OPERATIONS RESEARCH
Vol. 60, No. 2, March–April 2012, pp. 313–334
ISSN 0030-364X (print) | ISSN 1526-5463 (online)

http://dx.doi.org/10.1287/opre.1110.1012

Estimating Primary Demand for Substitutable Products from Sales Transaction Data

Gustavo Vulcano
Leonard N. Stern School of Business, New York University, New York, New York 10012,
gvulcano@stern.nyu.edu

Garrett van Ryzin
Graduate School of Business, Columbia University, New York, New York 10027,
gjv1@columbia.edu

Richard Ratliff
Sabre Holdings, Southlake, Texas 76092, richard.ratliff@sabre-holdings.com

We propose a method for estimating substitute and lost demand when only sales and product availability data are observable, not all products are displayed in all periods (e.g., due to stockouts or availability controls), and the seller knows its aggregate market share. The model combines a multinomial logit (MNL) choice model with a nonhomogeneous Poisson model of arrivals over multiple periods. Our key idea is to view the problem in terms of primary (or first-choice) demand; that is, the demand that would have been observed if all products had been available in all periods. We then apply the expectation-maximization (EM) method to this model, and we treat the observed demand as an incomplete observation of primary demand. This leads to an efficient, iterative procedure for estimating the parameters of the model. All limit points of the procedure are provably stationary points of the incomplete data log-likelihood function. Every iteration of the algorithm consists of simple, closed-form calculations. We illustrate the effectiveness of the procedure on simulated data and two industry data sets.

Subject classifications: demand estimation; demand untruncation; choice behavior; multinomial logit model; EM method.
Area of review: Revenue Management.
History: Received July 2008; revisions received December 2009, April 2011; accepted July 2011.

1. Introduction

Two important problems in retail demand forecasting are estimating turned-away demand when items are sold out and properly accounting for substitution effects among related items. For simplicity, most retail demand forecasts rely on time-series models of observed sales data, which treat each stock keeping unit (SKU) as receiving an independent stream of requests. However, if the demand lost when a customer's first choice is unavailable (referred to as *spilled* demand) is ignored, the resulting demand forecasts might be negatively biased; this underestimation can be severe if products are unavailable for long periods of time. Concurrently, *stockout-based substitution* will increase sales in substitute products that are available (referred to as *recaptured* demand); ignoring recapture in demand forecasting leads to an overestimation bias among the set of available SKUs. Correcting for both spill and recapture effects is important in order to establish a good estimate of the true underlying demand for products.

A similar problem arises in forecasting demand for booking classes in the airline industry. One common heuristic used in practice to correct for spilled demand is to assume that the demand turned away is proportional to the degree of "closedness" of a product (an itinerary-fare-class combination). For instance, suppose a booking class is open (available for sale) during 10 days of a month with 30 days. If 20 bookings are observed for the month, then this heuristic approach will estimate a demand of $20 \times 30/10 = 60$ for this booking class. However, because the observed 20 bookings might include some recapture from other, closed itinerary-fare-classes, this (uncorrected) approach can lead to a "double counting" problem; namely, spill is estimated on unavailable products but also counted as recapture on alternate, available products.

Empirical studies of different industries show that stockout-based substitution is a common occurrence. For airline passengers, recapture rates are acknowledged to be in the range of 15%–55% (e.g., Ja et al. 2001), while Gruen et al. (2002) report recapture rates of 45% across 8 categories at retailers worldwide.

Because spilled and recaptured demand are not directly observable from sales transactions, various statistical techniques have been proposed to estimate them. Collectively, these techniques are known as *demand untruncation* or *uncensoring* methods. One of the most popular such methods is the expectation-maximization (EM) algorithm. EM procedures ordinarily employ iterative methods to estimate the underlying parameters of interest; in our case, demand by SKU across a set of historical data. The EM method works by using alternating steps of computing conditional expected values of the parameter estimates to obtain an expected log-likelihood function (the E-step) and maximizing this function to obtain improved estimates (the M-step). Traditionally, retail forecasts that employ the EM approach have been limited to untruncating sales history for individual SKUs and disregard recapture effects from substitute products.

Classical economic theory on substitution effects (e.g., see Nicholson 2004) provides techniques for estimating demand shifts due to changes in prices of alternative offerings. However, an important practical problem is how to fit such demand models when products are out of stock or otherwise unavailable, and how to do so using only readily available data, which in most retail settings consist of sales transactions, product attributes (brand, size, price, etc.), and on-hand inventory quantities by SKU. Our work helps address this problem.

A widely used approach for estimating demand for different SKUs within a set of similar items is to use discrete choice models, such as the multinomial logit (MNL) (e.g., see Ben-Akiva and Lerman 1994 and Train 2003). Choice models predict the likelihood of customers purchasing a specific product from a set of related products based on their relative attractiveness. A convenient aspect of the MNL model is that the likelihood of purchase can be readily recalculated if the mix of available related products changes (e.g., due to another item being sold out or restocked).

In this paper, we propose a novel method of integrating customer choice models with the EM method to untruncate demand and correct for spill and recapture effects across an entire set of related products. Our model of demand combines a multinomial logit (MNL) choice model with a nonhomogeneous Poisson model of arrivals over multiple periods. The problem we address is how to jointly estimate the preference weights of the products and the arrival rates of customers. The only required inputs are observed historical sales, product availability data, and market share information. The key idea is to view the problem in terms of primary (or first-choice) demand and to treat the observed sales as incomplete observations of primary demand. We then apply the EM method to this primary demand model and show that it leads to an efficient, iterative procedure for estimating the parameters of the choice model. All limit points of the procedure are provably stationary points of the associated incomplete data log-likelihood function. Because our estimates are maximum likelihood estimates (MLEs), they inherit the statistical properties of a MLE: they are consistent (i.e., they converge in probability to the true parameter values), asymptotically normal, and asymptotically efficient (i.e., asymptotically unbiased and attaining equality of the Cramér–Rao lower bound for the variance, asymptotically).

Our EM method also provides an estimate of the number of lost sales—that is, the number of customers who would have purchased if all products were in stock—which is critical information in retailing. The approach is also remarkably simple, fast, and effective, as illustrated on simulated data and two industry data sets.

2. Literature Review

There are related papers in the revenue management literature on similar estimation problems. Talluri and van Ryzin (2004, §5) develop an EM method to jointly estimate arrival rates and parameters of a MNL choice model based on consumer level panel data under unobservable no-purchases. Vulcano et al. (2010) provide empirical evidence of the potential of that approach. Ratliff et al. (2008) provide a comprehensive review of the demand untruncation literature in the context of revenue management settings. They also propose a heuristic to jointly estimate spill and recapture across numerous flight classes, by using balance equations that generalize the proposal of Andersson (1998). A similar approach was presented before by Ja et al. (2001).

Another related stream of research is the estimation of demand and substitution effects for assortment planning in retailing. Kök and Fisher (2007) identify two common models of substitution:

1. The utility-based model of substitution, where consumers associate a utility with each product (and also with the no-purchase option) and choose the highest utility alternative available. The MNL model belongs to such class. The single-period assortment planning problem studied by van Ryzin and Mahajan (1999) is an example of the applicability of this model.
2. The exogenous model of substitution, where customers choose from the complete set of products, and if the item they choose is not available, they may accept another variant as a substitute according to a given substitution probability (e.g., see Netessine and Rudi 2003).

Other papers in the operations and marketing science literature also address the problem of estimating substitution behavior and lost sales. Anupindi et al. (1998) present a method for estimating consumer demand when the first choice variant is not available. They assume a continuous time model of demand and develop an EM method to uncensor times of stockouts for a periodic review policy, with the constraint that at most two products stock out in order to handle a manageable number of variables. They find maximum likelihood estimates of arrival rates and substitution probabilities.

Swait and Erdem (2002) study the effect of temporal consistency of sales promotions and availability on consumer choice behavior. The former encompasses variability of prices, displays, and weekly inserts. The latter also influences product utility, because the uncertainty of a SKU's presence in the store might lead consumers to consider the product less attractive. They solve the estimation problem via simulated maximum likelihood and test it on fabric softener panel data, assuming a variation of the MNL model to explain consumer choice; but there is no demand uncensoring in their approach.

Campo et al. (2003) investigate the impact of stockouts on purchase quantities by uncovering the pattern of within-category shifts and by analyzing dynamic effects on incidence, quantity, and choice decisions. They propose a modification of the usual MNL model to allow for more general switching patterns in stockout situations, and they formulate an iterative likelihood estimation algorithm. They then suggest a heuristic two-stage tracking procedure to identify stockouts: in a first stage, they identify potential stockout periods; in stage two, these periods are further screened using a sales model and an iterative outlier analysis procedure (see Appendix A therein).

Borle et al. (2005) analyze the impact of a large-scale assortment reduction on customer retention. They develop models of consumer purchase behavior at the store and category levels, which are estimated using Markov chain Monte Carlo (MCMC) samplers. Contrary to other findings, their results indicate that a reduction in assortment reduces overall store sales, decreasing both sales frequency and quantity.

Chintagunta and Dubé (2005) propose an estimation procedure that combines information from household panel data and store level data to estimate price elasticities in a model of consumer choice with normally distributed random coefficients specification. Their methodology entails maximum likelihood estimation (MLE) with instrumental variables regression (IVR) that uses share information of the different alternatives (including the no-purchase option). Different from ours, their model requires no-purchase store visit information.

Kalyanam et al. (2007) study the role of each individual item in an assortment, estimating the demand for each item as well as the impact of the presence of each item on other individual items and on aggregate category sales. Using a database from a large apparel retailer, including information on item specific out-of-stocks, they use the variation in a category to study the entire category sales impact of the absence of each individual item. Their model allows for flexible substitution patterns (beyond MNL assumptions), but stockouts are treated in a somewhat ad hoc way via simulated data augmentation. The model parameters are estimated in a hierarchical Bayesian framework, also through a MCMC sampling algorithm.

Bruno and Vilcassim (2008) propose a model that accounts for varying levels of product availability. It uses information on aggregate availability to simulate the potential assortments that consumers might face in a given shopping trip. The model parameters are estimated by drawing multivariate Bernoulli vectors consistent with the observed aggregate level of availability. They show that neglecting the effects of stockouts leads to substantial biases in estimation.

More recently, Musalem et al. (2010) also investigate substitution effects induced by stockouts. Different from ours, their model allows for partial information on product availability, which could be the case in a periodic review inventory system with infrequent replenishment. However, their estimation algorithm is much more complex and computationally intensive than ours because it combines MCMC with sampling using Bayesian methods.

The aforementioned paper by Kök and Fisher (2007) is close to ours. They develop an EM method for estimating demand and substitution probabilities under a hierarchical model of consumer purchase behavior at a retailer. This consumer behavior model is similar to the one in Campo et al. (2003) and is standard in the marketing literature; see e.g., Bucklin and Gupta (1992) and Chintagunta (1993). In their setting, upon arrival, a consumer decides: (1) whether or not to buy from a subcategory (purchase-incidence), (2) which variant to buy given the purchase incidence (choice), and (3) how many units to buy (quantity). Product choice is modeled with the MNL framework. Unlike our aggregate demand setting, they analyze the problem at the individual consumer level and assume that the number of customers who visited the store but did not purchase anything is negligible (see Kök and Fisher 2007, §4.3). The outcome of the estimation procedure is combined with the parameters of the incidence purchase decision, the parameters of the MNL model for the first choice, and the coefficients for the substitution matrix. Due to the complexity of the likelihood function, the EM procedure requires the use of nonlinear optimization techniques in its M-step.

Closest to our work is that of Conlon and Mortimer (2009), who develop an EM algorithm to account for missing data in a periodic review inventory system under a continuous time model of demand, where for every period they try to uncensor the fraction of consumers not affected by stockouts. They aim to demonstrate how to incorporate data from short-term variations in the choice set to identify substitution patterns, even when the changes to the choice set are not fully observed. A limitation of this work is that the E-step becomes difficult to implement when multiple products are simultaneously stocked out, because it requires estimating an exponential number of parameters (see Conlon and Mortimer 2009, Appendix A.2).

In summary, there has been a growing field of literature on estimating choice behavior and lost sales in the context of retailing for the last decade. This stream of research also includes procedures based on the EM method. Our main contribution to the literature in this regard is a remarkably simple procedure that consists of a repeated sequence of closed-form expressions. The algorithm can be readily implemented in any standard procedural computer language, and it requires only minimal computation time.

3. Model, Estimation, and Algorithm

3.1. Model Description

A set of n substitutable products is sold over T purchase periods, indexed $t = 1, 2, \ldots, T$. No assumption is made about the order or duration of these purchase periods. For example, a purchase period might be a day, and we might have data on purchases over T (not necessarily consecutive) days; or it might be a week, and we might have purchase observations for T weeks. Periods could also be of different lengths, and the indexing need not be in chronological order.

The only data available for each period are actual purchase transactions (i.e., how many units we have sold of each product in each period) and a binary indicator of the availability of each product during the period. (We assume products are either always available or always unavailable in a period; see discussion below.) The number of customers arriving and making purchase choices in each period is not known; equivalently, we do not observe the number of no-purchase outcomes in each period. This is the fundamental incompleteness in the data, and it is a common limitation of transactional sales data in retail settings in which sales transactions and item availability are frequently the only data available.

The full set of products is denoted $\mathcal{N} = \{1, \ldots, n\}$. We denote the number of purchases of product i observed in period t by z_{it} and define $\mathbf{z}_t = (z_{1t}, \ldots, z_{nt})$. We will assume that $z_{it} \geqslant 0$ for all i, t; that is, we do not consider returns. Let $m_t = \sum_{i=1}^{n} z_{it}$ denote the total number of observed purchases in period t. We will further assume without loss of generality that for every product i, there exists at least one period t such that $z_{it} > 0$; else, we can drop product i from the analysis.

We assume the following underlying model generates these purchase data: the number of arrivals in each period (i.e., number of customers who make purchase decisions) is denoted A_t. A_t has a Poisson distribution with mean λ_t (the arrival rate). Let $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_T)$ denote the vector of arrival rates. It could be that some of the n products are not available in certain periods due to temporary stockouts, limited capacity, or controls on availability (e.g., capacity controls from a revenue management system, or deliberate scarcity introduced by the seller). Hence, let $S_t \subset \mathcal{N}$ denote the set of products available for sale in period t. We assume S_t is known for each t and that the products in S_t are available throughout period t. Whenever $i \notin S_t$, for notational convenience we define the number of purchases to be zero, i.e., $z_{it} = 0$.

