COLUMBIA UNIVERSITY
Center for Pricing and Revenue Management

WORKING PAPER SERIES: NO. 2009-2

Estimating primary demand for substitutable products from sales transaction data

Gustavo Vulcano,
New York University

Garrett J. van Ryzin,
Columbia University

Richard Ratliff,
Sabre Holdings

2009

http://www.cprm.columbia.edu

Estimating primary demand for substitutable products from sales transaction data

Gustavo Vulcano * Garrett van Ryzin † Richard Ratliff ‡

Original version: July 2008. Revisions: May and December 2009.

Abstract

We consider a method for estimating substitute and lost demand when only sales and product availability data are observable, not all products are available in all periods (e.g., due to stock-outs or availability controls imposed by the seller), and the seller knows its market share. The model combines a multinomial logit (MNL) choice model with a non-homogeneous 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 were available in all periods. We then apply the expectation-maximization (EM) method to this model, and 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, which provably converges to a stationary point of the incomplete data log-likelihood function. Every iteration of the algorithm consists of simple, closed-form calculations. We illustrate the procedure on simulated data and two industry data sets.

Key words: Demand estimation, demand untruncation, choice behavior, multinomial logit model, EM method.

*Leonard N. Stern School of Business, New York University, New York, NY 10012, gvulcano@stern.nyu.edu.
†Graduate School of Business, Columbia University, New York, NY 10027, gjv1@columbia.edu.
‡Sabre Holdings, Southlake, Texas 76092, Richard.Ratliff@sabre-holdings.com.

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 may be negatively biased; this underestimation can be severe if products are unavailable for long periods of time. Additionally, if products are unavailable, *stockout-based substitution* will lead to increased sales in substitute products which 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. A heuristic to correct for spilled demand, implemented in many airline revenue management systems, is to assume that the percentage of demand turned away is proportional to the degree of "closedness" of a product (an itinerary-fare-class combination).[1] But this can lead to a "double counting" problem, whereby 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. [17]), while Gruen et al. [15] report recapture rates of 45% across eight categories at retailers worldwide.

While 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 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 [25]) 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 do so using only readily-available data, which in most retail settings

[1] For instance, if a booking class is open during 10 days of a month, and 20 bookings are observed, then this heuristic approach will estimate a demand of $20 \times 30/10 = 60$ for this booking class.

consists of sales transactions, product attributes (brand, size, price, etc.) and on-hand inventory quantities by SKU. Our work helps address this problem.

A convenient and 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 [3] and Train [29]). 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 these models 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 product sales history. 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 which provably converges to a stationary point of the associated incomplete data log-likelihood function. 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, practical 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 [28, Section 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. [31] provide empirical evidence of the potential of that approach. Ratliff et al. [26] 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 Anderson [1]. A similar approach was presented before by Ja et al. [17].

Another related stream of research is the estimation of demand and substitution effects for assortment planning in retailing. Kök and Fisher [19] 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 [30] 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 [23]).

Other papers in the operations and marketing science literature also address the problem of estimating substitution behavior and lost sales. Anupindi et al. [2] 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 stock-outs 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 [27] 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 may 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. [8] 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 stock-out situations, and formulate an iterative likelihood estimation algorithm. They then suggest a heuristic two-stage tracking procedure to identify stock-outs: 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. [5] 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 [10] 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. [18] 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 allow for flexible substitution patterns (beyond MNL assumptions), but stock-outs 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 [6] propose a model that accounts for varying levels of product availability. It uses information on aggregate availability to simulate the potential assortments that consumers may 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 stock-outs leads to substantial biases in estimation.

More recently, Musalem et al. [22] also investigate substitution effects induced by stock-outs. 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 since it combines MCMC with sampling using Bayesian methods.

The aforementioned paper by Kök and Fisher [19] 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. [8], and is quite standard in the marketing literature; see e.g. Bucklin and Gupta [7], and Chintagunta [9]. 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 visit the store but did not purchase anything is negligible (see Kök and Fisher [19, Section 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 non-linear optimization techniques in its M-step.

Closest to our work is that of Conlon and Mortimer [11], 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 stock-outs. 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, since it requires estimating an exponential number of parameters (see Conlon and Mortimer [11, 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 requires minimal computational 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 may be a day and we might have data on purchases over T (non-necessarily consecutive) days, or it may be a week and we 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 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 $\boldsymbol{z}_t = (z_{1t}, \ldots, z_{nt})$. We will assume that $z_{it} \geq 0$ 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 all 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 may be that some of the n products are not available in certain periods due to temporary stock-outs, 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, though 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 $\boldsymbol{v} \in \mathcal{R}^n$, $\boldsymbol{v} > 0$, that indicates the customer "preference weights" or "attractiveness" for the different products.[2] This vector, together with a normalized, no-purchase preference weight $v_0 = 1$, determine a customer's choice probabilities as follows: let $P_j(S, \boldsymbol{v})$ denote the probability that a customer chooses product $j \in S$ when S is offered and preference weights are given by vector $\boldsymbol{v}$. Then,

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

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

We denote the no-purchase probability by $P_0(S, \boldsymbol{v})$. 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*):

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

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

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

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:

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

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 $\lambda_t \sum_{i \in S_t} P_i(S_t, \boldsymbol{v})$, viz

$$\mathbb{P}(m_t \text{ customers buy in period } t | \boldsymbol{v}, \boldsymbol{\lambda}) = \frac{\left[\lambda_t \sum_{i \in S_t} P_i(S_t, \boldsymbol{v})\right]^{m_t} e^{-\lambda_t \sum_{i \in S_t} P_i(S_t, \boldsymbol{v})}}{m_t!}. \tag{3}$$

One could take the log of (2) and attempt to maximize this log-likelihood function with respect to $\boldsymbol{v}$ and $\boldsymbol{\lambda}$. However, it is clear that this is a complex likelihood function without much

[2] A further generalization of this MNL model to the case where the preference weights are functions of the product attributes is provided in Section 6.2.

structure, so maximizing it (or its logarithm) directly is not an appealing approach. Indeed, our attempts in this regard were not promising (see Section 5).

3.3 Multiple optima in the MLE and market potential

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

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

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

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

for all j and t. Since this product of the arrival rate and purchase probability is unchanged, by inspection of the form of (2) and (3), the solution $(\boldsymbol{v}^0, \boldsymbol{\lambda}^0)$ has the same likelihood and therefore is also a maximum. Since this holds true for any $\alpha > 0$, there are a continuum of maxima.[3]

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

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

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 Section 3.5 we look at a simpler and more efficient approach based on viewing the problem in terms of *primary* – or first-choice – demand.

3.4 Discussion of the 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 $\boldsymbol{v}$). 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. But the MNL has significant restrictions in terms of modeling choice behavior, most importantly the property of independence from irrelevant alternatives (IIA). Briefly, the property is that that the ratio of purchase probabilities for two alternatives is constant regardless of the consideration set containing them. Other choice models are more flexible in modeling substitution patterns (e.g., see Train [29, Chapter 4]). Among

[3]Of course, this observation holds more generally: For any pair of values $(\boldsymbol{v}, \boldsymbol{\lambda})$, there is a continuum of values $\alpha(\boldsymbol{v}, \boldsymbol{\lambda}), \alpha > 0$, such that $\mathcal{L}_I(\boldsymbol{v}, \boldsymbol{\lambda}) = \mathcal{L}_I(\alpha \boldsymbol{v}, \alpha \boldsymbol{\lambda})$.

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 [16], marketing researchers have found that the MNL model works quite 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.[4] 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 which 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 7am departure would be recaptured onto a same day, open 7pm departure in accordance with the IIA principle. Hence, it makes sense to restrict the choice 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 consideration in 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) competitor's 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, and $\boldsymbol{\lambda}$ refers to the volume of purchases in the market, including the retailer under consideration 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

[4]For example, Sabre has been implementing the single-segment MNL model for a large domestic US airline for more than two years, and has been observing very significant revenue improvements.

and empirical industrial organizations applications (e.g., see Berry et al. [4] for an empirical study of the U.S. automobile industry, and Nevo [24] for an empirical study of the ready-to-eat cereal industry). Henceforth, given an indicator s (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 higher participation in the market.

Also, note we work with store-level data (as opposed to household panel data). Chintagunta and Dubé [10] discuss the advantages of using store-level data to compute the mean utility associated with products.

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 (though that 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 [12] 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".

