

Demand Estimation from Censored Observations with Inventory Record Inaccuracy

Adam J. Mersereau

Kenan-Flagler Business School, University of North Carolina
ajm@unc.edu

A retailer cannot sell more than it has in stock; therefore its sales observations are a censored representation of the underlying demand process. When a retailer forecasts demand based on past sales observations, it requires an estimation approach that accounts for this censoring. Several authors have analyzed inventory management with demand learning in environments with censored observations, but the authors assume that inventory levels are known and hence that stockouts are observed. However, firms often do not know how many units of inventory are available to meet demand, a phenomenon known as inventory record inaccuracy. We investigate the impact of this unknown on demand estimation in an environment with censored observations. When the firm does not account for inventory uncertainty when estimating demand, we discover and characterize a systematic downward bias in demand estimation under typical assumptions on the distribution of inventory record inaccuracies. We propose and test a heuristic prescription that relies on a single error statistic and that sharply reduces this bias.

This version: December 19, 2014.

1. Introduction

There has been a surge of interest in recent years in the problem of “inventory record inaccuracy.” For a variety of reasons, firms’ electronic inventory records do not necessarily match the quantity of customer-available stock. This uncertainty brings several costs (DeHoratius et al. 2008). Firms must deploy resources to count stock and reconcile inventory records. Uncertain inventory records hinder the firm’s ability to match supply with demand, increasing inventory and stockout costs. The current paper exposes another source of cost due to record inaccuracy, namely its confounding of the firm’s demand estimation.

Even though much of the operations management literature assumes a known probability distribution describing demand, the reality is that firms’ forecasts of demand are perpetually evolving as data is accumulated. In many settings (e.g., retail) the firm observes sales data that are censored representations of demand; that is, sales in a period are the minimum of demand in the period and the amount of stock available. Sales are always less than or equal to demand, and Conrad (1976) demonstrates that a firm that treats sales and demand as equivalent will stock its shelves suboptimally. To correctly learn demand from sales data, the firm requires an estimation approach

that properly accounts for the censoring. But it is difficult for a firm to do this when erroneous inventory records prevent the firm from accurately assessing when censoring occurs.

In this paper, we investigate the impact of inventory record inaccuracy on demand estimation in an environment with censored demand. We assume that the firm is sophisticated with respect to demand estimation, and we borrow the parametric Bayesian demand learning model of Lariviere and Porteus (1999) and Ding et al. (2002) to account for censoring. However, most of our analysis assumes that the firm does not account for record inaccuracy when estimating demand. This appears to be common practice. While we are aware of several retailers who use methods to correct for censoring when estimating demand, it seems uncommon for retailers to account for record inaccuracy in these procedures. We examine the estimation biases resulting from such a mischaracterization, and we study a process wherein the firm uses these misspecified demand estimates when making inventory decisions. Even though the firm thinks it is properly correcting for censoring, we characterize a systematic downward bias in the firm’s demand estimates, assuming (as is typically done in the literature) that inventory errors are symmetric or tend to reduce physical stock relative to recorded stock. This underestimation by the firm occurs even while service levels may appear to the firm to meet or exceed targets. An implication is that ignoring record inaccuracy when estimating demand can negatively impact the profitability of a firm even as the firm may fail to detect the problem.

Tracking technologies such as Radio Frequency Identification (RFID), by providing the seller better visibility into inventory levels and censoring, may yield improved demand estimates. Our work is relevant for the evaluation of such technologies. On the other hand, item-level RFID implementations remain prohibitively expensive for many applications and often do not provide perfect visibility into inventory positions due to technological limitations and other practical considerations. In the absence of an RFID implementation, we propose a heuristic prescription for the problem that relies on a single inventory inaccuracy statistic (namely, the probability of a stock-reducing error) and that substantially reduces the estimation bias.

Our work investigates the interaction between demand censoring and inventory inaccuracy, two phenomena studied separately in previous research. By identifying a new cost of inventory record inaccuracy and its drivers, our work complements existing research by arguing for better understanding and mitigation of record inaccuracy. Much of the academic research on record inaccuracy and RFID has focused on their implications on replenishment policies. Our work is relevant to a broader set of firms—for example, fashion retailers who engage in limited replenishment but for whom learning about demand trends is important. To the literature on demand censoring, which

has analyzed the implications of ignoring censoring altogether and of accounting for (known) censoring events, our work contributes by analyzing the important practical case in which censoring is unreliably observed.

The paper is organized as follows. We review related literature in Section 2. Section 3 introduces our notation and model setting and reviews some background on demand learning from censored observations. We present in Section 4 a basic result suggesting a downward bias in demand estimates under record inaccuracy under fairly general assumptions. In Section 5, we provide a detailed analysis for the case of exponential demand, where we make use of concise demand updating equations under censoring established by Braden and Freimer (1991). We characterize fixed points of the firm’s estimation and stocking process given this misspecification, study their dependence on model parameters, and suggest a heuristic prescription that partially corrects for the misspecification. Section 6 presents a numerical framework for studying the estimation bias under discrete demand distributions. Using this framework, we corroborate many of our earlier findings for Poisson demand. We discuss future research opportunities in Section 7.

2. Related Literature

Our research lies at the intersection of two distinct research streams, both reviewed in Chen and Mersereau (2014). The first concerns estimating demand in inventory systems based on censored observations. Wecker (1978), Nahmias (1994), and Agrawal and Smith (1996) present methods for estimating demand parameters from sales observations when stocking levels are known. Especially relevant to our work are parametric Bayesian models for learning demand from censored observations, which date to Harpaz et al. (1982). Lariviere and Porteus (1999), making use of the “newsvendor distribution” framework of Braden and Freimer (1991) and assuming that inventory is perishable, show that an optimal forward-looking seller in a censored-demand setting will stock more than a myopic seller. Ding et al. (2002) extend this result to a more general set of continuous distributions (see also notes by Lu et al. 2005, 2008 and Bensoussan et al. 2009b). Our work shares the parametric Bayesian framework assumed in these papers for learning from censored demand. We review this model and the newsvendor distribution framework in Section 3.1. For the special case of exponential demand that we assume in Section 5, Bisi et al. (2011) and Bensoussan et al. (2009a) offer detailed results on policies and estimation consistency.