Customers choose among the alternatives in S_t according to a MNL model, which is assumed to be the same in each period (i.e., preferences are time homogeneous, although this assumption can be relaxed as discussed below). Under the MNL model, the choice probability of a customer is defined based on a preference vector $\mathbf{v} \in \mathcal{R}^n$, $\mathbf{v} > 0$, that indicates the customer "preference weights" or "attractiveness" for the different products.[1] This vector, together with a normalized, no-purchase preference weight $v_0 = 1$, determines a customer's choice probabilities as follows: let $P_j(S, \mathbf{v})$ denote the probability that a customer chooses product $j \in S$ when S is offered and preference weights are given by vector $\mathbf{v}$. Then,

$$P_j(S, \mathbf{v}) = \frac{v_j}{\sum_{i \in S} v_i + 1}. \tag{1}$$

If $j \notin S$, then $P_j(S, \mathbf{v}) = 0$.

We denote the no-purchase probability by $P_0(S, \mathbf{v})$. It accounts for the fact that when set S is offered, a customer may either buy a product from a competitor, or not buy at all (i.e., buys the *outside alternative*):

$$P_0(S, \mathbf{v}) = \frac{1}{\sum_{i \in S} v_i + 1}.$$

The no-purchase option can be treated as a separate product (labeled zero) that is always available. Note that by total probability, $\sum_{j \in S} P_j(S, \mathbf{v}) + P_0(S, \mathbf{v}) = 1$.

The statistical challenge we address is how to estimate the parameters of this model—namely, the preference vector $\mathbf{v}$ and the arrival rates $\boldsymbol{\lambda}$—from the purchase data $\mathbf{z}_t$, $t = 1, 2, \ldots, T$.

3.2. The Incomplete Data Likelihood Function

One can attempt to solve directly this estimation problem using maximum likelihood estimation (MLE). The incomplete data likelihood function can be expressed as follows:

$$\begin{aligned}
&\mathcal{L}_I(\mathbf{v}, \boldsymbol{\lambda}) \\
&= \prod_{t=1}^{T} \Bigg(\mathbb{P}(m_t \text{ customers buy in period } t \mid \mathbf{v}, \boldsymbol{\lambda}) \frac{m_t!}{z_{1t}! z_{2t}! \cdots z_{nt}!} \\
&\qquad \cdot \prod_{j \in S_t} \left[\frac{P_j(S_t, \mathbf{v})}{\sum_{i \in S_t} P_i(S_t, \mathbf{v})} \right]^{z_{jt}} \Bigg),
\end{aligned} \tag{2}$$

where the probabilities in the inner product are the conditional probabilities of purchasing product j given that a customer purchases something. The number of customers that purchase in period t, m_t, is a realization of a Poisson random variable with mean $\lambda_t \sum_{i \in S_t} P_i(S_t, \mathbf{v})$, viz

$$\begin{aligned}
&\mathbb{P}(m_t \text{ customers buy in period } t \mid \mathbf{v}, \boldsymbol{\lambda}) \\
&\quad = \frac{[\lambda_t \sum_{i \in S_t} P_i(S_t, \mathbf{v})]^{m_t} e^{-\lambda_t \sum_{i \in S_t} P_i(S_t, \mathbf{v})}}{m_t!}.
\end{aligned} \tag{3}$$

One could take the log of (2) and attempt to maximize this log-likelihood function with respect to $\mathbf{v}$ and $\boldsymbol{\lambda}$. However, it is clear that this is a complex likelihood function without much structure, so maximizing it (or its logarithm) directly is not an appealing approach. Indeed, our attempts in this regard were not promising as reported later in §5.

3.3. Multiple Optima in the MLE and Market Potential

A further complication is that the likelihood function (2) has a continuum of maxima. To see this, let $(\mathbf{v}^*, \boldsymbol{\lambda}^*)$ denote a maximizer of (2). Let $\alpha > 0$ be any real number and define a new preference vector $\mathbf{v}^0 = \alpha \mathbf{v}^*$. Define new arrival rates

$$\lambda_t^0 = \frac{\alpha \sum_{i \in S_t} v_i^* + 1}{\alpha (\sum_{i \in S_t} v_i^* + 1)} \lambda_t^*.$$

Then, it is not hard to see from (1) that

$$\lambda_t^* P_j(S_t, \mathbf{v}^*) = \lambda_t^0 P_j(S_t, \mathbf{v}^0),$$

for all j and t. Because this product of the arrival rate and purchase probability is unchanged, by inspection of (2) and (3), the solution $(\mathbf{v}^0, \boldsymbol{\lambda}^0)$ has the same likelihood and therefore is also a maximum. Because this holds true for any $\alpha > 0$, there is a continuum of maxima. Of course, this observation holds more generally: for any pair of values $(\mathbf{v}, \boldsymbol{\lambda})$, there is a continuum of values $\alpha(\mathbf{v}, \boldsymbol{\lambda})$, $\alpha > 0$, such that $\mathcal{L}_I(\mathbf{v}, \boldsymbol{\lambda}) = \mathcal{L}_I(\alpha\mathbf{v}, \alpha\boldsymbol{\lambda})$.

One can resolve this multiplicity of optimal solutions by imposing an additional constraint on the parameter values related to market share. Specifically, suppose we have an exogenous estimate of the preference weight of the outside alternative relative to the total set of offerings. Let's call it r, so that

$$r := \frac{1}{\sum_{j=1}^{n} v_j}. \quad (4)$$

Then fixing the value of r resolves the degree of freedom in the multiple maxima. Still, this leaves the need to solve a complicated optimization problem. In §3.5 we look at a simpler and more efficient approach based on viewing the problem in terms of *primary* demand. Before doing so, however, we briefly discuss the demand model itself.

3.4. Discussion of the Demand Model

Our model uses the well-studied MNL for modeling customer choice behavior in a homogeneous market (i.e., customer preferences are described by a single set of parameters $\mathbf{v}$). As mentioned, a convenient property of the MNL is that the likelihood of purchase can be readily recalculated if the availability of the products changes. However, the MNL has significant restrictions in terms of modeling choice behavior, most notably the property of independence from irrelevant alternatives (IIA). Briefly, this property says that the ratio of purchase probabilities for two available alternatives is constant regardless of the choice set containing them. Other choice models are more flexible in modeling substitution patterns (e.g., see Train 2003, Chapter 4). Among them, the nested logit (NL) model has been widely used in the marketing literature. While less restrictive, the NL requires more parameters and therefore a higher volume of data to generate good estimates.

Despite the limitation of the IIA property, MNL models are widely used. Starting with Guadagni and Little (1983), marketing researchers have found that the MNL model works well when estimating demand for a category of substitutable products (in Guadagni and Little's study, regular ground coffee of different brands). Recent experience in the airline industry also provides good support for using the MNL model.[2] According to the experience of one of the authors, there are two major considerations in real airline implementations: (i) the range of fare types included, and (ii) the flight departure time proximity. Regarding (i), in cases where airlines are dealing with dramatically different fare products, then it is often better to split the estimation process using two entirely separate data pools. Consider the following real-world example. An international airline uses the first four booking classes in their nested fare hierarchy for international point-of-sales fares that have traditional restrictions (i.e., advance purchase, minimum stay length, etc.); these are the highest-valued fare types. The next eight booking classes are used for domestic travel with restriction-free fares. Because there is little (or no) interaction between the international and domestic points-of-sales, the airline applies the MNL model to two different data pools: one for international sales and the other for domestic sales. Separate choice models are fit to the two different pools. Regarding (ii), it would be somewhat unrealistic to assume that first-choice demand for a closed 7:00 A.M. departure would be recaptured onto a same-day, open 7:00 P.M. departure in accordance with the IIA principle. Hence, it makes sense to restrict the consideration set to departure times that are more similar. Clearly, some customers will refuse to consider the alternative flight if the difference in departure times is large. Some recently developed revenue management systems with which the authors are familiar still use the MNL for such flight sets, but they implement a correction heuristic to overcome the IIA limitation.

Another important aspect of our model is the interpretation of the outside alternative, and the resulting interpretation of the arrival rates $\boldsymbol{\lambda}$. For instance, if the outside alternative is assumed to be the (best) competing product, then

$$s = 1/(1+r) = \frac{\sum_{j=1}^{n} v_j}{\sum_{j=1}^{n} v_j + 1}$$

defines the retailer's market share, including the retailer and its competitor(s). Alternatively, if the outside alternative is considered to consist of both the competitor's best product and a no-purchase option, then s gives the retailer's market potential, and $\boldsymbol{\lambda}$ is then interpreted as the total market size (number of customers choosing). This later interpretation is found in marketing and empirical industrial organizations applications (e.g., see Berry et al. 1995 for an empirical study of the U.S. automobile industry and Nevo 2001 for an empirical study of the ready-to-eat cereal industry). Henceforth, given a value s (retailer's market share or potential), we set the attractiveness of the outside alternative as $r = (1-s)/s$, which is equivalent to (4). Low values of r imply high market share or potential.

Note that we work with store-level data (as opposed to household panel data). Chintagunta and Dubé (2005) discuss the advantages of using store-level data to compute the mean utility associated with products (in our MNL case, $\log v_i$ is the mode of the random utility of product i). We also assume that for every product j, there is a period t for which $z_{jt} > 0$ (otherwise, that product can be dropped from the analysis). In this regard, our model can accommodate assortments with slow-moving items for which $z_{jt}=0$ for several (but not all) periods. It's worth noting that for retail settings, having zero sales in many consecutive periods could be a symptom of inventory record error. DeHoratius and Raman (2008) found that 65% of 370,000 inventory records of a large public U.S. retailer were inaccurate, and that the magnitude of the inaccuracies was significant (of around 35% of the inventory level on the shelf per SKU). A possible misleading situation is that the IT system records a SKU as being in stock even though there are no units on the shelf, and hence no sales will be observed despite the fact that the product is tagged as "available."

Furthermore, if a period t has no sales for any of the products, then that period can be dropped from the analysis. Note that for that period, $m_t = 0$ in Equations (2) and (3), and therefore $\lambda_t^* = 0$. Intuitively, this is because our model assumes that the market participation s is replicated in every single period, and hence the most likely arrival rate to produce no sale in a period is an arrival rate of zero.

Regarding the information on product availability, as mentioned above we assume that a product is either fully available or not available throughout a given period t. Hence, the time partitioning should be fine enough to capture short-term changes in the product availability over time.[3] However, in contrast to other approaches (e.g., Musalem et al. 2010), we do not require information on inventory levels; all we require is a binary indicator describing each item's availability.

Finally, note that our model assumes homogeneous preferences across the whole selling horizon but a non-homogeneous Poisson arrival process of consumers. The assumption of homogeneous preferences can be relaxed by splitting the data into intervals where a different choice model is assumed to apply over each period. The resulting modification is straightforward, so we do not elaborate on this extension. The estimates $\hat{\boldsymbol{\lambda}}$ can be used to build a forecast of the volume of demand to come by applying standard time series analysis to project the values forward in time.

3.5. Log-Likelihood Based on Primary Demand

By primary (or first-choice) demand for product j, we mean the demand that would have occurred for product j if all n alternatives were available. The (random) number of purchases, Z_{jt}, of product j in period t might be greater than the primary demand because it could include purchases from customers whose first choice was not available and who bought product j as a substitute (i.e., Z_{jt} includes demand that is spilled from other unavailable products and recaptured by product j). More precisely, the purchase quantity Z_{jt} can be split into two components: the primary demand, X_{jt}, which is the number of customers in period t that have product j as their first choice; and Y_{jt}, the *substitute demand*, which is the number of customers in period t that decide to buy product j as a substitute because their first choice is unavailable. Thus,

$$Z_{jt} = X_{jt} + Y_{jt}. \tag{5}$$

Clearly, $X_{jt} \geqslant 0$, but the equation remains true when $Z_{jt}=0$ and $Y_{jt} \leqslant 0$, as explained below. Our focus is on estimating the primary demand X_{jt}. While this decomposition seems to introduce more complexity in the estimation problem, in fact it leads to a considerably simpler estimation algorithm.

3.5.1. Basic Identities. Based on the purchase observations $\mathbf{z}_t$, we have that $\mathrm{E}[Z_{jt} \mid \mathbf{z}_t] = z_{jt}$. Let $\hat{X}_{jt} = \mathrm{E}[X_{jt} \mid \mathbf{z}_t]$ and $\hat{Y}_{jt} = \mathrm{E}[Y_{jt} \mid \mathbf{z}_t]$ denote, respectively, the conditional expectation of the primary and substitute demand given the purchase observations. We seek to determine these two quantities. In what follows, assume that the preference vector $\mathbf{v}$ is given.

Case 1. Consider first products that are unavailable in period t, that is $j \notin (S_t \cup \{0\})$. For these items, we have no observation z_{jt}, and for completeness we set $z_{jt}=0$. To determine $\hat{X}_{jt}$ for these items, note that

$$\mathrm{E}[X_{jt} \mid \mathbf{z}_t] = \frac{v_j}{\sum_{i=1}^n v_i + 1} \mathrm{E}[A_t \mid \mathbf{z}_t],$$

and

$$\sum_{h\in S_t} \mathrm{E}[Z_{ht} \mid \mathbf{z}_t] = \frac{\sum_{h\in S_t} v_h}{\sum_{h\in S_t} v_h + 1} \mathrm{E}[A_t \mid \mathbf{z}_t].$$

Combining these expressions to eliminate $\mathrm{E}[A_t \mid \mathbf{z}_t]$ yields

$$\mathrm{E}[X_{jt} \mid \mathbf{z}_t] = \frac{v_j}{\sum_{i=1}^n v_i + 1} \frac{\sum_{h\in S_t} v_h + 1}{\sum_{h\in S_t} v_h} \sum_{h\in S_t} \mathrm{E}[Z_{ht} \mid \mathbf{z}_t],$$

or equivalently,

$$\hat{X}_{jt} = \frac{v_j}{\sum_{i=1}^n v_i + 1} \frac{\sum_{h\in S_t} v_h + 1}{\sum_{h\in S_t} v_h} \sum_{h\in S_t} z_{ht}, \quad j \notin (S_t \cup \{0\}). \tag{6}$$

For this case, in view of (5), we have $\hat{Y}_{jt} = -\hat{X}_{jt}$, meaning that customers are "substituting out of" product j because j is not available.

