



WORKING PAPER

Working Paper Number (16-06)

October 10, 2016

Title: Estimating Demand for Substitutable Products when Inventory Records are Unreliable

Authors, Affiliations:

Daniel Steeneck

Department of Operational Sciences, Air Force Institute of Technology (corresponding author - steeneck@gmail.com)

Fredrik Eng-Larsson

Stockholm Business School, Stockholm University

Francisco Jauffred MIT Center for

Transportation and Logistics, Massachusetts Institute of Technology

MIT Global Scale Network
Working Paper Series

Estimating Demand for Substitutable Products when Inventory Records are Unreliable

Daniel Steeneck
Department of Operational Sciences, Air Force Institute of Technology

Fredrik Eng-Larsson
Stockholm Business School, Stockholm University

Francisco Jauffred
MIT Center for Transportation and Logistics, Massachusetts Institute of Technology

Abstract

We present a procedure for estimating demand for substitutable products when the inventory record is unreliable and only validated infrequently and irregularly. The procedure uses a structural model of demand and inventory progression, which is estimated using a modified version of the Expectation Maximization-method. The procedure leads to asymptotically unbiased estimates without any restrictive assumptions about substitution patterns or that inventory records are periodically known with certainty. The procedure converges quickly also for large product categories, which makes it suitable for implementation at retailers or manufacturers that need to run the analysis for hundreds of categories or stores at the same time. We use the procedure to highlight the importance of considering inventory reliability problems when estimating demand, first through simulation and then by applying the procedure to a data set from a major US retailer. The results show that for the product category in consideration, ignoring inventory reliability problems leads to demand estimates that on average underestimate demand by 5%. It also results in total lost sales estimates that account for only a fraction of actual lost sales.

Keywords: demand estimation; inventory uncertainty; choice behavior; multinomial logit model; EM method.

Acknowledgements:

Many thanks to our corporate partner who shared outstanding domain knowledge and insights with team. For confidentiality reasons, the name of the corporate partner is withheld.

Corresponding Author:

Name: Daniel Steeneck

Center/University: Department of Operational Sciences, Air Force Institute of Technology

Address: 2950 Hobson Way, Building 641, 250C, Wright Patterson Air Force Base, OH 45433

Email: steeneck@gmail.com

Estimating Demand for Substitutable Products when Inventory Records are Unreliable

Daniel Steeneck*, Fredrik Eng-Larsson†, Francisco Jauffred ‡

October 21, 2016

Abstract

We present a procedure for estimating demand for substitutable products when the inventory record is unreliable and only validated infrequently and irregularly. The procedure uses a structural model of demand and inventory progression, which is estimated using a modified version of the Expectation Maximization-method. The procedure leads to asymptotically unbiased estimates without any restrictive assumptions about substitution patterns or that inventory records are periodically known with certainty. The procedure converges quickly also for large product categories, which makes it suitable for implementation at retailers or manufacturers that need to run the analysis for hundreds of categories or stores at the same time. We use the procedure to highlight the importance of considering inventory reliability problems when estimating demand, first through simulation and then by applying the procedure to a data set from a major US retailer. The results show that for the product category in consideration, ignoring inventory reliability problems leads to demand estimates that on average underestimate demand by 5%. It also results in total lost sales estimates that account for only a fraction of actual lost sales.

*Department of Operational Sciences, Air Force Institute of Technology (corresponding author - steeneck@gmail.com)

†Stockholm Business School, Stockholm University

‡MIT Center for Transportation and Logistics, Massachusetts Institute of Technology

1 Introduction

Retail inventory records are notoriously unreliable. For instance, DeHoratius and Raman (2008) showed that across a sample of 37 stores at a major U.S. retailer, 65% of all inventory records were inaccurate. There are several reasons for this including theft, misplacement, cash register errors, inventory stuck in the backroom, and improper registration of returns. Even if the inventory records are accurate about in-store availability, shelf replenishment execution problems are estimated to result in between 8% and 40% of on-shelf stockouts (Gruen and Corsten, 2007).

Unreliable inventory records at a retailer create both acute and chronic problems in its supply chain. The *acute* problem is that the inventory management system may fail to recognize that a product's inventory level has reached its reorder point. This leads to late replenishment orders, which reduces the on-shelf availability in the store and leads lost sales. In the worst case scenario, a product is out of stock while still having an inventory record above its reorder point. In these cases, all potential demand is lost until the product's record is manually corrected and/or the product is manually ordered. The *chronic* problem is the effect these unknown inventory levels have on upstream planning. Production and distribution of products sold in a retail store are based on forecasts, and these forecasts are primarily based on sales data adjusted for stockouts. But when stockouts are not recorded in the inventory system, days with zero purchases of a specific product are easily mistaken for zero demand days, especially for slow moving items, which leads to biased forecasts and suboptimal inventory levels throughout the supply chain. The end result is either higher inventory holding costs than optimal, or lower service levels than optimal, depending on the direction of the bias.

In this work we develop a procedure that estimates daily demand for substitutable products when inventory records are unreliable and only validated infrequently and irregularly. The daily demand estimated may be analyzed using time-series techniques to be incorporated in forecasting models. As such the model is useful for establishing "base rate" demands, i.e., demand in the absence of promotional events. The research was motivated by work conducted in collaboration with a large consumer goods manufacturer. The manufacturer had noticed that service levels across its customers' stores were, in general, much lower than specified in the inventory management systems. Audits revealed that inventory records often failed to cap-

ture stockout events. In fact, over the categories we investigated, stockouts seemed to occur almost four times as often as indicated by the inventory records (80 percent of stockouts are unobserved!). Audits were, however, expensive. Large stores were audited every week, but for cost reasons only a very small subset of products could be audited each time. There were also limited incentives for store managers to restore the records of the stocked out items identified by the auditors, since it meant admitting that products had “disappeared” from the store. Auditors could sometimes restore the actual inventory levels in the store by, for instance, locating missing boxes within the store, but often the only option was to take a note that there was a discrepancy between a product’s inventory record and its on-shelf inventory. Having an appreciation of the size of the problem, but little possibility to directly fight its root causes, the manufacturer needed a way to integrate the unreliable records into the supply chain planning.

The main strength and novelty of our procedure is that it requires very little inventory validation data to work well; only sparse audit data is required to calibrate the model. The procedure leads to unbiased estimates without any restrictive assumptions about substitution patterns, or that inventory records are periodically known with certainty. The procedure converges quickly also for large product categories, which makes it practically suitable for manufacturers and retailers that need to perform the analysis for hundreds of product categories or stores at the same time.

We use our procedure to show the impact of ignoring inventory record reliability issues when estimating demand for substitutable products. Since unobserved stockouts is a known problem, manufacturers and retailers sometimes use a multiplier to adjust demand forecasts based on sales data, or simply remove all stockout days from the sales data. However, simple heuristics like these fail to account for substitution among products, which leads to biased demand estimates. Equally, if not more, problematic is when substitution *is* accounted for, but not the fact that inventory records are unreliable. This also leads to biased estimates, and it is many times difficult to know the extent or even the direction of this bias. In this work we compare our procedure to procedures that ignore substitution effects and/or inventory record reliability, and see that our procedure is both more accurate and less biased. We illustrate this point by applying our procedure to a data set covering half a year’s worth of sales of laundry care products in a large U.S. retail store over a random sample of 75 stores. We see that when considering inventory record reliability issues, demand is underestimated by 5 percent and lost sales due to unobserved stockouts are indeed a significant source of

lost sales, being on average 5 times observed lost sales.

The remainder of the paper is organized as follows. In the next section we review related literature. In Section 3 we describe our demand model and the estimation procedure, and prove some basic properties of the estimation procedure. In Section 4 we use the estimation procedure on simulated data to illustrate its capabilities as well as its performance compared to other state-of-the art procedures. Lastly, in Section 5 we apply the procedure to data from a U.S. retailer as an example of its applicability. Section 6 concludes the paper with a short discussion about possible future research directions.

2 Literature review

The problem of unreliable inventory records has been tackled in various ways in the inventory control literature. Iglehart and Morey (1972) present one of the earliest treatments of the problem. They model how to determine the frequency of inventory audits and adjust inventory policy under an unobserved inventory error process, with the objective of minimizing cost while maintaining a particular service level. Their model is later extended by K ok and Shang (2007) to account for backlogged demand. Kang and Gershwin (2005) suggest and evaluate various techniques of accounting for inventory uncertainty, including (i) decrementing inventory by average stock loss, (ii) increasing manual inspections which correct inventory records, and (iii) resetting the inventory record to zero automatically after a period of zero sales days. In a similar vein, DeHoratius et al. (2008) use a Bayesian update of the probability distribution of the inventory level following daily sales observations and develop a heuristic inventory management policy around this distribution. Chen (2014) considers only the stock-loss to be unobserved and develops an optimal seeking dynamic programming formulation to determine the inventory policy.

Rather than adjusting inventory policies, another vein of research considers continuous monitoring to trigger audits to fix inventory inaccuracy and improve store execution. Chuang et al. (2015) empirically investigates the effectiveness of monitoring retail store sales in real-time to detect strings of consecutive zero store-SKU sales days indicating unobserved stockout states (as suggested by e.g. Fisher and Raman, 2010). They present an experiment where external auditors check the on-shelf availability status of the products at the stores suspected of having unobserved stockouts of some products. If possible, the auditors also restore

the availability or adjust inventory records. They show that over a 12 week time-frame, the percentage of total audits reporting no stockouts, for a particular product category, are reduced from 8% to just under 2%.

Common to all of the above research is the focus on how inaccurate inventory records can be managed through the operations of the manufacturer and/or the retailer to avoid limited on-shelf availability and, as result, lost sales. However, the immediate impact of lost sales from low on-shelf availability is only one part of the problem. When inventory records are inaccurate, demand estimates based on sales also become inaccurate, since days with no on-shelf availability may be mistakenly treated as zero-sales days. This leads to inaccurate forecasts for the coming periods. With much of the upstream operations based on these forecasts, there is a serious risk that the errors permeate production planning at the manufacturer and store replenishment at the retailer, leading to suboptimal policies and more lost sales.