Further, if a period t has no sales for any of the products, then that period can be dropped from the analysis – or equivalently, it can be established ex-ante that all primary demands X_{jt} and substitute demands Y_{jt}, $j = 1, \ldots, n$, as well as the arrival rate λ_t will be zero. This is because our model assumes that the market participation s is replicated in every single period, and hence no sale in a period is a signal of no arrival in that period.

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. However, in contrast to other approaches (e.g., Musalem et al. [22]), 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 i, we mean the demand that would have occurred for product i if all n alternatives were available. The (random) number of purchases, Z_{it}, of product i in period t may be greater than the primary demand because it could include purchases from customers whose first choice was not available and bought product i as a substitute (i.e., Z_{it} includes demand that is spilled from other unavailable products and recaptured by product i). More precisely, the purchase quantity Z_{it} can be split into two components: the primary demand, X_{it}, which is the number of customers in period t that have product i as their first choice; and Y_{it}, the *substitute demand*, which is the number of customers in period t that decide to buy product i as a substitute because their first choice is unavailable. Thus:

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

We focus on estimating the primary demand X_{it}. While this decomposition seems to introduce more complexity in the estimation problem, we show below that in fact leads to a considerably simpler estimation algorithm.

3.5.1 Basic identities

Let $\hat{X}_{jt} = \mathrm{E}[X_{jt}|\boldsymbol{z}_t]$ and $\hat{Y}_{jt} = \mathrm{E}[Y_{jt}|\boldsymbol{z}_t]$ denote, respectively, the conditional expectation of the primary and substitute demand given the purchase observations $\boldsymbol{z}_t$. We seek to determine these quantities.

Consider first products that are unavailable in period t, that is $j \notin S_t \bigcup \{0\}$. For these items, we have no observation z_{jt}. To determine $\hat{X}_{jt}$ for these items, note that

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

and

$$\sum_{h\in S_t} \mathrm{E}\left[Z_{ht}|\boldsymbol{z}_t\right] = \frac{\sum_{h\in S_t} v_h}{\sum_{h\in S_t} v_h + 1}\mathrm{E}[A_t|\boldsymbol{z}_t].$$

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

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

or equivalently,

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

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

$$\begin{aligned}\mathbb{P}\{\text{Product } j \text{ is a first choice}|\text{Purchase } j\} &= \frac{\mathbb{P}\{\text{Product } j \text{ is a first choice}\}}{\mathbb{P}\{\text{Purchase } j\}} \\ &= \frac{v_j}{\sum_{i=1}^n v_i + 1} \Bigg/ \frac{v_j}{\sum_{h\in S_t} v_h + 1} \\ &= \frac{\sum_{h\in S_t} v_h + 1}{\sum_{i=1}^n v_i + 1}.\end{aligned}$$

Therefore, since $\hat{X}_{jt} = z_{jt}\mathbb{P}\{\text{Product } j \text{ is a first choice}|\text{Purchase } j\}$, we have

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

Lastly, 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., $\hat{X}_{0t} = \mathrm{E}[X_{0t}|\boldsymbol{z}_t]$. 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

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

In addition, the following identity should hold:

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

Conditioning on the observed purchases we have that

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

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, $\hat{Y}_{0t} = \mathrm{E}[Y_{0t}|\boldsymbol{z}_t]$:

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

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, $\hat{N}_j = \mathrm{E}[N_j|\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T]$, which is positive because $\hat{X}_{jt} \geq 0$, for all j and t, and for at least one period t, $\hat{X}_{jt} > 0$.[5]

[5]This is due to our assumption that $\boldsymbol{v} > 0$, and that for at least one period t, $z_{jt} > 0$, for each $j = 1, \ldots, n$.

3.5.2 Overview of our approach

The key idea behind our approach is to view the problem of estimating $\boldsymbol{v}$ 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 quite simple, namely

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

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, since we don't observe N_j, $j = 0, 1, \ldots, n$, directly, we use the EM method of Dempster et al. [13] 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 $\boldsymbol{v}$ 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 $\hat{\boldsymbol{v}} > 0$ (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 $\hat{\boldsymbol{v}}$ (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 $\boldsymbol{v}$:

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

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 $\hat{\boldsymbol{v}}$, 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 $\hat{\boldsymbol{v}}$ implies an estimate of r.

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

$$\begin{aligned}\mathcal{L}(\boldsymbol{v}) &:= \mathrm{E}[L(v_1,\ldots,v_n)|\hat{v}_1,\ldots,\hat{v}_n] \\ &= \sum_{j=1}^{n}\hat{N}_j\left\{\ln\left(\frac{v_j}{\sum_{i=1}^{n}v_i+1}\right)+r\ln\left(\frac{1}{\sum_{i=1}^{n}v_i+1}\right)\right\} \\ &= \sum_{j=1}^{n}\hat{N}_j\ln\left(\frac{v_j}{\sum_{i=1}^{n}v_i+1}\right)+r\ln\left(\frac{1}{\sum_{i=1}^{n}v_i+1}\right)\sum_{j=1}^{n}\hat{N}_j. \qquad (13)\end{aligned}$$

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

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

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 only involve 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

$$\hat{\lambda}_t \equiv \mathrm{E}[A_t|\boldsymbol{z}_t] = \hat{X}_{0t} + \sum_{i=1}^{n}\hat{X}_{it}. \qquad (14)$$

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 period 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:

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

Repeat

[E-step]:

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

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

If $j \notin S_t$, then set

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

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

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

EndIf

EndFor

Set

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

EndFor

[M-step]:

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 is 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 of our experiments we observed very quick convergence, so it would appear that the exact stopping criteria is 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.,

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

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

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

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:

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

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

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

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

$$v_j^* = \frac{\hat{N}_j}{r\sum_{i=1}^{n}\hat{N}_i}, j = 1,\ldots,n. \tag{15}$$

The second cross partial derivatives are:

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

where

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

and

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

Let H be the Hessian of $\mathcal{L}(v_1,\ldots,v_n)$. In order to check that our critical point (15) is a local maximum, we compute for $\boldsymbol{x} \in \mathcal{R}^n, \boldsymbol{x} \neq 0$,

$$\boldsymbol{x}^\mathsf{T} H(v_1,\ldots,v_n)\,\boldsymbol{x} = \frac{(1+r)\left(\sum_{i=1}^n \hat{N}_i\right)\left(\sum_{i=1}^n x_i\right)^2}{\left(\sum_{i=1}^n v_i + 1\right)^2} - \sum_{i=1}^n \hat{N}_i \frac{x_i^2}{v_i^2}. \tag{16}$$

The second order sufficient conditions are $\boldsymbol{x}^\mathsf{T} H(v_1^*,\ldots,v_n^*)\,\boldsymbol{x} < 0$, for all $\boldsymbol{x} \neq 0$. Plugging in the expressions in (15), we get

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

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

$$\left(\sum_{i=1}^n x_i\right)^2 - \left(\sum_{i=1}^n \hat{N}_i\right)\sum_{i=1}^n \frac{x_i^2}{\hat{N}_i} \leq 0,\ \forall \boldsymbol{x} \neq 0. \tag{17}$$

By the Cauchy-Schwartz inequality, i.e., $|\boldsymbol{y}^\mathsf{T}\boldsymbol{z}|^2 \leq ||\boldsymbol{y}||^2\,||\boldsymbol{z}||^2$, defining $y_i = \frac{x_i}{\sqrt{\hat{N}_i}}$ and $z_i = \sqrt{\hat{N}_i}$, we get:

$$\begin{aligned}
\left(\sum_{i=1}^n x_i\right)^2 &= \left(\sum_{i=1}^n \frac{x_i}{\sqrt{\hat{N}_i}} \times \sqrt{\hat{N}_i}\right)^2 \\
&\leq \left(\sqrt{\sum_{i=1}^n \frac{x_i^2}{\hat{N}_i}}\right)^2 \left(\sqrt{\sum_{i=1}^n \hat{N}_i}\right)^2 \\
&= \left(\sum_{i=1}^n \frac{x_i^2}{\hat{N}_i}\right)\left(\sum_{i=1}^n \hat{N}_i\right),
\end{aligned}$$

and therefore inequality (17) holds.