Case 2. Next, consider the available products $j \in S_t$. For each such product, we have z_{jt} observed transactions, which according to (5) can be split into

$$z_{jt} = \hat{X}_{jt} + \hat{Y}_{jt}, \quad j \in S_t.$$

Note that

$$\begin{aligned}
&\mathbb{P}\{\text{product } j \text{ is a first choice} \mid \text{purchase } j\} \\
&\quad = \frac{\mathbb{P}\{\text{product } j \text{ is a first choice}\}}{\mathbb{P}\{\text{purchase } j\}} \\
&\quad = \frac{v_j}{\sum_{i=1}^n v_i + 1} \Big/ \frac{v_j}{\sum_{h\in S_t} v_h + 1} \\
&\quad = \frac{\sum_{h\in S_t} v_h + 1}{\sum_{i=1}^n v_i + 1}.
\end{aligned}$$

Therefore, because $\hat{X}_{jt} = z_{jt}\mathbb{P}\{\text{product } j \text{ is a first choice} \mid \text{purchase } j\}$, we have

$$\hat{X}_{jt} = \frac{\sum_{h\in S_t} v_h + 1}{\sum_{i=1}^n v_i + 1} z_{jt}, \quad \text{and} \quad \hat{Y}_{jt} = \frac{\sum_{h\notin (S_t\cup\{0\})} v_h}{\sum_{i=1}^n v_i + 1} z_{jt}. \tag{7}$$

Case 3. Last, for the no-purchase option (i.e., $j = 0$), we are also interested in estimating its primary demand in period t conditional on the transaction data, i.e., $\hat{X}_{0t} = \mathrm{E}[X_{0t} \mid \mathbf{z}_t]$. Recall that A_t is the total (random) number of arrivals in period t, including the customers that do not purchase. Again, we do not observe A_t directly but note that

$$\mathrm{E}[X_{0t} \mid \mathbf{z}_t] = \frac{1}{\sum_{i=1}^n v_i + 1} \mathrm{E}[A_t \mid \mathbf{z}_t]. \tag{8}$$

In addition, the following identity must hold:

$$A_t = X_{0t} + \sum_{i=1}^n X_{it}.$$

Conditioning on the observed purchases we have that

$$\mathrm{E}[A_t \mid \mathbf{z}_t] = \hat{X}_{0t} + \sum_{i=1}^n \hat{X}_{it}. \tag{9}$$

Substituting (9) into (8), we obtain

$$\hat{X}_{0t} = \frac{1}{\sum_{i=1}^n v_i} \sum_{i=1}^n \hat{X}_{it}. \tag{10}$$

Interestingly, we can also get the lost sales in period t, given by the conditional expectation of the substitute demand for the no-purchase option, $\hat{Y}_{0t} = \mathrm{E}[Y_{0t} \mid \mathbf{z}_t]$:

$$\hat{Y}_{0t} = \frac{1}{\sum_{i\in S_t} v_i + 1} \sum_{h\notin (S_t\cup\{0\})} \hat{X}_{ht}.$$

Next, define N_j, $j = 0, \ldots, n$, as the total primary demand for product j over all periods (including the no-purchase option $j = 0$). Thus, $N_j = \sum_{t=1}^T X_{jt}$, giving an estimate

$$\hat{N}_j := \sum_{t=1}^T \hat{X}_{jt}, \tag{11}$$

where, consistent with our other notation, $\hat{N}_j = \mathrm{E}[N_j \mid \mathbf{z}_1, \ldots, \mathbf{z}_T]$, which is positive because $\hat{X}_{jt} \geqslant 0$, for all j and t, and for at least one period t, $\hat{X}_{jt} > 0$.[4]

3.5.2. Overview of Our Approach. The key idea behind our approach is to view the problem of estimating $\mathbf{v}$ and $\boldsymbol{\lambda}$ as an estimation problem with incomplete observations of the primary demand X_{jt}, $j = 0, 1, \ldots, n$, $t = 1, \ldots, T$. Indeed, suppose we had complete observations of the primary demand. Then the log-likelihood function would be simple, namely

$$L(\mathbf{v}) = \sum_{j=1}^n N_j \ln\left(\frac{v_j}{\sum_{i=1}^n v_i + 1}\right) + N_0 \ln\left(\frac{1}{\sum_{i=1}^n v_i + 1}\right),$$

where N_j is the total number of customers selecting product j as their first choice (or selecting not to purchase, $j = 0$, as their first choice). We show below this function has a closed-form maximum. However, because we do not observe N_j, $j = 0, 1, \ldots, n$, directly, we use the EM method of Dempster et al. (1977) to estimate the model. This approach drastically simplifies the computational problem relative to maximizing (2). It also has the advantage of eliminating $\boldsymbol{\lambda}$ from the estimation problem and reducing it to a problem in $\mathbf{v}$ only. (An estimate of $\boldsymbol{\lambda}$ can be trivially recovered after the algorithm runs, as discussed below.)

The EM method is an iterative procedure that consists of two steps per iteration: an expectation (E) step and a maximization (M) step. Starting from arbitrary initial estimates of the parameters, it computes the conditional expected value of the log-likelihood function with respect to these estimates (the E-step) and then maximizes the resulting expected log-likelihood function to generate new estimates (the M-step). The procedure is repeated until convergence. While technical convergence problems can arise, in practice the EM method is a robust and efficient way to compute maximum likelihood estimates for incomplete data problems.

In our case, the method works by starting with estimates $\hat{\mathbf{v}} > 0$ (the E-step). These estimates for the preference weights are used to compute estimates for the total primary demand values $\hat{N}_0, \hat{N}_1, \ldots, \hat{N}_n$, by using the formulas in (6), (7), and (10), and then substituting the values of $\hat{X}_{jt}$ in (11). In the M-step, given estimates $\hat{\mathbf{v}}$ (and therefore, given estimates for $\hat{N}_0, \hat{N}_1, \ldots, \hat{N}_n$), we then maximize the conditional expected value of the log-likelihood function with respect to $\mathbf{v}$:

$$\mathrm{E}[L(\mathbf{v}) \mid \hat{\mathbf{v}}] = \sum_{j=1}^n \hat{N}_j \ln\left(\frac{v_j}{\sum_{i=1}^n v_i + 1}\right) + \hat{N}_0 \ln\left(\frac{1}{\sum_{i=1}^n v_i + 1}\right). \tag{12}$$

Just as in the likelihood function (2), there is a degree of freedom in our revised estimation formulation. Indeed, consider the first iteration with arbitrary initial values for the estimates $\hat{\mathbf{v}}$, yielding estimates $\hat{N}_j$, $j = 0, 1, \ldots, n$. From (10), r defined in (4) must satisfy $\hat{N}_0 = r \sum_{j=1}^n \hat{N}_j$. As above, r measures the magnitude of outside alternative demand relative to the alternatives in $\mathcal{N}$. We will prove later, in Proposition 1, that this relationship is preserved across different iterations of the EM method. So the initial guess for $\hat{\mathbf{v}}$ implies an estimate of r.

Expanding (12), the conditional expected, complete data log-likelihood function is

$$\begin{aligned}\mathscr{L}(\mathbf{v}) &:= \mathrm{E}[L(v_1, \ldots, v_n) \mid \hat{v}_1, \ldots, \hat{v}_n] \\ &= \sum_{j=1}^n \hat{N}_j \left\{ \ln\left(\frac{v_j}{\sum_{i=1}^n v_i + 1}\right) + r \ln\left(\frac{1}{\sum_{i=1}^n v_i + 1}\right)\right\}\end{aligned}$$

$$= \sum_{j=1}^{n} \hat{N}_j \ln\left(\frac{v_j}{\sum_{i=1}^{n} v_i + 1}\right) + r \ln\left(\frac{1}{\sum_{i=1}^{n} v_i + 1}\right) \sum_{j=1}^{n} \hat{N}_j. \quad (13)$$

This expected log-likelihood function is then maximized to generate new estimates $\hat{v}_j^*$, $j = 1, \ldots, n$. We show below this a simple maximization problem, with closed-form solution

$$v_j^* = \frac{\hat{N}_j}{r \sum_{i=1}^{n} \hat{N}_i}, \quad j = 1, \ldots, n. \quad (14)$$

In the E-step of the next iteration, the EM method uses these maximizers to compute updated estimates $\hat{X}_{jt}$ in (6), (7), and (10), leading to updated values $\hat{N}_j$. These two steps are repeated until convergence.

Note that both the expectation and maximization steps in this procedure involve only simple, closed-form calculations. Also note that the whole EM procedure can be described only in terms of the preference weight estimates $\hat{v}_j$, $j = 1, \ldots, n$. The optimal first-choice estimates $\hat{X}_{jt}$ are returned by applying (6), (7), and (10) using the estimates $\hat{v}_j$ of the final iteration. Estimates of $\boldsymbol{\lambda}$ can also be recovered from (9) by simply noting that

$$\hat{\lambda}_t \equiv \mathrm{E}[A_t \mid \mathbf{z}_t] = \hat{X}_{0t} + \sum_{i=1}^{n} \hat{X}_{it}. \quad (15)$$

That is, the arrival rate is simply the sum of the primary demands of all n products plus the primary demand of the no-purchase alternative. Intuitively, this is why viewing the problem in terms of primary demand eliminates the arrival rate from the estimation problem; the arrival rate is simply the sum of primary demands.

3.5.3. Summary of the EM Algorithm. We next summarize the EM algorithm for estimating primary demand using pseudocode.

EM Algorithm for Estimating Primary Demand

[*Initialization*]: Given a market participation s, let $r := (1-s)/s$. For all product j and periods t, set $X_{jt} := z_{jt}$, with $X_{jt} := 0$ if $j \notin S_t$. Then, initialize variables $N_0, N_1, \ldots, N_n$, as follows:

$$N_j := \sum_{t=1}^{T} X_{jt}, \quad j = 1, \ldots, n, \quad N_0 := r \sum_{j=1}^{n} N_j,$$

$$X_{0t} := N_0/T, \quad \text{and} \quad v_j := N_j/N_0, \quad j = 1, \ldots, n.$$

Repeat

For $t := 1, \ldots, T$ do

For $j := 1, \ldots, n$ do

If $j \notin S_t$, then set

$$X_{jt} := \frac{v_j}{\sum_{i=1}^{n} v_i + 1} \frac{\sum_{h \in S_t} v_h + 1}{\sum_{h \in S_t} v_h} \sum_{h \in S_t} z_{ht}, \quad \text{and} \quad Y_{jt} = -X_{jt},$$

else (i.e., $j \in S_t$), then set

$$Y_{jt} := \frac{\sum_{h \notin (S_t \cup \{0\})} v_h}{\sum_{i=1}^{n} v_i + 1} z_{jt}, \quad \text{and} \quad X_{jt} := z_{jt} - Y_{jt}.$$

EndIf

EndFor

Set

$$X_{0t} := \frac{1}{\sum_{i=1}^{n} v_i} \sum_{i=1}^{n} X_{it}, \text{ and } Y_{0t} := \frac{1}{\sum_{i \in S_t} v_i + 1} \sum_{h \notin (S_t \cup \{0\})} X_{ht}.$$

EndFor

Set $N_0 := \sum_{t=1}^{T} X_{0t}$.

For $j := 1, \ldots, n$ do

Set $N_j := \sum_{t=1}^{T} X_{jt}$.

Set $v_j := N_j/N_0$.

EndFor

until Stopping criteria are met.

A few remarks on implementation: The initialization of X_{jt}, $j = 1, \ldots, n$, is arbitrary; we merely need starting values different from zero if $j \in S_t$. The stopping criteria can be based on various measures of numerical convergence, e.g., that the difference between all values X_{jt} from two consecutive iterations of the algorithm is less than a small constant ϵ, or on a maximum number of iterations. In all our experiments we observed very quick convergence, so it would appear that the precise stopping criteria are not critical.

4. Properties of the EM Algorithm

We start by noting some properties of the algorithm with respect to the retailer's market-participation-related parameter r (recall that $s = 1/(1+r)$). First, note that the function $\mathcal{L}$ in (13) is linearly decreasing as a function of r, for all $r > 0$. Second, as claimed above, the value r remains constant throughout the execution of the algorithm.

Proposition 1. *The relationship* $\hat{N}_0 = r \sum_{j=1}^{n} \hat{N}_j$, *is preserved across iterations of the EM algorithm, starting from the initial value of* r.

Proof. In the E-step of an iteration, after we compute the values $\hat{X}_{it}$, we use formula (10) with the v_js replaced by the optimal values obtained in the M-step of the previous iteration, i.e.,

$$\hat{X}_{0t} = \frac{1}{\sum_{i=1}^{n} (\hat{N}_i'/r \sum_{h=1}^{n} \hat{N}_h')} \sum_{i=1}^{n} \hat{X}_{it} = r \sum_{i=1}^{n} \hat{X}_{it},$$

where $\hat{N}_i'$ stand for the volume estimates from the previous iteration. The new no-purchase estimate is

$$\hat{N}_0 = \sum_{t=1}^{T} \hat{X}_{0t} = \sum_{t=1}^{T} r \sum_{i=1}^{n} \hat{X}_{it} = r \sum_{i=1}^{n} \sum_{t=1}^{T} \hat{X}_{it} = r \sum_{i=1}^{n} \hat{N}_i,$$

and hence the relationship $\hat{N}_0 = r \sum_{j=1}^{n} \hat{N}_j$, is preserved. □

Our next result proves that the complete data log-likelihood function $\mathcal{L}(v_1, \ldots, v_n)$ is indeed unimodal.

Proposition 2. *The function* $\mathscr{L}(v_1, \ldots, v_n)$, *with* $\mathbf{v} > 0$, *and* $\hat{N}_j > 0$, $\forall j$, *is unimodal, with unique maximizer* $v_j^* = \hat{N}_j / r \sum_{i=1}^n \hat{N}_i$, $j = 1, \ldots, n$.

Proof. Taking partial derivatives of function (13), we get

$$\frac{\partial}{\partial v_j} \mathscr{L}(v_1, \ldots, v_n) = \frac{\hat{N}_j}{v_j} - \frac{(1+r)\sum_{i=1}^n \hat{N}_i}{\sum_{i=1}^n v_i + 1}, \quad j = 1, \ldots, n.$$

Setting these n equations equal to zero leads to a linear system with unique solution

$$v_j^* = \frac{\hat{N}_j}{r \sum_{i=1}^n \hat{N}_i}, \quad j = 1, \ldots, n. \tag{16}$$

The second cross-partial derivatives are

$$\frac{\partial^2}{\partial^2 v_j} \mathscr{L}(v_1, \ldots, v_n) = -\frac{\hat{N}_j}{v_j^2} + \gamma(v_1, \ldots, v_n),$$

where

$$\gamma(v_1, \ldots, v_n) = \frac{(1+r)\sum_{i=1}^n \hat{N}_i}{(\sum_{i=1}^n v_i + 1)^2},$$

and

$$\frac{\partial^2}{\partial v_j \partial v_i} \mathscr{L}(v_1, \ldots, v_n) = \gamma(v_1, \ldots, v_n), \quad j \neq i.$$

Let H be the Hessian of $\mathscr{L}(v_1, \ldots, v_n)$. To check that our critical point (16) is a local maximum, we compute for $\mathbf{x} \in \mathscr{R}^n$, $\mathbf{x} \neq 0$,

$$\mathbf{x}^\top H(v_1, \ldots, v_n)\mathbf{x} = \frac{(1+r)(\sum_{i=1}^n \hat{N}_i)(\sum_{i=1}^n x_i)^2}{(\sum_{i=1}^n v_i + 1)^2} - \sum_{i=1}^n \hat{N}_i \frac{x_i^2}{v_i^2}. \tag{17}$$

The second-order sufficient conditions are $\mathbf{x}^\top H(v_1^*, \ldots, v_n^*)\mathbf{x} < 0$, for all $\mathbf{x} \neq 0$. Plugging in the expressions in (16), we get

$$\mathbf{x}^\top H(v_1^*, \ldots, v_n^*)\mathbf{x} = r^2 \left(\sum_{i=1}^n \hat{N}_i\right)\left(\frac{(\sum_{i=1}^n x_i)^2}{1+r} - \left(\sum_{i=1}^n \hat{N}_i\right)\sum_{i=1}^n \frac{x_i^2}{\hat{N}_i}\right).$$

Note that because $r > 0$, and $\hat{N}_j > 0$, $\forall j$, it is enough to check that

$$\left(\sum_{i=1}^n x_i\right)^2 - \left(\sum_{i=1}^n \hat{N}_i\right)\sum_{i=1}^n \frac{x_i^2}{\hat{N}_i} \leqslant 0, \quad \forall \mathbf{x} \neq 0. \tag{18}$$

By the Cauchy-Schwartz inequality, i.e., $|\mathbf{y}^\top \mathbf{z}|^2 \leqslant \|\mathbf{y}\|^2 \|\mathbf{z}\|^2$, defining $y_i = x_i / \sqrt{\hat{N}_i}$ and $z_i = \sqrt{\hat{N}_i}$, we get

$$\left(\sum_{i=1}^n x_i\right)^2 = \left(\sum_{i=1}^n \frac{x_i}{\sqrt{\hat{N}_i}} \times \sqrt{\hat{N}_i}\right)^2 \leqslant \left(\sqrt{\sum_{i=1}^n \frac{x_i^2}{\hat{N}_i}}\right)^2 \left(\sqrt{\sum_{i=1}^n \hat{N}_i}\right)^2 = \left(\sum_{i=1}^n \frac{x_i^2}{\hat{N}_i}\right)\left(\sum_{i=1}^n \hat{N}_i\right),$$

and therefore inequality (18) holds.