In this work we focus on the *estimation of demand* when inventory records are unreliable. Anupindi et al. (1998) is one of the first works that consider a similar problem. In their model they assume consumers have a first product choice and, if that product is not available, a fraction of customers chooses a second product. If the second product is not available, the demand is lost. They use their model to estimate the demand for products sold in vending machines, where on-shelf availability is registered periodically. More recent work follows a similar idea. Musalem et al. (2010) use a structural model with partial information on product availability. Conlon and Mortimer (2013) use a random utility model with partial information on product availability, and estimate demand using an Expectation-Maximization (EM) algorithm. Lee et al. (2015) consider a single period model with two products, and estimate the model through standard maximum likelihood techniques. Just like Anupindi et al. (1998), these works assume on-shelf availability is known periodically. However, our experience is that retail inventory records are rarely validated. Moreover, the validation does not follow a predetermined period length and varies greatly between products in the same category. In this research we collaborated with a large manufacturer, who on average audits each store-SKU less than two times per year. As a result, for the vast majority of the time, the manufacturer does not know the on-shelf availability of a product. Our approach overcomes this problem, and effectively estimates demand even for a large number of products with infrequently and irregularly validated inventory records.

The approach we propose involves modeling demand using a multinomial logit (MNL) choice model.

The MNL model has been widely used in the marketing and operations literature to estimate consumer demand and substitution effects (Conlon and Mortimer 2013; Vulcano et al. 2012; Kalyanam et al. 2007; Campo et al. 2003). The MNL model provides a convenient way to capture the effect of availability on both the underlying demand for the product and the spill-over demand to other products when a product stocks out. In particular, it allows us to remove restrictive assumptions about the number of products a consumer considers. Modeling choice in this way has been shown to work quite well when estimating demand for a category of substitutable products, and estimating the model is normally computationally convenient. Within this large body of research, the work by Vulcano et al. (2012) is of particular importance to our work. In their work, they construct an MNL model to capture the demand for each product, and then uses an EM algorithm to estimate the model. We will follow a similar process. However, their model uses an indicator of product availability to construct the choice set faced by each customer. This means they assume it is known with certainty whether or not a product is available at a certain point in time. As discussed above, this is not reasonable in many settings. In our work we extend this research to account for imperfect information about product availability by assuming that for any point in time, the only information on availability is a stockout probability.

A major challenge in incorporating stockout probabilities into an MNL model is that the potential number of choice sets grow exponentially in the number of choices. This combinatorial problem is similar to the problem observed by Anupindi et al. (1998), Musalem et al. (2010), and Conlon and Mortimer (2013). Anupindi et al. address this by limiting the “stages” of the choice model: a consumer is assumed to have two top choices and if none of those are available the demand is lost. Musalem et al. (2010) assume periodic knowledge of product availability, which allows them use Gibbs sampling. Since our setting does not allow for this assumption, we choose an approach similar to that suggested by Conlon and Mortimer (2013), i.e., instead of invoking restrictions in the behavioral assumptions, we develop a nested procedure that estimates demand by maximizing a simulated sample mean. This leads to an asymptotically unbiased estimator, that converges quickly also for large product categories.

3 Model and estimation

In the following section we present our model and the estimation procedure. To estimate the demand, we develop a consumer choice model based on a multinomial logit (MNL) formulation. Often, such models start by specifying an individual consumer’s utility from purchasing a number of substitutable products. However, for our purposes, individual utility is not of interest *per se*. Instead, following Vulcano et al. (2012), we let the choice probabilities in our model be determined by a nominal preference, v_j , for each product, j , and assume these preferences are homogeneous across consumers and time. The choice probabilities are then aggregated to find the expected underlying demand for all products. We develop a nested estimation procedure to estimate demand based on sales data from a category of products and their inventory record. The key problem is that inventory records are unreliable, which makes it difficult to estimate substitute demand. Our estimator handles this by sequentially computing stockout probabilities through a hidden Markov Chain model and estimating demand through a simulated sample mean maximization technique. We show this leads to asymptotically unbiased estimates of demand.

3.1 Demand model

In our model, a store carries n substitutable products over T time periods. While the time period t is one day in the context of our problem, the model and estimation procedure accommodates any period length. We denote the full set of substitutable products by $\mathcal{N} = \{1, 2, \dots, n\}$. This captures, for instance, a product category, or a group of similar products within the store. Since we focus on the manufacturer’s problem in the context of our work, competitors’ products sold in the same store are not part of the set \mathcal{N} .

Consumers arrive to the store according to a Poisson process with rate λ_t . We denote the number of arrivals from this process A_t . Note that this process is not observable in the data; nor are the actual choices made by individual consumers. From the data, we can only observe the aggregated purchases of product j in each period, z_{jt} , and the corresponding end of the day inventory record, \tilde{I}_{jt} . This is the record reported in the inventory management system and it may or may not match the actual available inventory on the shelf, which we denote I_{jt} . We will assume that for at least one period, $z_{jt} > 0$ for each product j , or we exclude product j from the analysis.

When a consumer arrives to the store, all products in \mathcal{N} may not be available. We denote the set of *available* products in period t as $S_t \subseteq \mathcal{N}$ and assume this set does not change during a period. Also available to the customer is the “outside option”, designated as product 0, which comprises the products of competing manufacturers and the no purchase option. We let $\mathbb{I}_j(S_t)$ be an indicator variable that is 0 if product $j \notin S_t$ and 1 otherwise, and let $\mathbb{I}(S_t) = (\mathbb{I}_1(S_t), \mathbb{I}_2(S_t), \dots, \mathbb{I}_n(S_t))$. Consumers choose from the set of available of products according to an MNL model, where the choice probabilities associated with each product is determined by a preference vector $\mathbf{v} = (v_1, v_2, \dots, v_n) > 0$. This vector attaches a preference weight, or a nominal utility, to each product, based on its perceived attractiveness versus similar products and the “outside option”. We let the normalized preference $v_0 = 1$ denote the attractiveness of the outside option. This, and all other preferences, are assumed to be constant across time and consumers. As noted by Vulcano et al. (2012), extending the model so that preferences vary over time is a straight-forward extension, which we will not elaborate upon here.

Using the MNL formulation, the probability that a consumer chooses product $j \in S_t$ is then given by

$$P_{jt}(S_t, \mathbf{v}) = \Pr\{\text{Product } j \text{ is first choice}\} = \frac{v_j}{\mathbb{I}(S_t)\mathbf{v}^T + 1}. \quad (1)$$

If $j \notin S_t$, then $P_{jt}(S_t, \mathbf{v}) = 0$. Similarly, the probability that a consumer chooses the outside option is

$$P_{0t}(S_t, \mathbf{v}) = \frac{1}{\mathbb{I}(S_t)\mathbf{v}^T + 1}. \quad (2)$$

3.2 Inventory model

While the set of available products, S_t , is observable to the consumer, it is not always observable in the data since \tilde{I}_{jt} does not necessarily match I_{jt} and if even these levels matched, it is still possible that the product is unavailable on the shelf. That is, just because we observe a positive inventory record, $\tilde{I}_{jt} > 0$, there is no guarantee there is available inventory.

For convenience, we drop the product index subscript. To handle the shelf-stock inventory uncertainty, we model the inventory progression for a given product as a Markov chain $\{M_t, t = 1, 2, \dots, T\}$, with transition probabilities $Q_{i,h}$ and initial state probabilities $\rho_i = \Pr\{M_1 = i\}$. For a given product, the chain has three

possible states in each period: either, the product is in an *in stock* state (state 1, $i = 1$), the product is in an *unobserved stockout* state (state 2, $i = 2$) or the product is in an *observed stockout* state (state 3, $i = 3$). We let the *stockout probability*, p_t , denote the probability that the product is in a stockout state (state 2 or state 3) in period t .

The true state is observable in a period in the data if (i) there are recorded purchases of the product in period t and thus it can be assumed that the product was in stock during (at least a portion of) period t or (ii) there are no sales and there is no recorded inventory at the end of the period and thus the product is in a stockout state. In all periods where the state is observable the stockout probability for the product is $p_t = \{0, 1\}$ and the set of available products is observable in the data. However, when $z_t = 0$ and $\tilde{I}_t > 0$, the true state is not observable. This could be the case for a single period or for several consecutive periods. We refer to any such block of periods as an “uncertainty block”. We consider each uncertainty block as a separate Markov chain $\{M^u, \tau = 1, 2, \dots, T^u\}$. That is, each such block starts at $\tau = 1$ corresponding to the actual period in which the uncertainty block begins, t^u . While the true state is observable in the period before the uncertainty block ($\tau = 0$, corresponding to $t^u - 1$) and in the period after the uncertainty block ($\tau = T^u + 1$, corresponding to $t^u + T^u + 1$), it is not observable in any of the periods within the uncertainty block. As a result, the inventory progression within an uncertainty block is a hidden Markov Chain (Ross, 2006) where, instead of observing the true chain state, we can only observe a signal, ς . We define this signal as the event, in period τ , that $z_\tau = 0$ and $\tilde{I}_\tau > 0$.

Following the hidden Markov Chain framework, the signal is emitted by the system each time a state is entered. However, the signal is emitted with different probabilities depending on the state of the system. Let the conditional probability of observing signal ς be $p(\varsigma|i)$, and denote a given sequence of signals $\varsigma_1, \varsigma_2, \dots, \varsigma_\tau$ by $\varsigma_k = (\varsigma_1, \varsigma_2, \dots, \varsigma_k)$, $k \leq T^u$. Now, to find the stockout probability for the product in a period k within an uncertainty block, we need to find the conditional probability of a product being in a stockout state given signals ς_k . We proceed according to Ross (2006) and let

$$F_\tau(i) = \Pr\{\varsigma_\tau, M_\tau = i\}. \quad (3)$$

Note that

$$\Pr\{M_\tau = i | \varsigma_\tau\} = \frac{\Pr\{\varsigma_\tau, M_\tau = i\}}{\Pr\{\varsigma_\tau\}} = \frac{F_\tau(i)}{\sum_h F_\tau(h)}. \quad (4)$$

From Ross (2006) , it is known that

$$F_\tau(i) = p(\varsigma_\tau | i) \sum_h F_{\tau-1}(h) Q_{h,i}, \quad (5)$$

which means we can find the conditional stockout probability from (4) recursively, starting from $F_1(i) = \rho_i p(\varsigma_1 | i)$. Note that the initial state probabilities for uncertainly blocks are trivially obtained from the transition probability matrix since, in period $\tau = 0$, we observe the true state of the system.

