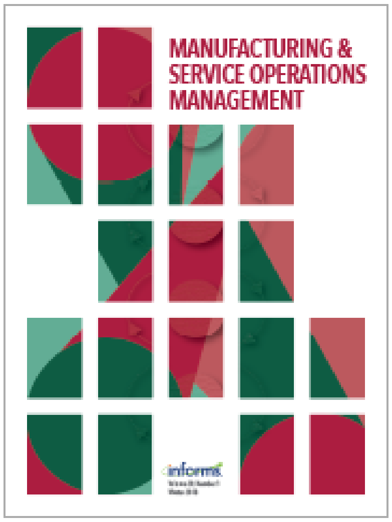


Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

A Hidden Markov Model to Detect On-Shelf Out-of-Stocks Using Point-of-Sale Data

Ricardo Montoya, Carlos Gonzalez

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Articles in Advance, pp. 1–17
http://pubsonline.informs.org/journal/msom/
ISSN 1523-4614 (print), ISSN 1526-5498 (online)

A Hidden Markov Model to Detect On-Shelf Out-of-Stocks Using Point-of-Sale Data

Ricardo Montoya,[a] Carlos Gonzalez[a]

[a] Industrial Engineering Department, University of Chile, 8370456 Santiago de Chile, Chile

Contact: rmontoya@dii.uchile.cl, http://orcid.org/0000-0002-0360-805X (RM); carlos.gonzalez.r@ing.uchile.cl (CG)

Received: November 9, 2016
Revised: September 27, 2017; January 25, 2018; April 7, 2018
Accepted: May 8, 2018
Published Online in Articles in Advance: February 6, 2019

https://doi.org/10.1287/msom.2018.0732

Abstract. We propose a hidden Markov model (HMM) approach to identifying on-shelf out-of-stock (OOS) by detecting changes in sales patterns resulting from unobserved states of the shelf. We calibrate our model using point-of-sale (POS) data from a big-box retailer. We validate our approach using visual inspections that monitor the state of the shelf and compare them to the HMM's predictions. We test the proposed approach on 14 products and 10 stores. We specify our model using a hierarchical Bayes approach and use a Monte Carlo–Markov chain methodology to estimate the model parameters. We identify three latent states in which one of them characterizes an OOS state. The results show that the proposed approach performs well in predicting out-of-stocks, combining high detection power (63.48%) and low false alerts (15.52%). Interestingly, the highest power of detection is obtained for medium-incidence products (77.42%), whereas the lowest false alarm rate is obtained for lower-incidence products (7.32%). Our HMM approach outperforms several benchmarks, particularly for lower-incidence products, which are not typically monitored using visual inspections. Using only POS data, our method uncovers useful information that provides actionable metrics that managers can use to evaluate the quality of demand forecasting and product replenishment at the store–product level.

Funding: R. Montoya gratefully acknowledges partial funding by Fondecyt [Project 1151395] and by Complex Engineering Systems Institute, ISCI [CONICYT–PIA–FB0816].
Supplemental Material: The online appendices are available at https://doi.org/10.1287/msom.2018.0732.

Keywords: out of stock • hidden Markov model • partially observable demand

1. Introduction

Product availability is a key component of service execution because, in addition to the location, customers decide from which store to buy based on the assortment of products and prices that the store offers. Thus, retailers need to monitor carefully that the proposed offer is available to consumers when they visit the store. However, big-box retailers carry thousands of products, and thus, monitoring the availability of their products on a daily basis is a complex task. On-shelf out-of-stock (OOS) is used in grocery retailing to describe the situation in which consumers do not find the products they wish to purchase on the shelf of a supermarket during a shopping trip. Retailers have been struggling with considerable OOS for decades with important consequences. Gruen and Corsten (2007) estimate an overall average OOS rate of 8.3%, which costs retailers, on average, 4% of their annual sales. Importantly, these authors indicate that 70%–75% of OOS are a direct result of retail store practices (either underestimating demand or having ordering processes/cycles that are too lengthy) and shelf-restocking practices (the product is at the store but not on the shelf). The (mis)alignment of the forecasting planning with the demand could determine the incidence of OOS, whereas the promptness of restocking procedures could determine the duration of each OOS event.

Despite recent developments in improving inventory systems the problems of shelf stockouts and inventory inaccuracy still remain unsolved. The complexity in keeping accurate inventories increases mainly with demand variability, poor performance of inventory and replenishment systems, number of products to monitor, product variety, inventory density, and lack of staffing among many other factors (DeHoratius and Raman 2008, Bensoussan et al. 2016). An additional problem that forecast and inventory systems need to deal with is the existence of phantom stockouts (Ton and Raman 2010, Chen 2014). Phantom stockouts occur when there is no product available in the store but the inventory system indicates a positive inventory. This creates immediate problems as the replenishment system does not order new products because of the existence of zero demand with (hypothetically) positive inventory. As the product is truly not available, customers cannot buy the product, creating a vicious circle. In addition, forecasting systems based on those data may consider these zero-sale periods as normal no-demand periods making it difficult to detect abnormal low levels of sales resulting from OOS.

On-shelf availability is a key performance indicator in the retail industry. Consequently, retailers use various techniques to verify their on-shelf product availability. These have evolved over time according to the number of products being monitored and the development of technological improvements. Visual inspections are used to measure the presence of a limited number of products at different periods of time manually, typically once a day only for the top products. These inspections help to adjust inventory systems and trigger replenishment orders when the product is not present. When implementing these visual inspections, retailers or manufacturers incur high costs because of the personnel involved in this task. More recently, image recognition and radio-frequency identification (RFID) technologies have been introduced with limited success because of the high costs of the required devices and technological limitations. Item-level RFID implementations remain prohibitively expensive for many applications and often do not provide perfect visibility into inventory positions because of technological limitations and other practical considerations (Mersereau 2015). Indeed, the costs of RFID are still often much larger than the costs of current identification technologies, such as visual inspections. Although RFID technologies are attractive for numerous companies, most of them still prefer to start with pilot projects and return-on-investment analyses to evaluate their costs and profits (Sarac et al. 2010). Consequently, the use of these technologies is limited to only the more important products. Because of the relatively high cost of a tag, it is not economically feasible to tag many fast-moving consumer goods individually (Hardgrave et al. 2011). Instead, in this research, we develop a methodology that can be applied to a large number of products because it requires only transactional point-of-sale data to determine on-shelf product availability.

We develop a hidden Markov model (HMM) that uncovers the underlying state of the demand for a product using the observed sales pattern. Indeed, the observed sales correspond to a partially observed state of the demand resulting from possible shortages in the offer. In our HMM, one of the hidden states corresponds to the OOS condition, whereas the other states capture the different states of the underlying demand. We exploit the information coming from different stores for the same product and specify a hierarchical Bayes model. We estimate the parameters using a Monte Carlo–Markov Chain simulation approach. We implement our methodology on 14 products and 10 stores using 15 months of daily sales data. We validate the model using data collected through visual inspection of the shelf for a period of 14 days for the same products and stores. For those periods, we contrast the model's prediction regarding the state of the shelf and the observed state of the shelf. Overall, the model identifies 63.48% of the OOS with a false alarm rate of 15.52%, outperforming several benchmarks, including the model then in use by the retailer. The HMM approach provides a higher advantage when considering lower- and medium-incidence products, which are not typically monitored using visual inspections. In addition, the estimated transition matrix from the HMM reflects how frequently the demand moves to an OOS state and how quickly each store reacts to stockouts, summarizing management performance regarding demand planning and product replenishment.

We hope to contribute to the operations management literature by the development of a method based on a well-grounded econometric model that is effective in detecting OOS using partially observable demand. Managerially, we provide an automatic method for detecting on-shelf OOS based on sales data. This offers an accurate view of the shelf availability to the retailer and the product supplier and uncovers interesting patterns regarding demand planning and product replenishment to compare store performance across products.

2. Literature Review

On-shelf OOS has been an active subject of research in recent decades. There are basically two streams of research. The first stream investigates consumer response to OOS (see, e.g., Campo et al. 2000, Fitzsimons 2000, Campo et al. 2003, and Che et al. 2012). Researchers in this area have proposed that consumers facing an OOS can switch to another product, buy the missing item in a competing store, defer the purchase to the next shopping occasion, or drop the purchase altogether (Corstjens and Corstjens 1995). The studies reveal that product switching is the predominant reaction followed by size switching (Campo et al. 2003, Van Woensel et al. 2007). The second stream of research investigates the causal effects of different factors on OOS. Among the many factors that have been studied are brand/store loyalty, urgency of need to buy the product, and consumer type (for a thorough review, see Aastrup and Kotzab 2010).

On-shelf OOS is closely related to the problem of inventory inaccuracy because the availability of a product on the shelf is directly related to the quality of the inventory management carried out by the retailer. Indeed, inventory inaccuracy is a significant cause of OOS. Inventory management with partial information is a classical topic of operations research (e.g., Azoury 1985, Lovejoy 1990), in which this partial information may be due to inventory inaccuracy. For instance, DeHoratius et al. (2008) developed an inventory management tool that accounts for record inaccuracy using a Bayesian belief of the physical inventory level. They model the interaction between the inventory records and sales explicitly. More recently, HMMs have been used to model unobserved inventory considering transactional data. Bayraktar and Ludkovski (2010) developed an analytical approach for inventory management with partially observed demand using an HMM. They analyzed the conditions for optimal inventory control numerically when demand is partially observed. Kök and Shang (2007) investigated inspection and replenishment policies for systems with inventory record inaccuracy. They proposed an inventory-based policy that triggers audits when the inventory records are below certain thresholds. Both the threshold levels and the base stock level depend on the level of inaccuracy. Chen (2014) proposed two partially observable Markov decision models for demand shrinkage and replenishment execution. The author derived a probabilistic belief model about the actual inventory level based on the system inventory records and historical sales data. Bhan (2015) proposed an HMM to represent the evolution of the inventory in a retail store. Similar to Chen (2014), Bhan (2015) investigated the properties of the developed model and the proposed estimation algorithm theoretically. In these HMM approaches to model inventory evolution, the hidden states correspond to the unobserved number of units in the inventory. Thus, these authors do not attempt to characterize the observed evolution in demand empirically because the number of states is, in principle, large, and the potential covariates involved in characterizing the transitions among the states could complicate the estimation of HMM with a large number of states. Indeed, despite the extensive research investigating optimal inspection policies for detecting on-shelf OOS, there is a lack of empirical studies. There are, however, a few exceptions. Papakiriakopoulos et al. (2009) developed a rule-based decision support system for automatically detecting OOS based on sales data, whereas Papakiriakopoulos and Doukidis (2011) proposed a machine-learning algorithm that learns from past OOS history and classifies the current state of the shelf. In these two cases, the algorithms developed are supervised in nature as they learn from history and, thus, require historical information of past OOS, which are typically unobserved or unrecorded.

