4.3) and decreases the value by  $\delta = 0.05$  in each iteration. If the true value of  $\mu$  is close to one (i.e., closer to the EM\_MNL), EM\_NMNL terminates after a lower number of iterations. For this reason, we changed the value of  $\mu$  from 0.1 to 1 (in increments of 0.1) and simulated the data for each  $\mu$ . We used the parameters  $\mathbf{v} = ((v_{A1}, v_{A2}, v_{A3}), (v_{B1}, v_{B2}, v_{B3})) = ((1, 0.5, 0.1), (0.9, 0.4, 0.05))$ ,  $\lambda = 50$ , and  $T = 15$  (note that the market share changes with  $\mu$ ). Figure 2 shows the computational time (in seconds) of the two algorithms with respect to  $\mu$ . The square marks represent the computational time of EM\_MNL. The cross marks represent the computational time of EM\_NMNL. For example, when  $\mu = 0.1$ , the computational time is 2.35 seconds for EM\_NMNL and 0.14 seconds for EM\_MNL. When  $\mu = 1$ , the computational time is 0.24 seconds for EM\_NMNL and 0.12 seconds for EM\_MNL. The line overlaid on the cross marks is a regression line with a slope of  $-2.29$ . When  $\mu = 1$ , the computational time of EM\_NMNL is twice that of EM\_MNL. In this case, the stopping criterion of Step 3 in Section 4.3 is met after the decrease of  $\mu$  from one (i.e., the initial value) to 0.95. As the slope of the regression line is  $-2.29$ , the computational time

increases approximately 0.23 seconds for each increase of  $\mu$  by 0.1 in this example.

### 5.2. Substitution patterns in NMNL model

The algorithm of EM\_NMNL predicts the choice hierarchy based on the group-wise substitution rather than the stockout pattern. To see this, assume that there are two brands (brand A and brand B) and two types (type 1 and type 2). The preference vector is given by  $(v_{A1}, v_{A2}, v_{B1}, v_{B2}) = (1, 0.5, 1, 0.5)$ . We considered a homogeneous arrival rate  $\lambda = 50$  over  $T = 15$ . We set exogenously that type 1 is more unavailable (i.e., one product type stocks out more frequently than others): product A1 is unavailable from period 10 to period 12 and product B1 is unavailable from period 13 to period 15. We simulated two data sets: one was from the brand-primary model with  $\mu = 0.3$  and the other was from the type-primary model with  $\mu = 0.3$ . Table 6 shows the two simulated data sets.

Different substitution patterns were observed in the two data sets depending on the underlying chosen hierarchy. We observed the strong substitution pattern within the same brand for the data set from the brand-primary process, whereas the strong substitution pattern within the same type for the data set from the type-primary process.

We applied EM\_NMNL to those simulated data sets, assuming that the chosen hierarchy was unknown. We first discuss the result of the data set from the brand-primary process in the upper part of Table 6. The incomplete data log-likelihood value is  $-118.5928$  under the brand-primary model and  $-134.5083$  under the type-primary model. Hence, the algorithm predicts that the data set follows the brand-primary process. Next, we discuss the result of the data set from the type-primary process in the lower part of Table 6. The incomplete data log-likelihood value is  $-131.4395$  under the brand-primary model and  $-124.8578$  under the type-primary model. Thus, the algorithm predicts that the data set is from the type-primary process.

### 5.3. Product assortment and primary demand estimation

Better accuracy in primary demand estimation would lead to a revenue increase through optimal assortment decisions and inventory management. In this section, we use the revenue model by Van Ryzin and Mahajan (1999) to illustrate the optimal assortment decision making.

We denote the number of purchases (or demand) of product  $j$  from group  $k$  as  $Z_{kj}(\mathbf{S})$  under the assortment of  $\mathbf{S}$ . We assume that  $Z_{kj}(\mathbf{S})$  is normally distributed with mean  $\lambda \times P_{kj}(\mathbf{S}; \mathbf{v}, \mu)$  and standard deviation  $\sqrt{\lambda} \times P_{kj}(\mathbf{S}; \mathbf{v}, \mu)$  (i.e., the Poisson distribution is approximated by a normal distribution).