We can now introduce the *random* choice set faced by consumers, $S \in \mathbb{P}(\mathcal{N})$, where $\mathbb{P}(\mathcal{N})$ is the power set of \mathcal{N} . The actual choice set S_t (as observed by the consumer) is the period t realization of the random set S . Given the stockout probabilities, and reintroducing the product index subscript, the probability of observing a choice set S in period t is given by

$$G_t(S) = \Pr\{S | \mathbf{p}_t\} = \left(\prod_{f \in S} (1 - p_{ft}) \right) \cdot \prod_{g \notin S} p_{gt}, \quad (6)$$

where $\mathbf{p}_t = (p_{1t}, p_{2t}, \dots, p_{nt})$ is the vector of stockout probabilities in period t . For periods in uncertainty blocks for a product, j , p_{jt} is given by $F_t(2)$; otherwise p_{jt} is known from the inventory and sales records. Note that we assume stockout probabilities are independent across products.

3.3 Estimation

In our estimation procedure, we seek to estimate the parameters of our demand model: the preference vector \mathbf{v} and the rate of the arrival process, $\boldsymbol{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_T)$. If we could observe the actual demands for all products over T time periods, we could estimate the parameters by maximizing the log-likelihood function,

$$\mathcal{L}(\mathbf{v}) = \sum_{j=1}^n N_j \ln \left(\frac{v_j}{\sum_{l=1}^n v_l + 1} \right) + N_0 \ln \left(\frac{1}{\sum_{l=1}^n v_l + 1} \right), \quad (7)$$

where N_j is the total demand for product j over all T periods. By specifying the store’s share of total demand in a period *ex-ante*, one can derive both the demand for the outside option, N_0 , and the rate of the arrival process, λ , from the parameter estimates. The problem is that we cannot observe demand directly. We only observe purchase quantities, and these quantities may be higher or lower than the demand because of stockout-based substitution. Moreover, we do not know if a product has stocked out or not. As stated in previous sections, the context we investigate is plagued by incorrect inventory records. The records are validated through audits, but these audits are infrequent, irregular, and inconsistent between product, which makes it difficult to know if the observed purchase quantities include spill over demand or not.

OVERVIEW OF THE ESTIMATION PROCEDURE. To estimate the demand under these circumstances, we will view demand as “missing data” that can be estimated from the observed purchase quantities. In previous literature, the Expectation-Maximization (EM) algorithm by Dempster et al. (1977) is often used for this: based on some initial parameter estimates and a model that specifies the relationship between observed and missing data, the missing data is estimated through its expected value, given the observed data. The likelihood function is then maximized using this estimated missing data to find new parameter estimates, which are used to compute new conditional expected values. The steps are repeated until convergence. As shown by e.g. Wu (1983), the algorithm tends to converge and, when it converges, leads to asymptotically unbiased estimates. However, for the algorithm to work in its basic form, one must compute the conditional expected values of the missing data. As we will see, since in our context we do not know whether a product is in stock or not, computing the conditional expected value requires us to consider all possible choice sets in every time period. With n products in the choice set, this means we may need to consider 2^n choice sets in every calculation. This is clearly infeasible for any practical application.

To solve these problems, we create a nested procedure consisting of an outer and an inner loop. In the outer loop, we compute stockout probabilities for all products and time periods. In the inner loop, we estimate the missing data using these probabilities, and optimize the likelihood function based on these estimates. Due to the combinatorial complexity of the inner loop, we use a simulated sample mean instead of the expected value. Like the original EM-algorithm this leads to asymptotically unbiased estimates, but with significantly fewer computations. After maximizing the likelihood function we get new parameter estimates, which are then used to generate new sample means. The inner loop repeats until convergence. Once the

inner loop converges, the outer loop generates new stockout probabilities based on the estimated missing data, and the inner loop starts over again. This nested procedure repeats until convergence. Below we explain these steps in greater detail.

THE OUTER LOOP – COMPUTING STOCKOUT PROBABILITIES. In the outer loop we compute the stockout probability for each product j in each period t , p_{jt} . Note that this is the probability that the product is in a stockout state in period t . Based on our Markov model of the inventory progression, the stockout probability is trivially found when the state is observed. We assume the state is observed in two cases.

1. If there are observed purchases of product j in period t , then $p_{jt} = 0$. Since we are investigating products on a store-product level, most items are rather slow moving, and the time of the day of a purchase has little impact on the amount of lost sales.
2. If the inventory record shows a stockout of product j in period t while we observe no purchases, then $p_{jt} = 1$. This means we assume that there is no “hidden” inventory, that is, more inventory on the shelf than in the record. This is motivated by the fact that if the inventory record is zero, the record is in fact normally correct, since most inventory discrepancies are due to there being less inventory than recorded (DeHoratius and Raman, 2008).

If the state is not observable, the stockout probability $p_{jt} \in (0, 1)$ needs to be estimated. In these cases we only observe the signal, ς , which means there is some positive probability that there is an *unobserved* stockout of product j . This is the probability we seek.

To estimate that probability, we use Equation (4), which can be seen as a type of Bayesian updating. To compute the updated stockout probability in each period for a product j , we need the conditional probabilities of observing the signal, $p_j(\varsigma|i)$, for all states i , and the transition probabilities $Q_{i,h}^j$. The conditional probability for the in stock state, $p_j(\varsigma|1)$, is simply the probability of zero demand. For the Poisson arrivals process, this is given by $p_j(\varsigma|1) = \exp(-\lambda_{jt})$, where λ_{jt} is the demand rate for product j in period t . The conditional probability for the unobserved stockout state is $p_j(\varsigma|2) = 1$, since we must observe zero purchases in this case. The conditional probability for the observed stockout state is $p_j(\varsigma|3) = 0$, since the state is always observed. The transition probabilities $Q_{i,h}^j$ are prior probabilities that need to be specified before running the algorithm. As such, these probabilities can be guessed based on knowledge about the

system or, as we do in Section 5, be identified from the purchase data and the audit data.¹

THE INNER LOOP – SAMPLE MEAN MAXIMIZATION. To model the relationship between the missing data and the observed data, we follow Vulcano et al. (2012) and say the (stochastic) purchase quantities, Z_{jt} consist of two components,

$$Z_{jt} = X_{jt} + Y_{jt}, \quad (8)$$

where X_{jt} is the (underlying) demand for product j in period t , and Y_{jt} is the substitute demand for product j . The demand X_{jt} is the number of purchases that would be observed if all products were available in period t . It is thus the missing data that we construct our likelihood function around. The observed purchase quantities, Z_{jt} , may be more or less than this demand due to limited availability. If product j is not available, then $Y_{jt} = -X_{jt}$, since some of the demand for product j will spill over to other products or to the outside option. We refer to this as *missed demand*. If on the other hand one or more *other* product than product j is not available, then the substitute demand $Y_{jt} \geq 0$, since the demand for the other products may spill over to purchases of product j . We refer to this as *recaptured demand*.

In a standard EM algorithm, the missing data is estimated through the expected value conditional on the observed purchases, that is, $\hat{X}_{jt} = E[X_{jt}|z_t]$. These estimates depend on the choice set in period t . The fact that we cannot observe the choice set S_t in period t makes it computationally difficult to find these estimates, as we need to take the expected value across all possible choice sets. We therefore find these estimates by first computing the conditional expected value for any given set S_t , and then simulate a sample of choice sets, from which \hat{X}_{jt} is computed as the sample mean of the conditional expected values.

Consider a choice set S_t . Given a choice set there are two possible cases for product j : either it is part of the choice set or it is not.

First, consider the case where product j is not available in period t ; that is, $j \notin S_t$. We thus have no

¹Note that without any additional information, the stockout rate could be a possible estimate of the stockout probabilities, p_{jt} . The problem with this strategy is that any period with zero purchases and positive inventory record will have the same stockout probability, leading to overestimation in zero demand periods and underestimation in real stockout periods. While the aggregate demand estimate may still be reasonable, the aggregate estimate will depend heavily on the prior estimate of the stockout rate. Therefore, to improve this estimate, we use the hidden Markov Chain model to find the stockout probabilities.

observed purchase quantities for product j , so we let $z_{jt} = 0$. It follows that

$$E[X_{jt}|z_t, S_t, j \notin S_t] = \frac{\mathbb{I}(S_t)\mathbf{v}^T}{\sum_{l=1}^n v_l + 1} E[A_t|z_t]. \quad (9)$$

Note also that

$$\sum_{l \in S_t} E[Z_{lt}|z_t, S_t] = \frac{\mathbb{I}(S_t)\mathbf{v}^T}{\mathbb{I}(S_t)\mathbf{v}^T + 1} E[A_t|z_t]. \quad (10)$$

Combining these expressions yields

$$E[X_{jt}|z_t, S_t, j \notin S_t] = \frac{v_j}{\sum_{l=1}^n v_l + 1} \frac{\mathbb{I}(S_t)\mathbf{v}^T + 1}{\mathbb{I}(S_t)\mathbf{v}^T} \sum_{l \in S_t} E[Z_{lt}|z_t, S_t]. \quad (11)$$

Now, consider the the case when product j is available in period t . In this case we have that

$$\begin{aligned} E[X_{jt}|z_t, S_t, j \in S_t] &= \frac{\Pr\{\text{Product } j \text{ is first choice}\}}{\Pr\{\text{Product } j \text{ is purchased}\}} z_{jt} \\ &= \frac{v_j}{\sum_{l=1}^n v_l + 1} \frac{\mathbb{I}(S_t)\mathbf{v}^T + 1}{v_j} z_{jt} \\ &= \frac{\mathbb{I}(S_t)\mathbf{v}^T + 1}{\sum_{l=1}^n v_l + 1} z_{jt}. \end{aligned} \quad (12)$$

The expected demand unconditional on the choice set can be found by combining (11) and (12) along with the probability of each case,

$$\begin{aligned} \hat{X}_{jt} &= E[X_{jt}|z_t] = \\ &= \sum_{S \in \mathbb{P}(\mathcal{N})} \left((1 - \mathbb{I}_j(S)) \frac{v_j}{\sum_{l=1}^n v_l + 1} \frac{\mathbb{I}(S)\mathbf{v}^T + 1}{\mathbb{I}(S)\mathbf{v}^T} (\mathbb{I}(S)z_t^T) + \mathbb{I}_j(S) \cdot \frac{\mathbb{I}(S)\mathbf{v}^T + 1}{\sum_{l=1}^n v_l + 1} z_{jt} \right) G_t(S), \end{aligned} \quad (13)$$

where $G_t(S)$ is the probability that consumers are facing choice set $S \in \mathbb{P}(\mathcal{N})$ in period t . This probability is given by (6). When working with retail data covering product categories with tens of products, it is easy to see that maximizing a likelihood function that uses (13) in its formulation will lead to a maximization problem that is computationally prohibitive in many cases. For instance, in this research we work with data of a product category originally consisting of $n = 70$ products observed over a year. Using the conditional

expected value for this data leads to computation times of more than 100 CPU years for one store alone.