Proceeding from first principles, we have a unique critical point for $\mathcal{L}(v_1,\ldots,v_n)$ 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

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

Hence, the function is unimodal. ∎

A few comments are in order. First, observe that equation (16) also shows that the function $\mathcal{L}(v_1,\ldots,v_n)$ is not jointly concave in general, since 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 $\boldsymbol{x}$, $\boldsymbol{x}^\mathsf{T} H(v_1,\ldots,v_n)\boldsymbol{x} > 0$.[6] Figure 1 illustrates one example where $\mathcal{L}(v_1,\ldots,v_n)$ is not concave. Second, due to the definition of v_j^*,

[6]For example, if we take $n = 2$, $\boldsymbol{v} = (1.5, 1.2)$, $\hat{N}_1 = 50, \hat{N}_2 = 3$, and $\boldsymbol{x} = (0.01, 1)$, then $r = 1/(v_1 + v_2) = 0.37$, and $\boldsymbol{x}^\mathsf{T} H(v_1, v_2)\boldsymbol{x} = 3.33$.

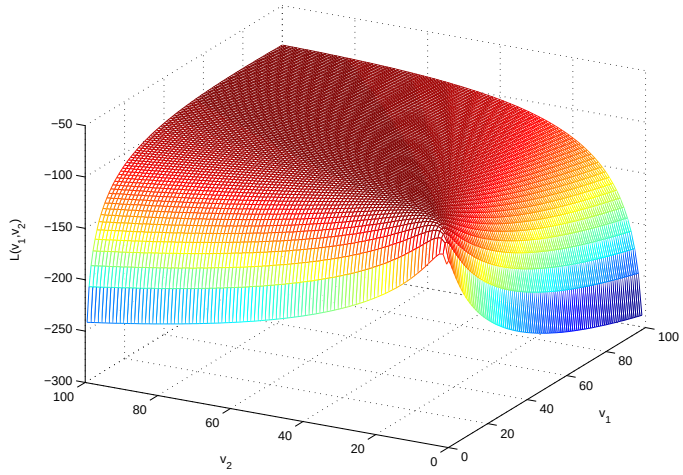

Figure 1: Example of log-likelihood function $\mathcal{L}(v_1, v_2)$, for $r = 0.02$, $\hat{N}_1 = 60$, and $\hat{N}_2 = 50$. The maximizer is $\boldsymbol{v}^* = (27.27, 22.73)$.

and since $\sum_{t=1}^{T} z_{jt} > 0$, then $\hat{N}_j > 0$ for every iteration of the EM method. Third, Theorem 1 proves that the M-step of the EM procedure is always well defined, and gives a unique global maximizer. This means it is indeed an EM algorithm, as opposed to the so-called Generalized EM algorithm (GEM). In the case of GEM, the M-step only requires that we generate an improved set of estimates over the current ones[7], and the conditions for convergence are more stringent (e.g., see McLachlan and Krishnan [21, Chapter 3] for further discussion.)

The following result shows that the EM method in our case satisfies a regularity condition that guarantees convergence of the log-likelihood values of the incomplete-data function to a stationary point:

Theorem 2 (Adapted from Wu [32]) [8] *The conditional expected value* $\mathrm{E}[L(v_1, \ldots, v_n)|\hat{v}_1, \ldots, \hat{v}_n]$ *in (13) is continuous both in* $\boldsymbol{v} > 0$ *and* $\hat{\boldsymbol{v}} > 0$*, and hence all the limit points of any instance* $\{\hat{\boldsymbol{v}}^{(k)}, \hat{\boldsymbol{\lambda}}^{(k)},\ k = 1, 2, \ldots\}$ *of the EM algorithm are stationary points of the corresponding incomplete-data log-likelihood function* $\mathcal{L}_I(\boldsymbol{v}, \boldsymbol{\lambda})$*, and* $\mathcal{L}_I(\hat{\boldsymbol{v}}^{(k)}, \hat{\boldsymbol{\lambda}}^{(k)})$ *converges monotonically to a value* $\mathcal{L}_I(\boldsymbol{v}^*, \boldsymbol{\lambda}^*)$*, for some stationary point* $(\boldsymbol{v}^*, \boldsymbol{\lambda}^*)$*.*

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 $\hat{X}_{jt}$ is continuous in $\hat{\boldsymbol{v}}$ according to equations (6), (7), and (10). Clearly, $\mathcal{L}$ is also continuous in $\boldsymbol{v}$. In addition, recall that the estimates $\hat{\boldsymbol{v}}$ imply a vector $\hat{\boldsymbol{\lambda}}$ once we fix a market participation r (through equation (14), and therefore the EM algorithm indeed generates an implied sequence $\{\hat{\boldsymbol{v}}^{(k)}, \hat{\boldsymbol{\lambda}}^{(k)},\ k = 1, 2, \ldots\}$. ∎

[7] In particular, the M-step of GEM just requires to find a vector $\bar{\boldsymbol{v}}$ such that $\mathrm{E}[L(\bar{\boldsymbol{v}})|\hat{\boldsymbol{v}}]] \geq \mathrm{E}[L(\hat{\boldsymbol{v}})|\hat{\boldsymbol{v}}]$. See McLachlan and Krishnan [21, Section 1.5.5].

[8] See also McLachlan and Krishnan [21, Theorem 3.2].

5 Numerical Examples

We next report on two sets of numerical examples. The first set is based on simulated data, which was used to get a sense of how well the procedure identifies a known demand system and how much data is necessary to get good estimates. Then, we provide two other examples based on real-world data sets, one for airlines and another for retail. In all the examples, we set a stopping criteria based on the difference between the matrices $\hat{X}$ from two consecutive iterations of the EM method, and stopped 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[9] procedural language, in which the algorithm detailed in Section 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. We provide the original generated data (observed purchases) and the final data (primary and substitute demands). Next, we look at the effect of data volume and quality on the accuracy of the estimates.

5.1.1 Preliminary estimation case

Given a known underlying MNL choice model (i.e., values for the preference weights $\boldsymbol{v}$) and assuming that the 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 $\boldsymbol{v} = (1, 0.7, 0.4, 0.2, 0.05)$ (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 retail data, showing the randomly generated purchases for each of the five products for each period and total number of no-purchases and arrivals. Here, period 1 represents the end of the selling horizon. Note 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 Section 3.5.3, i.e.,

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

We also assume perfect knowledge of the 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 values. Note that the results suggest an apparent bias in

[9] MATLAB is a trademark of The MathWorks, Inc. We used version 6.5.1 for Microsoft Windows.

Table 1: Purchases and no-purchases for the Preliminary Example

Observable data

Purchases	Periods															Total
	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	
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
Non observable 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 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

the estimates, which is not unexpected since 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 [21, 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 0.01 significance level.[10] 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 towards the end of the horizon (i.e., towards the right of the table) and the most preferred products become less available, the shifted demand tends to explain an increasing fraction of the sales and no-purchases.

From this information, we can also compute another important performance measure in

[10]The quasi-t statistic is computed as the ratio between the estimated value of the parameter and the ASE. Recall that for a two-tailed test, the critical values of this statistic are ± 1.65, ± 1.96, and ± 2.58 for the 0.10, 0.05, and 0.01 significance levels, respectively.

retail operations: the percentage of lost sales, defined as

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

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.,

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

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

Prod.	Periods															Total
	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	(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

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

Product	Periods															Total
	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	
1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	0.0	0.0	0.0	0.0	5.7	4.5	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	10.2
3	0.0	0.0	0.0	0.0	1.1	1.4	7.1	3.6	5.6	0.0	0.0	0.0	0.0	0.0	0.0	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	0.0	0.0	0.0	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

Figure 2 illustrates the aggregate first-choice demands (expected and calculated), sales and a *naïve* estimator of the primary demand. The expected first-choice demand is derived by

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

and the calculated first-choice demand $\hat{N}_j$ is the outcome of the EM method (i.e., last column of Table 3). The *naïve* estimator of the primary demand is computed by amplifying the observed

sales proportionally to the number of periods where the product was available.[11]

The figure shows an inverse relationship between sales and first-choice demand; while sales are determined by availability, the first-choice demand uncovered by our procedure is quite different for the least available but most preferred products (i.e., products 1 and 2). The expected number of no-purchases from the underlying true MNL model is remarkably close to the number estimated by our method: 223.9 vs. 219.0. A χ^2-test between the primary demand estimated by the EM algorithm and the true expected primary demand from the MNL gives a p-value=0.71, providing very strong justification not to reject the null hypothesis that the underlying choice demand model indeed follows the discrete distribution estimated via the EM procedure.

We also see that our first-choice demand estimates more closely track the true underlying primary demand relative to the *naïve* estimate. In particular, the error of the *naïve* estimates seems to be more significant for the less popular products.