Our work is related to Chuang et al. (2016) who investigated the effectiveness of using external audits to detect OOS. Similar to their work, we offer an unsupervised method that requires only POS data, which allows monitoring a large number of products. In contrast to Chuang et al. (2016), who used a negative binomial model to describe demand and a statistical process-control approach to trigger alerts, we use an HMM that models sales evolution and triggers alerts simultaneously when the system reaches the OOS state. Furthermore, following previous research, we model demand as being driven by an underlying Markov process that represents the state of the world (Song and Zipkin 1993, Chen and Song 2001, Treharne and Sox 2002). However, we take advantage of the HMM approach to uncover unobserved OOS.

3. Data

We use two sources of information: sales data and on-shelf product availability data. The first source of information is used to develop the methodology and calibrate the HMM model, and the second is used to evaluate the proposed methodology using various performance metrics. Note that we did not have access to inventory data. In our multiple conversations with the managers, they acknowledged that the use of their inventory data were ineffective for OOS detection, and therefore, they used another approach (the one described in Section 4.8.1) to tackle this problem. Indeed, past research shows that the number of existing units available in the store (shelf + back room) does not always correlate with the number of units on the shelf and may not be informative (unless both are zero) (DeHoratius and Raman 2008, Fisher and Raman 2010, Ton and Raman 2010). Therefore, our proposed model does not consider inventory data. However, in Section 5.4.1, we describe how this information can be incorporated, if available, such that its contribution can be tested empirically.

3.1. Sales Data

Our point-of-sale (POS) data contain information for all transactions from February 18, 2013, to June 1, 2014, of a big-box supermarket located in Latin America that offers more than 70,000 SKUs per store. We used data from February 18, 2013, to May 18, 2014, to calibrate the model and from May 19, 2014, to June 1, 2014, to validate the model and aggregate the information at a daily level. We aggregated the information on a daily basis for two reasons: (i) Managerially, this retailer replenished most of the products in a period that was longer than one day. Thus, it makes less sense to have shorter periods if managers cannot react to them. (ii) In the case of shorter periods, less information is available, and therefore, it is more difficult to discern when a zero-sale period corresponds to an OOS or if it actually reflects normal behavior. We also investigated a half-day period; however, the performance of the models was significantly worse.

We considered the number of transactions that involve product i at store j and day t (n_{ijt}) and the total number of transactions at store j and day t (N_{jt}). We used incidence data (whether the ticket contained a product) instead of sales data (the number of units purchased) because incidence summarizes the availability of the product on the shelf. We studied $P = 14$ products in $J = 10$ stores in the same market area. We selected the products with the help of the company to

Table 1. Descriptive Statistics: Daily Incidence of Products, Heterogeneity, and Observed Out-of-Stock Across Stores

Product	Average incidence (%)	Heterogeneity ratio	Average observed out-of-stock (%)
Canned mackerel	0.226	3.889	9.375
Canned tuna	0.297	2.136	13.043
Cheese	0.308	3.222	2.439
Chocolate	0.282	4.273	47.561
Dish detergent	0.766	4.031	4.800
Frozen potatoes	0.151	2.750	0.820
Lasagna	0.177	3.444	9.735
Milk	0.679	2.529	3.968
Paper towels	0.975	3.451	0.833
Juice powder	0.388	6.286	5.769
Shrimp	0.282	2.353	4.425
Soda	0.337	3.467	0.769
Sugar	0.322	4.333	4.839
Tea	0.746	2.780	3.419
Mean	0.424	2.803	7.985

Note. Heterogeneity ratio = max(incidence)/min(incidence) across stores.

cover a wide purchase-incidence range of products. The stores were selected by the company so that all of them were present in the same city (market area). In this way, the same distribution center supplies all these stores. In addition, unobserved factors could be better controlled (those that would most likely affect all stores, making the results more comparable). All 10 stores represent a subset of the total number of stores within the market area with comparable sizes and assortments.

Table 1 describes the average daily incidence ($\frac{n_{ijt}}{N_{jt}} \times 100$) for the products studied. It can be seen that, among these products, the product with the highest average incidence is paper towels with 0.975, which means that approximately one of every 100 customers purchased the product. At the other extreme, among the products studied, the product with the lowest average incidence corresponds to frozen potatoes; approximately 1.5 of every 1,000 customers purchased the product.

In addition, Table 1 shows some degree of heterogeneity across stores for some products. The column for heterogeneity ratio shows the ratio between the maximum and minimum incidence across stores for each product. It can be seen that, in some cases, the maximum incidence is about six times larger than the minimum across stores (juice powder). However, on average across products, the store with the highest incidence is about three times larger than the store with the lowest incidence. Finally, the column for average observed OOS reports the observed OOS information obtained through visual inspection, which is described in the next section. It corresponds to the percentage of the measured occasions in which the product was not found on the shelf. It can be seen that, on average, the OOS rate is about 8% but with substantial heterogeneity. For instance, chocolate presented 48% OOS, whereas frozen potatoes showed as low as an 0.8% OOS on average across stores during the validation period.

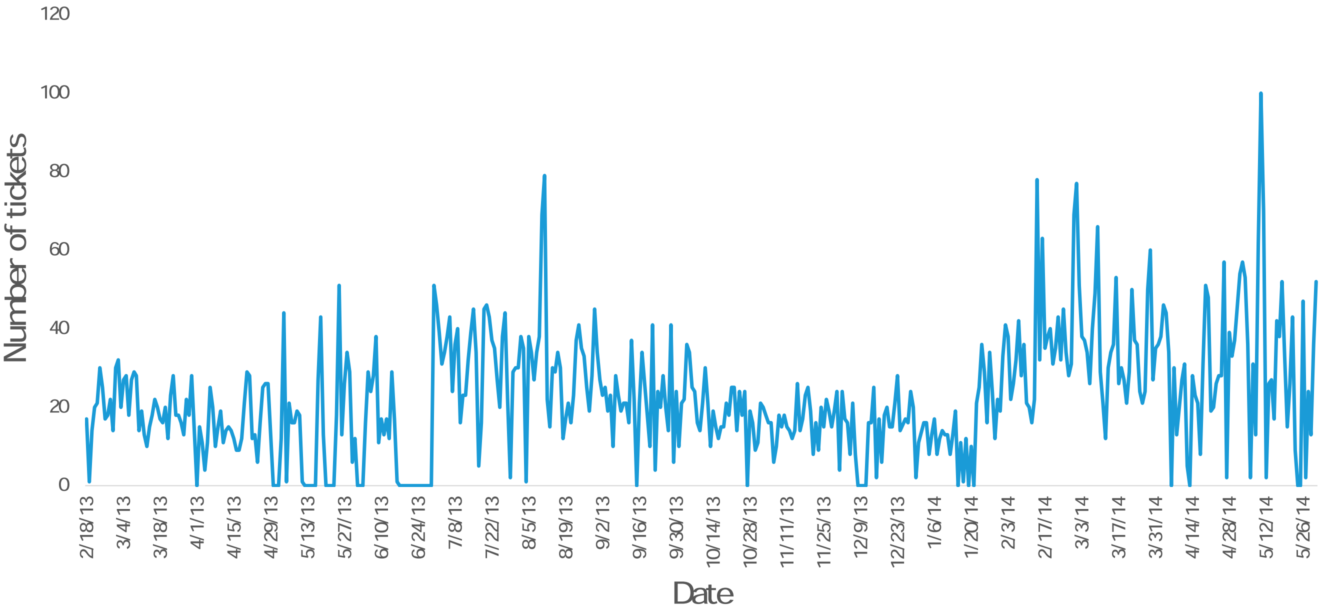
Before proceeding with a further description of the data and model, it is instructive to take a look at the sales data pattern for a particular SKU.

Figure 1 shows the sales time series for an illustrative product and store. It can be seen that sales show variability over time, including periods of zero sales. These periods of zero sales can vary in their duration and can be preceded by periods of either low or high demand. The goal of the HMM is to detect such fluctuations in demand and uncover the underlying state that may motivate such observed sales' behavior. It is also important to control for other factors that may affect sales but are not directly related to OOS, such as seasonality, trends, and price promotions. There are other factors that can also affect sales, such as space allocation or on-shelf product popularity (scarcity effect). Indeed, several studies have shown previously that the space allocated to the products, the location on the shelf, or how popular the products are (inferred from the percentage of units available on the shelf) can induce the purchase of the product (see Gierl et al. 2008 and the references therein). Because we do not have access to such information, these factors are not included in the model.

3.2. On-Shelf Availability Data

We have daily on-shelf availability information collected for the same products and stores from May 19, 2014, to June 1, 2014. Note that because the company did not register this information on a regular basis, we collected this information to evaluate the performance of the model. This information, based on visual inspections, determines whether the product was available during the day at each particular store. Specifically, 10 different inspectors checked exactly the

Figure 1. (Color online) Sales Time Series for an Illustrative Product



same SKUs visually across the 10 stores every day in the same range of time during the inspection period. To avoid partial OOS, when a SKU has multiple locations in the store and there is OOS in some of these locations, the chosen products were not on promotional display during the inspection period. The collected OOS information will be used to validate the predictive ability of the developed approach to detect those shelf OOS.

4. Hidden Markov Model for OOS

In this section, we describe the use of an HMM to capture dynamics in sales of a product resulting from partially observable demand implied by on-shelf out-of-stocks. An HMM is a Markov process with unobserved states. In our application, the hidden states represent the underlying state of the demand that can be implied by the state of the shelf. That is, different states capture the different unobserved levels of demand, and one of the states represents the truncated demand implied by OOS. For instance, let us assume three states. The first state represents the OOS condition, and we expect to observe no sales if the demand is in such state. The other two states may capture low and high levels of demand that do not suffer from OOS.