To circumvent the computational challenges we use the probability vector $G_t(S)$ to simulate a limited sample of choice sets, $S_{t,\eta} \in \mathbb{P}(\mathcal{N})$, in each period, and compute $E[X_{jt}|z_t, S_{t,\eta}]$ for each set. The basic idea is that many of the choice sets are very unlikely. For instance, if the stockout rate is 10% for all products, the probability that all products are stocked out in a period is only $10^{-(n+1)}$. Simulating the sets takes this into account, since the probabilities of generating a set is determined by the stockout probabilities.

The demand is then estimated as

$$\hat{X}_{jt} = \frac{1}{N_\eta} \sum_{\eta=1}^{N_\eta} E[X_{jt}|z_t, S_{t,\eta}], \quad (14)$$

where N_η is the sample size. The estimated demand for the outside option is given by

$$\hat{X}_{0t} = \frac{1}{\sum_{l=1}^n v_l + 1} \sum_{i=1}^n \hat{X}_{jt}, \quad (15)$$

while the estimated total demand for product j over the T periods is

$$\hat{N}_j = \sum_{t=1}^T \hat{X}_{jt}. \quad (16)$$

We can now construct the likelihood function using our estimated missing data. We have that

$$\mathcal{L}(\mathbf{v}|\hat{\mathbf{v}}) = \sum_{j=1}^n \hat{N}_j \ln \left(\frac{v_j}{\sum_{l=1}^n v_l + 1} \right) + \hat{N}_0 \ln \left(\frac{1}{\sum_{l=1}^n v_l + 1} \right), \quad (17)$$

where $\hat{\mathbf{v}}$ is the estimate of our preference vector. Once the inner loop converges, the estimated demands, \hat{N}_j , are saved and used by the outer loop to compute new stockout probabilities.

RECOVERY OF IMPORTANT VALUES. Once the procedure terminates, the demand rate can be recovered by

$$\hat{\lambda}_t = \hat{X}_{0t} + \sum_{j=1}^n \hat{X}_{jt}. \quad (18)$$

We can also use the estimates to compute the estimated lost sales. Denote by \hat{L}_{jt} the estimated lost

sales of product j at time t ; that is, the demand for product j that does not spill over to any other product, given that product j is not available. We can get this from

$$\hat{L}_{jt} = \sum_{S \in \mathbb{P}(\mathcal{N})} \left(\frac{1}{\sum_{l \in S} v_l + 1} \right) \hat{X}_{jt} \cdot \mathbb{I}_j(S) \cdot G_t(S). \quad (19)$$

\hat{L}_{jt} can be decompose into lost sales resulting from observed and unobserved stockouts. Let O_{jt} be 1 if $p_{jt} = 1$ and 0, otherwise. Also, let U_{jt} be 1 if $0 < p_{jt} < 1$ and 0, otherwise. The estimated lost sales to unobserved stockouts, or *unobserved lost sales*, is given by

$$\hat{J}_{jt} = U_{jt} \hat{L}_{jt} \quad (20)$$

while the estimated lost sales to observed stockouts, or *observed lost sales*, is given by

$$\hat{K}_{jt} = O_{jt} \hat{L}_{jt}. \quad (21)$$

The total lost sales in period t is given by

$$\hat{L}_t = \sum_{j=1}^n \hat{L}_{jt} = \sum_{j=1}^n \hat{K}_{jt} + \sum_{j=1}^n \hat{J}_{jt}. \quad (22)$$

4 Numerical example

In the following section we report a number of numerical examples to better understand the performance of our suggested estimation procedure. We start by analyzing a simulated instance of purchases and inventory data that mimics a small category of products sold in a store over a year to get a sense of the performance of the procedure. After this, we conduct sensitivity analyses to understand how environmental factors (number of products, time horizon, demand rate) and specification errors (market potential, transition probabilities) affect the performance of the procedure.

4.1 A simulated case

We first generate a year’s worth of daily data ($T = 364$) for a single store with $n = 10$ substitutable products. The arrival process, λ_t , is generated as a series of realizations from a stationary Poisson process with rate $\lambda = 50$. Note that this is the overall demand arriving at the store, which includes the outside option. The market potential for the store is set to $s = 78\%$, meaning that on average 78% of the arriving consumers prefer one of the n considered products rather than the outside option. We randomly generate on-shelf availability via the Markov Chain model detailed in Section 3.2. The transition probabilities are set so that the probability of entering either stockout state is 1% and the probability of staying in either stockout state is 82%. We do not consider transitions between the observed stockout state and the unobserved stockout state. We use the same probabilities for all products in the category. Note that by analyzing the limiting probabilities of the states, these numbers imply a true long-run stockout rate of

$$SR = \frac{1}{1 + \frac{1-0.82}{0.02}} = 10\%. \quad (23)$$

Also, since the probabilities of entering the stockout states are the same whether the stockout state is observed or not, 50% of all stockouts will, on average, be unobservable in the generated inventory record. As such, the recorded stockout rate is 5%. Demands for the different products in each period are then a realization of a multinomial distribution over the available alternatives.

To better understand the relative performance of our suggested procedure, we run four estimation algorithms on the generated data. For all four algorithms, the transition probabilities and the market potential are assumed to be known. The first algorithm is one that ignores both substitution effects and inventory record reliability problems, and simply estimates demand based on observed purchases during days with a positive inventory record. The demand during a stockout day is thus assumed to be the average demand over all observed non-stockout days. We refer to this as the Censored Demand Estimation (CDE) algorithm. The second algorithm is one that takes substitution effects into account, but ignores inventory record reliability problems. This is essentially the estimation algorithm proposed by Vulcano et al. (2012). We refer to this algorithm as the Observed Stockouts only EM (OSEM) algorithm. The third algorithm is the same as our procedure, except that it estimates the missing data through the full conditional probability (i.e. maximiza-

tion of (13) to estimate the v_j 's) instead of the simulated sample mean. We refer to this as the Nested EM (NEM) algorithm. The last algorithm is our proposed procedure, which we refer to as the Simulated Sample Mean Maximization (SSMM) algorithm. In our procedure, we simulate 100 choice sets in every period in the inner loop whenever the potential number of choice sets exceeds 100. The convergence criteria for the OSEM algorithm as well as for both loops of the NEM and SSMM procedures are set to the maximum difference between the preference vectors being less than 0.001.

The final estimates from the four algorithms are found in Table 1. The preference weights and the simulated demands are shown in columns 2 and 3. Columns 4-7 show the estimated preference weights and their standard errors, while columns 8-11 show the estimated yearly underlying demands as derived from these estimates, as given by algorithm CDE, OSEM, NEM, and SSMM, respectively.. All estimates are significant on a 99% level.

The CDE algorithm does comparatively well for some items, but overestimates demand for most products. While it might be tempting to conclude that this makes the algorithm suitable for those items, this is mostly a random effect. In fact, when using the CDE, two sources of error may cancel each other out in some instances. The first error is from ignoring substitution effects, which leads to overestimation of demand. The second error is from ignoring inventory record reliability issues, which leads to underestimation of demand. For some items in our data, these errors cancel each other out, whereas for other items, ignoring the substitution effects leads to overestimation. Ignoring these sources of error also has a strong impact on the estimated lost demand. In our example, the percent lost demand is estimated to be 5.6%, which is close to the observed stockout rate, but much higher than the true lost demand of 2.2%. Note that the true lost demand is much lower than the stockout rate (10%) due to substitution. The OSEM algorithm does consider substitution effects, but since the algorithm ignores inventory record reliability issues, it underestimates the spill-over demand and sometimes overestimates the direct demand. As can be seen in the table, it is not always clear in which direction the results will be biased owing to this. The OSEM algorithm also underestimates the lost demand. The managerial implication is that ignoring inventory record reliability issues not only leads to more lost sales because the wrong demand parameters are used in the inventory management system, but also that any post facto analysis of the demand leads the manager to believe lost sales are not as big of a problem as it is. Our proposed procedure avoids these problems by incorporating inventory record reliability

Table 1: Parameter estimation results from simulated data.

j	v_j	N_j	\hat{v}_j				\hat{N}_j			
			CDE	OSEM	NEM	SSMM	CDE	OSEM	NEM	SSMM
1	1.0	3,993	-	1.0529 (0.0068)	1.0369 (0.0154)	1.0369 (0.0147)	4,343	4,116	4,072	4,071
2	0.5	1,997	-	0.4865 (0.0021)	0.4864 (0.0106)	0.4864 (0.0103)	1,862	1,902	1,910	1,910
3	0.1	395	-	0.1045 (0.0021)	0.1051 (0.0052)	0.1051 (0.0053)	484	409	413	413
4	0.3	1,175	-	0.3171 (0.0016)	3.168 (0.0088)	0.3168 (0.0084)	1,067	1,239	1,244	1,244
5	0.1	402	-	0.0939 (0.0009)	0.0947 (0.0052)	0.0947 (0.0051)	357	367	372	372
6	0.5	1,941	-	0.4343 (0.0030)	0.4419 (0.0131)	0.4419 (0.0130)	1,994	1,698	1,735	1,735
7	0.1	406	-	0.1111 (0.0016)	0.1124 (0.0058)	0.1124 (0.0058)	410	434	441	441
8	0.2	792	-	0.2161 (0.0025)	0.2159 (0.0070)	0.2160 (0.0076)	800	845	848	848
9	0.2	849	-	0.1993 (0.0038)	0.1997 (0.0069)	0.1997 (0.0076)	907	779	784	784
10	0.6	2,337	-	0.5844 (0.0034)	0.5902 (0.0124)	0.5902 (0.0128)	2,404	2,284	2,317	2,317
% Lost Demand		2.2					5.6	1.5	1.9	1.9

Note. Bootstrapped standard errors in parenthesis. Overall demand rate $\bar{\lambda}_t = \lambda = 50$. Market potential $s = 78\%$. $T = 364$. $SR = 10\%$. The last row shows the amount of the total underlying demand lost to the outside option because of stockouts (% of total demand). SSMM based on 100 draws in each simulation.

issues in the estimation. As evident from both the preference estimates and the demand estimates in Table 1, our procedure (SSMM) leads to estimates very close to the estimates from the full Nested EM (NEM) algorithm. The annual underlying demands differ by only a few of units between these two algorithms, despite the large difference in computational complexity. The estimates are also close to the simulated data, and show no discernible bias. With our proposed procedure, the estimated percent lost demand (1.9%) is more in line with the actual percent lost demand.