Proceeding from first principles, we have a unique critical point for $\mathscr{L}(v_1, \ldots, v_n)$, which is a local maximum. The only other potential maxima can occur at a boundary point. But close to the boundary of the domain the function is unbounded from below; that is,

$$\lim_{v_j \downarrow 0} \mathscr{L}(v_1, \ldots, v_n) = -\infty, \quad j = 1, \ldots, n.$$

Hence, the function is unimodal. □

A few comments are in order. First, due to the definition of v_j^* and because $\sum_{t=1}^T z_{jt} > 0$, then $\hat{N}_j > 0$ for every iteration of the EM method. Second, observe that Equation (17) shows that the function $\mathscr{L}(v_1, \ldots, v_n)$ is not jointly concave in general, because there could exist a combination of values $\hat{N}_1, \ldots, \hat{N}_n$, and the vector $(v_1, \ldots, v_n)$ such that for some $\mathbf{x}$, $\mathbf{x}^\top H(v_1, \ldots, v_n)\mathbf{x} > 0$. For example, if we take $n = 2$, $\mathbf{v} = (1.5, 1.2)$, $\hat{N}_1 = 50$, $\hat{N}_2 = 3$, and $\mathbf{x} = (0.01, 1)$, then $r = 1/(v_1 + v_2) = 0.37$, and $\mathbf{x}^\top H(v_1, v_2)\mathbf{x} = 3.33$. In this regard, this is different from the usual linear-in-parameter MNL formulation, for which the complete data log-likelihood function is jointly concave in most cases (e.g., see Talluri and van Ryzin 2004, §5). However, from a computational point of view, what matters is that it is unimodal, and even better, the optimal solution has a closed form, which leads to our third observation: our procedure is indeed an EM algorithm, as opposed to the so-called generalized EM algorithm (GEM). In the case of GEM, the M-step requires only that we generate an improved set of estimates over the current ones (i.e., it requires to find a vector $\bar{\mathbf{v}}$ such that $\mathrm{E}[L(\bar{\mathbf{v}}) \mid \hat{\mathbf{v}}]] \geqslant \mathrm{E}[L(\hat{\mathbf{v}}) \mid \hat{\mathbf{v}}]$), and the conditions for convergence are more stringent (e.g., see McLachlan and Krishnan 1996, Chapter 3, for further discussion).

Because our EM method satisfies a mild regularity condition, we have the following convergence result due to Wu (1983).[5]

Theorem 1. *The conditional expected value*

$$\mathrm{E}[L(v_1, \ldots, v_n) \mid \hat{v}_1, \ldots, \hat{v}_n]$$

in (13) *is continuous both in* $\mathbf{v} > 0$ *and* $\hat{\mathbf{v}} > 0$, *and hence all the limit points of any instance* $\{\hat{\mathbf{v}}^{(k)}, \hat{\boldsymbol{\lambda}}^{(k)}, k = 1, 2, \ldots\}$ *of the EM algorithm are stationary points of the corresponding incomplete-data log-likelihood function* $\mathscr{L}_I(\mathbf{v}, \boldsymbol{\lambda})$, *and* $\mathscr{L}_I(\hat{\mathbf{v}}^{(k)}, \hat{\boldsymbol{\lambda}}^{(k)})$ *converges monotonically to a value* $\mathscr{L}_I(\mathbf{v}^*, \boldsymbol{\lambda}^*)$, *for some stationary point* $(\mathbf{v}^*, \boldsymbol{\lambda}^*)$.

Proof. The result simply follows from the fact that $\hat{N}_j = \sum_{t=1}^T \hat{X}_{jt}$, $j = 0, 1, \ldots, n$, and $\hat{X}_{jt}$ are continuous in $\hat{\mathbf{v}}$ according to Equations (6), (7), and (10). Clearly, $\mathscr{L}$ is also continuous in $\mathbf{v}$. In addition, recall that the estimates $\hat{\mathbf{v}}$ imply a vector $\hat{\boldsymbol{\lambda}}$ once we fix a market participation r (through Equation (15)), and therefore the EM algorithm, given the unique maximizer found in the M-step as proved in Proposition 2, indeed generates an implied sequence $\{\hat{\mathbf{v}}^{(k)}, \hat{\boldsymbol{\lambda}}^{(k)}, k = 1, 2, \ldots\}$. □

As pointed out by (Wu 1983, §2.2), the convergence of $\{\mathscr{L}_I(\hat{\mathbf{v}}^{(k)}, \hat{\boldsymbol{\lambda}}^{(k)})$ to $\mathscr{L}_I(\mathbf{v}^*, \boldsymbol{\lambda}^*)$, for some stationary point $(\mathbf{v}^*, \boldsymbol{\lambda}^*)$, does not automatically imply the convergence of $\{(\hat{\mathbf{v}}^{(k)}, \hat{\boldsymbol{\lambda}}^{(k)})\}$ to a point $(\mathbf{v}^*, \boldsymbol{\lambda}^*)$. Nevertheless, the convergence of the sequence of points $\{(\hat{\mathbf{v}}^{(k)}, \hat{\boldsymbol{\lambda}}^{(k)})\}$ can be checked numerically as part of the EM procedure. In our experiments reported in §5, we consistently observed that the sequence of estimates converged. Another caveat is the fact that the stationary point of $\mathscr{L}_I(\mathbf{v}, \boldsymbol{\lambda})$ is not guaranteed to be a global maximum, but this drawback is also shared by any standard nonlinear optimization method working directly on the original incomplete-data log-likelihood function.

5. Numerical Examples

We next report on two sets of numerical examples. The first set is based on simulated data, which are used to get a sense of how well the procedure identifies a known demand system and how much data are necessary to get good estimates. Then, we report results on two real-world data sets, one for airlines and another for retail. In all the examples, we set a stopping criterion based on the difference between the matrices $\hat{X}$ from two consecutive iterations of the EM method, halting the procedure as soon as the absolute value of all the elements of the difference matrix was smaller than 0.001. The algorithm was implemented using the MATLAB[6] procedural language, in which the method detailed in §3.5.3 is straightforward to code.

5.1. Examples Based on Simulated Data

Our first example is small and illustrates the behavior of the procedure on a known demand system. We provide the original generated data (observed purchases) and the final data (primary and substitute demands), as well as comparative results with two benchmark procedures. Next, we look at the effect of input data volume on the accuracy of the estimates. Finally, we run an exhaustive set of comparisons between our procedure and three benchmarks to get a broader sense of the relative performance of our method.

Table 2. Output parameters for preliminary example.

Parameter	True value	Est. value	Bias (%)	ASE	t-stat
$\hat{v}_1$	1.00	0.948	−5.25	0.092	10.32
$\hat{v}_2$	0.70	0.759	8.49	0.078	9.72
$\hat{v}_3$	0.40	0.371	−7.35	0.048	7.69
$\hat{v}_4$	0.20	0.221	10.25	0.035	6.28
$\hat{v}_5$	0.05	0.052	3.80	0.016	3.28

5.1.1. Preliminary Estimation Case. Given a known underlying MNL choice model (i.e., values for the preference weights $\mathbf{v}$) and assuming that arrivals follow a homogeneous Poisson process with rate $\lambda = 50$, we simulated purchases for $n = 5$ different products. Initially, we considered a selling horizon of $T = 15$ periods and preference weights $\mathbf{v} = (1, 0.7, 0.4, 0.2, 0.05)$ (recall that the weight of the no-purchase alternative is $v_0 = 1$). Note w.l.o.g. we index products in decreasing order of preference. These preference values give a market potential $s = \sum_{j=1}^n v_j / (\sum_{j=1}^n v_j + 1) = 70\%$.

Table 1 describes the simulated data, showing the randomly generated purchases for each of the five products for each period and the total number of no-purchases and arrivals. Here period 1 represents the end of the selling horizon. A label "NA" in position (j, t) means that product j is not available in period t. The unavailability was exogenously set prior to simulating the purchase data.

For the estimation procedure, the initial values of $\hat{v}_j$ are computed following the suggestion in §3.5.3, i.e.,

$$\hat{v}_j = \frac{\sum_{t=1}^T z_{jt}}{r \sum_{t=1}^T \sum_{i=1}^n z_{it}}, \quad j = 1, \ldots, n, \tag{19}$$

with $r = 0.4286$ (equivalently with a market share/potential of $s = 0.70$); we also assume perfect knowledge of this market potential. The output is shown in Table 2.

The second column includes the true preference weight values for reference. The third column reports the estimates computed by the EM method. The fourth column reports the percentage bias between the estimated and true

Table 1. Purchases and no-purchases for the preliminary example.

Observable data: Purchases and nonavailability (NA)																
	Periods															
Product	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	Total
1	10	15	11	14	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	50
2	11	6	11	8	20	16	NA	NA	NA	NA	NA	NA	NA	NA	NA	72
3	5	6	1	11	4	5	14	7	11	NA	NA	NA	NA	NA	NA	64
4	4	4	4	1	6	4	3	5	9	9	6	9	NA	NA	NA	64
5	0	2	0	0	1	0	1	3	0	3	3	5	2	3	3	26
Nonobservable data																
No purchases	8	17	15	12	29	24	40	35	32	37	40	32	48	45	52	466
λ	38	50	42	46	60	49	58	50	52	49	49	46	50	48	55	742

Table 3. Primary demand output $\hat{X}_{jt}$ and arrival rate output $\hat{\lambda}_t$ for $n = 5$ products and for the no-purchase option $j = 0$.

	Periods															
Product	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	Total (N_j)
1	10.0	15.0	11.0	14.0	15.0	12.1	13.0	10.8	14.5	15.9	11.9	18.5	11.5	17.2	17.2	207.5
2	11.0	6.0	11.0	8.0	14.3	11.5	10.4	8.7	11.6	12.7	9.5	14.8	9.2	13.8	13.8	166.3
3	5.0	6.0	1.0	11.0	2.9	3.6	6.9	3.4	5.4	6.2	4.7	7.2	4.5	6.7	6.7	81.2
4	4.0	4.0	4.0	1.0	4.3	2.9	1.5	2.5	4.4	3.4	2.3	3.4	2.7	4.0	4.0	48.3
5	0.0	2.0	0.0	0.0	0.7	0.0	0.5	1.5	0.0	1.1	1.1	1.9	0.6	0.9	0.9	11.4
No-purch.	12.8	14.0	11.5	14.5	15.9	12.8	13.7	11.4	15.3	16.7	12.5	19.5	12.1	18.1	18.1	219.0
$\hat{\lambda}_t$	42.8	47.0	38.5	48.5	53.1	42.8	46.0	38.3	51.1	56.0	42.0	65.4	40.5	60.8	60.8	733.7

values. Note that the results suggest an apparent bias in the estimates, which is not unexpected because the MLE is only asymptotically unbiased. The fifth column shows the asymptotic standard error (ASE) of the corresponding estimate (e.g., see McLachlan and Krishnan 1996, Chapter 4, for details on ASE calculation). Note that for all the coefficients we can reject the null hypothesis that the true value is zero at the 0.005 significance level.[7] The average estimated $\hat{\lambda}$ in this small example is 48.91, showing a small bias with respect to the mean rate: −2.18%.

Table 3 shows the uncensored primary demands obtained by the EM method (i.e., the estimates $\hat{X}_{jt}$, $j = 1, \ldots, n$, and $\hat{X}_{0t}$, $t = T, \ldots, 1$) as well as the estimate of the arrival rate in each period, $\hat{\lambda}_t$ (the sum of all primary demand estimates). Table 4 shows the substitute demand estimates $\hat{Y}_{jt}$, $j = 1, \ldots, n$, and $\hat{Y}_{0t}$, $t = T, \ldots, 1$. By inspection of the latter, observe that as we move toward the end of the horizon (i.e., toward the right of the table) and the most preferred products become less available, the substitute demand tends to explain an increasing fraction of the sales and no-purchases. As a simple validation, note that the total first-choice demand (i.e., $\sum_{j=1}^{n} N_j = 514.7$) matches the total number of bookings (i.e., $\sum_{t=1}^{T}\sum_{j=1}^{n} z_{jt} = 276$) plus the total substitute demand (i.e., 238.7). We also observe negative values of $\hat{Y}_{jt}$ for $j \notin S_t$, representing the total primary demand $\hat{X}_{jt}$ that shifted to another product or to the no-purchase alternative.