The results of Lariviere and Porteus (1999) and Ding et al. (2002) have been extended by several authors. Chen and Plambeck (2008), Lu et al. (2007), Chen (2010), and Bisi et al. (2011) consider non-perishable inventory. A number of works, including Burnetas and Smith (2000), Huh

and Rusmevichientong (2009), and Huh et al. (2011), consider learning of censored demand in non-parametric settings. Besbes and Muharremoglu (2013) show the importance of censoring information through an asymptotic analysis, particularly when demand takes discrete values. Nevertheless, in all the works we are aware of on learning from censored demand, a key, often implicit assumption is that inventory positions are known so that stockouts are accurately observed. By relaxing the assumption that the firm receives accurate information about when stockouts occur, our work presents a theoretically novel and practically relevant extension.

The second research stream most relevant to our work concerns inventory management with inventory record inaccuracy and inventory misplacement, which dates to Iglehart and Morey (1972). Recent interest in the problem follows empirical studies of inventory inaccuracy and misplacement (e.g., DeHoratius and Raman 2008, Ton and Raman 2006) and accompanies the emergence of new tracking technologies such as RFID (Lee and Özer 2007). While DeHoratius and Ton (2009) distinguish between “inventory record inaccuracy” (where the quantity of inventory in a firm’s facility does not match its inventory record) and “misplaced products” (where some on-hand inventory is in the facility but unavailable to customers), we model an error process that may be a combination of both types of errors, and we refer to it as “inventory record inaccuracy” for simplicity.

Several analytical studies have looked at the implications of inventory inaccuracy in supply chain settings (e.g., Heese 2007, Gaukler et al. 2007, Camdereli and Swaminathan 2010, Kök and Shang 2014), but much of the work in this area concerns inventory and inspection policies for single-item, single-location systems with inventory inaccuracy (Kang and Gershwin 2005, Kök and Shang 2007, Bensoussan et al. 2007, DeHoratius et al. 2008, Rekik et al. 2008, Atali et al. 2011, Mersereau 2013, Chen 2013). These papers differ from each other largely in the details underlying their inventory, error, and information models. Many of these papers estimate the value of inventory visibility by comparing a naive or ignorant seller (who is oblivious to inventory inaccuracies), an informed seller (who knows the distribution but not the realizations of inventory inaccuracies) and an idealized RFID-enabled seller (who has full inventory visibility). The papers then measure the value of visibility as the improvement gain from better information. For example, the value of RFID is often measured as the performance gap between an RFID-enabled seller and a naive or informed seller.

These papers differ from our work most notably in that they all assume that demand probability distributions (and typically also error distributions) are known. In our model, a parameter of the demand distribution is unknown and learned from (censored) observations. We model a naive seller who is oblivious to inventory record inaccuracy. We compare this naive seller with a “tracking” (or perfect RFID-enabled) seller who observes physical inventory positions, and we propose a heuristic

for an informed seller who is aware of discrepancies between recorded and physical inventory and has limited knowledge of their distribution. Our focus is the impact of inventory record inaccuracy on the seller's ability to estimate demand, in contrast with previous works that all measure the impact of inventory record inaccuracy on matching supply with demand given known demand processes. DeHoratius and Ton (2009) survey the literature on inventory record inaccuracy and identify an "opportunity to incorporate execution problems into models that estimate demand and assess forecast accuracy in the presence of stockouts" (p. 67). We believe we are the first to investigate this interaction between inventory inaccuracy and demand estimation.

Our naive seller assumes a misspecified demand process. A small literature has emerged on misspecified demand models in operations management (Cachon and K  k 2007, Cooper et al. 2006, 2009). Lee et al. (2012) study a repeated newsvendor setting in which the seller stocks according to a critical fractile-based policy and demand subsequently depends on the stocked quantity. This bears some resemblance to our problem, although the misspecification in their model involves an endogenous demand-generating process rather than uncertainty around stocking levels.

3. Preliminaries

We indicate random variables by upper-case letters, their realizations by lower-case letters, and vectors by superscripted right-arrows. For example, we will use S_t to indicate random sales in period t , s_t to indicate realized sales, and \vec{s}_t to indicate the vector of sales realizations (s_1, \dots, s_t) .

3.1. Review of Demand Estimation under Censoring

We first review the Bayesian censored demand model and introduce some notation without the complication of inventory record inaccuracy. We will add inventory record inaccuracy to the model in Section 3.2. We assume that a seller stocks a quantity j_t of units in period t to satisfy random demand. We assume that demand D_t in period t is a random variable with stationary probability distribution $F(d|\theta)$. The functional form $F(d|\theta)$ is known, and θ is a fixed but unknown parameter. Conditional on θ , demand is independent across periods. We denote the realization of D_t as d_t . Sales S_t in period t are $S_t = \min\{D_t, j_t\}$. At the end of period t , the seller observes the realization s_t of S_t , and any unfulfilled demand is unobserved. Therefore, if the seller observes $s_t < j_t$ then she assumes that $d_t = s_t$; if the seller observes $s_t = j_t$ then she assumes that $d_t \geq s_t$.

The seller begins with a prior density π_0 on θ that she updates periodically in a Bayesian fashion. For the purpose of computing a posterior distribution of θ at the end of period t , the seller requires not only the history of sales observations $\vec{s}_t = (s_1, \dots, s_t)$ but also a vector $\vec{c}_t = (c_1, \dots, c_t)$

of indicators indicating whether or not there was a stockout in each time period. We let $c_t = 1$ if $s_t = j_t$ and $c_t = 0$ if $s_t < j_t$.

Given this history, we can compute the posterior density $\pi_t(\cdot|\vec{s}_t, \vec{c}_t)$ for θ at the end of period t . Assuming that demand takes discrete, integer values, let $f(d|\theta)$ indicate the probability mass function corresponding with distribution $F(d|\theta)$. The posterior density $\pi_t(\cdot|\vec{s}_t, \vec{c}_t)$ is then:

$$\begin{aligned} \pi_t(\theta|\vec{s}_t, \vec{c}_t) &\propto \pi_0(\theta) \prod_{\tau=1}^t [\mathbb{1}\{c_\tau = 1\}(1 - F(s_\tau - 1|\theta)) + \mathbb{1}\{c_\tau = 0\}f(s_\tau|\theta)] \\ &\propto \pi_0(\theta) \prod_{\tau=1}^t (1 - F(s_\tau - 1|\theta))^{c_\tau} f(s_\tau|\theta)^{1-c_\tau}. \end{aligned} \quad (1)$$

If demand instead follows a continuous distribution with density $f(d|\theta)$, the posterior density is:

$$\begin{aligned} \pi_t(\theta|\vec{s}_t, \vec{c}_t) &\propto \pi_0(\theta) \prod_{\tau=1}^t [\mathbb{1}\{c_\tau = 1\}(1 - F(s_\tau|\theta)) + \mathbb{1}\{c_\tau = 0\}f(s_\tau|\theta)] \\ &\propto \pi_0(\theta) \prod_{\tau=1}^t (1 - F(s_\tau|\theta))^{c_\tau} f(s_\tau|\theta)^{1-c_\tau}. \end{aligned} \quad (2)$$

We denote the predictive demand distribution (after unconditioning on θ) after period t by $\hat{F}_t(d|\vec{s}_t, \vec{c}_t)$.