Table 2 shows three performance metrics for each estimation algorithm: the Mean Absolute Percent Error (MAPE), the Mean Percent Error (MPE), and the time to convergence in seconds. The MAPE and the

MPE measure accuracy and bias respectively, and are computed based on the error between the estimated daily underlying demand \hat{X}_{jt} and the simulated daily demand, X_{jt} . The MAPE is computed as

$$MAPE = \frac{1}{n} \sum_{j=1}^n \frac{\sum_{t=1}^T |\hat{X}_{jt} - X_{jt}|}{\sum_{t=1}^T X_{jt}} \quad (24)$$

while the MPE is computed as

$$MPE = \frac{1}{n} \sum_{j=1}^n \frac{\sum_{t=1}^T \hat{X}_{jt} - X_{jt}}{\sum_{t=1}^T X_{jt}}. \quad (25)$$

A negative MPE indicates that the estimate is, on average, lower than the simulated demand. The time to convergence is reported to give an idea of how computational complexity is indeed a possible issue with the full NEM algorithm.

The metrics reported in Table 2 corroborate the observations from Table 1. CDE is the least accurate method and leads to upwardly biased estimates, on average. It is, however, faster than any of the other algorithms. The OSEM is also relatively fast, and converges in less than a second for our simulated data. While it is on average more accurate than the CDE algorithm, it is instead downwardly biased, as expected. Our proposed procedure, the SSMM, creates estimates with accuracy and bias almost identical to the full NEM algorithm. However, it does so in roughly 1 second, instead of the 1.7 seconds it takes for the full NEM algorithm to converge: a small absolute difference, but a large relative difference considering the very small set of products. More importantly, both the SSMM algorithm and the NEM algorithm are *both* more accurate *and* less biased than the OSEM algorithm. This is promising, and highlights the importance of including substitution effects and inventory record reliability issues in the estimation procedure.

For practical application, it should be noted how switching from OSEM to SSMM leads to estimates very close to that of NEM, while only increasing the time to convergence about three times. Clearly, one can reduce this run time by reducing the number of draws in the simulation. This should be considered when one wants to estimate a large number of categories or stores.

Table 2: Performance of the estimation algorithms on our simulated set of data (single run).

	CDE	OSEM	NEM	SSMM
MAPE	78.3213	68.3775	66.6194	66.6195
MPE	1.9293	-0.9930	-0.2022	-0.2025
Time to convergence (seconds)	0.0671	0.3435	1.7150	0.9866

Note. Mean Absolute Percent Error (MAPE) and Mean Percent Error (MPE) are calculated based on daily errors in the estimate of the underlying demand. Time to convergence when using a CPU with Intel CORE i7 processor and 8GB RAM. Overall demand rate $\bar{\lambda}_t = \lambda = 50$. Market potential $s = 78\%$. $T = 364$. $SR = 10\%$. $n = 10$. SSMM based on 100 draws in each simulation.

4.2 Sensitivity analysis

We next analyze how sensitive the performance is to environmental factors; i.e., the number of products considered; the number of time periods; and the overall demand for the category; as well as how sensitive the performance is to specification errors; in this case, misspecification of the prior probabilities; and the market potential.

For the first analysis, we generate 30 scenarios using three product category sizes ($n = \{5, 10, 15\}$), three time horizon lengths ($T = \{100, 500, 1000\}$), and two levels of the overall demand for the category ($\lambda_t = \{50, 100\}$). In Table 3, the MPE (our metric for bias) for the low-demand scenarios ($\lambda_t = 50$) is presented. From the table, we make two interesting observations. First, as the number of time periods increases, the bias, in general, decreases. This is seen to some extent across all algorithms. However, the effect is much more pronounced for the NEM and the SSMM. For both these algorithms, the bias approaches zero as the time horizon increases. This is to be expected, since we know that the NEM and the SSMM are the only algorithms that are asymptotically unbiased when the inventory record is uncertain. Second, even though computational complexity increases significantly as the number of products increases, the relative performance of the algorithms is unchanged: the NEM and SSMM provide estimates with much lower bias than the other two algorithms as the number of products increases. The table also provides a clear indication on how the direction of the bias differs between the algorithms. The CDE tends to overestimate the demand, while the OSEM tends to underestimate the demand. The NEM and the SSMM show no such patterns. For the NEM and the SSMM, the MPE is never significantly different from zero despite the low variance.

Table 3: Average MPE (std error) for the estimation algorithms over 50 simulations for low demand, $\lambda_t = 50$.

Number of products, n	Time periods, T	MPE			
		CDE	OSEM	NEM	SSMM
5	100	4.7579 (4.9602)	-0.2109 (3.7549)	0.0394 (2.8401)	0.0393 (2.8403)
	500	4.0700 (2.2519)	-0.3901 (1.5613)	0.1499 (1.2525)	0.1502 (1.2525)
	1,000	3.3801 (1.8143)	-0.8252 (1.2370)	-0.1792 (0.8163)	-0.1787 (0.8162)
10	100	3.5118 (4.3669)	-0.6004 (2.6739)	0.1948 (2.6494)	0.1938 (2.6495)
	500	2.9759 (1.8020)	-1.1372 (1.1710)	-0.2949 (1.1207)	-0.2951 (1.1208)
	1,000	3.4377 (1.1542)	-0.7588 (0.8064)	0.0577 (0.7939)	0.0576 (0.7940)
15	100	2.7873 (3.4393)	-0.8946 (3.1199)	0.0582 (2.7866)	0.0571 (2.7865)
	500	3.4772 (1.5922)	-0.6609 (1.3387)	0.4449 (1.2580)	0.4448 (1.2579)
	1,000	3.0837 (0.9030)	-0.9766 (0.8149)	0.1904 (0.8216)	0.1903 (0.8215)

Note. Table reports mean and standard error of the MPE across 50 simulated realizations for each scenario. Market potential $s = 78\%$. $SR = 10\%$. SSMM based on 100 draws in each simulation.

In Table 4, the MPE for the high-demand scenarios ($\lambda_t = 100$) is presented. The insights from Table 3 still apply, but the difference between the algorithms is now much larger: the NEM and the SSMM display much smaller bias than the alternatives. While both the CDE and the OSEM display significant bias, the NEM and the SSMM has a mean MPE close to zero with a very small variance.

We next investigate our procedure’s sensitivity to misspecification of the prior probabilities associated with transitioning between different states, and the market potential of the product category. The sensitivity to the choice of the prior probabilities is investigated by considering the same setup as in Section 4.1. First, the probability of entering the stockout state is fixed at 1% while the probability of staying in the stockout state varies (Figure 1a). Note that this means the (long-run) stockout rate, SR , also varies accordingly. Next, we fix the probability of staying in the stockout state at 82% while the probability of entering the stockout state varies (Figure 1b). This also means the stockout rate varies. Lastly, we fix the stockout rate and vary the transition probabilities (Figure 1c). The sensitivity to the choice of market potential, \hat{s} , is investigated through five assumed market potentials at three actual market potentials. For each scenario, 50 simulations are run. The results are reported in Figure 2. In all figures, the edges of the box span from the 25th percentile to the 75th percentile, and the whiskers extend to the largest numbers not considered outliers. Outliers are plotted separately.

Table 4: Average MPE (std error) for the estimation algorithms over 50 simulations for high demand, $\lambda_t = 100$.

Number of products, n	Time periods, T	MPE			
		CDE	OSEM	NEM	SSMM
5	100	3.6500 (5.4294)	-0.7268 (3.3372)	-0.1325 (1.9554)	-0.1343 (1.9559)
	500	3.4213 (2.1784)	-0.9651 (1.3154)	0.0051 (0.7332)	0.0054 (0.7330)
	1,000	3.9334 (1.5983)	-0.7689 (1.0009)	0.0064 (0.6323)	0.0072 (0.6324)
10	100	3.5816 (3.3819)	-0.7047 (1.9845)	-0.1455 (1.7380)	-0.1461 (1.7370)
	500	2.9279 (1.4540)	-0.9803 (1.2277)	0.0595 (0.7999)	0.0594 (0.7999)
	1,000	3.0304 (1.1631)	-1.1260 (0.6747)	-0.0894 (0.4397)	-0.0894 (0.4397)
15	100	3.0227 (3.0232)	-1.0451 (1.9758)	0.0475 (1.6728)	0.0472 (1.6724)
	500	2.8551 (1.3854)	-1.1600 (1.0553)	-0.1060 (0.8132)	-0.1060 (0.8134)
	1,000	3.0326 (0.8199)	-1.1843 (0.7047)	0.0044 (0.5573)	0.0043 (0.5576)

Note. Table reports mean and standard error of the MPE across 50 simulated realizations for each scenario. Market potential $s = 78\%$. $SR = 10\%$. SSMM based on 100 draws in each simulation.

In Figure 1 we see how the bias in the demand estimates, as measured by the MPE, is affected by the assumption about the transition probabilities. The first two plots display misspecification of each probability when the stockout rate is not known. In both cases, the demand estimates are downwardly biased when the probabilities are underestimated, and upwardly biased when probabilities are overestimated. Consequently, if one probability is overestimated while the other is underestimated, the final demand estimate may, in fact, not be very biased. This is also seen in Figure 1c, which shows the MPE for when the stockout rate is known and the transition probabilities are set to vary so that the (long-run) stockout rate is unaffected. As seen, the bias is reduced, especially when the probability of staying in the stockout state is overestimated (and hence the probability of entering the stockout state is underestimated). In general, the results indicate that if the stockout rate is known, the SSMM algorithm leads to less biased demand estimates than the CDE algorithm and the OSEM algorithm even when the the transition probabilities are severely misspecified.