From Tables 3 and 4, we can also compute another important performance measure: the percentage of lost sales, defined as

$$\mathbb{P}(\text{lost sales}) = \frac{\sum_{t=1}^{T} Y_{0t}}{\sum_{j=1}^{n} N_j} = \frac{238.7}{514.7} = 46.38\%.$$

The total aggregate recapture rate is computed as the ratio of the total substitute demand across the n products to the total primary demand, i.e.,

$$\text{Recapture rate} = \frac{\sum_{t=1}^{T}\sum_{j\in S_t} Y_{jt}}{\sum_{j=1}^{n} N_j} = \frac{70.04}{514.7} = 13.61\%.$$

In this case, it took 31 iterations of the EM method to meet the stopping criteria in just 0.03 seconds of computation time. As a benchmark, we also optimized the incomplete-data log-likelihood function (i.e., the logarithm of function (2))—which we call *direct max* for short. We used the built-in MATLAB function "fminsearch" that implements the simplex search method of Lagarias et al. (1998). This is a direct search method that does not use numerical or analytic gradients. The initial point $(\mathbf{v}, \boldsymbol{\lambda})$ was based on the observed bookings as in the EM method. The tolerance was set at 0.001. For this small example, the MATLAB algorithm took 11,176 iterations to converge, requiring 14,063 evaluations of the log-likelihood function and 8.26 seconds of computational time. It converged to a point of a slightly higher level set of $\log \mathcal{L}_I(\mathbf{v}, \boldsymbol{\lambda})$ compared to the one obtained by our EM method: −92.38 versus −92.63. However, the two-orders-of-magnitude difference in computation time between the methods, especially considering the small size of the problem, is noteworthy.

A possible concern of the EM method is the sensitivity of the final result with respect to the starting point of the procedure. In the reported results, the initial point was the proportion of sales of each product (see Equation (19)). We also tried as starting point the values $\hat{v}_1 = \cdots = \hat{v}_5 = 1$, which led to a very close (although lower) log-likelihood

Table 4. Substitute demand output $\hat{Y}_{jt}$ for $n = 5$ products and for the no-purchase option $j = 0$.

	Periods															
Product	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	Total in S_t
1	0.0	0.0	0.0	0.0	−15.0	−12.1	−13.0	−10.8	−14.5	−15.9	−11.9	−18.5	−11.5	−17.2	−17.2	0.0
2	0.0	0.0	0.0	0.0	5.7	4.5	−10.4	−8.7	−11.6	−12.7	−9.5	−14.8	−9.2	−13.8	−13.8	10.2
3	0.0	0.0	0.0	0.0	1.1	1.4	7.1	3.6	5.6	−6.2	−4.7	−7.2	−4.5	−6.7	−6.7	18.8
4	0.0	0.0	0.0	0.0	1.7	1.1	1.5	2.5	4.6	5.6	3.7	5.6	−2.7	−4.0	−4.0	26.3
5	0.0	0.0	0.0	0.0	0.3	0.0	0.5	1.5	0.0	1.9	1.9	3.1	1.4	2.1	2.1	14.5
No-purch.	0.0	0.0	0.0	0.0	6.3	5.0	14.3	11.9	15.8	27.3	20.5	31.9	26.4	39.6	39.6	238.7

value of −92.86. Finally, we randomly generated 1,000 starting points, where the value of each $\hat{v}_i$ was Unif[0, 1]. The log-likelihood was -92.65 ± 0.07. So, even though the specific terminal estimates differed when starting from different points, the log-likelihood reached was very similar in all cases.

We next contrast the performance of our EM method with two benchmarks other than direct max. The first benchmark is a *naïve* estimate that sets the primary demand of closed periods at the average of the demand observed in open periods. As mentioned above, this is a traditional unconstraining method used by airlines (called "*Naïve* 2" by Weatherford and Pölt 2002).[8] While these *naïve* estimates are straightforward to compute, their quality is lower than our EM-based estimates. In particular, in this case they belong to a significantly lower level set $\log \mathcal{L}_I(\mathbf{v}, \boldsymbol{\lambda}) = -113.43$.

The second benchmark is the double exponential smoothing (*DES*) or Holt's method, reported by Queenan et al. (2007, §3) as more successful than four other common unconstraining methods, including an EM-related algorithm based on an underlying normal demand assumption (see the appendix of that article). This benchmark takes slightly longer to compute (1.6 seconds) because it has to optimize five quadratic programs (one per product) to find the corresponding base and trend smoothing constants. For this example, the DES estimates belong to an even lower level set $\log \mathcal{L}_I(\mathbf{v}, \boldsymbol{\lambda}) = -115.84$.

Figure 1 (left) illustrates true and estimated primary demands for the preliminary example. The true expected primary demand is described by

$$E[N_j] = \lambda \times T \times \frac{v_j}{\sum_{i=1}^{n} v_i + 1}. \tag{20}$$

The graph shows the decreasing unconstrained, original demand from product 1 (the most preferred according to Table 2) to product 5. Clearly, the primary demand inferred by our EM algorithm is more accurate than the estimates produced by both benchmarks. In particular, the errors of the *naïve* and DES estimates are significantly larger for the least preferred but most available products. This result is intuitive because substitution effects are ignored in the benchmark estimates. The root-mean-square errors (RMSEs) of estimates are 9.41 for EM, and 26.90 and 50.68, respectively, for the *naïve* and DES estimates, providing strong evidence for the relative quality of the EM estimates.

Figure 1 (right) compares the predicted and observed sales per product for the EM and benchmark estimates across the 15 periods. For each method, given estimates $\hat{\mathbf{v}}$, $\hat{\boldsymbol{\lambda}}$ and availability information of the different products, we compute the predicted sales per product per period:

$$E[\text{sales of product } j \text{ in period } t] = \hat{\lambda}_t \frac{\hat{v}_j I\{j \in S_t\}}{\sum_{i \in S_t} \hat{v}_i + 1}, \tag{21}$$

and then for each j we sum these values over t. The RMSEs are 4.42 for EM-based, and 23.44 and 24.90, respectively, for *naïve* and DES estimates, also confirming the strong support in favor of the former.

Figure 2 illustrates the behavior of the estimation methods in two extreme cases. Product 1 (left graph) is the most preferred product and becomes unavailable sooner (cumulative sales are steady from period 12 onward; see Table 1 above). It does not get substitute demand because when it is available, so are the other products. Because in our example $(\mathbf{v}, \boldsymbol{\lambda})$ are homogeneous across time, the cumulative primary demand follows a linear trend. This is tracked closely by our adaptive EM and conforms with the linear proration assumed by the *naïve* heuristic. In contrast, as seen in Figure 2 (left), the DES estimator takes the exponentially smoothed increasing trend and diverges from the true primary demand by overestimating it.

Product 5 (Figure 2, right graph) is the least preferred product and is always available. Because it is always offered, the *naïve* estimate coincides with actual sales

Figure 1. Primary demand (left) and realized sales (right), for the preliminary example, and for estimates under EM, *naïve*, and DES methods.

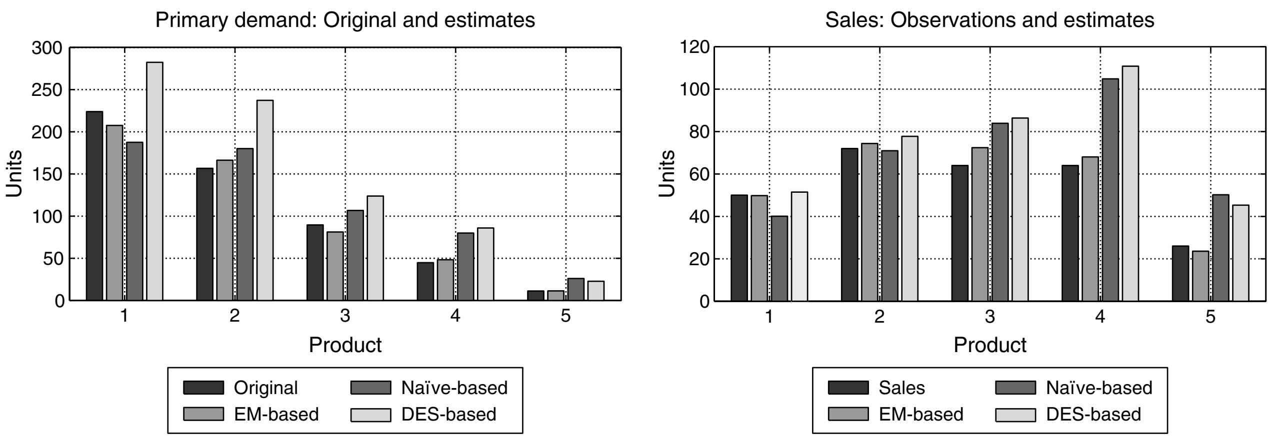
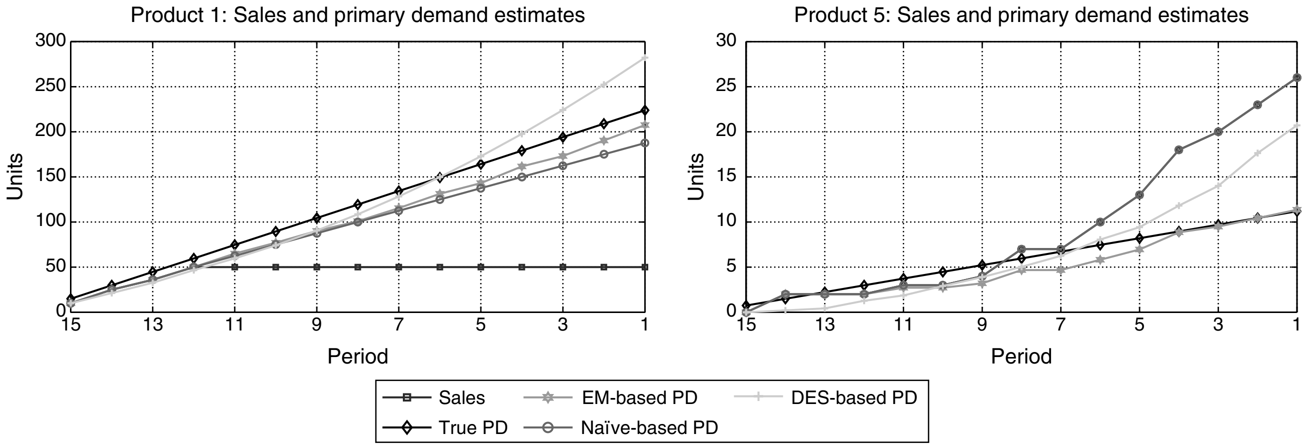


Figure 2. Cumulative observed sales and primary demand (PD) for product 1 (left) and product 5 (right), for the true parameters of the Preliminary Example, and for estimates under EM, *naïve*, and DES methods.



throughout the sales horizon. However, because both the *naïve* and DES estimators are based on the assumption of independent demand for each product, they do not deduct product 5 substitute demand from the observed sales, and therefore both overestimate the true primary demand. This phenomenon is related to the double counting problem discussed in §1. If we consider all products, the expected aggregate primary demand per period in this case is: $\lambda \times \sum_{j=1}^{n} v_j/(\sum_{j=1}^{n} v_j + 1) = 35.07$. Despite the small sample size of this preliminary example, the average of the cumulative first-choice demand per period estimated by the EM algorithm is close: 34.31.

5.1.2. Effects of Data Volume and Quality. In this section, we report on a test of the accuracy of estimates produced by our procedure under different volumes and quality of input data. As in the previous example, given an underlying MNL choice model and assuming that customers arrive according to a homogeneous Poisson process with rate $\lambda = 50$, we used Monte Carlo simulation to generate purchases for $n = 10$ different products. Here, unlike in the previous example, we randomly generated the availability of products: in each period, each product is available independently with probability 0.70. We then tested various volumes of simulated data, ranging from 10 to 5,000 periods.

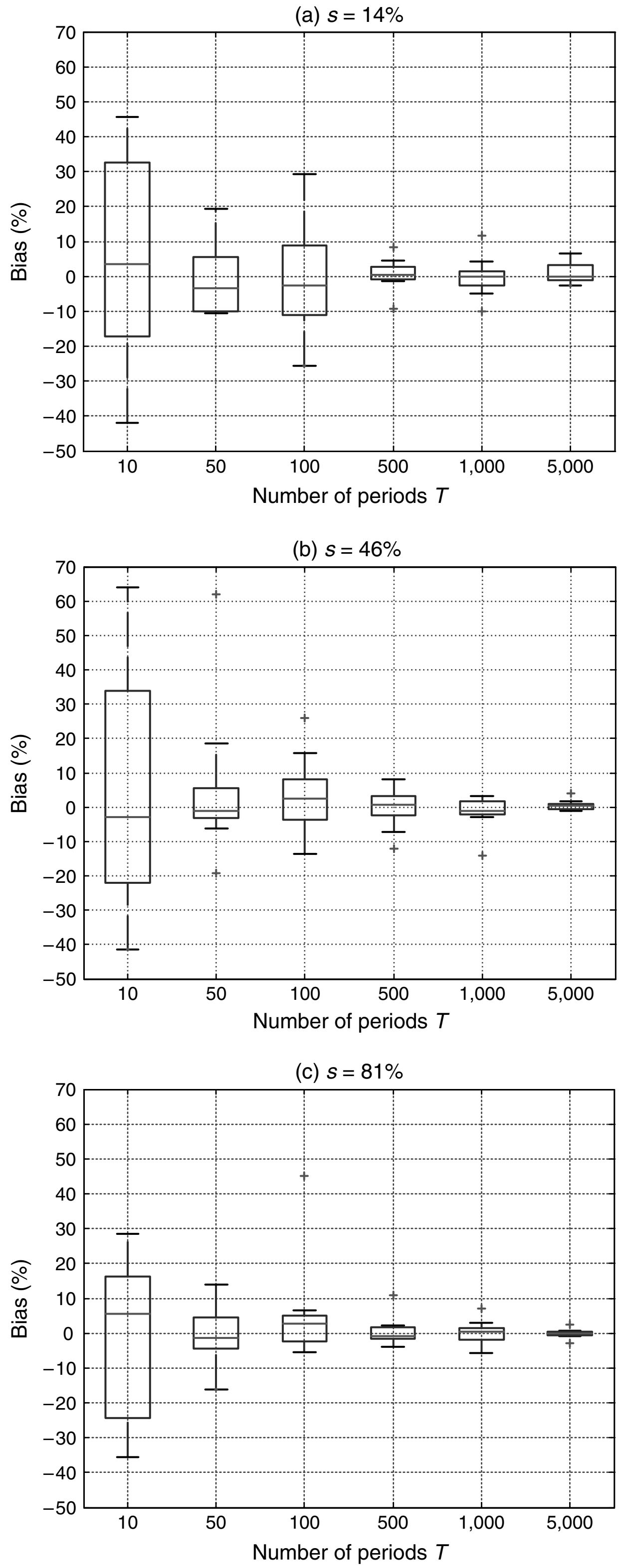
We further considered three different market potential scenarios: a weak market position where $s = 14\%$, an intermediate market position where $s = 46\%$, and a dominant position where $s = 81\%$. Figure 3 shows the box plot of the biases of the estimates $\hat{\mathbf{v}}$ under the different market potential conditions. On each box, the central mark is the median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. The average of the estimates $\hat{\lambda}_t$ was always very close to the mean 50, consistently exhibiting a very small bias compared with the bias for the $\hat{\mathbf{v}}$ (generally within $[-2\%, 2\%]$), so we did not include it in the box plot.