An especially tractable set of special cases arises when we assume the demand distribution falls within the family of “newsvendor distributions” identified by Braden and Freimer (1991) (see also Lariviere and Porteus 1999). A newsvendor distribution is a continuous distribution of the form $F(d|\theta) = 1 - \exp\{-\theta g(d)\}$ for $d \in \mathbb{R}^+$, where the function $g(\cdot)$ is positive, differentiable, and increasing for $d \geq 0$; also, $\lim_{d \rightarrow 0^+} g(d) = 0$ and $\lim_{d \rightarrow \infty} g(d) = \infty$. The exponential distribution is a newsvendor distribution with $g(d) = d$.

The gamma distribution is a conjugate prior for all newsvendor distributions. If we let a_0 and b_0 indicate the shape and scale parameters, respectively, of a gamma prior density on θ , $\pi_0(\theta|a_0, b_0) = b_0^{a_0} \theta^{a_0-1} e^{-b_0\theta} / \Gamma(a_0)$, then the posterior distribution of θ given observations \vec{s}_t and \vec{c}_t is gamma with scale and shape parameters

$$a_t = a_0 + \sum_{\tau=1}^t (1 - c_\tau), \quad b_t = b_0 + \sum_{\tau=1}^t g(s_\tau), \quad (3)$$

mean a_t/b_t , and variance a_t/b_t^2 .

An important implication of the update (3) is that the matching of the stockout observation c_t with its corresponding sales observation s_t becomes unimportant when we assume a newsvendor demand distribution. The seller only needs to keep track of the total number of censored observations and an aggregate measure of sales.

3.2. Model of Inventory Record Inaccuracy

We model inventory record inaccuracy by distinguishing between the quantity j_t the seller believes is in stock and the quantity I_t that is in fact customer-available. A discrepancy between I_t and j_t may arise from a variety of sources: merchandise theft or damage, recording or counting errors, or misplacement. We assume that $I_t = \{j_t - \epsilon_t\}^+$, where ϵ_t is a random variable drawn from a (possibly non-stationary) distribution H_t representing possible discrepancies between physical and recorded inventory levels. We assume that the ϵ_t 's are statistically independent of demand and that any changes to inventory discrepancies occur before demand within a period. These assumptions lend tractability to the model and are common in the analytical literature on inventory record inaccuracy. We allow H_t to be continuous, discrete, or a mixture of continuous and discrete distributions, although we assume that H_t is non-degenerate (i.e., $\Pr\{\epsilon_t = 0\} < 1$) for ease of exposition.

We will consider two sellers operating with different information sets:

1. A “naive” seller knows j_t and thinks that j_t units are available to the customer in period t . That is, the seller ignores inventory discrepancies or is oblivious to them. We assume that the naive seller maintains a database of sales observations $\vec{s}_t^n = (s_1^n, \dots, s_t^n)$ and of stockout indicators $\vec{c}_t^n = (c_1^n, \dots, c_t^n)$, where $c_t^n = 1$ if $s_t^n \geq j_t$ and $c_t^n = 0$ if $s_t^n < j_t$, and that she computes posterior densities $\pi_t^n(\cdot | \vec{s}_t^n, \vec{c}_t^n)$ for θ using equations (1) or (2). The naive seller does not directly observe the inventory position at the beginning or end of the period.

We permit $\epsilon_t < 0$, in which case it is possible for the naive seller to observe $s_t^n > j_t$. This is a zero-probability event given her belief that j_t is the true customer-available inventory level. We assume that the seller interprets such an observation as a stockout with sales s_t^n . Compared with other reasonable responses to this situation, our assumption maximizes the stockout indicator c_t^n and is a conservative assumption in light of the results to come in Sections 4 and 5.

2. A “tracking” seller knows the true customer-available inventory level I_t , which is plausible under a perfect implementation of a tracking technology such as RFID. We assume that the tracking seller maintains a database of observed sales $\vec{s}_t^r = (s_1^r, \dots, s_t^r)$ and of stockout indicators $\vec{c}_t^r = (c_1^r, \dots, c_t^r)$, where $c_t^r = 1$ if $s_t^r = I_t$ and $c_t^r = 0$ if $s_t^r < I_t$, and that she computes posterior densities $\pi_t^r(\cdot | \vec{s}_t^r, \vec{c}_t^r)$ for θ using equations (1) or (2).

4. Comparison when Stocking Quantities are Common

We explore the impact of inaccurate inventory records on demand estimation by comparing the performances of the naive and tracking sellers. Table 1 and Figure 1 present the possible observations of the two sellers in period t , broken down into six cases. In cases (A), (B), and (D), the

Case	Definition	s_t	c_t^r	c_t^n
(A)	$i_t \geq j_t > d_t$	d_t	0	0
(B)	$j_t > i_t > d_t$	d_t	0	0
(C)	$i_t > d_t \geq j_t$	d_t	0	1
(D)	$d_t \geq i_t \geq j_t$	i_t	1	1
(E)	$j_t > d_t \geq i_t$	i_t	1	0
(F)	$d_t \geq j_t > i_t$	i_t	1	0

Table 1: Sales and stockout observations seen by tracking and naive sellers.