Let  $p$  be the unit revenue for each sale and  $c$  be the unit cost. For computational simplicity for inventory decision making, we assumed that  $c = p/2$ . We denote the number of units stocked as  $x_{kj}$  (inventory level). Then, the expected profit under assortment  $\mathbf{S}$  is

$$\pi(\mathbf{S}, \mathbf{v}, \mu) = \max_{x_{kj} \geq 0} \sum_{k=1}^K \sum_{j \in S_k} E[p \times \min\{x_{kj}, Z_{kj}\} - c \times x_{kj}].$$

**Table 6.** Simulated purchases and no-purchases over 15 periods (NA denotes non-availability) by brand-primary process and type-primary process

		Brand-primary choice														
Brand	Type	Observable data (purchases)														
		Periods														
A	1	11	11	13	13	4	9	8	12	11	NA	NA	NA	17	10	8
	2	4	4	5	5	5	8	6	4	1	11	19	17	6	6	3
B	1	11	7	9	8	12	11	6	12	10	9	20	12	NA	NA	NA
	2	8	4	5	5	11	5	5	5	3	4	7	4	16	10	14
No-purchases		14	15	11	17	16	15	16	13	13	15	21	21	15	19	14
$\lambda$		48	41	43	48	48	48	41	46	38	39	67	54	54	45	39
		Type-primary choice														
Type	Brand	Observable data (purchases)														
		Periods														
1	A	10	11	10	11	4	5	7	9	11	NA	NA	NA	21	15	10
	B	7	5	9	8	6	13	7	10	1	14	23	20	NA	NA	NA
2	A	7	2	3	5	9	7	4	6	9	4	13	7	9	5	7
	B	10	7	10	7	13	8	7	6	4	6	11	6	9	6	8
No-purchases		14	16	11	17	16	15	16	15	13	15	20	21	15	19	14
$\lambda$		48	41	43	48	48	48	41	46	38	39	67	54	54	45	39

When  $c = p/2$ , the optimal procurement level is simply  $x_{kj}^* = \lambda \times P_{kj}(S; \mathbf{v}, \mu)$  (see Van Ryzin and Mahajan (1999)). Applying the optimal procurement level, we have

$$\pi(S, \mathbf{v}, \mu) = \sum_{k=1}^K \sum_{l \in S_k} \frac{p\lambda}{2} P_{kj}(S; \mathbf{v}, \mu) - \frac{p\sqrt{\lambda}}{\sqrt{2\pi}} \sqrt{P_{kj}(S; \mathbf{v}, \mu)}.$$

The structure of the optimal assortment for the MNL model was studied in Van Ryzin and Mahajan (1999), and that for the NMNL model was performed in Kok and Xu (2011). For the MNL model, the optimal assortment is to simply choose the best  $i$  alternatives in terms of the preference weights. Similarly, the optimal assortment for the NMNL model is to choose the best  $i$  alternatives for each product group. For the numerical study, we used the simulated data set in Table 2. We assumed that  $p = \$1$ . The optimal assortment under the MNL model is (A1, A2, B1, B2) with the procurement level of  $(x_{A1}, x_{A2}, x_{B1}, x_{B2}) = (13.7, 7.7, 11.4, 6.3)$ . The optimal assortment under the NMNL model is (A1, B1) with the procurement level of  $(x_{A1}, x_{B1}) = (15.5, 14.6)$ .

Using those procurement plans, we computed the expected profits under the true model parameter of  $\nu = ((v_{A1}, v_{A2}, v_{A3}), (v_{B1}, v_{B2}, v_{B3})) = ((1, 0.5, 0.1), (0.9, 0.4, 0.05))$ ,  $\mu = 0.3$ , and  $\lambda = 50$ . In Table 7, we summarize the assortment plans (i.e., the expected number of purchases), the expected sales, and the expected remaining inventory after sales using EM\_MNL