The demand transitions stochastically among these states through a Markovian first-order process. The transitions represent the probability of switching from one state to another. These probabilities capture intrinsic shifts in demand (e.g., from low demand to high demand) as well as how quickly the store reacts from an OOS (e.g., from OOS to low or high states). In addition, the probabilities of moving from low or high states to the OOS state may capture the performance of the store regarding how frequently it moves from regular demand (either low or high) to an OOS. This may reflect the performance of the replenishment or demand-forecasting systems. To separate intrinsic shifts between the states of the conditional demand and shifts caused by price promotions, we study the effect of price promotions in the transition matrix. Demand can also increase because of short-term promotions. These short-term changes are captured in the conditional component of the HMM.

Let n_{ijt} be the number of tickets that contain the product i at store j on day t. In the HMM, the joint probability of a sequence of observations up to time t $\{n_{ij1}, \ldots, n_{ijt}\}$ is a function of three main components: (i) the initial hidden states probabilities (π_{ij}), (ii) a sequence of transition probabilities among the states of demand (Q_{ijt}), and (iii) a set of incidence probabilities conditioned on the (truncated) demand states (M_{ijt}). We describe our formulation of each of these components next.

4.1. Initial State Probabilities

Let s denote a demand state ($s = 1, \ldots, S$). Let π^s_{ij} be the probability that the demand for product i at store j is initially in state s, where $\pi^s_{ij} \geq 0$ and $\sum_{s=1}^{S} \pi^s_{ij} = 1$. This probability can represent, for example, the typical performance of the store regarding in-stock and fill rates. One can prespecify these probabilities or estimate them from the POS data. For instance, one can specify equal probabilities for each state or another distribution based on a theoretical basis. Another option is to impose the steady-state probabilities derived from the transition matrix as the initial probabilities. We tried different specifications, but all of them yield similar results as is expected in cases of long time series (MacDonald and Zucchini 1997). Therefore, we assumed equal initial probabilities across states.

4.2. Transition Matrix

The transition matrix Q_{ijt} describes transitions among the states after the first period. Let $X_{ijt} \in \{1, \ldots, S\}$ denote product i's demand state at time t and store j. Then each element of the transition matrix can be written as

$$q_{ijt}^{s's} = P(X_{ijt} = s \mid X_{ijt-1} = s', \text{price}_{ijt}), \tag{1}$$

where $q_{ijt}^{s's} \geq 0, \sum_{s=1}^{S} q_{ijt}^{s's} = 1$ and price_{ijt} corresponds to the price of product i in store j at time t. Thus, the propensity to transition from one state to another is a function of unobserved factors that can be captured by a random-effect coefficient and the effect of price (price_{ijt}). The inclusion of price in the transition matrix helps us to disentangle the transition pattern resulting from unobserved factors, such as replenishment and fill-rate policies that affect the probability of moving from one state to another and the transition caused by price promotions. To estimate such effects properly, we mean center the variable price_{ijt} to capture the effect of deviations from the regular price. (Alternatively, we can center this variable with respect to the most frequently observed price. This change does not provide significant differences in the results.)

We follow Netzer et al. (2008) and Montoya et al. (2010) in parameterizing the nonhomogeneous hidden-state transitions as an ordered logit model. In our formulation, the nonhomogeneous HMM is affected by the product's price, which affects the dynamic behavior observed for this product. Thus, the transition probabilities are given by

$$\begin{aligned}
q_{ijt}^{s1} &= \frac{\exp(\hat{\tau}_{ij}^{s1} - \rho_{ij}^{s}\text{price}_{ijt})}{1 + \exp(\hat{\tau}_{ij}^{s1} - \rho_{ij}^{s}\text{price}_{ijt})}\\
q_{ijt}^{s2} &= \frac{\exp(\hat{\tau}_{ij}^{s2} - \rho_{ij}^{s}\text{price}_{ijt})}{1 + \exp(\hat{\tau}_{ij}^{s2} - \rho_{ij}^{s}\text{price}_{ijt})} - \frac{\exp(\hat{\tau}_{ij}^{s1} - \rho_{ij}^{s}\text{price}_{ijt})}{1 + \exp(\hat{\tau}_{ij}^{s1} - \rho_{ij}^{s}\text{price}_{ijt})}\\
&\vdots\\
q_{ijt}^{sS} &= 1 - \frac{\exp(\hat{\tau}_{ij}^{sS-1} - \rho_{ij}^{s}\text{price}_{ijt})}{1 + \exp(\hat{\tau}_{ij}^{sS-1} - \rho_{ij}^{s}\text{price}_{ijt})},
\end{aligned} \tag{2}$$

where $\{\hat{\tau}_{ij}^{ss'}, s' = 1, \ldots, S\}$ are thresholds that represent the areas of switching and capture regular transition patterns and ρ_{ij}^{s} captures the effect of price on the propensity to transition from state s to other states. To properly characterize the thresholds, we impose $\hat{\tau}_{ij}^{s1} \leq \hat{\tau}_{ij}^{s2} \cdots \leq \hat{\tau}_{ij}^{sS}$ by $\hat{\tau}_{ij}^{s1} = \tau_{ij}^{s1}, \hat{\tau}_{ij}^{s2} = \hat{\tau}_{ij}^{s1} + \exp(\tau_{ij}^{s2}), \ldots, \hat{\tau}_{ij}^{sS} = \hat{\tau}_{ij}^{sS-1} + \exp(\tau_{ij}^{sS})$.

4.3. Conditional Observed Incidence

Conditional on being in state s in period t, we assume that the number of tickets that contain the product i at store j (n_{ijt}) follows a binomial distribution with parameters N_{jt} and p_{ijts}. That is,

$$P_{ijts}(n_{ijt} | X_{ijt} = s, \mathbf{z}_{ijt}) = \binom{N_{jt}}{n_{ijt}} p_{ijts}^{n_{ijt}} (1 - p_{ijts})^{N_{jt} - n_{ijt}}, \tag{3}$$

where N_{jt} corresponds to the total number of tickets at store j and period t, and N_{jt} helps to capture the overall demand effects at the store level and, thus, is treated as exogenous. To capture the different levels of (partially) observed demand, we model the incidence conditional probability p_{ijts} as

$$p_{ijts} = \frac{1}{1 + \exp(-(\hat{\alpha}_{ijs} + \beta_{ijs}'\mathbf{z}_{ijt}))}, \tag{4}$$

where $\hat{\alpha}_{ijs}$ is an intercept that corresponds to the probability that a random customer includes product i in the customer's basket during the customer's visit to the store j at time t if the conditional demand is in state s and $\mathbf{z}_{ijt}$ includes covariates that affect demand, such as day of the week, holidays, month, and price. Following standard notation in HMMs (MacDonald and Zucchini 1997, Netzer et al. 2008, Montoya et al. 2010), we write the vector of state-dependent incidence probabilities as a diagonal matrix $\mathbf{M}_{ijt}$.

To ensure the identification of different states and mitigate the label-switching problem (Jasra et al. 2005), we imposed a nondecreasing order of the intercepts ($\hat{\alpha}_{ij1} \leq \hat{\alpha}_{ij2} \leq \ldots \leq \hat{\alpha}_{ijS}$) by setting $\hat{\alpha}_{ij1} = \alpha_{ij1}, \hat{\alpha}_{ij2} = \hat{\alpha}_{ij1} + \exp(\alpha_{ij2}), \ldots, \hat{\alpha}_{ijS} = \hat{\alpha}_{ijS-1} + \exp(\alpha_{ijS})$.

There are several advantages for using the binomial distribution in the current application. First, accounting for store volume (N_{jt}) allows us to control for exogenous variations in customer visits to the store that affect overall demand unrelated to variations in demand for specific products. For example, consider an increase in overall demand because of a holiday season. A volume model, such as a Poisson model, would attribute such an increase to a change in the state of demand or to marketing actions specific to the product and might overestimate such an impact. Second, store volume helps to control for seasonal or time-specific effects that may affect the market. Third, the binomial distribution can easily handle zero-incidence situations ($n_{ijt} = 0$), which provides key information in our application. Indeed, empirically, an equivalent HMM with conditional Poisson distribution performs consistently worse than the chosen binomial distribution.

4.4. Out-of-Stock State

Recall that the hidden states represent different levels of partially observable demand. To identify the OOS, we constrain one of the states to represent such a condition (we denote this OOS state by $s = 0$). The rationale behind this is that we expect that an OOS implies a serious impact on the product's sales, substantially reducing sales from their regular levels. In the extreme case, we expect to observe zero sales when the product is unavailable. However, given that an OOS may occur at any time during the day and because it is possible to observe sales during the day, therefore, the probability of observing tickets that include the product is not necessarily equal to zero. Consequently, we allow for nonzero probability (albeit very small) in the OOS state. Thus,

$$P(n_{ijt}|X_{ijt}=0)=\binom{N_{jt}}{n_{ijt}}\varepsilon^{n_{ijt}}(1-\varepsilon)^{N_{jt}-n_{ijt}}, \tag{5}$$

where ε is a very small probability (yet not equal to zero) to be determined empirically. That is, we estimate the models for different values of ε and choose the one that is small enough to discriminate among the states and large enough to avoid numerical underflow (see Section 5.3.1). In addition, we tested our constrained specification of an OOS state empirically and estimate an unconstrained HMM in which the OOS state is not imposed. In such a specification, all states are free (see Section 5.1).