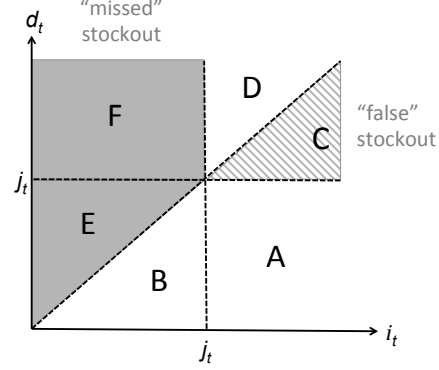


Figure 1: Graphical representation of “false stockout” and “missed stockout”.

naive seller registers the same information as the tracking seller. Case (C) represents a situation (a “false stockout”) in which the naive seller perceives a stockout even though there is no physical stockout. Cases (E) and (F) represent situations (“missed stockouts”) in which there is a physical stockout that goes unseen by the naive seller.

Our first result is that if the naive and tracking sellers stock the same quantity in a period and if the error distribution is such that stock-reducing errors ($\epsilon_t > 0$) are at least as likely as stock-increasing errors, the naive seller perceives too few stockouts in the period compared with the tracking seller. In other words, a missed stockout is more likely than a false stockout under typical error distributions.

PROPOSITION 1. Suppose the discrepancy distribution H_t is such that $\Pr\{\epsilon_t > 0\} \geq \Pr\{\epsilon_t < 0\}$ and suppose that the naive and tracking sellers stock the same quantity j_t in period t . Then $\Pr\{C_t^n = 1\} \leq \Pr\{C_t^r = 1\}$.

Proof. Observe from Table 1 that $\Pr\{C_t^n = 1\} - \Pr\{C_t^r = 1\} = \Pr\{I_t > D_t \geq j_t\} + \Pr\{D_t \geq I_t \geq j_t\} - \Pr\{D_t \geq I_t\} = \Pr\{(D_t \geq j_t) \cap (I_t \geq j_t)\} - \Pr\{D_t \geq I_t\} = \Pr\{D_t \geq j_t\} \Pr\{\epsilon_t \leq 0\} - \Pr\{D_t \geq I_t\} \leq \Pr\{D_t \geq j_t\} \Pr\{\epsilon_t \geq 0\} - \Pr\{D_t \geq I_t\} = \Pr\{D_t \geq j_t \geq I_t\} - \Pr\{D_t \geq I_t\} \leq \Pr\{D_t \geq I_t\} - \Pr\{D_t \geq I_t\} = 0$, where the third equality relies on the independence of D_t and I_t given j_t , and the first inequality makes use of the assumption $\Pr\{\epsilon_t > 0\} \geq \Pr\{\epsilon_t < 0\}$. ■

Proposition 1 requires no assumptions on the demand distribution. Furthermore, the condition $\Pr\{\epsilon_t > 0\} \geq \Pr\{\epsilon_t < 0\}$ of Proposition 1 is consistent with the audit results observed by DeHoratius and Raman (2008), who found positive discrepancies in 59% of inaccurate records identified in a retail chain’s inventory audit. Furthermore, this condition covers many of the assumptions made in the analytical literature on inventory record inaccuracy. It subsumes nonnegative discrepancies

(i.e., $\Pr\{\epsilon_t \geq 0\} = 1$), which are reasonable to assume for certain error sources—for example, theft, damage, or misplacement. This assumption is made in Kang and Gershwin (2005) and K  k and Shang (2014). Proposition 1 also applies to discrepancy distributions that are symmetric about zero, which is a plausible model of clerical errors and assumed by K  k and Shang (2007). Finally, Proposition 1 also applies to mixtures of nonnegative and symmetric discrepancy distributions, which is an appropriate model when multiple sources of errors are assumed (e.g., Atali et al. 2011).

Proposition 1 implies that the naive seller sees too few stockouts on average compared with the tracking seller, assuming they stock the same quantities. Given that they see the same sales observations in this case, it is intuitive that the naive seller would therefore underestimate demand over time relative to the tracking seller. We prove a corollary in Appendix A illustrating this intuition in a repeated newsvendor setting in which demand follows a newsvendor distribution and the naive and tracking sellers stock the same sequence of quantities. (All appendices are included in an online companion to this paper.) In practice, however, we would expect the two sellers to stock different quantities based on their evolving beliefs about demand. In the following sections, we study the behavior of this process and the magnitude of the naive seller’s bias in steady state.

5. Exponential Demand

In this and subsequent sections, we assume that the naive seller chooses a stocking quantity j_t in period t that depends on her own demand estimates, engaging in an iterative process of estimate-stock-estimate. We assume throughout this section that demand follows an exponential distribution, which enables us to express fixed points of the estimate-stock-estimate process as the solution to a single equation.

5.1. A Model Of Accumulating Errors

We seek to characterize fixed points of the naive seller’s stocking and estimation process: pairs $(\tilde{\theta}, \tilde{j})$ such that (1) if the seller’s demand estimate is fixed at $\tilde{\theta} > 0$ then the seller will stock $\tilde{j} > 0$, and (2) if the seller stocks \tilde{j} repeatedly then the seller’s demand parameter estimate converges to $\tilde{\theta}$. We choose to characterize fixed points rather than engage in a detailed period-by-period convergence analysis for analytical tractability and because we seek insights that are independent of time and of initial prior distributions. Numerical experiments demonstrate (see Appendix B) that the fixed points we characterize here describe the long-run behavior of the naive seller’s estimate-stock-estimate process. We therefore believe that this analysis is appropriate whenever decisions are based on long data histories—for example, for long lifecycle products or for products whose initial forecasts are based on historical data gathered in similar environments.

In order to obtain closed-form expressions for these fixed points, we impose an exponential demand distribution and a more detailed model of inventory discrepancies. We assume that the seller has a $\text{gamma}(a_0, b_0)$ prior on the parameter θ of the exponential demand distribution. This is to simplify our analysis, although the Bernstein-von Mises theorem (see Le Cam 1986) suggests that the choice of prior is irrelevant as the number of observations becomes large and thus has no bearing on our results. We also assume that, given an estimate $\tilde{\theta}$ of the demand parameter, the naive seller employs a critical fractile policy, choosing a nominal stocking quantity \tilde{j} to achieve a particular pre-defined service level $\gamma \in (0, 1)$ with respect to the assumed exponential demand distribution $F(\cdot|\tilde{\theta})$, i.e.,

$$\tilde{j} = -\frac{1}{\tilde{\theta}} \ln(1 - \gamma). \quad (4)$$

The firm replenishes each period with zero lead time. The quantity (4) is optimal for a newsvendor with known $\theta = \tilde{\theta}$ and no inventory errors; therefore it specifies reasonable limiting behavior of a seller who is learning θ while oblivious to errors.