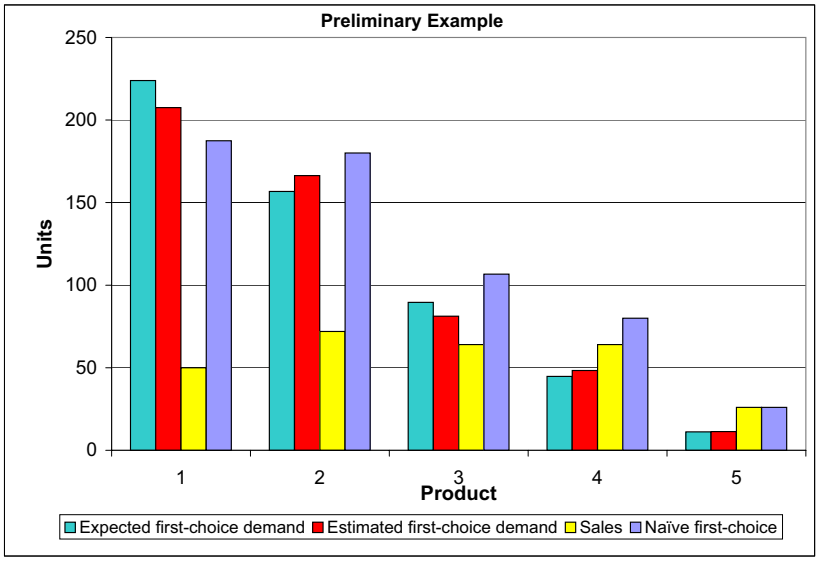

Figure 2: First-choice demand (expected and calculated), sales and *naïve* estimates for the Preliminary Example, disaggregated per product.

Figure 3 shows the total purchases per period, across the 5 products under consideration. Note that due to the high market potential (and related high recapture rate), even for the periods with less available products (i.e., periods 1-3), some purchases for the least preferred products still occur. The expected aggregate demand per period for products $j = 1, \ldots, 5$, in

[11] For instance, based on Table 1, product 1 shows 50 sales in 4 out of 15 periods, so this ad-hoc estimator gives $50 \times 15/4 = 187.5$. This is a standard, single-class untruncation method used by airlines on booking curves under the independent demand paradigm.

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, the average of the cumulative first-choice demand per period estimated by the EM algorithm is close: 34.31.

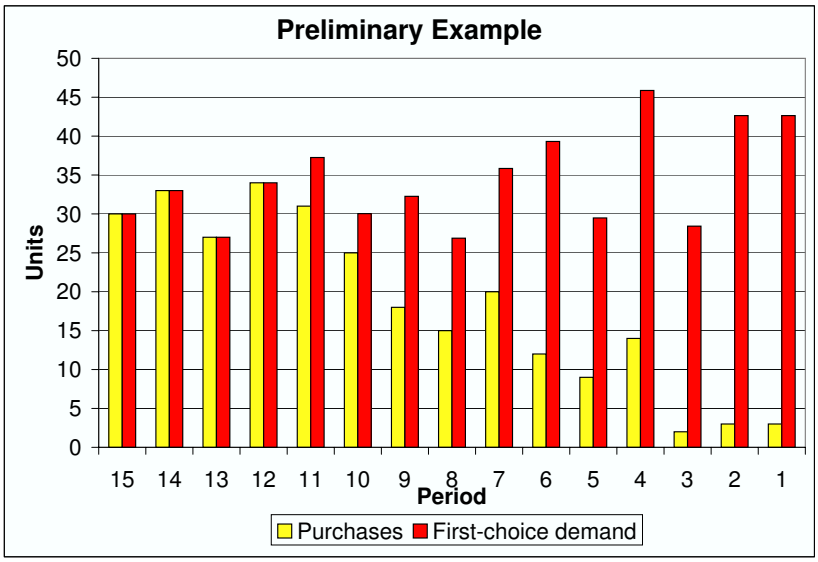

Figure 3: Observed purchases and estimated first-choice demand for the Preliminary Example, disaggregated per period.

In this case, it took 23 iterations of the EM method to meet the stopping criteria for the current example in just a fraction of a second of computational time. As a benchmark, we also optimized the incomplete data log-likelihood function (i.e., the logarithm of function (2)). We used the built-in MATLAB function "fminsearch" that implements the simplex search method of Lagarias et al. [20]. This is a direct search method that does not use numerical or analytic gradients. The initial point $(\boldsymbol{v}, \boldsymbol{\lambda})$ was based on the observed bookings as in the EM method. The tolerance was also set at 0.001. For this example, the MATLAB algorithm took 10,672 iterations to converge, requiring 13,382 evaluations of the log-likelihood function and several minutes of computational time. It converged to a point of the same level set of $\log \mathcal{L}_I(\boldsymbol{v}, \boldsymbol{\lambda})$ as the one obtained by our EM method. However, the order of magnitude difference in computation time between the two methods, especially considering the small size of the problem, is noteworthy.

5.1.2 Effects of data volume and quality

In this section, we report on the performance of our procedure under different volumes and quality of the input data. As in the previous example, given a known 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 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 4 shows the box plot of the biases of the estimates $\hat{\boldsymbol{v}}$ 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 $\hat{\boldsymbol{v}}$ (generally within [-2%, 2%]), and 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 seem to be enough to get most biases smaller than 10%. At the same time, as we increase the market potential (and hence, more purchases per period are observed), we note a trend towards more accuracy with fewer periods.

One potential concern of our procedure is the need to get an exogenous estimate of market share and the impact this estimate has on the quality of the estimation. To test this sensitivity, we used the same inputs for generating data as above (i.e., $\lambda = 50$, $n = 10$, products available with probability 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 $\hat{\boldsymbol{v}}$ and the average $\hat{\lambda}$ (Figure 5, left), and of the estimates of the primary demand $\hat{N}_j, j = 1, \ldots, n$, and the average $\hat{\lambda}$ (Figure 5, right). Note that a perturbation of the market potential generally amplifies the biases of the estimated parameters $\hat{\boldsymbol{v}}$ 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 Section 6.1 for further discussion), yet Figure 5 (right) shows a relatively small bias in the resulting primary demand estimates.

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. We present it to illustrate the practical feasibility of our approach and to show the impact of the choice set design on the estimation outcome.

We analyzed bookings data for the last seven selling days prior to departure for each consecutive Monday from January to March of 2004 (eleven departure days total). There were eleven