Figure 2 shows how sensitive the demand estimates are to the assumption about the market potential. While the box plots show that there might be bias when the market potential is not correctly specified, this bias is insignificant and without any clear pattern. These results combined with those in Figure 1 indicate that our procedure is relatively robust to specification errors.

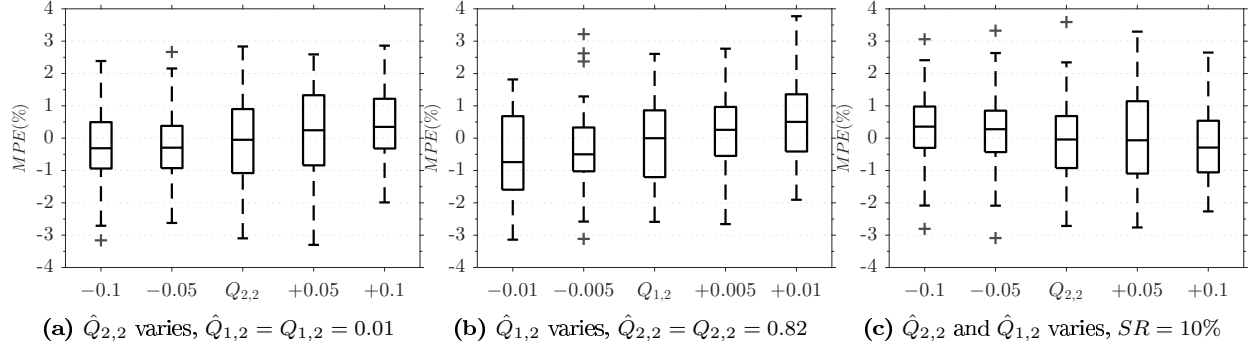


Figure 1: The effect of errors in the assumed transition probabilities on MPE.

Note. Scenario values relative to the actual value. For instance, “-0.1” means the assumed probability is given by $\hat{Q}_{i,h} = Q_{i,h} - 0.1$, where $Q_{i,h}$ is the true transition probability. $T = 364$, $n = 10$

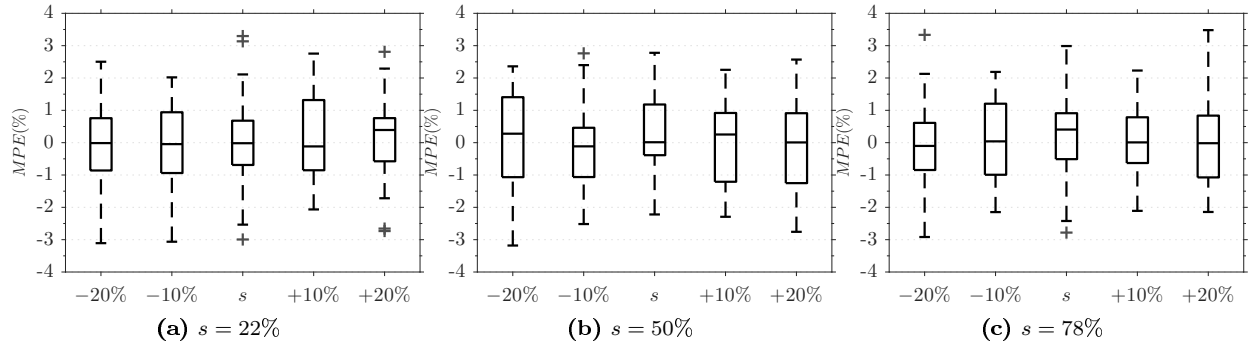


Figure 2: The effect of errors in the assumed market size, \hat{s} , on MPE .

Note. Scenario values relative to the actual value. For instance, “-10%” means the algorithm uses market size $\hat{s} = 0.9s$, where s is the true market size. $T = 364$, $n = 10$

5 Application: laundry detergent

In the following section, we report the results from applying our procedure to a real data set from a major U.S. retailer. The data set covers 148 non-promotional days of point of sales data and inventory records for a category of substitutable laundry detergent products. The category consists of 29 stock keeping units (SKUs) of two different brands. First, we illustrate how our procedure (SSMM) is applied to the data to estimate the demand for the products. Next, in order to get a sense of the real-world importance of considering inventory record reliability issues when estimating demand, we replicate the study for 75 randomly selected stores from the same retailer and compare the result to that of the OSEM algorithm.

5.1 Audit data and transition probability matrix estimation

To calibrate the transition probabilities that are used in the procedure, we use audit data in combination with the point of sales data. The manufacturer of the SKUs in consideration hires an independent audit company to verify the on shelf availability of a list of SKUs at each store on a weekly basis. The generated list of SKUs comprises both non-randomly selected and randomly selected SKUs, and the process by which the list is generated was hidden to the authors. As such, the list is biased in some unknown manner; nevertheless, it is the best information available to estimate instock and stockout rates and, as seen in previous sections, all that is needed for the procedure to work well is a “good enough” estimate of the probabilities. The audit data consists of responses by the auditors at the store-SKU-day level related on-shelf-availability of the SKU. The responses are detailed enough to ascertain whether a product is (a) instock, (b) out of stock with positive inventory record, or (c) out of stock with inventory record of zero. For product j , we can thus identify the instock probability, α_j , the unobserved stockout probability, β_j , and the observed stockout probability, δ_j .

From the point of sale data we can estimate several of the store level transition probabilities directly from their observable frequencies. We estimate $\hat{Q}_{1,1}^j$ as the number of instock days that transition to instock days divided by the number of days instock; $\hat{Q}_{1,3}^j$ as the number of instock days that transition to observable stockout days divided by the number days instock; $\hat{Q}_{3,1}^j$ as the number of observable stockout days which transition to instock days divided by the number of observable stockouts; and $\hat{Q}_{3,3}^j$ as the number of observable stockout days which transition to observable stockout days divided by the number of observable

Table 5: Descriptive statistics for store estimates of transition probabilities

	$\hat{Q}_{1,1}$	$\hat{Q}_{1,2}$	$\hat{Q}_{1,3}$	$\hat{Q}_{2,1}$	$\hat{Q}_{2,2}$	$\hat{Q}_{2,3}$	$\hat{Q}_{3,1}$	$\hat{Q}_{3,2}$	$\hat{Q}_{3,3}$
Store mean	0.9589	0.0338	0.0073	0.3700	0.6120	0.0181	0.6267	0.0404	0.3330
Store std.	0.0128	0.0097	0.0051	0.1508	0.1096	0.0624	0.1284	0.0125	0.1346

stockout days. In case of days transitioning to uncertainty blocks, we assume an $\alpha/(\alpha + \beta)$ probability of transitioning to an instock day. Now, by noting that the instock probability, α_j , the unobserved stockout probability, β_j , and the observed stockout probability, δ_j are the limiting probabilities for states 1, 2, and 3, respectively, it follows that

$$\hat{Q}^j = \begin{bmatrix} \hat{Q}_{1,1}^j & 1 - \hat{Q}_{1,1}^j - \hat{Q}_{1,3}^j & \hat{Q}_{1,3}^j \\ \frac{\alpha_j(1 - \hat{Q}_{1,1}^j) - \delta_j \hat{Q}_{3,1}^j}{\beta_j} & 1 - \frac{\alpha_j(1 - \hat{Q}_{1,1}^j - \hat{Q}_{1,3}^j) + \delta_j(1 - \hat{Q}_{3,1}^j - \hat{Q}_{3,3}^j)}{\beta_j} & \frac{\delta_j(1 - \hat{Q}_{3,3}^j) - \alpha_j \hat{Q}_{1,3}^j}{\beta_j} \\ \hat{Q}_{3,1}^j & 1 - \hat{Q}_{3,1}^j - \hat{Q}_{3,3}^j & \hat{Q}_{3,3}^j \end{bmatrix}. \quad (26)$$

Based on audit data from the half year period under consideration, across all the SKUs we find the instock rate, α , is 0.9, the unobservable stockout rate, β , is 0.09, and the observable stockout rate, δ , is 0.01.² The store level mean and standard deviation of the estimated transition probabilities are reported in Table 5.

5.2 Store level analysis

We first apply our procedure (SSMM) to half a year's worth of data from a single store. The combined total market share for both brands was known to be 58%. Table 6 reports the summary statistics of the data set: the total and average daily sales (columns 3 and 4) and the total and average daily observed stockouts (columns 5 and 6) of all SKUs in the category sold at the store. Note that not all 29 SKUs in the category are sold in all stores. We observe that total sales range between 50 for Brand B's SKU 26 and 423 for Brand A's SKU 15, corresponding to a range of 0.3 to 2.9 units per day with an average of 1.5. Total observed stockouts range from 0 to 3 with an average of 0.7. Finally, the observed stockout rate for both Brand A

²Note that the total stockout rate is given by $SR = \beta + \delta = 1 - \alpha = 0.1$.

Table 6: Summary statistics

SKU #	Brand	Sales		Observed Stockouts	
		Total	Average Daily	Total	Average Daily
1	A	294	2.0	0	0.000
2	A	277	1.9	0	0.000
3	A	334	2.3	0	0.000
7	A	212	1.4	2	0.014
9	A	211	1.4	1	0.007
10	A	165	1.1	0	0.000
12	A	353	2.4	0	0.000
14	A	250	1.7	1	0.007
15	A	423	2.9	0	0.000
16	A	231	1.6	0	0.000
17	A	194	1.3	2	0.014
18	A	93	0.6	0	0.000
20	A	143	1.0	1	0.007
26	B	50	0.3	3	0.020
28	B	186	1.3	1	0.007
29	B	149	1.0	0	0.000
Brand A SKU Average		244.6	1.7	0.5	0.004
Brand B SKU Average		128.3	0.9	1.3	0.009
Overall Average		222.8	1.5	0.7	0.005

and Brand B are 0.004 and 0.009 stockouts per day per SKU, respectively.