We assume that “discrepancies” between physical inventory levels and recorded inventory positions arise as accumulations of a hidden, stationary “error” process that perturbs physical inventory levels each period. We assume that errors are independent over time, but generally discrepancies will not be. The seller in our model periodically inspects physical inventory and resets the inventory record accordingly. Such a practice is known as “cycle counting” and is typical in many warehousing and retailing settings. Specifically, we assume that inspections, which set the inventory record equal to physical inventory, occur at the end of every T th period (an “inspection epoch”) where T is a pre-specified inspection frequency.

We make two further assumptions to simplify the fixed point analysis. First, we assume that the seller has the option at inspection epochs to return merchandise to her supplier with no financial penalty. This greatly simplifies the analysis of situations in which an inspection reveals that there is more physical inventory than the seller desires to stock. Modeling these situations is complicated and is not core to our research questions. Appendix C provides evidence that this assumption has little impact on the results. Second, the naive seller’s demand estimation remains “naive” with respect to inventory record inaccuracy; the naive seller does not incorporate the results of the inspections into her demand estimation procedure. For example, in theory the naive seller could somehow revise her stockout observations in periods $1, \dots, T$ based on the result of the count in period T . We assume this does not happen.

An implication of these assumptions is that, in steady state, the seller will stock the same nominal quantity \tilde{j} each period. We index inspection cycles by $m \in \{0, 1, 2, \dots\}$, and we let $m(t) = \lfloor (t-1)/T \rfloor$

indicate the cycle index corresponding to period t and $w(t) = [(t-1) \bmod T] + 1$ indicate the index of period t within its cycle. Assume that a count occurs at the end of period mT . Nature draws error e_{mT+1} and replaces nominal inventory \tilde{j} with physical inventory $I_{mT+1} = \tilde{j} - R_{mT+1}$, where $R_{mT+1} = \min\{\tilde{j}, e_{mT+1}\}$. In subsequent periods $t \in \{mT+2, \dots, (m+1)T\}$, the seller replenishes so that nominal stock is \tilde{j} . Physical inventory in period $t \in \{mT+2, \dots, (m+1)T\}$ is $I_t = \tilde{j} - R_t$, where $R_t = \min\{\tilde{j}, e_t + R_{t-1}\}$. We let $K_{t,\tilde{j}}$ denote the distribution of the random variable R_t given nominal stock \tilde{j} .

We can write expected sales and stockout indicators as a function of \tilde{j} and the true demand parameter θ . Let the random variables \tilde{I}_t , \tilde{S}_t , and \tilde{C}_t^n be the customer-available inventory, sales, and stockout indicator in period t if the naive seller stocks \tilde{j} . The naive seller perceives a stockout with probability

$$\mathbb{E}[\tilde{C}_t^n] = \Pr\{\tilde{I}_t \geq \tilde{j}\} \Pr\{D_t \geq \tilde{j}\} = \Pr\{R_t \leq 0\} \exp\{-\theta\tilde{j}\}, \quad (5)$$

and sees average sales

$$\begin{aligned} \mathbb{E}[\tilde{S}_t] &= \int_{-\infty}^{\infty} \mathbb{E}_{D_t}[\min\{D_t, \tilde{j} - r\}] dK_{t,\tilde{j}}(r) = \frac{1}{\theta} - \frac{1}{\theta} \int_{-\infty}^{\infty} \exp\{-\theta(\tilde{j} - r)\} dK_{t,\tilde{j}}(r) \\ &= \frac{1}{\theta} - \frac{1}{\theta} \exp\{-\theta\tilde{j}\} M_{R_t}(\theta, \tilde{j}), \end{aligned} \quad (6)$$

where $M_{R_t}(\theta, \tilde{j}) \equiv \mathbb{E}[e^{\theta R_t} | \tilde{j}]$ is the moment-generating function of the random variable R_t evaluated at θ . We assume this function exists at θ for all $\tilde{j} > 0$.

The process regenerates following each inspection epoch; therefore, $K_{t,\tilde{j}} = K_{w(t),\tilde{j}}$ for all t , and total sales and stockout indicators over a cycle are i.i.d. across cycles. Summing over an inspection cycle (and choosing the first cycle in the interest of indexing simplicity), we have:

$$\begin{aligned} \mathbb{E}\left[\sum_{u=1}^T \tilde{C}_u^n\right] &= \exp\{-\theta\tilde{j}\} \sum_{u=1}^T \Pr\{R_u \leq 0 | K_{u,\tilde{j}}\}, \\ \mathbb{E}\left[\sum_{u=1}^T \tilde{S}_u\right] &= \frac{T}{\theta} - \frac{1}{\theta} \exp\{-\theta\tilde{j}\} \sum_{u=1}^T M_{R_u}(\theta, \tilde{j}). \end{aligned}$$

Now we write down conditions for the pair $(\tilde{\theta}, \tilde{j})$ to be a fixed point. Suppose the naive seller stocks \tilde{j} each period and that $\mathbb{E}[\tilde{j} - R_t] > 0$ for all t , then the expected sales $\mathbb{E}[\sum_{u=1}^T \tilde{S}_u]$ each cycle is constant and positive and b_t/t converges almost surely to this expectation by the strong law of large numbers. This enables us to show that the naive seller's posterior belief converges to $\mathbb{E}[T - \sum_{u=1}^T \tilde{C}_u^n] / \mathbb{E}[\sum_{u=1}^T \tilde{S}_u]$. In particular, letting θ_t denote a random variable distributed according to the seller's posterior belief at time t ,

$$\mathbb{E}[\theta_t] \rightarrow \lim_{t \rightarrow \infty} \frac{(1/t)a_t}{(1/t)b_t} = \lim_{t \rightarrow \infty} \frac{(1/t)a_t}{(1/t)b_t} = \frac{\lim_{m \rightarrow \infty} (1/mT) \sum_{u=1}^{mT} (1 - c_u^n)}{\lim_{m \rightarrow \infty} (1/mT) \sum_{u=1}^{mT} s_u} = \frac{\mathbb{E}[T - \sum_{u=1}^T \tilde{C}_u^n]}{\mathbb{E}[\sum_{u=1}^T \tilde{S}_u]},$$

$$\text{Var}[\theta_t] \rightarrow \lim_{t \rightarrow \infty} \frac{(1/t^2)a_t}{(1/t^2)b_t^2} = \frac{\mathbb{E}[T - \sum_{u=1}^T \tilde{C}_u^n]}{\mathbb{E}[\sum_{u=1}^T \tilde{S}_u]} \lim_{t \rightarrow \infty} \frac{1/t}{(1/t)b_t} = 0.$$