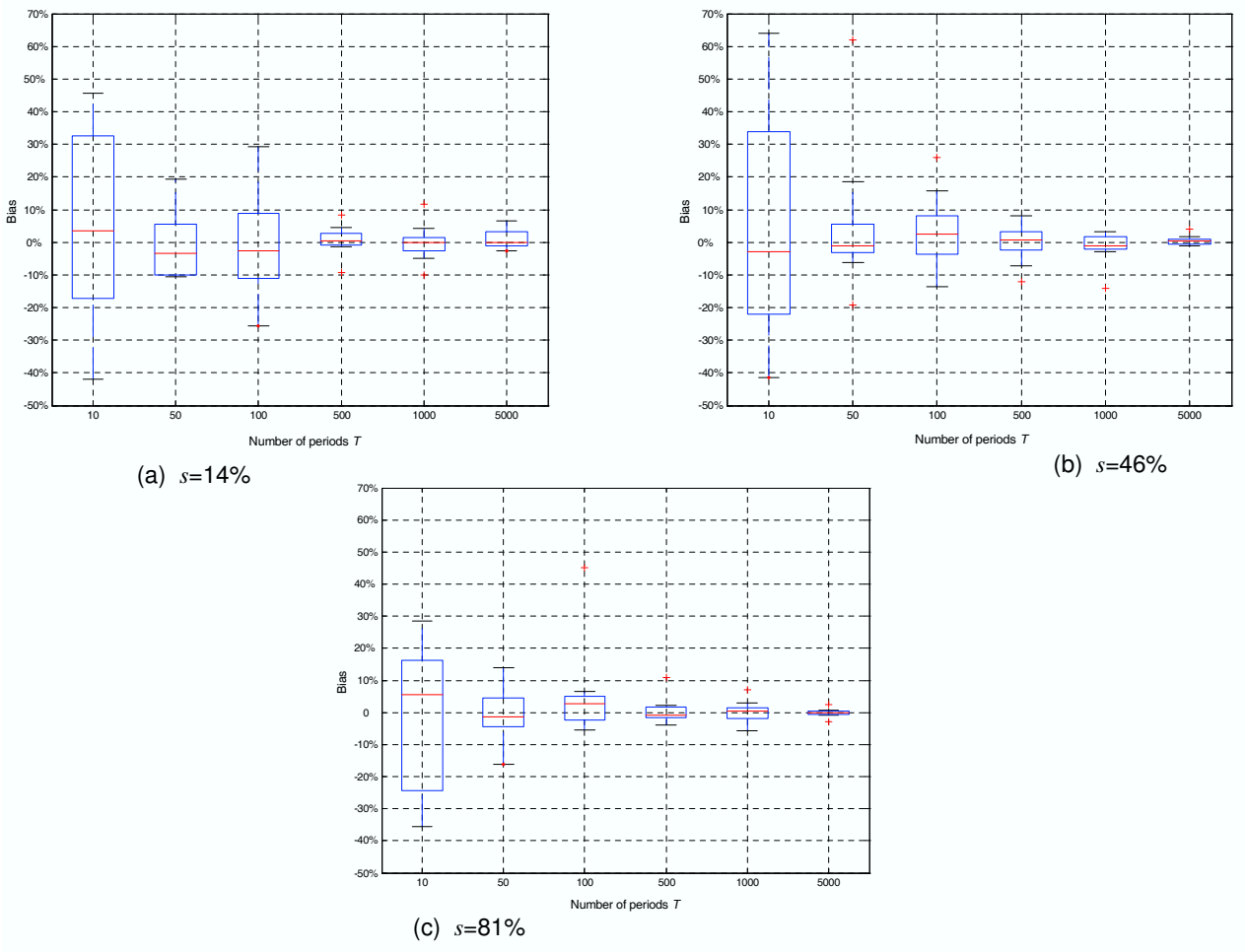

Figure 4: Biases of the preference weights $\hat{\boldsymbol{v}}$ under different market potentials: (a) $s = 14\%$, (b) $s = 46\%$, and (c) $s = 81\%$, for different selling horizon lengths.

classes per flight, and each class has a different fare value, which were constant during the eleven departure days under consideration. The market share of the airline for this particular O-D pair was known to be approximately 50%. As explained in Section 3.3, we use this information as a first order approximation for the market potential s.

We define a *product* as a flight-number-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 non-zero bookings for 15 out of the 22 products, so we focused our analysis on these 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 estimates for the demand volume, under different assumptions: In the

multi-flight case we assumed customers chose between both flights in the day, so the choice 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 choice sets, one for each of the flights with 7 and 8 products respectively, and with a market share of 25% per flight. It took 31 iterations of the EM method to compute the multi-flight estimates, and 24 and 176 iterations for each of the independent flights. In all cases, the total computational time was only a few seconds.

Figure 6 shows the observed and predicted bookings, and EM-based and *naïve* estimates of the first-choice demand for the 15 products under consideration, for the multi-flight case. The predicted bookings are computed based on the EM estimates $\hat{\boldsymbol{v}}$ and $\hat{\boldsymbol{\lambda}}$, and on the availability information of the different products:

$$\text{E[Bookings of product } j \text{ on day } t] = \lambda_t \frac{v_j \mathbf{I}\{j \in S_t\}}{\sum_{i \in S_t} v_i + 1}. \tag{18}$$

The label in the horizontal axis represent the fares of the corresponding products (e.g., "F1, \$189" means "Flight 1, bucket with fare \$189"). Figure 7 shows similar statistics for the independent flight case. While the total number of bookings observed is 337, the total estimated volume for the first-choice demand of the 15 products is 613 for the multi-flight case and 1,597 for the independent-flight case. As one might expect, the most significant mismatches occur at the low fares, where price sensitive customers could either buy-up, buy from a competitor, or not buy at all. Note that the procedure leads to significantly different volume estimates for some of the alternatives (e.g., product "F2, \$279" has a primary demand of 128 units for the multi-flight case and 735 units for the independent-flight case). Comparing the two cases, the multi-flight case has more degrees of freedom in fitting the product demands because it includes relative attractiveness across more options. The significant differences between the two approaches suggests that the definition of choice set can have a profound impact on the demand volume estimates. Hence, how best to construct these sets is an important area of future research (e.g., see Fitzsimons [14] for an analysis of the impact of choice set design on stockouts).

Figure 8 shows the preference weights $\hat{\boldsymbol{v}} = (\hat{v}_1, \ldots, \hat{v}_{15})$ computed for both the multi-flight and the independent flight cases via the EM and the *naïve* methods. Even though there are differences between the preference weights derived under both choice set definitions, these differences are not as extreme as the difference in the first-choice demand volume estimates.

Using the estimates $\hat{Y}_{0t}$ and $\hat{N}_j$, we then computed the fraction of lost sales. For the multi-flight case the estimate was 42.4%, and for the two-independent flight case, the estimates are 33.1% and 86.1% for each flight, respectively. In addition, for the multi-flight case, in Figure 9 we plot the fraction of lost sales as a fraction of the number of unavailable products. This fraction is almost linearly increasing in the number of stocked-out products.[12]

[12]This pattern is different from the one observed by Musalem et al. [22] for a mass consumer good (shampoo in their case), for which the increase is exponential.

We conducted two in-sample tests to check the goodness-of-fit of our estimation procedure. [13] First, we ran our estimation procedure to generate $\hat{\boldsymbol{v}}$ and $\hat{\boldsymbol{\lambda}}$. Then, using information on the availability of each product across the 77 periods, we estimated the number of bookings per period. We aggregated these expected bookings at the weekly level, producing a 15×11 grid of estimated bookings. The resulting root mean square error (RMSE) for the multi-flight case was 5.10. For both independent flight cases, the RMSEs were 5.82 and 2.90, respectively.

Next, we aggregated the observed bookings and the expected bookings across all the 77 periods and ran a χ^2-test for the multi-flight case (giving a p-value=0.72) and for the independent flight cases (with both p-values=0.99), giving strong justification not to reject the null hypothesis that the observed bookings follow the distribution estimated via the EM procedure.[14] The accuracy measures obtained suggest that considering both flights separately seems to better explain the choice behavior of customers.

Table 5 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.

Table 5: Estimation results for the Airline Market Example

Parameter	Product	Multi-flight demand			Independent-flight demand		
		Coefficient	ASE	t-statistic	Coefficient	ASE	t-statistic
v_1	F1, \$189	0.0832	0.0121	6.8760	0.0695	0.0100	6.9500
v_1	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

Finally, we again tried to optimize the incomplete data log-likelihood function for this example using the MATLAB built-in function "fminsearch", but the attempt failed. For example, for the multi-flight case, the MATLAB function was running for several minutes, taking 15,241

[13] We also tried out-of-sample tests, but the amount of data was very limited and highly volatile to have representative enough fit and testing periods.

[14] We needed to do this global aggregation to guarantee a number of expected bookings greater or equal than 5 for all the products.

iterations and 17,980 evaluations of the function $\log \mathcal{L}_I(\boldsymbol{v}, \boldsymbol{\lambda})$, but it converged to a meaningless point involving negative arrival rates. While one could attempt to stabilize this procedure and come up with better starting points, the experience attests to the simplicity, efficiency and robustness of our method relative to brute-force MLE.

5.2.2 Retail Market Example

This next example illustrates our methodology applied to sales data from a retail chain. We consider sales observed during eight weeks over a sample selling season. We assume a unique choice set defined by 6 substitutable products within the same small subcategory of SKUs. The market share of this retail location is estimated to be 48%. In this example, it took 120 iterations of the EM method to reach convergence; again, the computation time was a few seconds at most.

Figure 10 shows the observed sales, the predicted sales based on the EM estimates, the EM-based primary demand, and the *naïve* primary demand for the 6 products under consideration. While the total number of observed sales was 93, the total estimated volume of first-choice demand for these products was 303. The observed and estimated sales seem to be quite close. In fact, a χ^2-test between the observed and estimated sales volume gives a p-value=0.97, providing very strong evidence not to reject the null hypothesis that the observed bookings follow the distribution estimated via the EM method. For this example, we observe a major discrepancy between the primary demands computed via EM and those computed via the *naïve* approach, much more so than in the airline market example.

Figure 11 shows the preference weights $\hat{\boldsymbol{v}} = (\hat{v}_1, \ldots, \hat{v}_6)$ for the retail example. Again, we also report the *naïve* weights. We also observe a major discrepancy between both sets of preference weights. Table 6 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.