As expected, we note that for each market potential scenario, as we increase the number of periods, the biases decrease. Having $T = 50$ periods seems to be enough data to drive most biases under 10%. At the same time, as the market potential increases (and hence, more purchases per period are observed), accuracy also increases.

One potential concern of our procedure is the need to get an exogenous estimate of market share and the resulting impact this estimate has on the quality of the estimates. To test this sensitivity, we used the same inputs for generating data as above (i.e., $\lambda = 50$, $n = 10$, and products available with probability 0.70) for the case of $T = 500$ periods. We then applied our EM procedure assuming inaccurate information about the market potential. Specifically, we perturbed s by $\pm 10\%$ and $\pm 20\%$, and plotted the biases of the estimates $\hat{\mathbf{v}}$ and the average $\hat{\lambda}$ (Figure 4, left) and of the estimates of the primary demand $\hat{N}_j$, $j = 1, \ldots, n$, and the average $\hat{\lambda}$ (Figure 4, right). Note that a perturbation of the market potential generally amplifies the biases of the estimated parameters $\hat{\mathbf{v}}$ and the average $\hat{\lambda}$ with respect to their original values. However, the algorithm adjusts these biases in such a way that it preserves the quality of the estimates of the primary demand volume for products $j = 1, \ldots, n$. In other words, the relative preferences across products are sensitive to the initial assumption made about market potential (see §6.1 for further discussion), yet Figure 4 (right) shows a relatively small bias in the resulting primary demand estimates.

5.1.3. Comparison with Three Benchmarks. Our last experiment on synthetic data assesses the performance of our EM method relative to the three aforementioned benchmarks: the direct max, *naïve*, and DES method. We generated transaction data for $n \in \{5, 15, 25\}$, $T \in \{30, 100, 300\}$ and two product availability settings: one with open availability, where each product is available in each period with probability 0.8, and one with limited availability, where this probability is set at 0.5, for a total of 18

Figure 3. Biases of the preference weights $\hat{\mathbf{v}}$ under different market potentials: (a) $s = 14\%$, (b) $s = 46\%$, and (c) $s = 81\%$, for different selling horizon lengths.



scenarios. We simulated 1,000 instances for each of those scenarios based on the following underlying MNL demand model: preference weights $v_j \sim \text{Unif}[0.05, 1]$, $j = 1, \ldots, n$, $v_0 = 1$, and arrival rates $\lambda_t \sim \text{Unif}[10, 100]$.

After generating an instance of data, we applied the four methods under consideration. For direct max, we again used the "fminsearch" MATLAB function, setting both the iteration and function evaluation limits at 1,000 and the tolerance (i.e., the difference between two consecutive function values) at 0.001. Tables 5 and 6 summarize the results for both availability settings, fixing the values provided by direct max as the baseline. We report the difference between the log-likelihood values and root mean squared errors (RMSEs) of EM, *naïve* and DES with respect to direct max; desirable outcomes are positive values in the log-likelihood difference and negative values in the RMSE difference. Note that EM consistently achieves this desirable performance, and the difference tends to be more significant when the problem is larger (large n and large T). Direct max in turn is consistently better than *naïve* and DES.

In terms of speed, direct max was clearly the most computationally intensive method. With the configuration described above for running the MATLAB function, it took about one minute to calculate each of the large cases, and the procedure frequently terminated due to the iteration limit. The other methods took only a couple of seconds to compute, except for DES, which occasionally ran longer due to its need to solve quadratic programs (minimizing squared error) during its execution. When relaxing the constraint on the number of iterations of direct max, the quality of the estimates increases and becomes closer (and even slightly better) than EM, but the computation time also escalates; for example, for the open availability case, $n = 15$ and $T = 100$, the average RMSE difference in favor of direct max is 1.97, but the procedure requires around 20 minutes of calculation to converge to the solution.

It is also noticeable that the quality of the EM and direct max estimates improve with respect to *naïve* and DES estimates as products become less available (i.e., Table 6 compared to Table 5). This is because the *Naïve* and DES estimates correct for demand censoring but do not adjust for double counting recaptured demand.

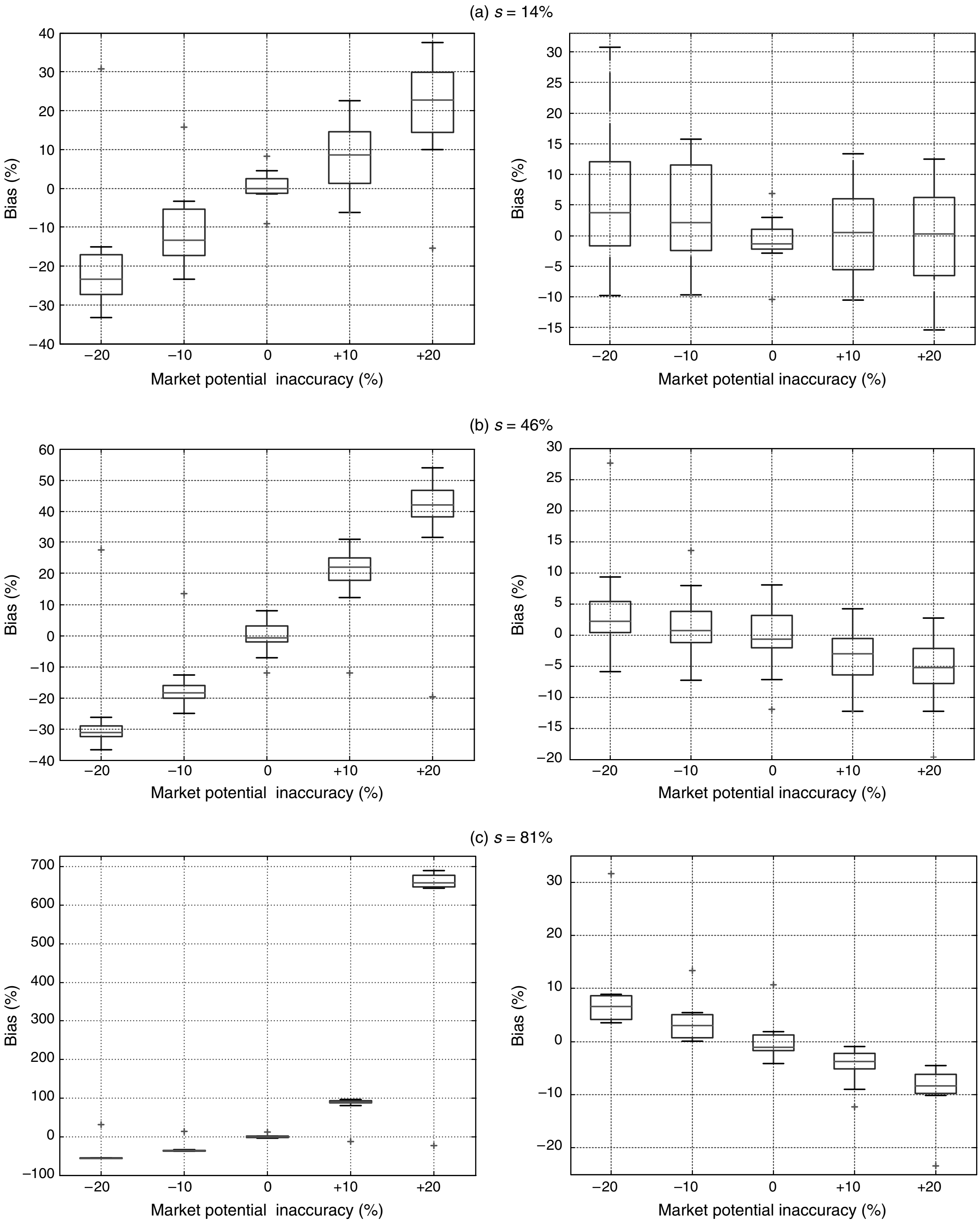
All in all, EM is clearly attractive relative to the benchmark methods—in terms of both estimation quality and computational speed.

5.2. Industry Data Sets

We next present results of two estimation examples based on real-world data sets, one for an airline market and one for a retail market.

5.2.1. Airline Market Example. This example is based on data from a leading commercial airline serving a sample *O–D* market with two daily flights. It illustrates the practical feasibility of our approach and shows the impact of the consideration set design on the estimation outcome.

Figure 4. Biases of the estimates $\hat{\mathbf{v}}$ and the average $\hat{\lambda}$ (left) and of the estimates of the primary demand $\hat{N}_j$ and the average $\hat{\lambda}$ (right) under noisy market potentials.



Notes. The raw data were generated based on the true market potentials: (a) $s = 14\%$, (b) $s = 46\%$, and (c) $s = 81\%$, and then the parameters were estimated assuming perturbed values: $\pm 1.2s$ and $\pm 1.1s$.

Table 5. Comparative results with respect to direct max for the open availability case.

		Difference in log-likelihood values						Difference in RMSEs					
		EM		*Naïve*		DES		EM		*Naïve*		DES	
Products n	Periods T	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
5	30	2.0	1.7	−25.6	10.2	−126.0	34.4	−6.5	3.6	26.1	8.5	32.0	15.3
	100	23.5	9.3	−64.4	28.2	−391.7	71.9	−42.9	6.7	59.4	23.1	64.9	32.5
	300	85.8	24.5	−178.3	78.0	−1,168.4	166.5	−155.8	18.2	153.0	70.4	160.4	79.2
15	30	0.5	0.5	−46.7	11.2	−165.4	34.4	−0.7	0.8	21.3	3.2	23.2	5.7
	100	4.3	1.6	−148.9	19.8	−535.3	62.1	−10.2	2.5	61.8	6.0	62.4	10.2
	300	15.8	4.0	−439.7	40.4	−1,584.5	104.9	−38.3	6.4	177.6	13.6	173.7	19.7
25	30	0.8	0.7	−51.7	10.3	−177.1	34.1	−0.7	0.7	14.6	2.0	16.1	3.4
	100	3.3	1.3	−165.0	18.3	−563.4	63.3	−5.6	1.6	44.4	3.6	44.3	6.2
	300	9.6	2.4	−491.8	31.9	−1,674.0	106.3	−19.8	3.3	130.5	6.9	125.2	11.0

Note. Mean and standard deviation (SD) of the differences of log-likelihood and RMSE values.

We analyzed bookings data for the last seven selling days prior to departure for each consecutive Monday from January to March of 2004 (11 departure days total). There were 11 classes per flight, and each class has a different fare value. Fares were constant during the 11 departure days under consideration. The market share of the airline for this particular O-D pair was known to be approximately 50%, which we used as the value for s (recall the discussion in §3.4).

We define a *product* as a flight-class combination, so we had $2 \times 11 = 22$ products. For each product, we had seven booking periods (of length 24 hours) per departure day, leading to a total of $7 \times 11 = 77$ observation periods. There were nonzero bookings for 15 out of the 22 products, so we focused our analysis on those 15 products. We note that in the raw data we occasionally observed a few small negative values as demand realizations; these negative values corresponded to ticket cancelations, and for our analysis we simply set them to zero.

We computed two sets of estimates for the demand, under different assumptions: in the multiflight case we assumed customers chose between both flights in the day, so the consideration set consisted of all 15 products; in the independent-flight case, we assumed customers were interested in only one of the two flights, implying there were two disjoint consideration sets (one for each of the flights with 7 and 8 products, respectively) and with a market share of 25% per flight.

Again we tested the performance of our EM-based estimates versus the performance of three alternative estimation methods: naïve, DES, and direct max. While both EM and direct max consider each day independently, both naïve and DES methods rely on a time series model of the demand. Therefore, for the latter two, we treated data at the week level; i.e., for each week and for each product, we came up with an estimate of the primary demand. Then for each product j, we aggregated the primary demand across the 11 weeks to get $\hat{N}_j$ and used it to compute $\hat{v}_j$ as in (14).

Table 7 shows the results. Besides checking the value of $\log \mathcal{L}_I(\mathbf{v}, \boldsymbol{\lambda})$ for each pair $(\hat{\mathbf{v}}, \hat{\boldsymbol{\lambda}})$, we conducted two in-sample tests.[9] After running our estimation procedure, we aggregated the observed bookings and the predicted bookings (computed as in (21)) across all the 77 periods and computed RMSEs and goodness-of-fit χ^2-tests for the multiflight and independent cases. We needed to do this global aggregation to ensure the number of expected bookings was

Table 6. Comparative results with respect to direct max for the limited availability case.

		Difference in log-likelihood values						Difference in RMSEs					
		EM		*Naïve*		DES		EM		*Naïve*		DES	
Products n	Periods T	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
5	30	17.4	7.2	−87.5	37.0	−153.5	68.7	−20.6	6.2	60.4	22.3	63.6	31.1
	100	126.3	22.0	−215.1	109.5	−460.0	177.6	−112.1	10.5	148.8	61.5	175.4	78.0
	300	424.7	56.6	−594.0	328.0	−1,355.8	463.7	−374.1	22.6	402.1	185.6	497.5	212.7
15	30	10.1	4.6	−304.6	54.4	−401.2	108.6	−6.7	2.0	66.7	9.1	70.6	15.3
	100	49.3	12.0	−980.0	120.9	−1,350.0	217.8	−34.9	4.6	202.1	20.2	223.8	32.0
	300	145.8	30.7	−2,938.8	301.9	−4,072.3	444.6	−110.7	13.0	597.6	52.7	668.6	67.5
25	30	9.2	4.5	−397.6	58.2	−510.7	110.7	−3.6	1.3	51.7	6.0	55.7	9.9
	100	30.5	7.7	−1,294.5	110.0	−1,714.2	214.2	−17.6	2.7	161.4	11.1	178.8	19.1
	300	73.3	13.0	−3,884.3	236.7	−5,172.1	383.2	−52.7	5.8	480.7	23.3	534.1	33.8

Note. Mean and standard deviation (SD) of the differences of log-likelihood and RMSE values.

Table 7. Estimation results for the airline market example.

		Estimation method			
Consideration set	Measure	EM	*Naïve*	DES	Direct max
Flight 1	$\log \mathscr{L}_I(\mathbf{v}, \boldsymbol{\lambda})$	−279.44	−300.15	−290.06	−378.67
	RMSE	2.18	3.64	14.32	7.39
	χ^2-test (p-value)	0.99	0.33	0.00	0.00
Flight 2	$\log \mathscr{L}_I(\mathbf{v}, \boldsymbol{\lambda})$	−276.22	−383.36	−450.36	−365.49
	RMSE	2.00	20.42	13.56	4.65
	χ^2-test (p-value)	0.99	0.00	0.00	0.01
Joint	$\log \mathscr{L}_I(\mathbf{v}, \boldsymbol{\lambda})$	−681.08	−805.27	−992.81	−756.39
	RMSE	6.31	21.88	15.03	7.44
	χ^2-test (p-value)	0.58	0.00	0.00	0.00

greater than or equal to 5 for all the products, to have meaningful χ^2-tests. Across all measures, EM clearly dominates the other methods. The relative performance among the other methods is mixed.