**Table 7.** Procurement level, expected sales, and remaining inventory by EM\_MNL and EM\_NMNL

Product	Procurement level	Expected sales	Remaining inventory
EM_MNL			
A1	13.7	11.1	2.6
A2	7.7	5.5	2.2
B1	11.4	10.2	1.2
B2	6.3	4.7	1.6
Sum	39.1	31.5	7.6
EM_NMNL			
A1	15.5	14.4	1.1
B1	14.6	13.7	0.9
Sum	30.1	28.1	2

and EM\_NMNL. The procurement levels of both product A1 and product B1 increase under the NMNL model, when compared with those under the MNL model. Under the NMNL model, the store can have a less diverse assortment plan (i.e., lower number of available products) by using the strong substitution within the product group. The total expected revenue decreases with lower diversity, but the inventory cost decreases more (i.e., the cost saving outweighs the revenue loss). The expected profit is \$12.0138 under the MNL model and \$13.1005 under the NMNL model. The store has a 9% profit increase by using the NMNL model rather than the MNL model.

### 5.4. Market share estimation

The demand estimation algorithm proposed in this article requires knowledge of the market share. In this section, we show the effect of market share on the estimation of the primary demand and discuss how to estimate the market share when it is unknown.

In Section 3.2, we showed that if  $(\mathbf{v}^*, \mu^*, \lambda^*)$  is a maximizer of the incomplete data likelihood function of Equation (3), we can generate another maximizer,  $(\mathbf{v}^0, \mu^*, \lambda^0)$ , where  $\mathbf{v}^0 = \alpha \mathbf{v}^*$  and

$$\lambda_t^0 = \frac{\sum_{k=1}^K (\sum_{l \in S_{kt}} \alpha v_{kl}^*)^{\mu^*} + 1}{\sum_{k=1}^K (\sum_{l \in S_{kt}} \alpha v_{kl}^*)^{\mu^*} + \alpha^{\mu^*}} \lambda_t^*$$

for any  $\alpha > 0$ . The market share is

$$s = 1 - P_0(\mathbf{B}; \mathbf{v}, \mu) = \frac{\sum_{m=1}^K (\sum_{l \in B_m} v_{ml}^*)^\mu}{\sum_{m=1}^K (\sum_{l \in B_m} v_{ml}^*)^\mu + 1}.$$

Thus, for any  $\alpha > 0$ , we define

$$s(\alpha) = \frac{\sum_{m=1}^K (\sum_{l \in B_m} \alpha v_{ml}^*)^\mu}{\sum_{m=1}^K (\sum_{l \in B_m} \alpha v_{ml}^*)^\mu + 1},$$

which is the market share for the parameter set of  $(\mathbf{v}^0, \mu^*, \lambda^0)$ . We let the parameter estimators in Table 3 be  $(\mathbf{v}^*, \mu^*, \lambda^*)$  and

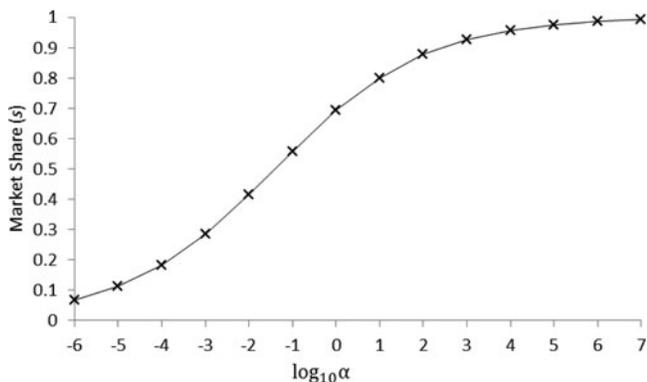


Figure 3. Market share with respect to  $\log_{10}\alpha$ .

depict the market share in terms of  $\log_{10}\alpha$  in Fig. 3. When  $\alpha = 1$ , the market share is  $s(1) = 0.6919$ . When  $\alpha = 2$ , the market share is  $s(2) = 0.7276$ . Thus, the effect of the market share can be equivalently analyzed by using  $\alpha$ .