The estimation results are presented in Table 7. Reported are the customer preference weights (column 3), the annual and daily demand derived from these estimates (columns 4 - 5), percentage of estimated demand missed and recaptured (columns 6 - 7), the percentage of estimated demand which are total, unobserved and observed lost sales (columns 8-10) and the ratio of unobserved to observed lost sales (column 11). From the table we make a number of observations. First, all estimates are significant on a 99% level, which is encouraging. Second, the demand rate is not always higher than the sales rate. SKU 15 Brand A is one such example, for which annual sales is 423 and the estimated annual demand is only 421.3. This is because of substitution. Sometimes the total recaptured demand is greater than the total missed demand, which means that ignoring substitution leads to overestimation of demand. We notice that in total, SKUs of Brand A and

B both miss, on average about 10.7 units of demand but Brand A recaptures about 6.4 units of other SKUs' demand while Brand B only recaptures on average 3.6 units. Unsurprisingly, Brand A has lower lost sales as percentage of demand than Brand B (2.2 and 3.4 percent of demand, respectively). Third, fast moving SKUs and slow moving SKUs show no discernible differences in the ratio between observed and unobserved lost sales. This is also encouraging, since such a pattern would indicate that the estimated unobserved lost sales rate is affected by the demand rate. This provides some confidence to the observation about the size of these ratios. For those SKUs with observed lost sales, unobserved lost sales are 3 times observed lost sales at a minimum (see SKU 17), while at a maximum they are 6.4 times larger (see SKU 20). On average, unobserved lost sales are about 4.9 and 5.2 times observed lost sales for Brand A and Brand B, respectively.

Note that while the table presents the annual aggregates, the procedure creates daily demand estimates that can be used by time-series techniques for forecasting in an online fashion by the store, in the same way reported sales are used. Our experience, however, is that the procedure is most useful as a tool to get the "base-demand" rates for the SKUs used for planning purposes; i.e. demand rates in absence of promotions. These base-demand rates tend to be fairly stable over time and thus do not need to be updated every day.

5.3 The impact of ignoring inventory record reliability issues

To highlight the importance of considering inventory record reliability issues, we replicate the analysis for a random selection of 75 stores from the same U.S. retailer, and use the OSEM algorithm on the same data. We report the result on store-SKU level, to convey a general idea about the size of the annual bias for a given item. Table 8 presents the average of the store-SKU level estimates of Demand, Missed Demand, Recaptured Demand, Total Lost Sales, Unobserved Lost Sales, and Observed Lost Sales. Column 2 indicates the method (SSMM or OSEM) used to create the estimates while columns 3 - 8 report the average store-SKU estimates. There were 1774 store-SKUs represented. We notice that, on average, the demand is greater for SSMM compared to OSEM. This is expected, since SSMM accounts for censored demand due to inventory record reliability issues, and is well in line with our previous analyses. Also, note that observed lost sales for SSMM are slightly higher than total lost sales for OSEM due to this increase in demand estimates. Most notable, however, are the unobserved lost sales, which amount to 4.2 percent of sales per store-SKU overall with a discrepancy between Brands A and B at 4.2 and 5.4 percent, respectively. Overall, unobserved lost sales

Table 7: Estimation results

SKU #	Brand	Preferences	Demand Estimate		Substitution Estimate		Lost Sales as % Demand			
			Annual	Daily	Missed	Recaptured	Total	Unobserved	Observed	Ratio
1	A	0.112 (0.0059)	295.9	2.0	9.2	7.4	1.4	1.4	0.0	-
2	A	0.109 (0.0100)	288.2	1.9	17.7	6.5	2.8	2.8	0.0	-
3	A	0.127 (0.0064)	333.8	2.3	8.1	8.3	1.1	1.1	0.0	-
7	A	0.083 (0.0055)	218.7	1.5	12.7	6.0	2.6	2.2	0.4	4.9
9	A	0.080 (0.0050)	211.3	1.4	6.6	6.3	1.4	1.2	0.2	5.9
10	A	0.066 (0.0053)	174.0	1.2	13.1	4.1	3.4	3.4	0.0	-
12	A	0.135 (0.0084)	355.9	2.4	11.5	8.5	1.5	1.5	0.0	-
14	A	0.097 (0.0054)	256.3	1.7	13.2	6.9	2.3	1.9	0.4	4.3
15	A	0.160 (0.0076)	421.3	2.8	9.7	11.4	1.1	1.1	0.0	-
16	A	0.088 (0.0052)	231.8	1.6	6.5	5.7	1.3	1.3	0.0	-
17	A	0.076 (0.0050)	200.2	1.4	12.0	5.9	2.7	2.0	0.7	2.9
18	A	0.039 (0.0040)	102.2	0.7	11.5	2.4	5.0	5.0	0.0	-
20	A	0.055 (0.0046)	145.4	1.0	6.8	4.4	2.1	1.8	0.3	6.4
26	B	0.020 (0.0026)	53.6	0.4	5.0	1.4	4.1	3.5	0.7	5.2
28	B	0.072 (0.0051)	190.3	1.3	9.2	4.9	2.1	1.9	0.3	-
29	B	0.062 (0.0070)	163.0	1.1	17.8	3.8	4.8	4.8	0.0	-
Brand A SKU Average		0.094	248.9	1.7	10.7	6.4	2.2	2.0	0.2	4.9
Brand B SKU Average		0.051	135.6	0.9	10.7	3.4	3.7	3.4	0.3	5.2
All		0.086	227.6	1.6	10.7	5.9	2.5	2.3	0.2	4.9

Note. Estimation using SSMM with 100 draws. Bootstrapped standard errors in parenthesis (50 bootstrap iterations)

Table 8: Comparison of results from SSMM and OSEM

Brand	Procedure	Average Demand			Average Lost Sales		
		Total	Missed	Recaptured	Total	Unobserved	Observed
Brand A	SSMM	136.5	9.1	4.9	4.2	3.9	0.2
	OSEM	130.3	1.4	0.7	0.6	-	-
Brand B	SSMM	194.7	11.7	7.0	5.4	5.3	0.1
	OSEM	188.3	1.4	1.1	0.7	-	-
Overall	SSMM	136.5	9.1	4.9	4.2	3.9	0.2
	OSEM	130.3	1.4	0.7	0.6	-	-

are over 16 times observed lost sales. Not only does not accounting for inventory record reliability issues reduce our estimate of demand, ignoring inventory uncertainty results in a total lost sales estimate which only account for a fraction of actual lost sales.

Due to substitution effects, the store-SKU demand estimate for SSMM is not always greater than of OSEM. Figure 3a presents the distribution of the percent difference in demand as estimated by SSMM and OSEM. Table 9 reports the summary statistics for the distribution of the percent difference in demand as estimated by SSMM and OSEM. The mean, median and standard deviation are reported in columns 2 - 4. Average demand estimates for SSMM are 6% greater than those for OSEM with a standard deviation of 4.3. This means the typical difference between the two demand estimates is between -1.7 and 10.3%. Additionally, we observe for Brand A and Brand B store-SKUs, the percent difference in demand is 6.1 and 4.4,

However, estimates of lost sales made by SSMM and OSEM are quite different since SSMM accounts for inventory uncertainty and OSEM does not. To compare OSEM and SSMM with respect to changes in estimates of lost sales, we must consider store-SKUs with observed stockouts and those without observed stockouts separately. For store-SKUs with observed stockouts the percent difference in estimated lost sales can be calculated, while for those store-SKUs without observed stockouts, the percent difference is undefined. Figure 3b presents the distribution of the percent difference in total lost sales as estimated by SSMM and OSEM for those store-SKUs with observed stockouts. Table 9 reports the summary statistics for the distribution of the percent difference in lost sales as estimated by SSMM and OSEM. The mean, median and

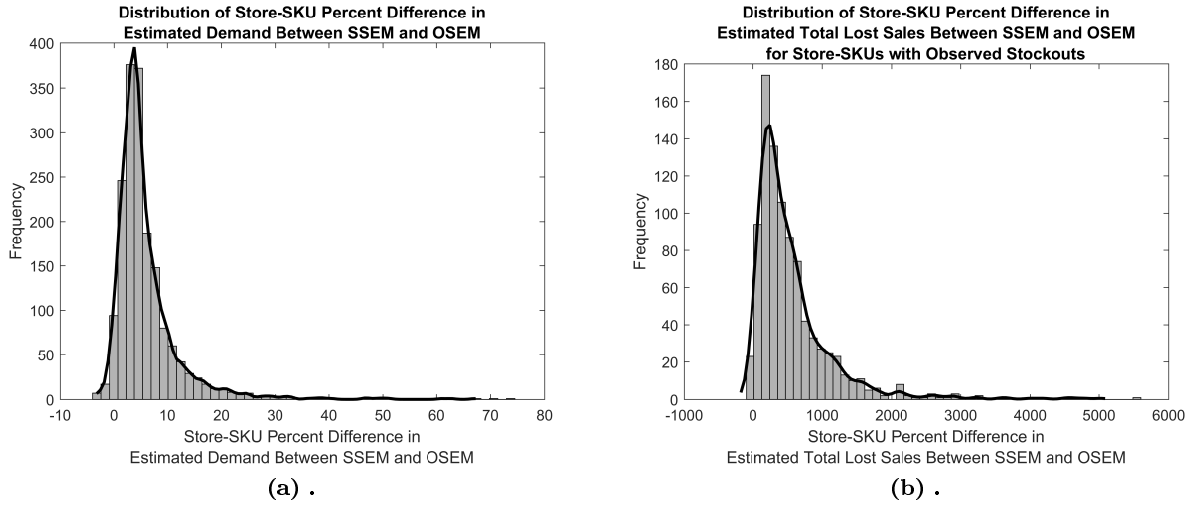


Figure 3: Distribution of the percent difference in (a) demand and (b) lost sales as estimated by SSMM and OSEM

standard deviation are reported in columns 5 - 7. We observe that SSMM estimates over 5.5 times greater lost sales for a store-SKU than OSEM for each brand and overall. For many store-SKUs (922 of 1774), SSMM estimates lost sales where OSEM does not.