Computationally, EM, naïve, and DES are fast, while direct max is considerably slower. In fact, despite the small size of the data set, when maximizing $\log \mathscr{L}_I(\mathbf{v}, \boldsymbol{\lambda})$ using the standard built-in MATLAB optimization function, it took 16 minutes for the multiflight case and 7 minutes and 12 minutes, respectively, for each of the flights in the independent case. Recall that this incomplete data function does not have much structure, and a standard optimization algorithm can get stuck in a local extremum or saddle point. While one could attempt to stabilize the MATLAB procedure and try different starting points, the experience on this example attests to the simplicity, efficiency, and robustness of our EM method relative to brute-force MLE (i.e., direct max). In fact, it took only 31 iterations of the EM method to compute the multiflight estimates and 24 and 176 iterations for each of the independent flights, taking only a fraction of a second. For a major airline estimating hundreds of thousands of *O–D* markets on a daily or even more frequent basis, such differences in computing time are significant.

Overall, the EM algorithm outperforms the three benchmarks in terms of both computational time and quality of output for this independent-flight case.

For the multiflight case, Figure 5 shows the observed bookings and predicted bookings for the 15 products under consideration. The labels in the horizontal axis represent the fares of the corresponding products (e.g., "F1, $189" means "Flight 1, bucket with fare $189"). Figure 6 shows a similar plot for the independent-flight case. In both figures, EM-based predictions track closely the observed sales. An exception is the first product "F1, $189," which according to our data is available throughout the whole horizon but experiences sales just in the last two weeks.

Comparing the two cases, the multiflight case offers more degrees of freedom in fitting the product demands because it includes relative attractiveness across more options, and therefore it is a harder estimation case. Moreover, the differences in predictions produced by the two approaches suggest that the definition of the consideration set can have significant impact on the quality of the estimates. Hence, how best to construct these sets is an important area of future research (e.g., see Fitzsimons 2000 for an analysis of the impact of choice set design on stockouts).

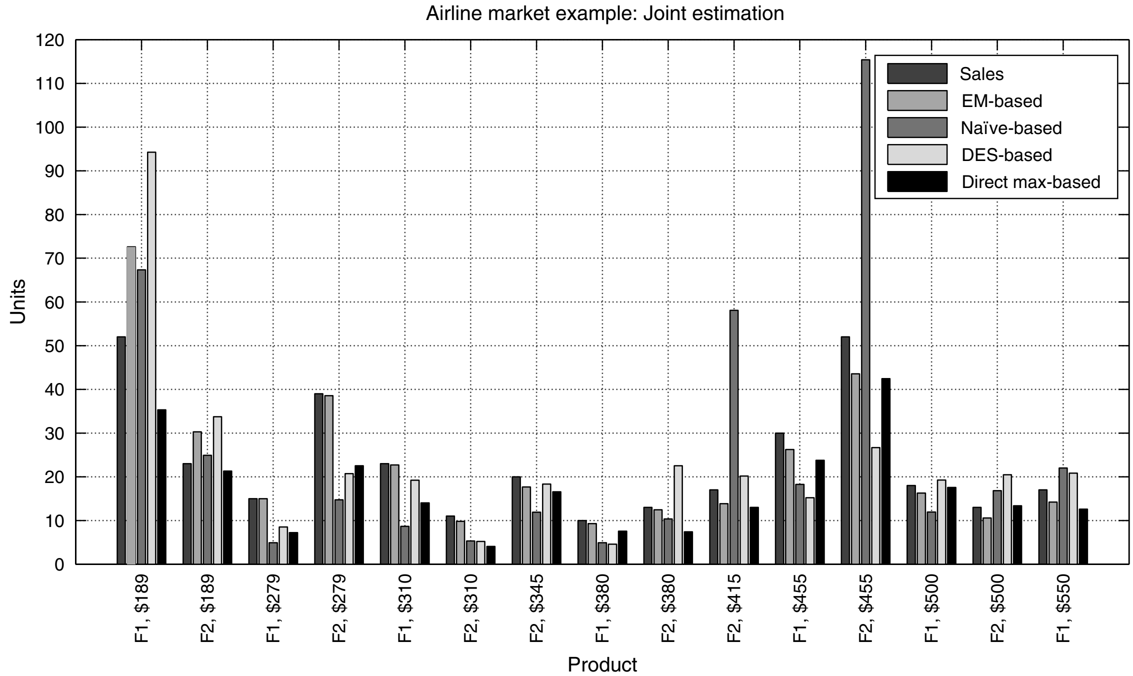
Finally, focusing on the EM-based estimates and using $\hat{Y}_{0t}$ and $\hat{N}_j$, we computed the fraction of lost sales. For the multiflight case the estimate was 42.4%, and for the two-independent-flight case, the estimates were 33.1% and 86.1% for each flight, respectively. Table 8 summarizes the estimation statistics for the output of the EM method under both market segmentation cases. The t-statistics indicate that we can reject the null hypothesis that the true value of any coefficient is zero at the 0.01 significance level.

5.2.2. Retail Market Example. This next example illustrates our EM method applied to sales data from a retail chain. We consider sales observed during eight weeks over a sample selling season. We assume a choice set defined by six substitutable products within the same small subcategory of SKUs. The market share of this retail location is estimated to be 48%. The first few products (P1–P3) had more limited availability, while product P6 was the most available. As in the previous example, we tested the performance of our EM-based estimates against the three benchmarks: naïve, DES, and direct max. For naïve and DES, we treated data at the week level and then aggregated the primary demand across the eight weeks to get $\hat{N}_j$, and next $\hat{v}_j$ as in (14).

Table 9 shows the results. Besides checking the value of $\log \mathscr{L}_I(\mathbf{v}, \boldsymbol{\lambda})$ for the estimates $(\hat{\mathbf{v}}, \hat{\boldsymbol{\lambda}})$, we conducted two in-sample tests. After running our estimation procedure, we aggregated the observed bookings and the predicted bookings (computed as in (21)) across all the 56 periods, and computed RMSEs and goodness-of-fit χ^2-tests. Across all measures, EM again clearly outperforms the others.

In terms of computation time for this example, EM, naïve, and DES are straightforward to compute and take less than a second (although, again, DES requires solving simple quadratic minimization problems during its execution). For instance, for EM it just took 120 iterations to

Figure 5. Comparison of observed and predicted bookings for EM-based, *naïve*-based, DES-based, and direct max-based estimates under the multiflight assumption for the airline market example.



reach convergence in only 0.3 seconds. In contrast, when running the MATLAB built-in function "fminsearch" to optimize the log-likelihood function for this example, its performance (in terms of the likelihood value) was worse than our EM method and it ran for over three minutes, taking 87,829 iterations and 97,862 evaluations of the function $\log \mathscr{L}_I(\mathbf{v}, \boldsymbol{\lambda})$. Again, for a large retailer estimating hundreds of categories across thousands of stores, such computation time differences matter.

Figure 7 shows the observed and predicted sales for the six products under consideration. Again, EM-based predictions closely tracked the observed sales. Naïve tends to

Figure 6. Comparison of observed and predicted bookings for EM-based, *naïve*-based, DES-based, and direct max-based estimates under the independent flight assumption for the airline market example.

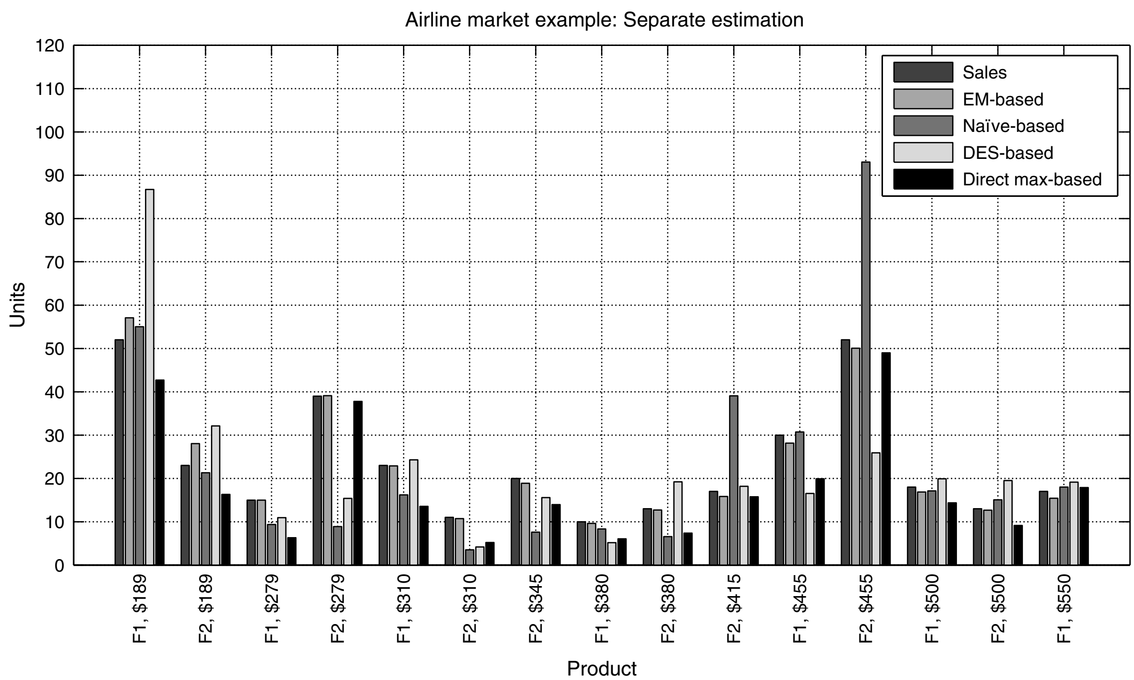


Table 8. Estimation results for the airline market example.

Parameter	Product	Multiflight demand Coefficient	Multiflight demand ASE	Multiflight demand t-statistic	Independent-flight demand Coefficient	Independent-flight demand ASE	Independent-flight demand t-statistic
v_1	F1, $189	0.0832	0.0121	6.8760	0.0695	0.0100	6.9500
v_2	F2, $189	0.0397	0.0082	4.8415	0.0105	0.0016	6.5625
v_3	F1, $279	0.1249	0.0151	8.2715	0.0658	0.0097	6.7835
v_4	F2, $279	0.2087	0.0203	10.2808	0.1814	0.0073	24.8493
v_5	F1, $310	0.1361	0.0159	8.5597	0.0747	0.0104	7.1827
v_6	F2, $310	0.0455	0.0088	5.1705	0.0353	0.0030	11.7667
v_7	F2, $345	0.0524	0.0095	5.5158	0.0379	0.0031	12.2258
v_8	F1, $380	0.0442	0.0087	5.0805	0.0289	0.0063	4.5873
v_9	F2, $380	0.0358	0.0078	4.5897	0.0248	0.0025	9.9200
v_{10}	F2, $415	0.0314	0.0073	4.3014	0.0183	0.0021	8.7143
v_{11}	F1, $455	0.0725	0.0113	6.4159	0.0488	0.0083	5.8795
v_{12}	F2, $455	0.0614	0.0103	5.9612	0.0227	0.0024	9.4583
v_{13}	F1, $500	0.0359	0.0078	4.6026	0.0268	0.0061	4.3934
v_{14}	F2, $500	0.0121	0.0045	2.6889	0.0024	0.0008	3.0000
v_{15}	F1, $550	0.0163	0.0052	3.1346	0.0188	0.0051	3.6863

Table 9. Estimation results for the retail market example.

Measure	Estimation method EM	*Naïve*	DES	Direct max
$\log \mathcal{L}_I(\mathbf{v}, \boldsymbol{\lambda})$	−132.63	−172.36	−232.51	−182.10
RMSE	1.86	8.19	7.48	5.12
χ^2-test (p-value)	0.97	0.00	0.00	0.00

underestimate the less available products (P1–P3) and overestimate the most available ones (P6). DES is more conservative, although it also overestimates the sales of product P6. The direct max procedure seemingly was trapped in a (bad) stationary point of the incomplete data log-likelihood function, producing poor estimates.

Table 10 summarizes the estimation statistics for the output of the EM method. The t-statistics indicate that we can reject the null hypothesis that the true value of all the coefficients is zero at the 0.01 significance level.

Figure 7. Comparison of observed and predicted sales for EM-based, *naïve*-based, DES-based, and direct max-based estimates for the retail market example.

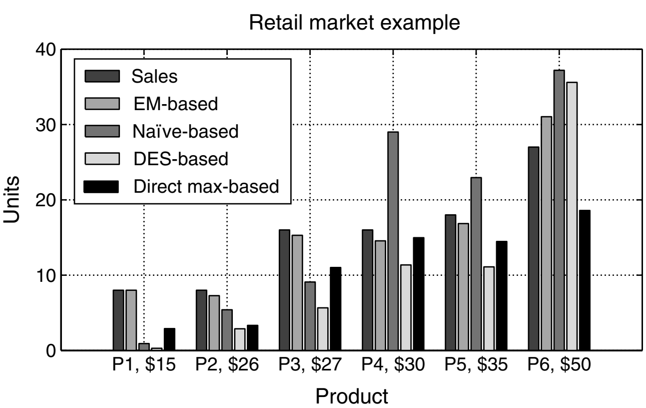


Table 10. Estimation results for the retail market example.

Parameter	Product and price	Coefficient	ASE	t-statistic
v_1	P1, $15	0.4342	0.0435	9.98
v_2	P2, $26	0.1366	0.0217	6.29
v_3	P3, $27	0.2093	0.0277	7.56
v_4	P4, $30	0.0541	0.0132	4.10
v_5	P5, $35	0.0313	0.0099	3.16
v_6	P6, $50	0.0576	0.0136	4.24

Finally, using the EM-based estimates, we compute the percentage of lost sales for this example, which turns out to be very significant:

$$\mathbb{P}(\text{lost sales}) = \frac{\sum_{t=1}^{T} Y_{0t}}{\sum_{j=1}^{n} N_j} = \frac{210}{303} = 69.3\%.$$

6. Implementation Issues and Extensions

6.1. Model Parameters

While the overall EM procedure as stated above is simple and efficient, there are several practical issues that warrant further discussion.[10] One issue we observed is that the estimates are sensitive to how consideration sets are defined. Hence, it is important to have a good understanding of the set of products that customers consider and to test these different assumptions.

We have also noticed that with some data sets, the method can lead to extreme estimates; for example, arrival rates that tend to infinity, or preference values that tend to zero. This is not a fault of the algorithm per se but rather the maximum likelihood criterion. In these cases, we have found it helpful to impose various ad hoc bounding rules to keep the parameter estimates within a plausible range. In markets where the seller has significant market power, we have found it reasonable to set a value s no larger than

90%; otherwise, one can get abnormally high recapture rates into the least preferred products.