4.5. Bayesian Hierarchy and Estimation

Let $(n_{ij1},\ldots,n_{ijt},\ldots,n_{ijT})$ be a sequence of T daily tickets for store j and product i. Given the structure of the model, the likelihood function for product i considering a set of J stores can be written as

$$L_i=\prod_{j=1}^{J}P(n_{ij1},\ldots,n_{ijt},\ldots,n_{ijT})=\prod_{j=1}^{J}\pi_{ij}'\mathrm{M}_{ij1}\prod_{t=2}^{T}\mathrm{Q}_{ijt}\mathrm{M}_{ijt}1, \tag{6}$$

where $\mathbf{1}$ is an S $\times$ 1 vector of ones. We take advantage of the fact that the same product is present in all stores of the same market, and consequently, we pool the information across the stores. However, we allow for heterogeneity at the store level to permit different baseline sales, sensitivity to promotions, and the effects of controls using a hierarchical Bayesian approach. Given that the products studied are quite diverse, we do not pool information across products but, instead, estimate a completely independent model for each product. To include heterogeneity across stores, we specify the HMM parameters at the store level and use a hierarchical Bayesian Markov chain Monte Carlo (MCMC) procedure for parameter estimation. Specifically, $\Phi_{ij}=\{\hat{\tau}_{ij}^{ss'},\rho_{ij}^{s},\alpha_{ijs},\beta_{ijs}\}\forall s,s'$ is the set of parameters for a product i and store j. We capture unobserved heterogeneity across stores with the distribution of Φ_{ij} by allowing a multivariate normal distribution with mean Φ_{i0} and variance $V_{i\Phi}$. That is, $\Phi_{ij}=\Phi_{i0}+\eta_{ij}$, where $\eta_{ij}\sim N(0,V_{i\Phi})$. The hyperparameters Φ_{i0} and $V_{i\Phi}$ have a multivariate normal and inverse Wishart distributions, respectively. Finally, we derive the full conditional distributions of the unknowns using the likelihood function (Equation (6)) and the specified prior distributions. As the full conditional distributions do not have closed forms, we use a Metropolis–Hasting procedure to derive the posterior distributions. See Online Appendix A for the specification of the priors and full conditional distributions.

4.6. Inferring Latent States

One key property of the HMM formulation is that, after observing sales, we can infer the state of the demand for product i and store j probabilistically at each time t. Specifically, given the parameter estimates and sales up to time t, we use the *filtering* approach (MacDonald and Zucchini 1997) to calculate the probability that demand of product i at store j is in state s at time t. This filtering probability is given by

$$P(X_{ijt}=s|n_{ij1},\ldots,n_{ijt})=\boldsymbol{\pi}_{ij}'\mathbf{M}_{ij1}\prod_{\tau=2}^{t}\mathbf{Q}_{ij\tau}\mathbf{M}_{ij\tau}^{s}/L_{ijt}, \tag{7}$$

where $\mathbf{M}_{ij\tau}^{s}$ is the sth column of the matrix $\mathbf{M}_{ij\tau}$ and $L_{ijt}=\boldsymbol{\pi}_{ij}'\mathbf{M}_{ij1}\prod_{l=2}^{t}\mathbf{Q}_{ijl}\mathbf{M}_{ijl}\mathbf{1}$ is the likelihood of the observed sequence of tickets that include the product i at store j up to time t.

4.7. Performance Metrics

We develop an HMM to identify OOS using POS data. To characterize the performance in identifying the OOS we use type I error, false alarms, and predictive power metrics. In addition, we evaluate the model fit and predictive ability to the sales data. For that purpose, we use log-marginal density (LMD) and validation log-likelihood, respectively. Explanations of all these metrics follow.

4.7.1. Identifying OOS. The focus of the current research is on identifying the unobserved OOS state. For this, we use Equation (7) to predict at which state the product is at each period of time. We assign the state of the demand to the state with maximum probability. We generate alarms when the system is predicted to be in the OOS state (we analyze alternative rules for generating alerts when the system is in the OOS state, see Online Appendix B). To evaluate the performance of the classification algorithm, we consider the confusion matrix illustrated in Figure 2.

From this matrix, we define the following metrics (Montgomery 2013):

Type I error is the percentage of observations that are not OOS but that the model classifies as such. We want it to be as close to zero as possible. This indicator is calculated as $\frac{\text{False Positives}}{\text{N}'}$.

False alarms are the percentage of instances that the model classifies as OOS even though they are not OOS.

Figure 2. Confusion Matrix to Evaluate the Classification Performance of the HMM Model

		prediction outcome		
		p	**n**	**total**
actual value	**p′**	True Positive	False Negative	P′
	n′	False Positive	True Negative	N′
	total	P	N	

We want it to be as close to zero as possible. This indicator is calculated as $\frac{\text{False Positives}}{\text{P}}$.

Power of detection is the percentage of OOS that the model identifies correctly. We want it to be as close to one as possible. This indicator is calculated as $\frac{\text{True Positives}}{\text{P}'}$.

These classification metrics are the key performance indicators for the proposed approach. Indeed, the retailer's goal is obtaining high detection power with low false alarms. However, there is an important trade-off between these two metrics because high detection power could imply low performance regarding false alerts. This occurs because, to detect as many OOS as possible, the system should send alerts even when the probability of such events is not high enough. This implies that, in many cases, the system could classify the state as an OOS incorrectly, which would increase the false-alert metric. In contrast, if the system focuses on low false-alert rates, the system would maximize this indicator by sending alerts only when there is a high probability of such an event. This conservative behavior could leave many OOS undetected.

4.8. Benchmarks

4.8.1. Approach Used by the Retailer. We consider the methodology the company was using to detect OOS during the time horizon that covers our data as a benchmark. This methodology uses a statistical process control approach that alerts when the system is "out of control." This occurs when sales fall below certain levels defined by specific thresholds (control limits). These thresholds are determined by first fitting a sales-incidence model using historical transactional data controlling for trend and seasonality and then using the model's residuals such that a low percentage of them fall below those thresholds (e.g., $\alpha = 1\%$ of the observations). Therefore, if the system is always in control, by design, this approach would have a type I error of α (Montgomery 2013).

To calibrate their forecasting model, the company uses the same data (POS, price, and controls for seasonality) as we do for the proposed HMM. And, similar to our approach, the company defines daily periods for monitoring the products at each store. One important difference between the company's approach and the methodology proposed in this paper is that, to implement their approach, the historical data need to be cleaned from OOS. This is because, to detect OOS in the validation sample, their forecast model needs to be calibrated with data that reflect a system that is always in control (one in which there are no OOS). A failure in doing so may cause their model to neglect identifying substantive drops in sales as OOS because such behavior would be considered to be normal. The complexity of eliminating OOS from historical data are that most OOS are unobserved and are rarely recorded by the company. This problem has been dealt with in various forms in previous research. For instance, Chuang et al. (2016) removed all historical data with zero sales. Such a procedure may overestimate OOS because it would probably assume that a zero-sale period corresponds to an OOS. Thus, applying that procedure will produce a high false-alert rate (see Chuang et al. 2016, table 4 on p. 943, which implies a false alarm rate of 40%, power 60%, and type I error 13%). The company implemented an ad hoc procedure in which they first used the historical purchase incidence (p) for each product and store to compute the probability of observing zero sales when N tickets have been received during the time period ($= (1-p)^N$). If this probability is lower than γ, the observation corresponding to such a period is eliminated from the historical data. If this condition is satisfied by considering more than one zero-sale period, all consecutive periods involved are removed. They evaluated different values for γ and used the one that yielded the best results, considering the performance metrics across products. In contrast, our HMM does not require a data-cleaning procedure as the model is built assuming the existence of unobserved OOS in the historical data at least for one of the stores that was analyzed for a particular product.

4.8.2. Other Benchmarks. In addition, we consider the following three benchmarks:

1. Zero-sale heuristic: we used a simple decision rule based on the assumption that zero sales are caused exclusively by on-shelf OOS. For this model, an alarm is triggered only if there are zero sales for the product–store during the day.

2. Binomial model: we estimated a one-state HMM, in which a binomial distribution characterizes the number of tickets containing a certain product. Given that this HMM has only one state (strictly speaking, there is no hidden state in this formulation), we needed a different criterion to trigger an alarm. In this case, and similar to Chuang et al. (2016), an alarm is

triggered if the observed *number of sales* corresponds to a low probability event ($p(\text{sales}) < \beta$). In particular, we implemented $\beta = 0.01$ and 0.05.

3. Binomial zero-sale (BZS) model: we followed the approach proposed by Chuang et al. (2016) and estimated a binomial model to determine the probability of observing *zero sales* during Z consecutive days. If this probability is lower than β for Z consecutive days, then we trigger the OOS alarm. In particular, we implemented $\beta = 0.01$ and 0.05 and $Z = 1$ and 2 days. Note that Chuang et al. (2016) use a negative binomial (NB) distribution to fit the sales data. The NB is also known as a gamma–Poisson model, in which the heterogeneity in the Poisson parameter follows a gamma distribution. Instead of incorporating heterogeneity in this way and estimating such a model using a maximum-likelihood method, we use a hierarchical Bayesian approach to allow for heterogeneity and an MCMC approach to estimate the models.

5. Results

In this section, we report the results of estimating the HMM model and benchmarks for detecting OOS using POS data. We used the historical data to calibrate the model and visual inspections for validation purposes. We ran the hierarchical Bayes estimation for 500,000 iterations for each product independently. The first 300,000 iterations were used as a "burn-in" period, and the last 200,000 iterations were used to estimate the conditional posterior distributions. Convergence was assessed by monitoring the trace plots of the MCMC output. Estimating the model's parameters for each product takes less than two hours and, as this estimation is independent for each product, this task can be performed in parallel. Once the model is estimated, inferring the state of the shelf for all products and stores during the validation period takes less than one minute.

5.1. Model Selection

We tested various specifications of the proposed HMM (i) regarding number of states, (ii) including an OOS state, and (iii) allowing for nonhomogeneous transitions. To infer the number of states that best capture the dynamic evolution of the demand and best identify the OOS states, we estimated models with one, two, three, and four states. In addition, we investigated relaxing the need for a constrained OOS state. This allows for a more flexible model that fits the data better but requires estimating more parameters. More importantly, the lack of structure of this flexible HMM formulation may hurt the identification of the OOS states. Finally, we also studied the contribution of allowing for nonhomogeneous transitions by removing price from the transition matrix (see Equation (2)). This resulted in 13 different model variants.