Table 6: Estimation results for the Retail Market Example

Parameter	Product	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

For this example, the percentage of lost sales is very significant:

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

Finally, when applying the MATLAB built-in function "fminsearch" to optimize the log-likelihood function for this example, the attempt failed once again. The MATLAB function ran for several minutes, taking 5,304 iterations and 6,809 evaluations of the function $\log \mathcal{L}_I(\boldsymbol{v}, \boldsymbol{\lambda})$, but it converged to a point with several negative arrival rates.

6 Implementation issues and extensions

6.1 Model inputs

While the overall procedure as stated above is quite simple and efficient, there are several practical issues that warrant further discussion. One issue we observed is that the estimates are sensitive to how choice 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, our experience is that we get abnormally high recapture rates into the least preferred products.

Lastly, there is 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 may not be available in every period.

Nevertheless, the following procedure avoids estimating an "all open"-based s: Recall from Section 3.3 that given MLE estimates $\boldsymbol{v}^*$ and $\boldsymbol{\lambda}^*$, we can scale this estimate by an arbitrary constant $\alpha > 0$ to obtain a new MLE of the form

$$\begin{aligned} \boldsymbol{v}(\alpha) &= \alpha \boldsymbol{v}^* \\ \boldsymbol{\lambda}(\alpha) &= \frac{\alpha \sum_{i \in S_t} v_i^* + 1}{\alpha \left(\sum_{i \in S_t} v_i^* + 1 \right)} \boldsymbol{\lambda}^*. \end{aligned}$$

The family of MLE estimates $\boldsymbol{v}(\alpha), \boldsymbol{\lambda}(\alpha)$ all lead to the same expected primary demand for the 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 (18)) that matches the total observed market share. This is a simple one-dimensional, closed-form search since the family of MLE's $\boldsymbol{v}(\alpha), \boldsymbol{\lambda}(\alpha)$ is a closed-form function of α.

6.2 Linear-in-parameters utility

In our basic setting, we focus on estimating a vector of preference weights $\boldsymbol{v}$. 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 $u_j = \boldsymbol{\beta}^T \boldsymbol{x}_j$ is the mean utility of alternative j, $\boldsymbol{x}_j$ 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 could try the following heuristic approach. The vector of coefficients $\boldsymbol{\beta}$ could be computed in a two-phase approach: In step 1, we run the EM algorithm as described here to estimate $\hat{\boldsymbol{v}}$. In step 2 we can look for a vector $\boldsymbol{\beta}$ that best matches these values using the fact that $\hat{v}_j = e^{u_j} = \boldsymbol{\beta}^T \boldsymbol{x}_j$, $j = 1, \ldots, n$. In most cases, this will be an over-determined system of equations, in which case we could a run least squares regression to fit $\boldsymbol{\beta}$.

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. In this paper, we propose a methodology for estimating demand when the seller knows her market share, only sales transaction data and product availability data are available, and the assortment changes from period to period.

Our approach combines a multinomial logit (MNL) demand model with a non-homogeneous Poisson model of arrivals over multiple periods. 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. Our key idea is to view the problem in terms of primary demand, and to treat the observed sales as incomplete observations of primary demand. We then apply the expectation-maximization (EM) method to this incomplete demand model, and show that this leads to a very simple, highly efficient iterative procedure for estimating the parameters of the model which provably converges to a stationary point of the incomplete data log-likelihood function. We provide numerical examples that illustrate the applicability of the approach on two industry data sets.

The methodology is very computationally efficient and simple to implement. In our experience, when applied over large volumes of data, our algorithm runs an order of magnitude faster than directly optimizing the corresponding incomplete data log-likelihood function (assuming that a maximum is found for the latter, which in our experience is not always feasible with standard optimization routines). 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.

Acknowledgements

We would like to thank John Blankenbaker at Sabre Holdings for his careful review and constructive suggestions on earlier drafts of this work, in particular his important finding that showed 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 comments on our work. Finally, we thank Marcelo Olivares (Columbia University), the associate editor, and three anonymous referees for their constructive feedback.

References

[1] S.E. Andersson. Passenger choice analysis for seat capacity control: A pilot project in Scandinavian Airlines. *International Transactions in Operational Research*, 5:471–486, 1998.

[2] R. Anupindi, M. Dada, and S. Gupta. Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science*, 17:406–423, 1998.

[3] M. Ben-Akiva and S. Lerman. *Discrete Choice Analysis: Theory and Applications to Travel Demand*. The MIT Press, Cambridge, MA, sixth edition, 1994.

[4] S. Berry, J. Levinsohn, and A. Pakes. Automobile prices in market equilibrium. *Econometrica*, 63:841–890, 1995.

[5] S. Borle, P. Boatwright, J. Kadane, J. Nunes, and S. Galit. The effect of product assortment changes on customer retention. *Marketing Science*, 24:616–622, 2005.

[6] H. Bruno and N. Vilcassim. Structural demand estimation with varying product availability. *Marketing Science*, 27:1126–1131, 2008.

[7] R. Bucklin and S. Gupta. Brand choice, purchase incidence, and segmentation: An integrated modeling approach. *Journal of Marketing Research*, 29:201–215, 1992.

[8] K. Campo, E. Gijsbrechts, and P. Nisol. The impact of retailer stockouts on wether, how much, and what to buy. *International Journal of Research in Marketing*, 20:273–286, 2003.

[9] P. Chintagunta. Investigating purchase incidence, brand choice and purchase quantity decisions on households. *Marketing Science*, 12:184–208, 1993.

[10] P. Chintagunta and J-P Dubé. Estimating a SKU-level brand choice model that combines household panel data and store data. *Journal of Marketing Research*, 42:368–379, 2005.

[11] C. Conlon and J. Mortimer. Demand estimation under incomplete product availability. Working paper, Department of Economics, Harvard University, Cambridge, MA, 2009.

[12] N. DeHoratius and A. Raman. Inventory record inaccuracy: An empirical analysis. *Management Science*, 54:627–641, 2008.

[13] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39:1–38, 1977.

[14] G. Fitzsimons. Consumer response to stockouts. *Journal of Consumer Research*, 27:249–266, 2000.

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

[16] P. Guadagni and J. Little. A logit model of brand choice calibrated on scanner data. *Marketing Science*, 2:203–238, 1983.

[17] S. Ja, B. V. Rao, and S. Chandler. Passenger recapture estimation in airline revenue management. In *AGIFORS 41st Annual Symposium*, Sydney, Australia, August 2001. Presentation.

[18] K. Kalyanam, S. Borle, and P. Boatwright. Deconstructing each item's category contribution. *Marketing Science*, 26:327–341, 2007.

[19] G. Kök and M. Fisher. Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research*, 55:1001–1021, 2007.

[20] J. Lagarias, J. Reeds, M. Wright, and P. Wright. Convergence properties of the Nelder-Mead simplex method in low dimensions. *SIAM Journal of Optimization*, 9:112–147, 1998.

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

[22] A. Musalem, M. Olivares, E. Bradlow, C. Terwiesch, and D. Corsten. Structural estimation of the effect of out-of-stocks. Working paper, Fuqua School of Business, Duke University, Durham, NC, 2009.

[23] S. Netessine and N. Rudi. Centralized and competitive inventory models with demand substitution. *Operations Research*, 51:329–335, 2003.

[24] A. Nevo. Measuring market power in the ready-to-eat cereal industry. *Econometrica*, 69:307–342, 2001.

[25] W. Nicholson. *Microeconomic Theory: Basic Principles and Extensions*. The Dryden Press, 5th edition, 1992.

[26] R. Ratliff, B. Rao, C. Narayan, and K. Yellepeddi. A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. *Journal of Revenue and Pricing Management*, 7:153–171, 2008.

[27] J. Swait and T. Erdem. The effects of temporal consistency of sales promotions and availability on consumer choice behavior. *Journal of Marketing Research*, 39:304–320, 2002.

[28] K. T. Talluri and G. J. van Ryzin. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50:15–33, 2004.

[29] K. Train. *Discrete choice methods with simulation*. Cambridge University Press, New York, NY, 2003.

[30] G. van Ryzin and S. Mahajan. On the relationships between inventory costs and variety benefits in retail assortments. *Management Science*, 45:1496–1509, 1999.

[31] G. Vulcano, G. van Ryzin, and W. Chaar. Choice-based revenue management: An empirical study of estimation and optimization. Leonard N. Stern School of Business, New York University, New York, NY. Forthcoming in *M&SOM*, 2009.

[32] C.F. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11:95–103, 1983.