Our method requires binary data to describe the availability of a product during a time period. However, a simple heuristic variation can accommodate partial availability, which is a common situation for airlines. The idea is to partition the continuous time within a period into a finite number of periods where the product is either fully available or unavailable, and assume that arrivals occurred uniformly during the original period. For instance, suppose that a product has been available during 60% of a period duration, and that there have been three purchases observed. Then, the period could be split into five periods. In three of them the product will be fully available, and there will be a single purchase in each. In the remaining two periods, the product would be unavailable.

There is also the issue of obtaining a good estimate of the market share or market potential s (recall that this depends on our interpretation of the outside alternative). In either case, note that this share is based on an implicit "all-open" product offering, i.e., $s = \sum_{i=1}^{n} v_i/(\sum_{i=1}^{n} v_i + 1)$. This is a difficult quantity to measure empirically in some environments, and indeed our entire premise is that products might not be available in every period. Nevertheless, the following procedure avoids estimating an "all open"-based s: recall from §3.3 that given MLE estimates $\mathbf{v}^*$ and $\boldsymbol{\lambda}^*$, we can scale these estimates by an arbitrary constant $\alpha > 0$ to obtain a new MLE of the form

$$\mathbf{v}(\alpha) = \alpha \mathbf{v}^*,$$

$$\boldsymbol{\lambda}(\alpha) = \frac{\alpha \sum_{i \in S_t} v_i^* + 1}{\alpha(\sum_{i \in S_t} v_i^* + 1)} \boldsymbol{\lambda}^*.$$

The family of MLE estimates $\mathbf{v}(\alpha), \boldsymbol{\lambda}(\alpha)$ all lead to the same expected primary demand for the firm's own products $j = 1, \ldots, n$ for all α, but they produce different expected numbers of customers who choose the outside alternative (i.e., buy a competitor's product or do not buy at all). Therefore, if we have a measure of actual market share over the same time periods from other sources (based on actual availability rather than on the "all open" assumption), one can simply search for a value of α that produces a total expected market share (using (21)) that matches the total observed market share. This is a simple one-dimensional, closed-form search because the family of MLEs $\mathbf{v}(\alpha), \boldsymbol{\lambda}(\alpha)$ is a closed-form function of α.

Finally, note that by correcting for both the censoring and double counting problems, our model and estimation approach provides an underlying independent-demand estimate of primary demand, because the Poisson arrivals are partitioned according to (full-availability) MNL probabilities. That is, primary demand for each product is nonhomogenous Poisson and independent across products. Thus, one can use standard time-series methods applied over these primary demand estimates to forecast future primary demand.

6.2. Linear-in-Parameters Utility

In our basic setting, we focus on estimating a vector of preference weights $\hat{\mathbf{v}}$. A common form of the MNL model assumes the preference weight v_j can be further broken down into a function of attributes of the form $v_j = e^{u_j}$ where $u_j = \boldsymbol{\beta}^T \mathbf{x}_j$ is the nominal utility of alternative j, $\mathbf{x}_j$ is a vector of attributes of alternative j, and $\boldsymbol{\beta}$ is a vector of coefficients (part worths) that assign a utility to each attribute. Expressed this way, the problem is one of estimating the coefficients $\boldsymbol{\beta}$.

Our general primary demand approach is still suitable for this MNL case. The only difference is that now there is no closed-form solution for the M-step of the EM algorithm, and one must resort to nonlinear optimization packages to solve for the optimal $\boldsymbol{\beta}$ in each iteration. Alternatively, one can use the following two-step approach: In step 1, run the EM algorithm as described here to estimate $\hat{\mathbf{v}}$. In step 2, look for a vector $\boldsymbol{\beta}$ that best matches these values using the fact that $\log \hat{v}_j = \boldsymbol{\beta}^T \mathbf{x}_j$, $j = 1, \ldots, n$. In most cases, this will be an over-determined system of equations, in which case we could run a least-squares regression to fit $\boldsymbol{\beta}$. The following proposition provides theoretical support for this procedure.

Proposition 3. *Suppose that the observed purchases are generated by an underlying linear-in-parameters MNL model, so that the preference weights v_i, $1 \leqslant i \leqslant n$ satisfy $v_i = \boldsymbol{\beta}'^{T} \mathbf{x}_i$ for some unknown vector $\boldsymbol{\beta}'$. For a given sample size $N = n \times T$, let the MLE estimate (e.g., a limit point of the EM algorithm) be denoted $\hat{\mathbf{v}}$. Now consider the least-squares problem*

$$\min_{\boldsymbol{\beta}} g(\boldsymbol{\beta}) = \sum_{i=1}^{n} (\hat{v}_i - e^{\boldsymbol{\beta}^{\mathrm{T}} \mathbf{x}_i})^2 .$$

Then $\boldsymbol{\beta}'$ converges in probability to an optimal solution of the least-squares problem as the sample size N increases.

Proof. Note that if we substitute the true value $\boldsymbol{\beta}'$ in $g(\cdot)$, then $g(\boldsymbol{\beta}') = \sum_{i=1}^{n} (\hat{v}_i - v_i)^2$. Because $\hat{v}_i$ is a MLE estimator for v_i, then it is consistent, and therefore $\hat{v}_i \Rightarrow v_i$, where "$\Rightarrow$" stands for *convergence in probability* (see Billingsley 1995, Theorem 25.3), and where the limit is taken over the number of periods T (so over the sample size N). Consider the continuous functions $h_i(x) = (x - v_i)^2$, $i = 1, \ldots, n$. From (Billingsley (1995, Corollary 2 of Theorem 25.7) we then have that for each i, $h_i(\hat{v}_i) = (\hat{v}_i - v_i)^2 \Rightarrow 0$. Hence, $g(\boldsymbol{\beta}') = \sum_{i=1}^{n} h_i(\hat{v}_i) \Rightarrow 0$. Because $g(\boldsymbol{\beta}) \geqslant 0$ for any $\boldsymbol{\beta}$, this means the true vector $\boldsymbol{\beta}'$ solves the least-squares problem asymptotically. □

Again, note from Theorem 1 that the EM procedure is not guaranteed to provide a limit point and moreover might provide only a local maximum. To ensure that the above procedure correctly estimates $\boldsymbol{\beta}$, care must be taken to check numerically that the sequence of EM estimates is convergent, and it might be necessary to try multiple starting points to ensure that the algorithm if finding a global maximum.

7. Conclusions

Estimating the underlying demand for products when there are significant substitution effects and lost sales is a common problem in many retail markets. Our approach combines a multinomial logit (MNL) demand model with a nonhomogeneous Poisson model of arrivals over multiple periods. It assumes realistic data: observed sales, product availability, and an aggregate estimate of the market share of the set of products. The problem we address is how to jointly estimate the parameters of this combined model; i.e., preference weights of the products and arrival rates. By viewing the problem in terms of primary demand and treating the observed sales as incomplete observations of primary demand, we are able to apply the expectation-maximization (EM) method to this incomplete demand model. This leads to a very simple, highly efficient iterative procedure for estimating the parameters of the model that provably converges to a stationary point of the incomplete data log-likelihood function. Numerical examples show that the method performs very well in terms of estimation quality and speed relative to other simple benchmark estimation methods and to direct maximization of the incomplete log-likelihood function. Given its simplicity to implement, the realistic input data needed, and the quality of the results, we believe that our EM algorithm has significant practical potential. The general strategy of considering demand estimation in terms of primary demand might also help improve estimation procedures in other cases. For example, it would be interesting to see if the approach could be adapted to a latent-segment, mixed MNL model or a nested logit model.

Acknowledgments

The authors thank John Blankenbaker at Sabre Holdings for his careful review and constructive suggestions on earlier drafts of this work, in particular for pointing out the existence of a continuum of maxima in the absence of a market potential parameter. Ross Darrow and Ben Vinod at Sabre Holdings also provided helpful feedback on this work. The authors are also grateful to Martín Gonzalez Rozada and Martín Solá from Torcuato di Tella University, Argentina, for their comments on the numerical examples. Finally, they thank Marcelo Olivares (Columbia University), the associate editor, and three anonymous referees for their constructive feedback.

Endnotes

1. A further generalization of this MNL model to the case where the preference weights are functions of the product attributes is provided in §6.2.
2. For example, Sabre has been running the single-segment MNL model for a large origin-destination airline for more than two years and has been observing very significant revenue improvements.
3. Later in §6.1 we discuss how to relax this requirement and accommodate partial availability of products per period, e.g., how to account for the fact that a product is available during 60% of the time within a period.
4. This is due to our assumption that $\mathbf{v} > 0$, and that for at least one period t, $z_{jt} > 0$, for each $j = 1, \ldots, n$.
5. See also McLachlan and Krishnan (1996, Theorem 3.2).
6. MATLAB is a trademark of The MathWorks, Inc. We used version 7.10 for Microsoft Windows 7 on a CPU with Intel Core i7 processor and 4 Gb of RAM.
7. The quasi-t statistic is computed as the ratio between the estimated value of the parameter and the ASE. The preference weights v_j are always nonnegative. Recall that for a one-tailed test, the critical values of this statistic are ±1.65, ±1.96, and ±2.58 for the 0.05, 0.025, and 0.005 significance levels, respectively.
8. This is the standard, single-class untruncation method used by airlines on booking curves under the independent demand model. For instance, based on Table 1, product 1 shows 50 sales in 4 out of 15 periods, so this ad-hoc estimator sets the average $50/4 = 12.5$ as the primary demand for periods where product 1 was not available. Weatherford and Pölt (2002) report better results for another averaging method, called "*Naïve* 3," that exploits partial closures during a period. Recall that our setting allows only full or no availability of a product during a period.
9. We also tried out-of-sample tests, but the quantity of data was very limited and too volatile to allow for good out-of-sample testing.
10. The comments in this section are based not only on our own experience but also on Sabre's experience, obtained through the use of a proprietary variation of this EM method that has been in production since 2008.

References

Andersson, S. E. 1998. Passenger choice analysis for seat capacity control: A pilot project in Scandinavian Airlines. *Internat. Trans. Oper. Res.* **5**(6) 471–486.
Anupindi, R., M. Dada, S. Gupta. 1998. Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Sci.* **17**(4) 406–423.
Ben-Akiva, M., S. Lerman. 1994. *Discrete Choice Analysis: Theory and Applications to Travel Demand*, 6th ed. MIT Press, Cambridge, MA.
Berry, S., J. Levinsohn, A. Pakes. 1995. Automobile prices in market equilibrium. *Econometrica* **63**(4) 841–890.
Billingsley, P. 1995. *Probability and Measure*, 3rd ed. John Wiley & Sons, New York.
Borle, S., P. Boatwright, J. Kadane, J. Nunes, S. Galit. 2005. The effect of product assortment changes on customer retention. *Marketing Sci.* **24**(4) 616–622.
Bruno, H., N. Vilcassim. 2008. Structural demand estimation with varying product availability. *Marketing Sci.* **27**(6) 1126–1131.
Bucklin, R., S. Gupta. 1992. Brand choice, purchase incidence, and segmentation: An integrated modeling approach. *J. Marketing Res.* **29**(2) 201–215.
Campo, K., E. Gijsbrechts, P. Nisol. 2003. The impact of retailer stockouts on whether, how much, and what to buy. *Internat. J. Res. Marketing* **20**(3) 273–286.
Chintagunta, P. 1993. Investigating purchase incidence, brand choice and purchase quantity decisions on households. *Marketing Sci.* **12**(2) 184–208.
Chintagunta, P., J.-P. Dubé. 2005. Estimating a SKU-level brand choice model that combines household panel data and store data. *J. Marketing Res.* **42**(3) 368–379.
Conlon, C., J. Mortimer. 2009. Demand estimation under incomplete product availability. Working paper, Department of Economics, Harvard University, Cambridge, MA.
DeHoratius, N., A. Raman. 2008. Inventory record inaccuracy: An empirical analysis. *Management Sci.* **54**(4) 627–641.
Dempster, A., N. Laird, D. Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *J. Royal Statist. Soc. B* **39**(1) 1–38.

Fitzsimons, G. 2000. Consumer response to stockouts. *J. Consumer Res.* **27** 249–266.

Gruen, T., D. Corsten, S. Bharadwaj. 2002. Retail out-of-stocks: A worldwide examination of causes, rates, and consumer responses. Grocery Manufacturers of America, Washington, DC.

Guadagni, P., J. Little. 1983. A logit model of brand choice calibrated on scanner data. *Marketing Sci.* **2**(3) 203–238.

Ja, S., B. V. Rao, S. Chandler. 2001. Passenger recapture estimation in airline revenue management. *AGIFORS 41st Annual Symposium,* AGIFORS, Sydney, Australia.

Kalyanam, K., S. Borle, P. Boatwright. 2007. Deconstructing each item's category contribution. *Marketing Sci.* **26**(3) 327–341.

Kök, G., M. Fisher. 2007. Demand estimation and assortment optimization under substitution: Methodology and application. *Oper. Res.* **55**(6) 1001–1021.

Lagarias, J., J. Reeds, M. Wright, P. Wright. 1998. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM J. Optim.* **9**(1) 112–147.

McLachlan, G., T. Krishnan. 1996. *The EM Algorithm and Extensions.* John Wiley & Sons, New York.

Musalem, A., M. Olivares, E. Bradlow, C. Terwiesch, D. Corsten. 2010. Structural estimation of the effect of out-of-stocks. *Management Sci.* **56**(7) 1180–1197.

Netessine, S., N. Rudi. 2003. Centralized and competitive inventory models with demand substitution. *Oper. Res.* **51**(2) 329–335.

Nevo, A. 2001. Measuring market power in the ready-to-eat cereal industry. *Econometrica* **69**(2) 307–342.

Nicholson, W. 2004. *Microeconomic Theory: Basic Principles and Extensions*, 9th ed. South-Western College Publishing, Mason, OH.

Queenan, C., M. Ferguson, J. Higbie, R. Kapoor. 2007. A comparison of unconstraining methods to improve revenue management systems. *Production Oper. Management* **16**(6) 729–746.

Ratliff, R., B. Rao, C. Narayan, K. Yellepeddi. 2008. A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. *J. Revenue Pricing Management* **7**(2) 153–171.

Swait, J., T. Erdem. 2002. Sales promotions consistency and consumer preferences. *J. Marketing Res.* **34**(3) 304–320.

Talluri, K. T., G. J. van Ryzin. 2004. Revenue management under a general discrete choice model of consumer behavior. *Management Sci.* **50**(1) 15–33.

Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press, New York.

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

Vulcano, G., G. van Ryzin, W. Chaar. 2010. Choice-based revenue management: An empirical study of estimation and optimization. *Manufacturing Service Oper. Management* **12**(3) 371–392.

Weatherford, L., S. Pölt. 2002. Better unconstraining of airline demand data in revenue management systems for improved forecast accuracy and greater revenues. *J. Revenue Pricing Management* **1**(3) 234–254.

Wu, C. F. 1983. On the convergence properties of the EM algorithm. *Ann. Statist.* **11**(1) 95–103.