We analyzed the performance of the models along two dimensions. First, we turned our attention to the focus of the current research regarding the identification of the OOS state. We want the model to be able to identify OOS states with high power and low type I error and false alerts. Next, we explored the models' performance regarding the sales data. We want the models to be able to capture dynamics in sales. This is not the focus of our research, but the models should be able to represent sales properly. Following Bayesian research, we used the log-marginal density computed on the calibration data to evaluate model fit to the sales data. The marginal density (likelihood) corresponds to the likelihood of the data marginalizing over the parameters (Sorensen and Gianola 2007). In Bayesian statistics, it is approximated using the MCMC draws (Chib and Jeliazkov 2001). The LMD is the log of such a function, and it is typically a negative value. Thus, the higher the value, the better the model represents the data. In addition, to evaluate the predictive ability of the different specifications, we included out-of-sample log-likelihood (Gelman et al. 2014). Similarly to LMD, a higher value indicates a higher predictive performance.

Table 2 summarizes the performance of the HMMs regarding the identification of the OOS as well as the fit to and the prediction ability of the sales data.

Table 2 shows several results. First, considering overall performance, we observe that the models that do not specify an OOS state perform badly regarding type I error and false alarms. This poor performance may prevent the use of these specifications. Next, allowing for nonhomogeneous transitions has mostly a positive effect on reducing type I error and false alarms and on increasing detection power (its biggest contribution is in detection power). Finally, as the number of states increases, the performance of fit and predictive ability generally improves. Within models with the same number of states, constraining the HMM to have an OOS tends to deteriorate penalized fit (LMD), but in contrast, it tends to improve predictive ability (out-of-sample log-likelihood). Based on all these performance metrics, we selected the model specification with three states that incorporates an OOS and has dynamic transitions. Such a model provides a type I error of 0.85%, a false-alarm rate of 15.12%, and a detection power of 63.48% on average across products. In summary, we identified three sources of gains for the selected specification: (i) imposing an OOS structure is essential to capture OOS; (ii) three states describe the demand well and simultaneously give flexibility to capturing transitions to the OOS state and detecting OOS; and (iii) allowing for prices to influence transitions allows moving to the OOS, separating common transitions from those motivated by promotions.

Table 2. Hidden Markov Model Selection

Number of states	Out-of-stock state	Nonhomogeneous transitions	Type I error	False alarms	Detection power	log-Marginal	Out-of-sample log-likelihood
1	No	No	23.51	86.30	49.57	–316,050	–10,843
2	No	No	43.35	92.85	44.35	–283,970	–10,404
2	No	Yes	52.32	91.52	64.35	–275,130	–10,770
2	Yes	No	2.03	28.97	66.09	–300,500	–13,334
2	Yes	Yes	1.96	28.57	65.22	–311,370	–13,675
3	No	No	26.92	88.39	46.96	–217,380	–8,928
3	No	Yes	27.90	88.38	48.70	–268,640	–9,507
3	Yes	No	0.98	17.86	60.00	–222,840	–8,094
3	Yes	Yes	0.85	15.12	63.48	–222,720	–8,424
4	No	No	25.34	93.48	23.48	–215,720	–8,867
4	No	Yes	17.94	94.16	14.78	–216,970	–8,915
4	Yes	No	0.92	17.50	57.39	–217,330	–8,488
4	Yes	Yes	0.85	16.25	58.26	–217,520	–8,444

Notes. Out-of-sample performance metrics: type I error, false alarms, detection power. Model fit and predictive ability of sales: log-marginal density and out-of-sample log-likelihood.

5.2. Benchmark Comparison by Purchase Incidence

To analyze the strengths and weaknesses of the different models, we segmented the 14 products into three different groups based on their average purchase incidence. The high group contains four products with an average incidence higher than 0.6%. The medium group has five products and an average incidence between 0.3% and 0.6%. Finally, the low group has five products with an average incidence lower than 0.3%. Among the top 1,000 products for this retailer (considering purchase incidence), high-, medium-, and low-incidence products represent 13%, 30%, and 57%, respectively. We compared the results of the chosen model to the benchmarks described in Section 4.8. Recall that, as described in Section 4.8.2, for the binomial and binomial zero sales models, we tested two different values of β (1% and 5%). In addition, for the BZS model we also tested two different values of Z (one and two), representing the number of consecutive days with zero sales. Table 3 presents the results obtained by all benchmarks across the three groups of products.

We note that despite its simplicity, the zero-sale rule detects a high percentage of OOS (54.78%). However, this power of detection is achieved by triggering a high number of false alarms (37%), making this approach unreliable. This situation is similar for the binomial model, whose false-alarm rate is higher than 75% across all groups.

The BZS model yields low type I error and false-alarm rates, which makes it a good candidate for OOS detection. However, the power of detection of OOS is lower than the best-performing models. Both versions of the model (Z = 1, 2) maintain a low percentage of errors and a power of detection between 30% and 40% across the three product groups—results that are consistent with the 37% obtained by Chuang et al. (2016) with a similar approach. The 5% threshold

Table 3. Benchmark Comparison

Performance metric	Incidence group	Zero sales	Binomial 1%	Binomial 5%	Binomial zero sales 1% Z = 1	Binomial zero sales 1% Z = 2	Binomial zero sales 5% Z = 1	Binomial zero sales 5% Z = 2	Statistical process control	Hidden Markov model
Type I error	High	1.27	32.42	78.39	0.21	0.21	0.42	0.21	0.42	0.42
	Medium	2.48	16.99	44.60	1.06	0.53	1.24	0.53	2.65	1.42
	Low	3.47	22.45	42.86	0.61	0.00	0.82	0.00	1.84	0.61
	All	2.42	23.51	54.49	0.65	0.26	0.85	0.26	1.70	0.85
False alarms	High	42.86	93.87	96.10	25.00	11.11	22.22	11.11	15.38	15.38
	Medium	38.89	87.27	92.99	31.58	14.29	28.00	13.64	38.46	25.00
	Low	34.00	76.92	83.33	13.64	0.00	16.00	0.00	23.08	7.32
	All	37.00	86.30	91.63	22.22	9.30	22.03	7.69	28.57	15.12
Power of detection	High	50.00	62.50	93.75	18.75	50.00	43.75	50.00	68.75	68.75
	Medium	70.97	45.16	61.76	41.94	58.06	58.06	61.29	77.42	77.42
	Low	48.53	48.53	72.34	27.94	19.12	30.88	30.88	44.12	55.88
	All	54.78	49.57	66.09	30.43	33.91	40.00	41.74	56.52	63.48

Note. Products are grouped based on purchase incidence.

seems to work better for both values of Z because it detects a higher percentage of OOS without compromising the false-alarm performance ($\beta = 5\%$ imposes a less-strict condition on sales behavior to be considered abnormal, which generates more alerts). In particular, for two consecutive days ($Z = 2$), the 5% threshold improves the results of the model, detecting a greater number of OOS and keeping the number of false alarms under control. Note that the power of detection of the BZS model has an upper bound given by the power of detection of the zero-sale model. This is because the BZS is more conservative with the alerts because it requires $p < \beta$ to trigger an alert. In contrast, the zero-sale model does not restrict the alerts and, therefore, is able to detect more OOS.

In the BZS model, the increase in the number of consecutive days with zero sales (Z) has two opposite effects. First, the model becomes more conservative as it requests two consecutive days of zero sales ($Z = 2$) instead of one ($Z = 1$) to generate a candidate of OOS that restricts the number of alerts and decreases, as a consequence, the number of false alerts. Second, in contrast, the model becomes less conservative regarding the threshold for generating the alarm for the candidate of OOS. To understand this fact, recall that in this model an alarm is triggered when P(observing two zero-sale days) $< \beta$. In this case, we assume that P(observing two zero-sale days) = P(zero-sale day 1) P(zero-sale day 2). Thus, this independence between days increases the probability that two consecutive days satisfy the criterion ($p < \beta$) that allows generating more alarms and, as a consequence, detecting more OOS. Empirically, the results of this model improve when we request two consecutive days with zero sales because it is more likely that this situation is produced by an on-shelf OOS.

The proposed HMM strictly dominates the results of the statistical process control model used by the company on average across products. In particular, we notice that this superior performance occurs in the medium- and low-incidence groups; for the high-incidence group, both approaches obtain the same results. In fact, the HMM's power of detection outperforms each of the benchmarks except for the binomial model with the 5% threshold. Compared with the statistical process control model, the largest gain of the proposed HMM approach (besides its decrease in false alarms) is obtained for the low-incidence group. Indeed, because of their low presence in daily tickets, it is harder to discriminate if low sales correspond to normal sales behavior or if they correspond to OOS situations. Therefore, an increase in the power of detection of this group is an important result because, as mentioned before, most of the products that the company offers belong to this group of products. Finally, for the HMM approach (and also for most benchmarks), the best performance in terms of detection of OOS is obtained for medium-incidence products. It is likely that these products present OOS earlier in the day, affecting the sales more strongly, which allows the models to detect such a change in sales. Indeed, in our data, the products with higher incidence tend to have a lower number of OOS.

Note that we have developed each approach independently with parameters typically used in past research. However, to get a direct comparison between the HMM and the BZS models, we explored different threshold parameters to classify OOS for both HMM and BZS such that the corresponding performance indicators line up on one of these metrics. The results show that for the same type I error (0.85%), the HMM outperforms the BZS ($Z = 1$) model on both false alarms (15.12% versus 23.21%) and power of detection (63.48% versus 37.39%). Similarly, for the same false alarm rate (14.29%), the HMM strongly outperforms the BZS ($Z = 2$) model on power of detection (62.61% versus 10.43%), whereas the BZS ($Z = 2$) model slightly outperforms the HMM on type I error (0.13% versus 0.79%). These results allow us to support the finding that the HMM provides a superior performance.