Hence, if the naive seller repeatedly stocks $\tilde{j} > 0$, her demand parameter estimate converges to $\tilde{\theta} = \mathbb{E}[T - \sum_{u=1}^T \tilde{C}_u^n] / \mathbb{E}[\sum_{u=1}^T \tilde{S}_u]$. We can also show that $\mathbb{E}[T - \sum_{u=1}^T \tilde{C}_u^n] / \mathbb{E}[\sum_{u=1}^T \tilde{S}_u]$ is the limiting, maximum likelihood estimator for $\tilde{\theta}$ as $t \rightarrow \infty$. We will exploit the maximum likelihood perspective in Section 6.

It follows that a fixed point must satisfy

$$\frac{\tilde{\theta}}{\theta} = \frac{1 - \left(\frac{1}{T} \sum_{u=1}^T \Pr\{R_u \leq 0 | K_{u,\tilde{j}}\} \right) \exp\{-\theta\tilde{j}\}}{1 - \left(\frac{1}{T} \sum_{u=1}^T M_{R_u}(\theta, \tilde{j}) \right) \exp\{-\theta\tilde{j}\}}, \text{ where } \tilde{j} = -\frac{1}{\tilde{\theta}} \ln(1 - \gamma). \quad (7)$$

A solution to equation (7) will have $\tilde{\theta} \geq \theta$ if $\sum_{u=1}^T M_{R_u}(\theta, \tilde{j}) \geq \sum_{u=1}^T \Pr\{R_u \leq 0 | K_{u,\tilde{j}}\}$. The following proposition gives an interpretable sufficient condition for this to be true. Recall that the parameter (rate) of an exponential distribution is the reciprocal of its mean. Therefore, $\tilde{\theta} \geq \theta$ implies that the seller underestimates demand.

PROPOSITION 2. If (7) has a finite solution $(\tilde{\theta}, \tilde{j})$ such that $\mathbb{E}[R_t | \tilde{j}] \geq 0$ for all t , then $\tilde{\theta} \geq \theta$.

Proof. Let $\xi_t \equiv \mathbb{E}[R_t | \tilde{j}]$, then $M_{R_t}(\theta, \tilde{j}) = \mathbb{E}[e^{\theta R_t} | \tilde{j}] = e^{\theta \xi_t} \mathbb{E}[e^{\theta(R_t - \xi_t)} | \tilde{j}] \geq \mathbb{E}[e^{\theta(R_t - \xi_t)} | \tilde{j}] \geq \mathbb{E}[e^0] = 1$, where the first inequality follows from the assumption $\mathbb{E}[R_t | \tilde{j}] = \xi_t \geq 0$ so $e^{\theta \xi_t} \geq 1$, and the second inequality follows from Jensen's inequality.

Because $\Pr\{R_t \leq 0\} \leq 1$, we have $M_{R_t}(\theta, \tilde{j}) \geq \Pr\{R_t \leq 0\}$ for all t and therefore the right-hand side of equation (7) is greater than one. (Note that $M_{R_t}(\theta, \tilde{j}) \exp\{-\theta\tilde{j}\} = \mathbb{E}[e^{-\theta(\tilde{j} - R_t)} | \tilde{j}] < 1$ so the right-hand side is positive.) Therefore $\tilde{\theta} \geq \theta$. \blacksquare

Equation (7) is challenging to analyze further (and even to solve numerically) because $M_{R_u}(\theta, \tilde{j})$ depends on the error distribution and $\tilde{\theta}$ (through \tilde{j}) in a complicated way. This motivates our study of an important special case of equation (7). Define $\epsilon_t = \sum_{u=Tm(t)+1}^t e_u$ and assume that errors are small relative to stocking quantities so that we can approximate $R_t \approx \epsilon_t$ for all t . We assume the individual errors e_t are i.i.d. according to distribution H (We sometimes use e to denote a generic random variable with distribution H), in which case the accumulated discrepancy ϵ_t at the beginning of period t follows a distribution H_t that is a $w(t)$ -fold convolution of H . Denote a u -fold convolution of H by $H^{(u)}$ and its moment-generating function by $M_\epsilon^{(u)}(\theta)$. Then equation (7) simplifies to

$$\frac{\tilde{\theta}}{\theta} = \frac{1 - \left(\frac{1}{T} \sum_{u=1}^T H^{(u)}(0) \right) \exp\{-\theta\tilde{j}\}}{1 - \left(\frac{1}{T} \sum_{u=1}^T M_\epsilon^{(u)}(\theta) \right) \exp\{-\theta\tilde{j}\}}, \text{ where } \tilde{j} = -\frac{1}{\tilde{\theta}} \ln(1 - \gamma). \quad (8)$$

Examining (8), we see that the seller’s estimation bias $\tilde{\theta}/\theta$ depends on just three factors: the service level parameter γ , the statistics $H^{(u)}(0) = \Pr\{\epsilon_u \leq 0\} = 1 - \Pr\{\epsilon_u > 0\}$, and an interaction between the true demand rate and the error distribution through $M_\epsilon^{(u)}(\theta)$. We examine the sensitivity of the estimation bias to all three factors in Section 5.4.

We have found equation (8) to be a good approximation of (7) for wide range of parameters—we provide evidence in Section 5.4—except when errors are large relative to demand. A simple modification to the proof of Proposition 2 shows that if $\mathbb{E}[\epsilon_t] \geq 0$ for all t then any solution $\tilde{\theta}$ to equation (8) satisfies $\tilde{\theta} \geq \theta$. In other words, assuming errors are nonnegative in expectation, the naive seller underestimates demand at the fixed point described by equation (8).

5.2. A Heuristic Prescription

Up until now, our focus has been on describing the estimation process of a seller operating under a misspecification of the sales-generating process. The question remains: what can a seller do about this misspecification?