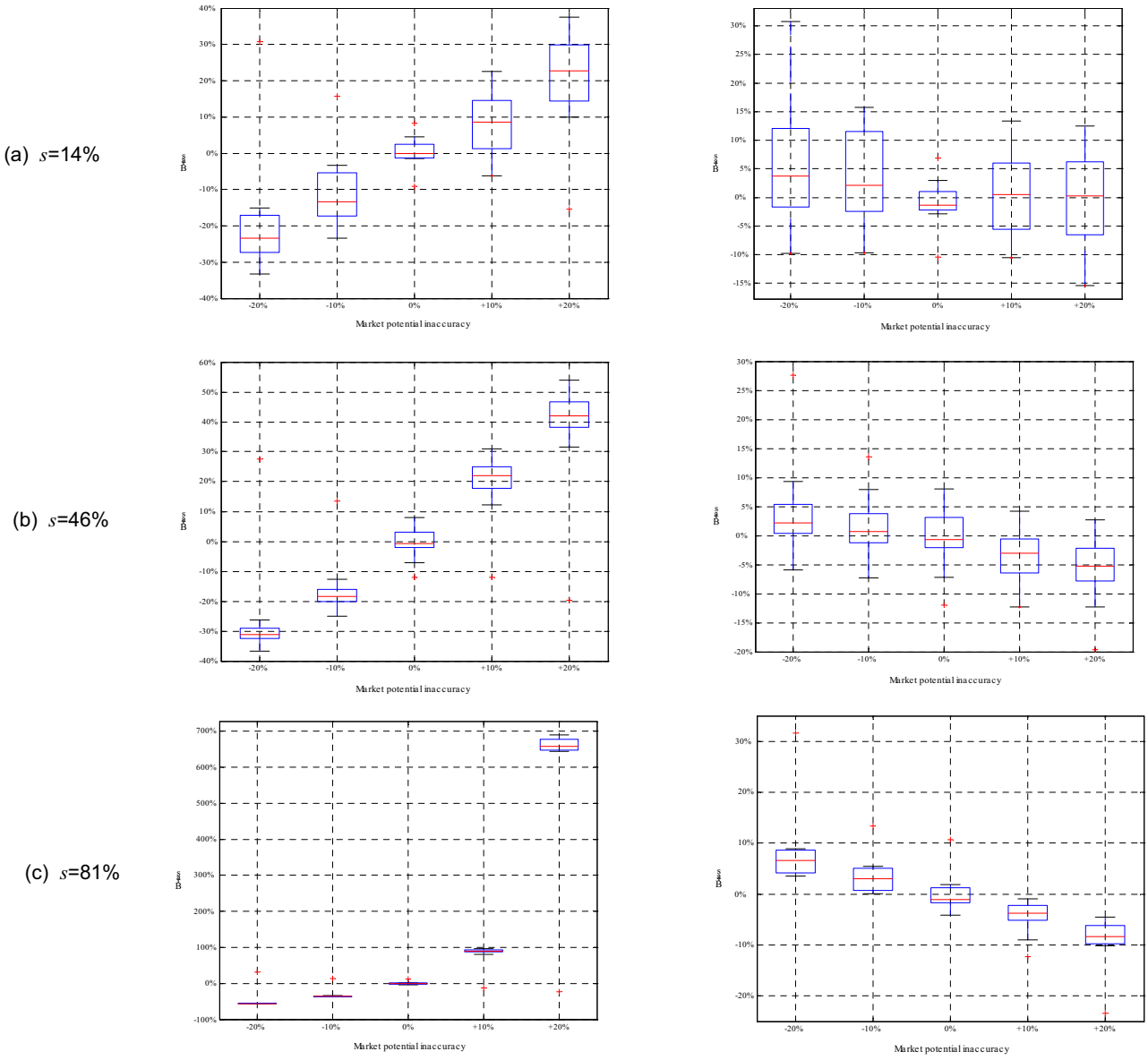

Figure 5: Biases of the estimates $\hat{\boldsymbol{v}}$ and the average $\hat{\lambda}$ (left), and of the estimates of the primary demand $\hat{N}_j$ and the average $\hat{\lambda}$ (right) under noisy market potentials. The raw data was generated based on the true market potentials: (a) $s = 14\%$, (b) $s = 46\%$, and (c) $s = 81\%$, and then the parameters were estimated assuming perturbed values: $\pm 1.2s$, and $\pm 1.1s$.

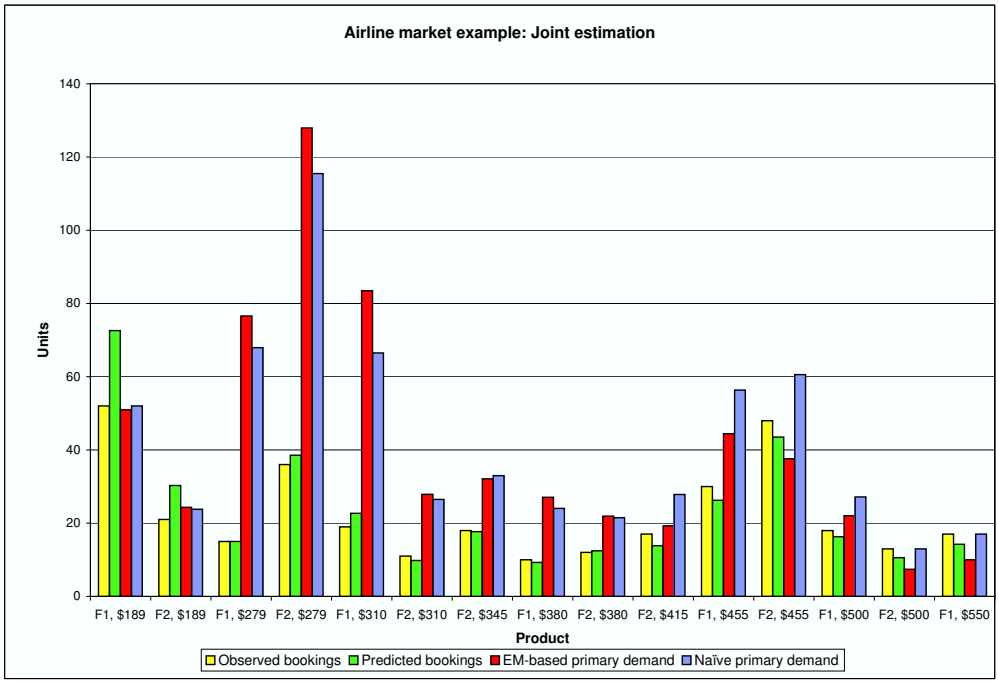

Figure 6: Comparison of observed and predicted bookings, EM-based and *naïve* first-choice demand under the multi-flight assumption for the Airline Market Example.

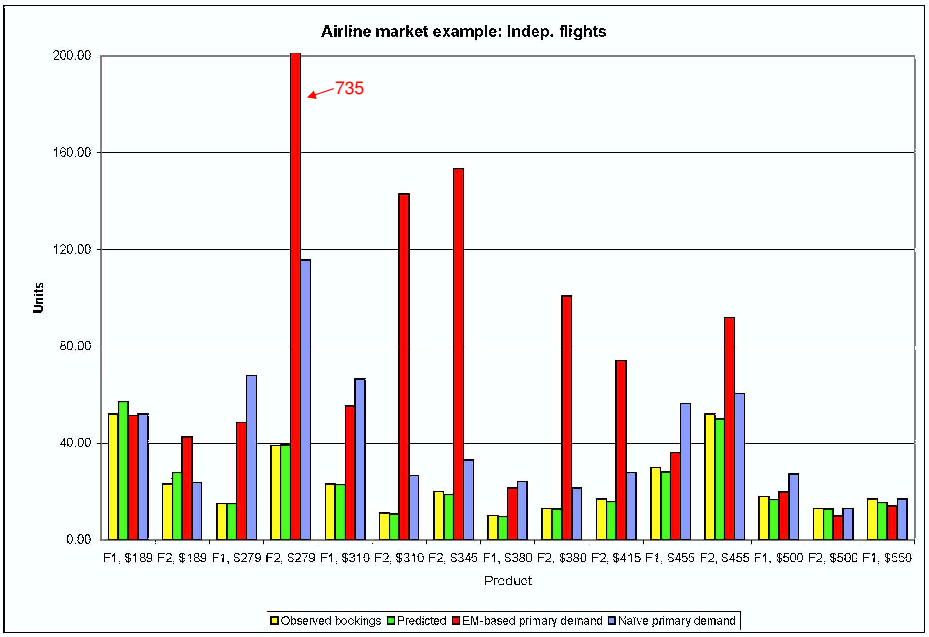

Figure 7: Comparison of observed and predicted bookings, EM-based and *naïve* first-choice demand under the independent flight assumption for the Airline Market Example.