The superior performance of the HMM may yield important gains for the retailer. First, let us analyze the power of detection. If 1,000 products are monitored, assuming an average OOS rate of 10% would imply 100 products OOS on a given day. A 1% extra detection would imply detecting one additional OOS instance. If we assume that the OOS would last only one day if detected, average daily sales of the entire chain would imply approximately \$6.9MM in a year for a retailer with 400 stores. For this computation, we considered an average incidence of 0.424 every 100 sales and an average price of \$2.24 for the analyzed products. Thus, for a store with 5,000 sales per day, the average daily revenue for one product is \$47.41 = \$2.24 × 5,000 × 0.424/100. Next, the annual revenues for the whole chain with 400 stores for an average product is \$6,921,552 = \$47.41 × 365 days × 400 stores. Consequently, as the proposed HMM yields an approximately 7% higher detection power than the model used by the company, this yields about \$48.45MM (= \$6.9MM × 7) in gains. Next, let us consider the type I error. Once again, if 1,000 products are monitored, assuming an average OOS of 10% would imply 100 products OOS (and 900 products not OOS) on a given day. A 0.1% type I error implies that 0.001 × 900 = 0.9 products are incorrectly classified as OOS. This would trigger an unnecessary replenishment for 0.9 products, which we assume involves 0.5 hours of the replenishment personnel. If we assume labor cost of \$10 per hour, a 0.1% lower type I error would yield \$0.66MM (0.9 products/store × \$5/(product-day) × 365 days × 400 stores) in savings. As the proposed HMM yields a 0.85% lower type I error than the model used by the company (0.85% versus 1.7%), this yields about \$5.58MM in savings. Finally, let us consider the false-alert error. First, the total alert rate (number of alerts/number of days) can be computed as follows. Let α, β, and γ be the type I, false alarm, and power of detection rates, respectively. Then, from the confusion matrix in Figure 2, it is easy to verify that the alert rate $= \frac{1}{\beta}/(\frac{1}{\gamma\beta} - \frac{1}{\gamma} + \frac{1}{\alpha})$. Then, if 1,000 products are monitored, considering an alert rate of 5.2% gives a total of 52 alerts. A 1% false alert implies that $0.01 \times 52 = 0.52$ products are incorrectly classified as OOS. This would trigger an unnecessary replenishment of about 0.52 products, which we assume involves 0.5 hours of the replenishment personnel. If we assume labor cost of \$10 per hour, a 1% lower false alert would yield \$0.38MM (0.52 products/store × \$5/(product-day) × 365 days × 400 stores) in savings. As the proposed HMM yields a 13.45% lower false alarm error than the model used by the company (15.12% versus 28.57%), this yields about \$5.14MM in savings. Therefore, considering all the performance indicators and because the false alerts are included in the alerts associated with the type I error, the HMM gives a total of \$54.03MM (= \$5.58MM + \$48.45MM) additional gains compared with the system in use by the company.

Note that we did not consider other related costs that affect the entire supply chain, such as retail or manufacturer substitution because of constant OOS (Corsten and Gruen 2003).

5.3. The HMM's Parameter Estimates

We now present the HMM estimates for the chosen model. Given that we estimate a separate model for each product, we report detailed results for only one product (dish detergent). In addition, to better illustrate the results, we report the analysis for only one store (store 1). We chose dish detergent mainly because its observed OOS level is close to the median across the analyzed products. In addition, its relatively high incidence may help us to illustrate the dynamics uncovered by the HMM. The analysis is equivalent for all products and stores.

As the raw parameter estimates are of less interest, we report the transformed parameters of the transition matrix given by Equation (2) and conditional state probabilities given by Equation (4). All covariates $\mathbf{z}_{ijt}$ in Equation (4) are fixed to zero, representing the base levels for the controls and mean price. In addition, we report the effect of price on the transition matrix and on the conditional probabilities (see Table 4).

5.3.1. Interpreting the States. Recall that the chosen model considers three states; the first corresponds to the OOS state, and thus, its purchase probability p_0 is not estimated. Accordingly, we tested different values for ε in Equation (5) ranging from 10^{-3} to 10^{-6}. The chosen value of 10^{-5} provides the best results for discriminating the OOS state from the other states and does not run into underflow problems when computing the likelihood in Equation (6). Next, in the low state, the purchase probability is 0.913%, whereas in the high state the purchase probability is 1.428%. We observe that the demand in the high state is 56% higher than the demand in the low state (0.913% versus 1.428%).

Table 5 shows the purchase probabilities for the different demand states at different prices. We note that the low demand state is more price sensitive than the high demand state. That is, when the system presents a low demand, an increase of 5% in the price would reduce the demand to close to zero, whereas a decrease of 5% in the price would leverage the demand to 29%. This might be because when dish detergent is at the low state (when we observe a lower number of sales at the regular price), an increase in demand because of a price discount may be explained by the arrival of more price-sensitive customers who are willing to buy the product when the price decreases. However, at the high state, when sales are already high at the regular price, there is less room to increase demand by decreasing the price. This may explain why some discounts are introduced when demand is low to boost demand.

5.3.2. Transition Dynamics. Table 6 shows the transition matrix for product dish detergent at store 1. We observe that all states are relatively sticky as the probability of staying at the state is higher than the probability of transitioning to another state. Specifically, the persistence probabilities are 0.744, 0.535, and 0.566 for the OOS, low, and high states, respectively. The probability of staying at the OOS state once the system is OOS is high although the probabilities of reaching such a state are low (0.022 and 0.013 from the low- and high-demand states, respectively). This may indicate relatively good precision in the forecasting system but poor performance of the restocking system. (We analyze these managerial aspects further in the next section). To analyze the effect of price on transition dynamics, we computed the transition matrix at different prices (see Table 6).

The top panel in Table 6 shows the transition matrix at the mean price. The transition matrix when there is a 5% reduction in price is in the middle panel. We observe that this decrease in price increases the probability of staying at the OOS state. However, if the system is at any of the other states, this price reduction increases dynamics among these states. Interestingly, if the system is at the high state, the price reduction increases the probability of moving to the OOS state. This is consistent with reports suggesting that promotions increase OOS. The bottom panel in Table 6 shows the transition matrix when there is a 5% increase in price. Price increases make the OOS slightly less sticky, whereas, if the system is at the other states, it tends to stay at the same demand states.

Table 4. Posterior Means, Standard Deviations, and 95% Confidence Intervals for Dish Detergent at Store 1

Parameter label	Posterior mean	Posterior standard deviation	2.5%	97.5%
Transition matrix				
q_{00}	0.744	0.046	0.640	0.827
q_{01}	0.162	0.028	0.115	0.230
q_{02}	0.093	0.039	0.027	0.185
q_{10}	0.022	0.009	0.010	0.044
q_{11}	0.535	0.040	0.454	0.612
q_{12}	0.441	0.040	0.365	0.521
q_{20}	0.013	0.004	0.006	0.022
q_{21}	0.421	0.044	0.344	0.514
q_{22}	0.566	0.043	0.473	0.643
Conditional probabilities				
p_0	0.00001	—	—	—
p_1	0.00913	0.00078	0.00770	0.01074
p_2	0.01428	0.00072	0.01293	0.01587
Effect of price on transitions				
ρ_0	0.232	0.532	−0.560	1.531
ρ_1	−0.684	1.240	−1.809	2.103
ρ_2	1.912	0.592	0.509	2.923
Effect of price on conditional probabilities				
β_1^p	−4.259	0.608	−9.282	−0.978
β_2^p	−0.290	0.327	−0.526	−0.130

Note. Recall that $s = 0$ represents the out-of-stock state.

5.4. Model Extensions

5.4.1. Inventory Data. Our HMM approach can accommodate additional information easily if it is available. For instance, one critical piece of information relates to the availability of inventory data that can complement the information already incorporated in the proposed HMM. To illustrate this extension, let δ_{ist} be a dummy variable that takes the value one if inventory is zero and zero otherwise. This variable can be incorporated in the transition matrix in Equation (2) similarly to the addition of other covariates. Replenishment data, if they are available, can be incorporated in the same way as inventory information.

5.4.2. Inspection Data. Inspection data can also be incorporated into our HMM framework in the following way. Given that this information reveals the true state of the system, one can incorporate such information by breaking the likelihood function into two parts: before and after the inspection. Thus, the first part of the sequence ends at the known state, and the second part of the sequence begins at the known state. Specifically, suppose that the visual inspection reveals that product i was OOS at store j on day t^*. Equation (6) can be modified to accommodate the known state at period t^*, in the hidden sequence, as follows

$$L_i = \prod_{j=1}^{J} \boldsymbol{\pi}'_{ij} \mathrm{M}_{ij1} \prod_{t=2}^{t^*-1} \mathrm{Q}_{ijt} \mathrm{M}_{ijt} \mathrm{q}_{ijt^*}^{\cdot 0} \mathrm{m}_{ijt^*}^{00} \mathrm{q}_{ijt^*+1}^{0 \cdot} \mathrm{M}_{ijt^*+1} \cdot \prod_{t=t^*+2}^{T} \mathrm{Q}_{ijt} \mathrm{M}_{ijt} 1, \tag{8}$$

where $\mathbf{q}_{ijt^*}^{\cdot 0}$ and $\mathbf{q}_{ijt^*+1}^{0 \cdot}$ indicate the first column and first row of the transition matrix, respectively, and $\mathbf{m}_{ijt^*}^{00}$ corresponds to the binomial distribution for the OOS state.

We have described the previous procedure for illustration purposes. However, the extension to more periods with known states can be done by introducing

Table 5. State-Specific Purchase Probabilities at Different Prices for Dish Detergent at Store 1

States		Purchase probability as a function of price		
		Mean price	5% decrease	5% increase
Out-of-stock	p_0	0.00001	0.00001	0.00001
Low	p_1	0.00913	0.29215	0.00020
		(0.00770, 0.01074)	(0.02145, 0.97694)	(0.00000, 0.00375)
High	p_2	0.01428	0.01850	0.01094
		(0.01293, 0.01587)	(0.01564, 0.02367)	(0.00904, 0.01307)

Note. In parentheses, 95% confidence intervals.