Item-level RFID ideally provides the seller full visibility into customer-available inventory positions, and has the potential to foster demand estimation on par with our tracking seller. We believe that our work provides a fuller understanding of the value of RFID. However, even though RFID tag costs have fallen in recent years, a full item-level RFID implementation remains expensive and currently is only cost-effective for high-margin products. Furthermore, a number of technological and other challenges have hindered the implementation of RFID (Kumar et al. 2009).

In the absence of RFID, a seller who knows the probability distributions H_t of inventory discrepancies could in theory incorporate them into her demand estimation procedure. This is conceptually straightforward to do, but it is computationally complicated, since the updating simplicity afforded by assuming a newsvendor demand distribution does not carry over to the case with assumed errors. More importantly, the seller must know the distributions H_t of discrepancies even as she does not know the demand distribution. For these reasons, we do not pursue this approach further.

Instead, we propose a heuristic approach that makes use of a specific statistic of the discrepancy distribution at time t , $H_t(0) = \Pr\{\epsilon_t \leq 0\}$ ($= 1 - \Pr\{\epsilon_t > 0\}$). In general, estimating a single discrepancy statistic is easier than estimating the whole distribution. We envision a firm estimating the statistic for each SKU-location using a limited history of audit observations by exploiting similarities among locations and SKUs. In the case of newsvendor demand distributions, the demand “update” using this approach can be expressed in closed form similar to the case without errors.

The heuristic is based on the observation from Table 1 that $\mathbb{E}[C_t^n] = \Pr\{I_t > D_t \geq j_t\} + \Pr\{D_t \geq I_t \geq j_t\} = \Pr\{D_t \geq j_t\} \Pr\{\epsilon_t \leq 0\}$. Therefore $\mathbb{E}[C_t^n]/\Pr\{\epsilon_t \leq 0\}$ is an unbiased estimator for $\Pr\{D_t \geq$

$j_t\}$, which is the expectation of the tracking seller's stockout indicator when $I_t = j_t$. Specifically, the heuristic is simply to base stocking decisions on a belief on θ that takes a gamma distribution with parameters (a_t^h, b_t) , where

$$a_t^h = \max \left\{ 1, \quad a_0 + \sum_{\tau=1}^t \left(1 - \frac{c_\tau^n}{\Pr\{\epsilon_\tau \leq 0\}} \right) \right\}, \quad (9)$$

and b_t is as defined in equation (3). The “max” in the expression for a_t^h is to preclude $a_t^h < 0$ for small t , and has no impact on the long-run performance of the heuristic. Beyond this, a comparison of equations (9) and (3) shows that we are simply dividing the stockout indicators c_τ^n by $\Pr\{\epsilon_\tau \leq 0\}$. This adjustment inflates the stockout indicators more when the probability of positive discrepancies is high. This is intuitive, since positive discrepancies tend to interfere with the naive seller's detection of stockouts.

We assume that errors are small so that we can make the approximation $R_t \approx \epsilon_t$ as we did in deriving equation (8). Assuming exponentially distributed demand and following a similar analysis to the one in Section 5.1, we can characterize fixed points of the estimate-stock-estimate process involving the suggested heuristic. We find that fixed points $(\tilde{\theta}, \tilde{j})$ solve

$$\frac{\tilde{\theta}}{\bar{\theta}} = \frac{1 - \exp\{-\theta\tilde{j}\}}{1 - \left(\frac{1}{T} \sum_{u=1}^T M_\epsilon^{(u)}(\theta) \right) \exp\{-\theta\tilde{j}\}}, \text{ where } \tilde{j} = -\frac{1}{\tilde{\theta}} \ln(1 - \gamma). \quad (10)$$

Valid $\tilde{\theta}$ solutions to (8) and (10) are such that the denominator of the right-hand-side is strictly between zero and one, corresponding to the range $(0, \bar{\theta})$ for $\tilde{\theta}$, where $\bar{\theta} \equiv -\theta \frac{\ln(1-\gamma)}{\ln\left(\frac{1}{T} \sum_{u=1}^T M_\epsilon^{(u)}(\theta)\right)}$. The following proposition establishes that at fixed points in this range, the heuristic seller underestimates demand, but the bias is smaller than that of the naive seller. We see from equation (10) that the bias remaining after applying the heuristic is caused by deviation of $M_\epsilon^{(u)}(\theta)$ from one.

PROPOSITION 3. Suppose that both equations (8) and (10) have solutions on $(0, \bar{\theta})$, and let $\tilde{\theta}^n$ and $\tilde{\theta}^h$ be the smallest such solutions, respectively. If $\mathbb{E}[\epsilon_t] \geq 0$ for all t then $\tilde{\theta}^n \geq \tilde{\theta}^h \geq \theta$.

Proof. Let $Z^n(\tilde{\theta})$ and $Z^h(\tilde{\theta})$ indicate the right-hand sides of equations (8) and (10), respectively. We first show that $\tilde{\theta}^n \geq \tilde{\theta}^h$. Because $H^{(u)}(0) \leq 1$ for all u ,

$$\begin{aligned} Z^n(\tilde{\theta}) &\equiv \frac{1 - \left(\frac{1}{T} \sum_{u=1}^T H^{(u)}(0) \right) \exp\left\{ (\theta/\tilde{\theta}) \ln(1 - \gamma) \right\}}{1 - \left(\frac{1}{T} \sum_{u=1}^T M_\epsilon^{(u)}(\theta) \right) \exp\left\{ (\theta/\tilde{\theta}) \ln(1 - \gamma) \right\}} \\ &\geq \frac{1 - \exp\left\{ (\theta/\tilde{\theta}) \ln(1 - \gamma) \right\}}{1 - \left(\frac{1}{T} \sum_{u=1}^T M_\epsilon^{(u)}(\theta) \right) \exp\left\{ (\theta/\tilde{\theta}) \ln(1 - \gamma) \right\}} \equiv Z^h(\tilde{\theta}) \end{aligned}$$

for all $\tilde{\theta}$. (Note that for all $\tilde{\theta} \in [0, \max\{\tilde{\theta}^h, \tilde{\theta}^n\}]$ we have $M_\epsilon^{(u)}(\theta) \exp\{-\theta/\tilde{\theta} \ln(1-\gamma)\} < 1$ due to the assumption that $\tilde{\theta}^h \in (0, \bar{\theta})$ and $\tilde{\theta}^n \in (0, \bar{\theta})$.) Because $\lim_{\tilde{\theta} \rightarrow 0^+} Z^h(\tilde{\theta}) = 1 > 0$, we know that $Z^h(\tilde{\theta}) > \frac{\tilde{\theta}}{\theta}$ for $\tilde{\theta} < \tilde{\theta}^h$. Therefore, $\frac{\tilde{\theta}^n}{\theta} = Z^n(\tilde{\theta}^n) \geq Z^h(\tilde{\theta}^n)$ implies $\tilde{\theta}^n \geq \tilde{\theta}^h$.