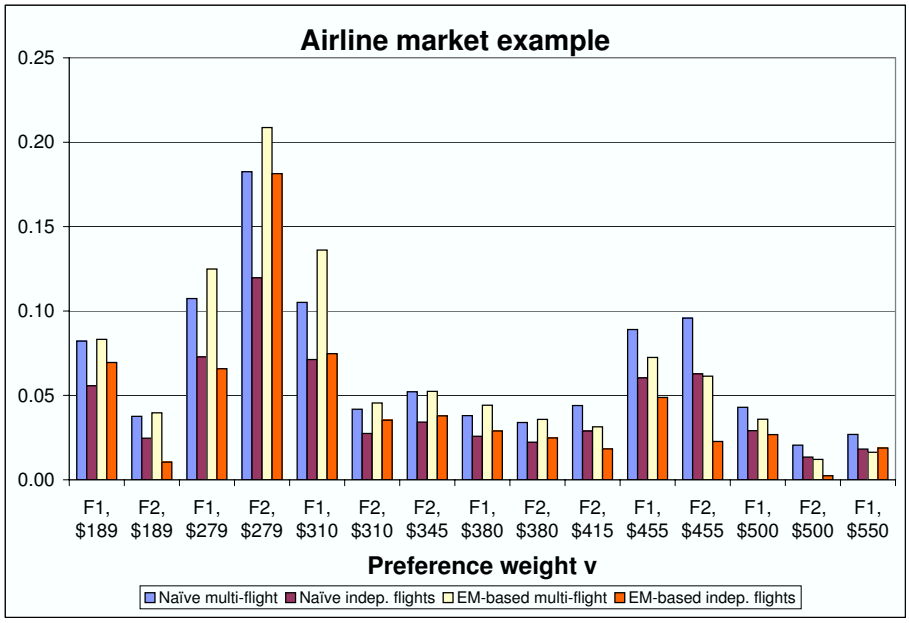

Figure 8: Comparison of EM-based and *naïve* preference weights $\hat{v}$ for multi-flight and independent flight demands for the Airline Market Example.

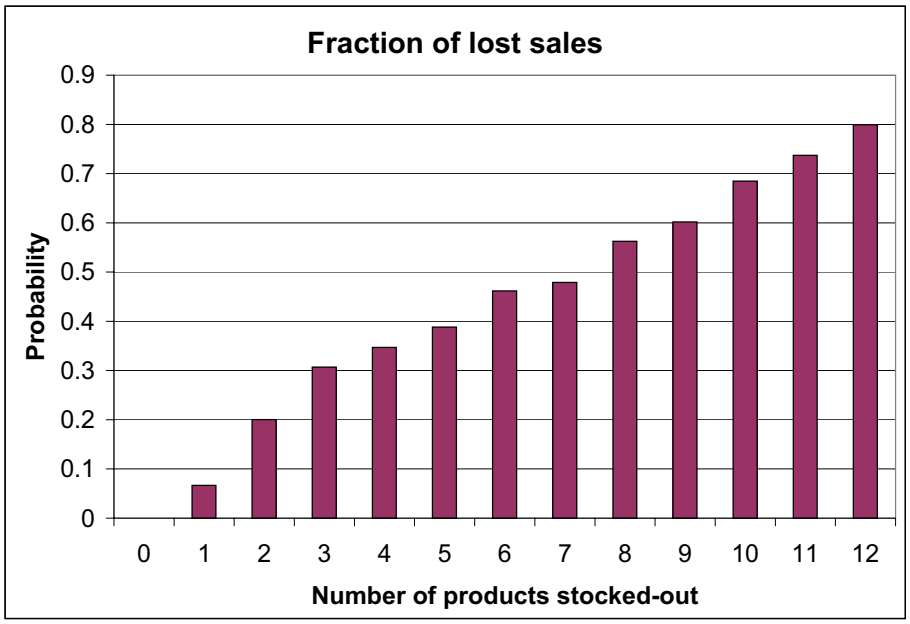

Figure 9: Fraction of lost sales as a fraction of total primary demand for own products, for the Airline Market Example.

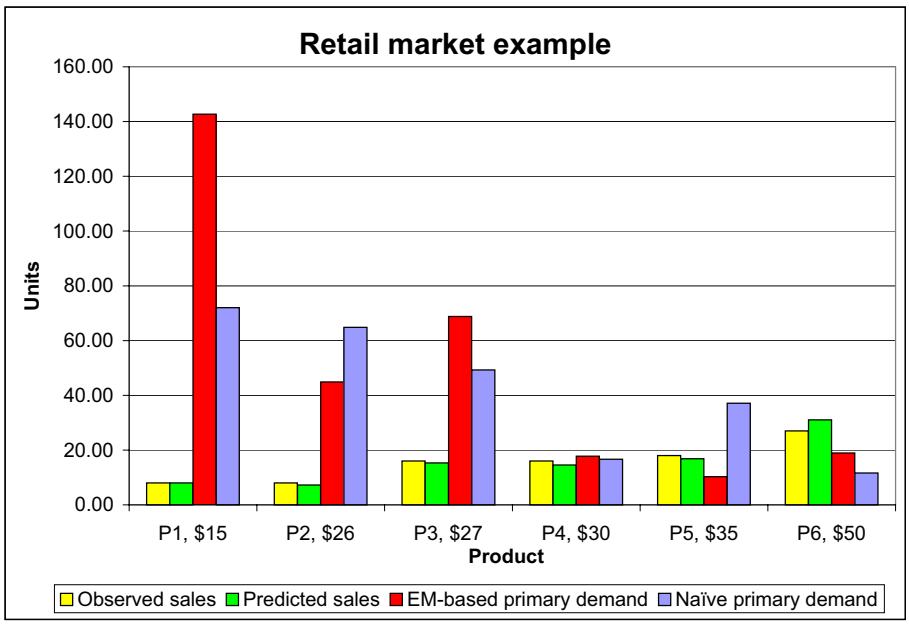

Figure 10: e Comparison of observed and predicted bookings, EM-based and *naïve* first-choice demand across 56 days for the Retail Market Example.

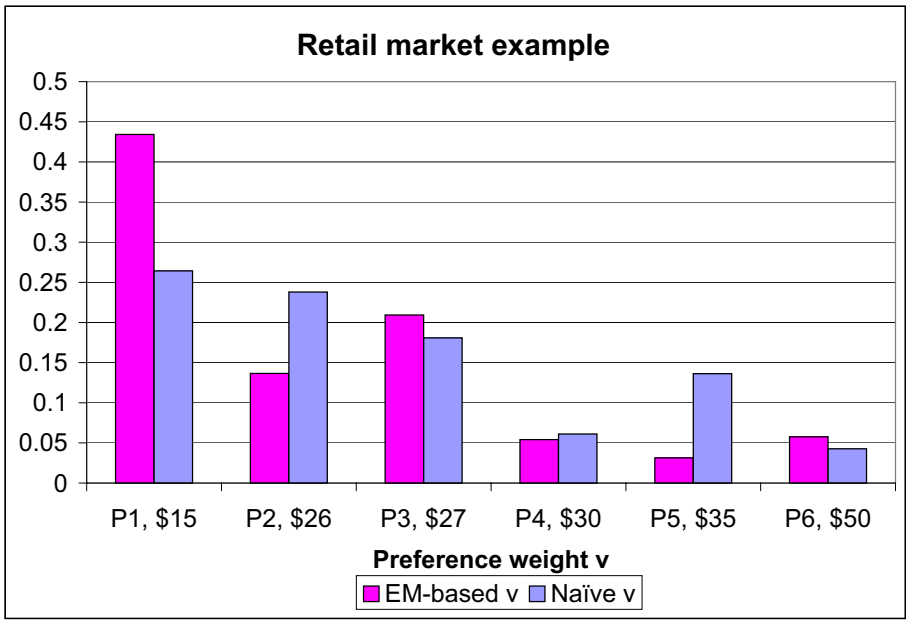

Figure 11: Comparison of preference weight $\hat{\boldsymbol{v}}$ computed via EM and the *naïve* approach for the Retail Market Example.