Table 6. Transition Probabilities at Different Prices for Dish Detergent at Store 1

	Out-of-stock	Low	High
Transition matrix at mean price			
Out-of-stock	0.74 (0.64, 0.83)	0.16 (0.11, 0.23)	0.09 (0.03, 0.19)
Low	0.02 (0.01, 0.04)	0.53 (0.45, 0.61)	0.44 (0.37, 0.52)
High	0.01 (0.01, 0.02)	0.42 (0.34, 0.51)	0.57 (0.47, 0.64)
Transition matrix at 5% decrease in price	Out-of-stock	Low	High
Out-of-stock	0.77 (0.61, 0.94)	0.15 (0.05, 0.22)	0.08 (0.01, 0.22)
Low	0.02 (0.00, 0.18)	0.40 (0.18, 0.81)	0.58 (0.10, 0.81)
High	0.07 (0.01, 0.21)	0.72 (0.52, 0.80)	0.19 (0.08, 0.47)
Transition matrix at 5% increase in price	Out-of-stock	Low	High
Out-of-stock	0.69 (0.53, 0.86)	0.19 (0.08, 0.35)	0.11 (0.05, 0.19)
Low	0.04 (0.00, 0.09)	0.63 (0.15, 0.80)	0.31 (0.12, 0.84)
High	0.00 (0.00, 0.01)	0.12 (0.05, 0.34)	0.88 (0.65, 0.95)

Note. In parentheses, 95% confidence intervals.

two additional matrices, A and B, such that we extract only the components of the transition matrix that are feasible given the information provided by the visual inspection. For instance, if the audit reveals that the product is OOS in period t, then $A_t = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix}$ and $B_t = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$ are such that $A_t \circ Q_t$ gives only the first column with nonzero values and $Q_{t+1} \circ B_{t+1}$ gives only the first row with nonzero values for the respective transition matrices; $\circ$ denotes element-wise multiplication. Note that, as seen in Equation (8), transitions after the visual inspection period also have to be constrained, such that only transitions from the known state are allowed. Alternatively, if the audit reveals that the product is available in period t, then $A_t = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{bmatrix}$ and $B_t = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$ because we know the product is not OOS, but the system can be at any of the other remaining states. Therefore, with these new matrices, the likelihood can be written as

$$L_i = \prod_{j=1}^{J} \pi_{ij}' \mathrm{M}_{ij1} \prod_{t=2}^{T} [\mathrm{A}_{ijt} \circ \mathrm{Q}_{ijt} \circ \mathrm{B}_{ijt}] \mathrm{M}_{ijt} 1. \qquad (9)$$

5.5. Managerial Insights

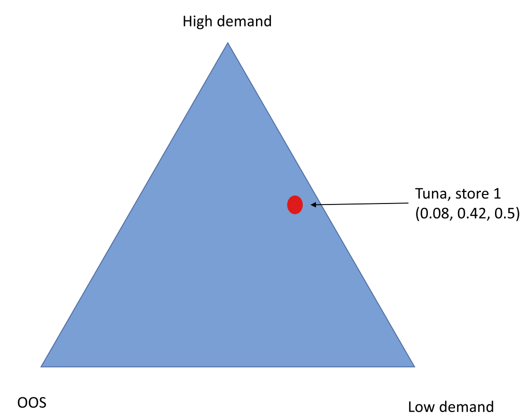
In this section, we demonstrate how the HMM estimates can be used to compare OOS performance across stores. Using the transition matrix, we can determine the long-term behavior of the demand for a product in a particular store. Specifically, at the average price, we can determine the probability that the conditional demand of the product is at any of the three states: OOS, low demand, or high demand. These probabilities correspond to the steady-state probabilities. Given that these probabilities add up to one, they can be represented in the simplex space. For instance, the steady-state probabilities for product tuna at store 1 are $(\pi_0, \pi_1, \pi_2) = (0.08, 0.42, 0.50)$. This implies that in 8% of the purchase occasions on random days it is expected that the product will be in the OOS state (see Figure 3).

These probabilities allow comparing the performance across stores to detect any abnormal behavior at particular stores for the same product. In addition, we can detect systematic failures at some stores. Figure 4 illustrates these probabilities for some other products. For instance, chocolate in store 3 has poor performance compared with that in the other stores because it is more likely that the product is OOS at that store compared with the other stores. Similar figures for the remaining products can be seen in Online Appendix C.

Similarly, for a given store, the store manager can analyze the behavior of the corresponding products. Figure 5 illustrates the underlying performance for different stores. For instance, at store 1, all products except cheese perform similarly. In the case of store 3, all products except chocolate show a relatively high variance between the corresponding low and high states but are homogeneous regarding the OOS state.

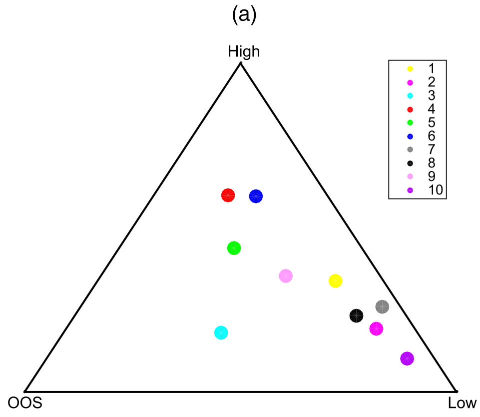
The steady-state probabilities summarize the probability of products being at each state. As shown earlier, the probability of being at the OOS state is of particular interest. However, the steady-state probabilities do not reveal if being at the OOS state is caused by a high probability of moving to the OOS because of constant shortages in stock or because the replenishment procedure is inefficient once the system is in the OOS state.

Figure 3. (Color online) Steady-State Probabilities



Notes. (Product, Store) = (Tuna, 1). {OOS, Low demand, high demand} = {0.08, 0.42, 0.50}.

Figure 4. (Color online) Steady-State Probabilities by Product



Notes. (a) Chocolate. (b) Powder juice. Each circle represents the steady-state probability of the product for a particular store.

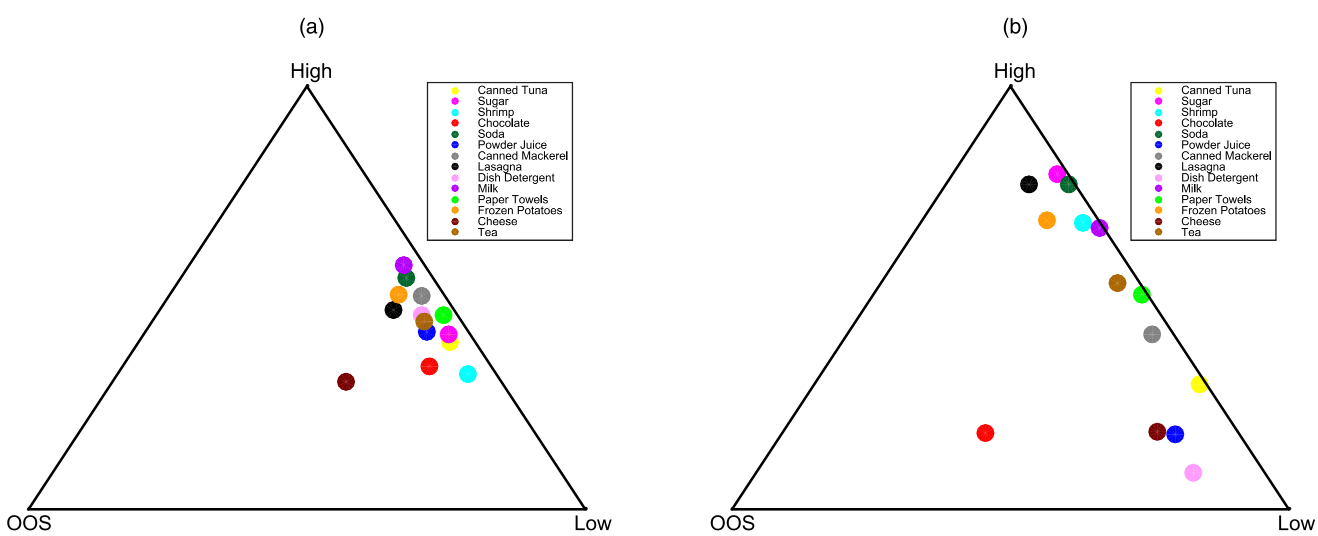
Disentangling these two possible causes could help store managers to expedite the solution. We explore these conditions next.

The transition matrix associated with each store summarizes the average performance regarding the frequency of products moving to the OOS state from any other state and how quickly they can move away from this state. Accordingly, to compare the performance of the store for a given product, we define two metrics related to demand planning and replenishment. To determine the performance regarding demand planning, we consider the probability of moving from any state other than OOS to the OOS state. These probabilities represent how frequently the conditional demand moves to the OOS state. We compute the metric as $\mathrm{DP} = 1 - \pi_1 \cdot q_{10} - \pi_2 \cdot q_{20}$, where π_1 and π_2 are the corresponding steady-state probabilities for low and high states, respectively. Similarly, to determine the performance regarding replenishment, we consider the probability of a product moving away from the OOS state. These probabilities represent how frequently the conditional demand moves from the OOS state to the other states. We compute the metric as $\mathrm{R} = 1 - q_{00}$, where q_{00} represents the probability of staying at the OOS state an extra day. Note that both metrics are constructed to be between zero and one; zero represents a poor performance, and one represents an excellent performance.

For example, consider the product tuna. For this product, we obtain the transition matrix for each store at the average price. Then, we compute the performance metrics demand planning and replenishment as described previously. The resulting comparison is illustrated in Figure 6(a).

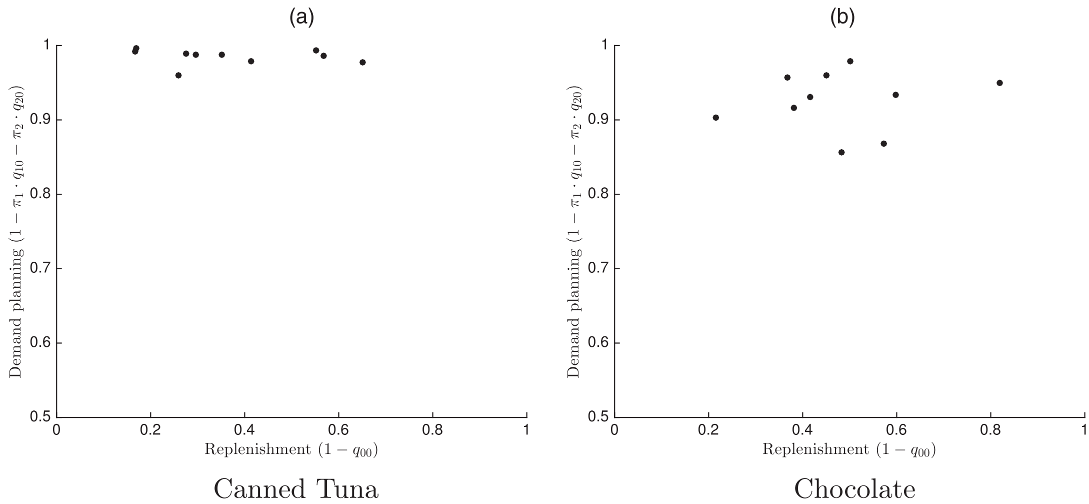
Figure 6(a) shows that stores differ significantly in their replenishment performance, ranging from 0.18 to

Figure 5. (Color online) Steady-State Probabilities by Store



Notes. (a) Store 1. (b) Store 3. Each circle represents the steady-state probabilities of a product at the store.