The primary demand estimation depends on the market share, and hence the value of  $\alpha$ . The parameter set of  $(v^0, \mu^*, \lambda^0)$  has the primary demand estimation

$$\hat{X}_{kjt}(\alpha) = \begin{cases} \frac{P_{kj}(\mathbf{B}; \alpha v^*, \mu^*)}{1 - P_0(S_t; \alpha v^*, \mu^*)} \sum_{m=1}^K \sum_{l \in S_{mt}} z_{mlt} & \text{for all } j \in B_k \setminus S_{kt} \\ \frac{P_{kj}(\mathbf{B}; \alpha v^*, \mu^*)}{P_{kj}(S_t; \alpha v^*, \mu^*)} \times z_{kjt} & \text{for all } j \in S_{kt} \end{cases} \quad (9)$$

The primary demand of no-purchase is given by

$$\hat{X}_{0t}(\alpha) = \frac{1 - s(\alpha)}{s(\alpha)} \sum_{k=1}^K \sum_{j \in B_k} \hat{X}_{kjt}(\alpha). \quad (10)$$

We now analyze Equations (9) and (10) under the full assortment and under the partial assortment.

Under the full assortment (i.e.,  $S_t = \mathbf{B}$ ), we have  $\hat{X}_{kjt}(\alpha) = \hat{X}_{kjt}(1)$ . This means  $\hat{X}_{kj1} = z_{kj1}$  and implies that demand estimation does not depend on  $\alpha$  and, thus, the market share. However, the no-purchase estimation of  $\hat{X}_{0t}(\alpha)$ , as can be seen in

Equation (10), depends on the market share. We depict the estimated values of no-purchase in period 1 in terms of the market shares in Fig. 4(a). It shows that the estimation indeed depends on the market share. If no-purchase in period 1 is observed by some means (e.g., CCTV, hired observers), Fig. 4(a) can be used to estimate the market share.

The demand estimation, unlike the full assortment, depends on the market share under the partial assortment. To show this, we estimated the primary demand of product B2 in period 15 and drew show a graph of it with respect to  $s$  as Fig. 4(b). For example, its estimate is 5.8 when  $s = 0.2854$  and it is 4.2 when  $s = 0.9266$ . Although the market share changes from 0.2854 to 0.9266 (an approximately 325% increase), the demand estimate for B2 decreases from 5.8 to 4.2 (approximately a 27% decrease). In this example, the estimation changes depending on the market share, but the sensitivity does not seem high.

If no-purchase under the full assortment is not available to estimate the market share (as discussed earlier), another way to estimate it is using the sales data in periods having the full assortment (i.e., no stockout). First, we estimate the number of purchases using the estimated parameters for each market share, using  $\hat{z}_{kjt} = \hat{\lambda}_t \times P_{kj}(\mathbf{B}; \hat{v}, \hat{\mu})$ . We then select the market share  $s$  at which the sum of squared errors is minimized. The sum of squared errors is

$$\sum_{t=1}^5 \sum_{k=1}^K \sum_{j \in B_k} (z_{kjt} - \hat{z}_{kjt})^2.$$

For example, the sales data in periods 1, 2, 3, 4, 5 of Table 2 were used to estimate  $s$ . In Fig. 4(c), the sum of squared errors is depicted with respect to  $s$ . That value is minimized when  $s = 0.99$  (the true market share is 0.6919). The graph shows that any market share greater than 0.4 does not make much difference in estimating the primary demands under the full assortment.