5.4 Comparison of observed and unobserved lost sales

Figure 4 presents the distributions of total, observed and unobserved lost sales as percentage of demand. Table 10 reports the summary statistics for these distributions with the mean, median and standard deviation

Table 9: Descriptive statistics of the percent difference in demand (for all store-SKUs) and lost sales (for store-SKUs with observed stockouts) as estimated by SSMM and OSEM

	Percent difference in estimated demand			Percent difference in lost sales (store-SKUs w/ observed stockouts)		
	Mean	Median	Standard Deviation	Mean	Median	Standard Deviation
Brand A	6.1	4.5	6.6	567	385	625
Brand B	4.4	3.0	6.2	593	397	630
Total	6.0	4.3	6.6	569	389	626

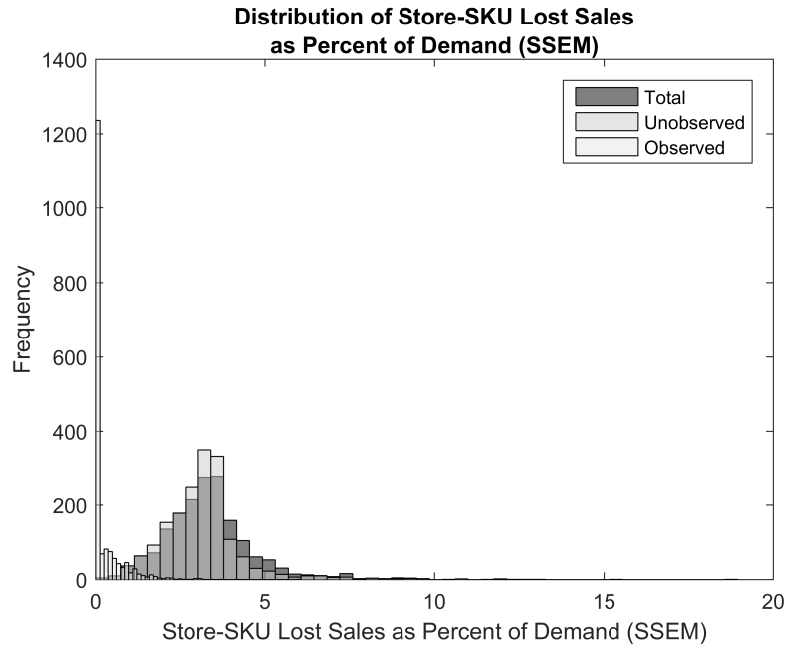


Figure 4: Distributions of total, observed and unobserved lost sales as percentage of demand as estimated by SSMM

of the total, unobserved and observed lost sales given in columns 2-4, 5-7, and 7-9, respectively. The average of the distributions of total, observed and unobserved lost sales as percentage of demand are 3.7, 3.1, and 0.6, respectively. For this product category, virtually all lost sales are unobserved.

Figure 5 presents a scatter plot of observed versus unobserved lost sales as percentage of demand grouped by brand with their marginal distributions shown. Note that there is no correlation between observed and unobserved lost sales. This particular data set does not support the hypothesis in the literature suggesting that there is a positive relationship between observed lost sales and unobserved lost sales (Gruen and Corsten, 2007). The managerial implication of this is that consideration of solely observed stockouts to estimate lost sales is effectively useless for estimating unobserved stock outs and, therefore, total stockouts. As such, a method such as the one herein must be used to have any understanding of the actual lost sales in a network of stores.

Table 10: Descriptive statistics of total, observed and unobserved lost sales as percentage of demand as estimated by SSMM

	Lost Sales as Percentage of Demand								
	Total			Unobserved			Observed		
	Mean	Median	Standard deviation	Mean	Median	Standard deviation	Mean	Median	Standard deviation
Brand A	3.4	3.3	1.5	3.2	3.2	1.3	0.2	0.0	0.5
Brand B	3.1	3.3	1.5	3.0	2.8	2.0	0.1	0.0	0.4
Total	3.4	3.3	1.5	3.2	3.2	1.3	0.2	0.0	0.5

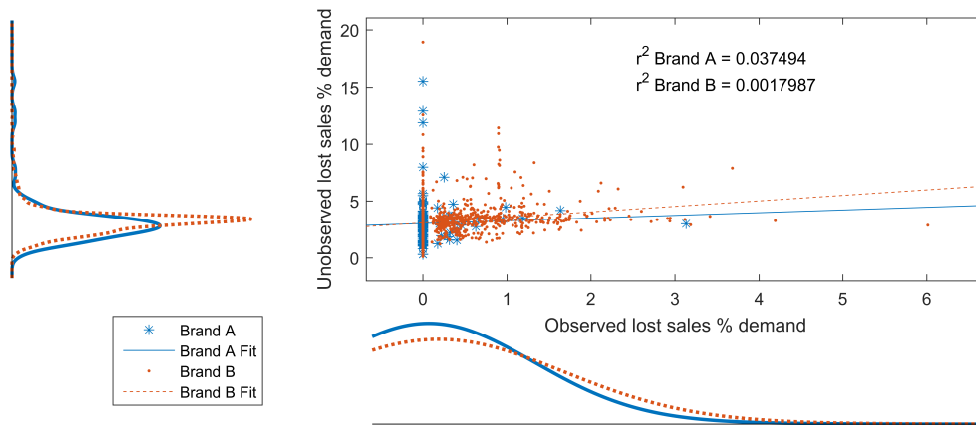


Figure 5: Observed vs unobserved lost sales as percentage of demand as estimated by SSMM

6 Conclusion

Unreliable inventory records is a problem that plagues retailers and manufacturers alike. As an acute problem, unobserved stockouts cannot be readily replenished and result in lost sales. As a chronic problem, demand planning systems are unaware of much censored demand which generally results in under-estimation of demand. Anecdotally and through scant audit information retailers and manufacturers are indeed aware of the problem it poses, but likely not its scope. Furthermore, no methods exist to date which allow retailers and manufacturers to use existing, available data to estimate demand and lost sales in the presence of inventory record reliability problems. Our work aims to fill this gap by introducing a nested procedure that estimates demand using a structural model of consumer demand and inventory progression. The estimation procedure iterates between (1) estimating the stockout probabilities based on a hidden Markov chain model of inventory progression and (2) estimating consumer preferences in a multinomial logit formulation of demand, by maximizing the mean of a weighted sample of possible choice sets. .

Simulation results show clearly that the suggested procedure is more accurate and less biased than other available methods. That is, ignoring inventory record reliability issues may lead to severely biased demand estimates, which in turn leads to improper parameters in replenishment and production policies. The procedure we suggest provides asymptotically unbiased estimates of demand and converges quickly also for large categories considered over long time horizons. While the procedure rests on the researchers ability to initiate it using reasonable estimates of transition probabilities and market potential, sensitivity analyses of the simulated data reveal that the Bayesian nature of the procedure makes it very robust to specification errors.

Application of the SSMM method to a dataset with two brands of SKUs in the laundry care category sold at a major U.S. retailer yielded some surprising results. Firstly, unobserved lost sales are many multiples of observed lost sales. Secondly, we could not support the current hypothesis in the literature that observed and unobserved lost sales are positively correlated (Gruen and Corsten, 2007). Additionally, we also clearly saw the real-life implications indicated in the simulation: not considering inventory uncertainty when estimating demand significantly underestimates demand.

In summary, this work provides a method of estimating demand in the presence of unreliable inventory

records that are only validated infrequently and irregularly. Furthermore, the method provides a new lens with which to view lost sales such that previously unobserved lost sales can now be quantified. Needless to say, solving the problem of inaccurate inventory records can only be accomplished if retailers work with detecting and fighting the root causes of the problem. However, record inaccuracies seem to be a persistent problems that decades of research have not yet resolved. While the root causes of the problem are being addressed and mitigated, we believe this work provide retailers and manufacturers with a powerful tool improve and analyze supply chain performance.

References

- Anupindi, R., Dada, M., and Gupta, S. (1998). Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science*, 17(4):406–423.
- Campo, K., Gijbrecchts, E., and Nisol, P. (2003). The impact of retailer stockouts on whether, how much, and what to buy. *International Journal of Research in Marketing*, 20(3):273–286.
- Chen, L. (2014). Fixing phantom stockouts: Optimal data-driven shelf inspection policies. *Available at SSRN 2689802*.
- Chuang, H. H.-C., Oliva, R., and Liu, S. (2015). On-shelf availability, retail performance, and external audits: A field experiment. *Production and Operations Management*.
- Conlon, C. T. and Mortimer, J. H. (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, 5(4):1–30.
- DeHoratius, N., Mersereau, A. J., and Schrage, L. (2008). Retail inventory management when records are inaccurate. *Manufacturing & Service Operations Management*, 10(2):257–277.
- DeHoratius, N. and Raman, A. (2008). Inventory record inaccuracy: an empirical analysis. *Management Science*, 54(4):627–641.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

- Fisher, M. and Raman, A. (2010). *The new science of retailing: how analytics are transforming the supply chain and improving performance*. Harvard Business Review Press.
- Gruen, T. W. and Corsten, D. S. (2007). A comprehensive guide to retail out-of-stock reduction in the fast-moving consumer goods industry.
- Iglehart, D. L. and Morey, R. C. (1972). Inventory systems with imperfect asset information. *Management Science*, 18(8):B-388.
- Kalyanam, K., Borle, S., and Boatwright, P. (2007). Deconstructing each item's category contribution. *Marketing Science*, 26(3):327-341.
- Kang, Y. and Gershwin, S. B. (2005). Information inaccuracy in inventory systems: stock loss and stockout. *IIE transactions*, 37(9):843-859.
- Kök, A. G. and Shang, K. H. (2007). Inspection and replenishment policies for systems with inventory record inaccuracy. *Manufacturing & service operations management*, 9(2):185-205.
- Lee, J., Gaur, V., Muthulingam, S., and Swisher, G. F. (2015). Stockout-based substitution and inventory planning in textbook retailing. *Manufacturing & Service Operations Management*, 18(1):104-121.
- Musalem, A., Olivares, M., Bradlow, E. T., Terwiesch, C., and Corsten, D. (2010). Structural estimation of the effect of out-of-stocks. *Management Science*, 56(7):1180-1197.
- Ross, S. M. (2006). Introduction to probability models.
- Vulcano, G., Van Ryzin, G., and Ratliff, R. (2012). Estimating primary demand for substitutable products from sales transaction data. *Operations Research*, 60(2):313-334.
- Wu, C. J. (1983). On the convergence properties of the em algorithm. *The Annals of statistics*, pages 95-103.