Figure 6. Comparison of the OOS Performance Across Stores



Notes. (a) Canned tuna. (b) Chocolate.

0.66 for the worst- and best-performing stores, respectively. However, there is not much difference regarding demand planning as the probability of moving to the OOS is relatively low for all stores. Similarly, Figure 6(b) illustrates the performance for chocolate across stores. In this case, stores show differences on both dimensions. In replenishment, the performance moves from 0.21 to 0.82 for the worst- and best-performing stores, respectively. For demand planning, the performance moves from 0.78 to 0.96 for the worst- and best-performing stores, respectively. Similar figures can be obtained for the other products (see Online Appendix D).

6. Conclusions and Future Research

This paper presents an HMM approach to identifying on-shelf OOS using point-of-sale data. For each product, the HMM model accounts for heterogeneity across stores and captures the dynamics in demand and the effect of pricing activities. We used a hierarchical Bayes modeling approach and an MCMC procedure to estimate the model parameters. We applied our modeling framework in the context of a big-box supermarket analyzing 14 products and 10 stores for a period of approximately 15 months. This application reveals several insights. First, we find that an HMM with three latent states performs best considering OOS detection and false alerts. The states relate to different demand levels that show different sensitivity to pricing decisions. Second, constraining the HMM to have an OOS state performs substantially better than allowing for flexible HMM states considering OOS detection metrics. The enhancement in performance greatly compensates for the slight deterioration in fit measures caused by a simpler HMM structure. Third, the chosen HMM outperforms several benchmarks, including the approach used by the sponsoring supermarket to detect OOS. Fourth, the HMM transition matrix reveals underlying dynamics among the states of demand and the unobserved OOS state. In particular, products differ in the probability of moving from the two demand states (either low or high) to the OOS state. A high probability may relate to poor performance of the forecasting system that results in frequent shortages in demand. Similarly, products differ in the probability of staying in the OOS state once the system is there. A high probability may relate to poor performance of the replenishment system that results in long periods of zero sales. Consequently, stores and products can be compared not only by considering OOS detection metrics, but also by using the product behavior uncovered by the HMM's transition matrix. It would be interesting to link this uncovered information with operational details that the company may have regarding, for example, forecasting and replenishment policies.

We note some limitations and some directions that future research could explore. Our model does not consider potential information of other stores to improve OOS detection in a particular store. In our application the correlation among stores regarding OOS is not significant. However, some stores could be more informative than others regarding OOS, which can be analyzed empirically. Second, to keep the parsimony of the model under control, we decided not to include the correlations among products, mainly because the products belong to different categories in which there is no clear theoretical complementarity or substitution patterns. However, one could add another hierarchy at a product category or department level to try to capture such correlations. Third, although we have applied our model in a supermarket setting, our approach could be used by other retailers who have difficulty in continuously monitoring the state of their shelves.

More generally, our HMM approach can be useful for detecting other retail execution failures with POS data. For instance, pricing errors or missing pricing information, planogram compliance, or staff failure can also be tracked with POS data. The fundamental idea is to use the signal given by abnormal sales patterns and link it to the execution problem under study through the HMM framework. Fourth, the proposed approach provides an alternative method to decensoring the data to be used for either improving forecasting models for demand planning or for detection by other OOS identification methodologies that require data processing such as the existing approach used by the retailer. Finally, because of data limitations, our model does not incorporate inventory data. This information may be helpful in fitting the sales data, but its contribution in detecting OOS is not clear. However, if this information is available, it would be interesting to test its contribution empirically because the inventory dynamics might help to better anticipate and diagnose OOS situations. An open question remains for future researchers as to how best to incorporate such data if available.

References

Aastrup J, Kotzab H (2010) Forty years of out-of-stock research–And shelves are still empty. *Internat. Rev. Retail Distribution Consumer Res.* 20(1):147–164.

Azoury KS (1985) Bayes solution to dynamic inventory models under unknown demand distribution. *Management Sci.* 31(9):1150–1160.

Bayraktar E, Ludkovski M (2010) Inventory management with partially observed nonstationary demand. *Ann. Oper. Res.* 176(1):7–39.

Bensoussan A, Cakanyildirim M, Li M, Sethi SP (2016) Managing inventory with cash register information: Sales recorded but not demands. *Production Oper. Management* 25(1):9–21.

Bhan N (2015) Inventory estimation from transactions via hidden markov models. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Campo K, Gijsbrechts E, Nisol P (2000) Towards understanding consumer response to stock-outs. *J. Retailing* 76(2):219–242.

Campo K, Gijsbrechts E, Nisol P (2003) The impact of retailer stockouts on whether, how much, and what to buy. *Internat. J. Res. Marketing* 20(3):273–286.

Che H, Chen X, Chen Y (2012) Investigating effects of out-of-stock on consumer stockkeeping unit choice. *J. Marketing Res.* 49(4):502–513.

Chen F, Song J-S (2001) Optimal policies for multiechelon inventory problems with Markov-modulated demand. *Oper. Res.* 49(2):226–234.

Chen L (2014) Fixing phantom stockouts: Optimal data-driven shelf inspection policies. Available at https://ssrn.com/abstract=2689802.

Chib S, Jeliazkov I (2001) Marginal likelihood from the Metropolis–Hastings output. *J. Amer. Statist. Assoc.* 96(453):270–281.

Chuang HH-C, Oliva R, Liu S (2016) On-shelf availability, retail performance, and external audits: A field experiment. *Production Oper. Management* 25(5):935–951.

Corsten D, Gruen T (2003) Desperately seeking shelf availability: An examination of the extent, the causes, and the efforts to address retail out-of-stocks. *Internat. J. Retail Distribution Management* 31(12):605–617.

Corstjens J, Corstjens M (1995) *Store Wars: The Battle for Mindspace and Shelfspace* (Wiley, West Sussex, UK).

DeHoratius N, Raman A (2008) Inventory record inaccuracy: An empirical analysis. *Management Sci.* 54(4):627–641.

DeHoratius N, Mersereau AJ, Schrage L (2008) Retail inventory management when records are inaccurate. *Manufacturing Service Oper. Management* 10(2):257–277.

Fisher M, Raman A (2010) *The New Science of Retailing: How Analytics Are Transforming the Supply Chain and Improving Performance* (Harvard Business Review Press, Boston).

Fitzsimons GJ (2000) Consumer response to stockouts. *J. Consumer Res.* 27(2):249–266.

Gelman A, Carlin JB, Stern HS, Rubin DB (2014) *Bayesian Data Analysis*, vol. 2 (Taylor & Francis, Boca Raton, FL).

Gierl H, Plantsch M, Schweidler J (2008) Scarcity effects on sales volume in retail. *Internat. Rev. Retail Distribution Consumer Res.* 18(1):45–61.

Gruen TW, Corsten DS (2007) *A Comprehensive Guide to Retail Out-of-Stock Reduction in the Fast-Moving Consumer Goods Industry* (Grocery Manufacturers of America, Washington, DC).

Hardgrave BC, Goyal S, Aloysius JA (2011) Improving inventory management in the retail store: The effectiveness of RFID tagging across product categories. *Oper. Management Res.* 4(1–2):6–13.

Jasra A, Holmes CC, Stephens DA (2005) Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statist. Sci.* 20(1):50–67.

Kök AG, Shang KH (2007) Inspection and replenishment policies for systems with inventory record inaccuracy. *Manufacturing Service Oper. Management* 9(2):185–205.

Lovejoy WS (1990) Myopic policies for some inventory models with uncertain demand distributions. *Management Sci.* 36(6):724–738.

MacDonald IL, Zucchini W (1997) *Hidden Markov and Other Models for Discrete-Valued Time Series*, vol. 110 (CRC Press, Boca Raton, FL).

Mersereau AJ (2015) Demand estimation from censored observations with inventory record inaccuracy. *Manufacturing Service Oper. Management* 17(3):335–349.

Montgomery DC (2013) *Introduction to Statistical Quality Control*, vol. 7 (Wiley, New York).

Montoya R, Netzer O, Jedidi K (2010) Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Marketing Sci.* 29(5):909–924.

Netzer O, Lattin JM, Srinivasan V (2008) A hidden Markov model of customer relationship dynamics. *Marketing Sci.* 27(2):185–204.

Papakiriakopoulos D, Doukidis G (2011) Classification performance for making decisions about products missing from the shelf. *Adv. Decision Sci.* 2011:1–13.

Papakiriakopoulos D, Pramatari K, Doukidis G (2009) A decision support system for detecting products missing from the shelf based on heuristic rules. *Decision Support Systems* 46(3):685–694.

Sarac A, Absi N, Dauzère-Pérès S (2010) A literature review on the impact of RFID technologies on supply chain management. *Internat. J. Production Econom.* 128(1):77–95.

Song J-S, Zipkin P (1993) Inventory control in a fluctuating demand environment. *Oper. Res.* 41(2):351–370.

Sorensen D, Gianola D (2007) *Likelihood, Bayesian, and MCMC Methods in Quantitative Genetics* (Springer Science & Business Media, New York).

Ton Z, Raman A (2010) The effect of product variety and inventory levels on retail store sales: A longitudinal study. *Production Oper. Management* 19(5):546–560.

Treharne JT, Sox CR (2002) Adaptive inventory control for nonstationary demand and partial information. *Management Sci.* 48(5):607–624.

Van Woensel T, Van Donselaar K, Broekmeulen R, Fransoo J (2007) Consumer responses to shelf out-of-stocks of perishable products. *Internat. J. Physical Distribution Logist. Management* 37(9):704–718.