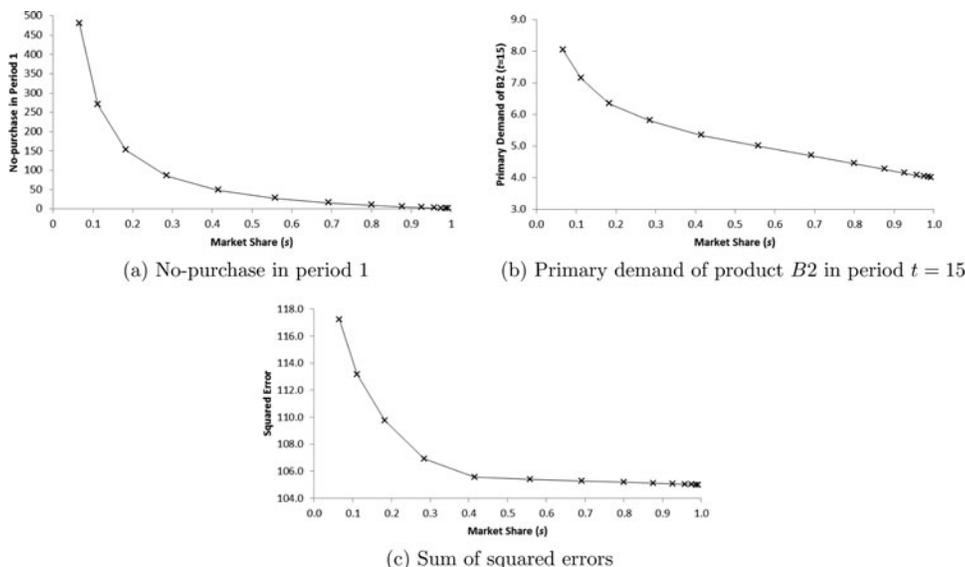


Figure 4. Demand estimation with respect to market share  $s$ : (a) no-purchase in period 1; (b) primary demand of product B2 in period  $t = 15$ ; and (c) sum of squared errors.

## 6. Extensions and further discussions

We have assumed that all consumers use the same hierarchical purchase decisions: either the brand-primary process or the type-primary process. This can be extended to the mixture of the two models. An arriving consumer follows the brand-primary hierarchy with probability  $q^B$  and the type-primary hierarchy with probability  $q^T = 1 - q^B$ . The primary demand approach can also be applied in this mixture model. The model description and the EM algorithm are in Online Supplement A. We present an interesting numerical example where a simple MNL model is better than the mixture model, although the data set is from the mixture model.

The primary demand estimation method for the NMNL model with two-stage choice hierarchy can be extended to the NMNL model with more than two stages. The model description and the EM algorithm for the NMNL model with three stages are in Online Supplement B. We present a numerical example with the simulated data set from the choice hierarchy with three stages.

The choice model based on utility  $U_{kj} = u_{kj} + \epsilon_k + \epsilon_{kj}$  in Section 3.1 results in the closed-form solution in the M-step. If there are factors influencing the utility, the utility can be written as  $u_{kjt} = \beta^T \mathbf{x}_{kjt}$ . For example, vector  $\mathbf{x}_{kjt}$  can contain  $p_{kjt}$  (the price of product  $j$  from group  $k$  in period  $t$ ), an individual consumer indicator variable, individual consumer characteristics, and product characteristics. The preference weight is given by  $v_{kjt} = e^{u_{kjt}}$ . The primary demand approach can still be applied, but the nonlinear optimization algorithm needs to be applied to find optimal coefficients of  $\beta_{kjt}$  in the M-step. If all of the factors are stationary (i.e., not dynamic so that  $u_{kj} = \beta^T \mathbf{x}_{kj}$ ), then we can apply the two-step approach proposed by Vulcano *et al.* (2012): In step 1, find the optimal preference weight vector of  $\hat{\mathbf{v}}$  in the M-step. In step 2, use the regression method to find  $\beta$  solving the equation  $\log \hat{\mathbf{v}}_{kj} = \beta^T \mathbf{x}$ . This approach is suitable when the system of equations are over-determined.

## 7. Conclusions

We studied the primary demand-estimation problem for a heterogeneous-group product category that is sold in a department store setting. For such products, the hierarchical consumer choice model better represents consumers' purchasing behaviors. We considered two versions of hierarchical choice model: the brand-primary choice model and the type-primary choice model. For demand estimation, the NMNL choice model of hierarchical choice is combined with the non-homogeneous Poisson arrivals over multiple periods. The NMNL model better represents the different substitution rates among product groups, which cannot be incorporated into the MNL model. Our approach addresses how to estimate the NMNL model parameters, which are the inter-product-group similarity and preference vector along with the arrival rates. In the estimation, we assumed that sales data along with product availability and market share information are available. We applied the EM algorithm to estimate the demand parameters. In the M-step, there is no closed-form solution maximizing the conditional expected log-likelihood function developed in the E-step. To overcome

this difficulty, we proposed a method whereby the EM algorithm is considered as a sub-module that is called repeatedly by an optimization problem in terms of the inter-product-group similarity. In this way, the revenue management system is relieved of the computational burden of using a nonlinear optimization package. In the present study, we compared the estimation result with that of the MNL model by Vulcano *et al.* (2012) for simulated data sets reflecting a high level of inter-product-group heterogeneity. With these simulated data, the NMNL model more correctly estimated the demand parameters. By contrast, in the primary demand estimation, the MNL model overestimated the demand for a product in the group of an unavailable product and underestimated the demand for a product in the other groups. This happened because the MNL model, unlike the NMNL model, does not consider the inter-product-group heterogeneity that leads to different substitution rates of product groups. The proposed approach also can be extended to the general NMNL model in Train (2009), where different levels of product-group scale parameters are used instead of the single scale parameter. Our preliminary analysis, not included in this article, showed that the optimal preference vector maximizing the conditional expected log-likelihood in the M-step can be obtained for a given set of scale parameters. However, a multivariate optimization problem thereby is incurred. The resolution of the multivariate optimization problem will be the focus of future research.

## Acknowledgements

The authors thank the referees for their constructive comments and suggestions.

## Funding

This research was supported in part by Global Research Laboratory Program (2013K1A1A2A02078326) through the NRF and also by the IITP grant funded by the Korea government (B0101-15-0557).

## Notes on contributors

**Haengju Lee** is a Visiting Research Professor in the Department of Information and Communication Engineering at DGIST. Her research interests include revenue management and intelligent transportation systems.

**Yongsoon Eun** is an Associate Professor in the Department of Information and Communication Engineering at DGIST. His research interests include state estimation in cyber-physical systems, resilient control systems, and intelligent transportation systems.

## References

- Anupindi, R., Dada, M. and Gupta, S. (1998) Estimation of consumer demand with stock-out based substitution: an application to vending machine products. *Marketing Science*, **17**, 406–423.
- Ben-Akiva, M. and Lerman, S.R. (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press, Cambridge, MA.
- Bruno, H. and Vilcassim, N. (2008) Structured demand estimation with varying product availability. *Marketing Science*, **27**, 1126–1131.
- Campo, K., Gijsbrechts, E. and Nisol, P. (2003) The impact of retailer stock-outs on whether, how much, and what to buy. *International Journal of Research in Marketing*, **20**, 273–286.

- Conlon, C. and Mortimer, J. (2009) Demand estimation under incomplete product availability. Working paper Department of Economics, Harvard University, Cambridge, MA.
- Fisher, M. and Vaidyanathan, R. (2014) A demand estimation procedure for retail assortment optimization with results from implementations. *Management Science*, **60**, 2401–2415.
- Gallego, G. and Topaloglu, H. (2014) Constrained assortment optimization for the nested logit model. *Management Science*, **60**, 2583–2601.
- Kalyanam, K., Borle, S. and Boatwright, P. (2007) Deconstructing each item's category contribution. *Marketing Science*, **26**, 327–341.
- Kannan, P.K. and Wright, G.P. (1991) Modeling and testing structured markets: a nested logit approach. *Marketing Science*, **10**, 58–82.
- Kök, G. and Fisher, M. (2007) Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research*, **55**, 1001–1021.
- Kök, G. and Xu, Y. (2011) Optimal and competitive assortments with endogenous pricing under hierarchical consumer choice models. *Management Science*, **57**, 1546–1563.
- Musalem, A., Olivares, M., Bradlow, E., Terwies, C. and Corsten, D. (2010) Structural estimation of the effect of out-of-stocks. *Management Science*, **56**, 1180–1197.
- Train, K. (2009) *Discrete Choice Methods with Simulation*, Cambridge University Press, New York, NY.
- Van Ryzin, G. and Mahajan, S. (1999) On the relationships between inventory costs and variety benefits in retail assortments. *Management Science*, **45**, 1496–1509.
- Vulcano, G., Van Ryzin, G. and Ratliff, R. (2012) Estimating primary demand for substitutable products from sales transaction data. *Operations Research*, **60**, 313–334.
- Wu, C.F. (1983) On the convergence properties of the EM algorithm. *The Annals of Statistics*, **11**, 95–103.