The inequality $\tilde{\theta}^h \geq \theta$ follows from equation (10) and the fact $M_\epsilon^{(u)}(\theta) = \mathbb{E}[e^{\theta\epsilon_u}] = e^{\theta E[\epsilon_u]} \mathbb{E}[e^{\theta(\epsilon_u - E[\epsilon_u])}] \geq \mathbb{E}[e^{\theta(\epsilon_t - E[\epsilon_t])}] \geq 1$. As in the proof of Proposition 2, the first inequality follows from the assumption $\mathbb{E}[\epsilon_u] \geq 0$ and the second follows from applying Jensen's inequality. ■

5.3. Stationary Discrepancy Model

Consider a simplified special case of the model in Section 5.1 in which $T = 1$. That is, inventory record discrepancies in each period are independent draws from a stationary distribution. This model can be viewed as a straightforward extension of the repeated newsvendor setting commonly used in studies of demand learning (e.g., Lariviere and Porteus 1999).

Assume that the seller stocks \tilde{j} each period and that the random variable ϵ_t is drawn i.i.d. each period from a distribution H . The corresponding inventory discrepancy is $R_t = \min\{\tilde{j}, \epsilon_t\}$, with distribution $K_{t,\tilde{j}}$. Let $M_\epsilon(\theta)$ and $M_R(\theta, \tilde{j})$ indicate the moment-generating functions corresponding to H and $K_{t,\tilde{j}}$, respectively. Equation (7) simplifies to

$$\frac{\tilde{\theta}}{\theta} = \frac{1 - H(0) \exp\{-\theta\tilde{j}\}}{1 - M_R(\theta, \tilde{j}) \exp\{-\theta\tilde{j}\}}, \text{ where } \tilde{j} = -\frac{1}{\tilde{\theta}} \ln(1-\gamma), \quad (11)$$

equation (8) simplifies to

$$\frac{\tilde{\theta}}{\theta} = \frac{1 - H(0) \exp\{-\theta\tilde{j}\}}{1 - M_\epsilon(\theta) \exp\{-\theta\tilde{j}\}}, \text{ where } \tilde{j} = -\frac{1}{\tilde{\theta}} \ln(1-\gamma), \quad (12)$$

and equation (10) simplifies to

$$\frac{\tilde{\theta}}{\theta} = \frac{1 - \exp\{-\theta\tilde{j}\}}{1 - M_\epsilon(\theta) \exp\{-\theta\tilde{j}\}}, \text{ where } \tilde{j} = -\frac{1}{\tilde{\theta}} \ln(1-\gamma). \quad (13)$$

Furthermore, Propositions 2 and 3 continue to hold.

Observe that these equations yield exactly the same solutions as the corresponding equations in Sections 5.1 and 5.2 if we view the stationary error parameters as time-averaged versions of the discrepancy parameters in Section 5.1. For example, equations (12) and (8) are equivalent if we let $H(0) = (1/T) \sum_{u=1}^T H^{(u)}(0)$ and $M_\epsilon(\theta) = (1/T) \sum_{u=1}^T M_\epsilon^{(u)}(\theta)$. Therefore, the stationary model is a useful abstraction of the accumulating error model for the purpose of generating insights.

Name	Discrepancy distribution	Mean	Variance	$M_\epsilon(1/20)$	$M_\epsilon(1/2)$	$1 - H(0)$
Normal(0,1)	$\epsilon \sim \text{normal}(0,1)$	0	1	1.00125	1.13315	0.5
Normal(0,4)	$\epsilon \sim \text{normal}(0,4)$	0	4	1.00501	1.64872	0.5
ZeroNormal	$\epsilon \sim \begin{cases} \delta_0 & \text{w.p. } 3/4 \\ \text{normal}(0,4) & \text{w.p. } 1/4 \end{cases}$	0	1	1.00125	1.16218	0.125
ZeroExponential	$\epsilon \sim \begin{cases} \delta_0 & \text{w.p. } 1/2 \\ \text{exponential}(1) & \text{w.p. } 1/2 \end{cases}$	0.5	0.75	1.02632	1.50000	0.5

Table 2: Stationary discrepancy distributions assumed for numerical evaluations. The notation δ_0 denotes the probability distribution with all its mass at zero.

5.4. Numerical Evaluations

In this section, we compute solutions to the fixed-point equations derived in the previous subsections. Our purpose is to understand the dependence of the estimation bias identified by Proposition 2 on model primitives such as the true demand parameter θ , the critical fractile γ , and the discrepancy distribution H . We also seek to understand the impact of the heuristic prescription described in Section 5.2. For presentation simplicity, our discussion assumes the stationary discrepancy model of Section 5.3, which we have argued is sufficient for generating insights. Appendix C demonstrates this sufficiency.

We first report on instances in which $\theta = 1/20$ so that the true demand distribution is exponential with mean 20. We compute fixed points for four stationary discrepancy distributions presented in Table 2, all of which have nonnegative means. Figure 2 plots, as a function of the critical fractile parameter γ , estimated demand means and (type 1) service levels implied by solutions to equation (12). Appendix B reports the data behind the figure. For all instances represented in Figure 2, we find two solutions to equation (12). In all such cases, one has $\tilde{\theta}$ in the vicinity of the true demand mean and one has $\tilde{\theta}$ much larger than the mean demand such that \tilde{j} is very small. We report the former solution, which simulations show to represent long-run behavior of the estimate-stock-estimate process (see Appendix B).

The left-hand panel of Figure 2 shows the mean demand estimate $1/\tilde{\theta}$ for each instance. These results conform to the finding of Proposition 2 in that the naive seller consistently underestimates demand. That is, we see $1/\tilde{\theta} < 1/\theta$ for every instance. The magnitude of the bias depends on the critical fractile parameter γ : the estimation error is decreasing in γ , approaching zero as γ approaches 1. This is intuitive given that stockouts are less frequent for both tracking and naive sellers when γ is high.

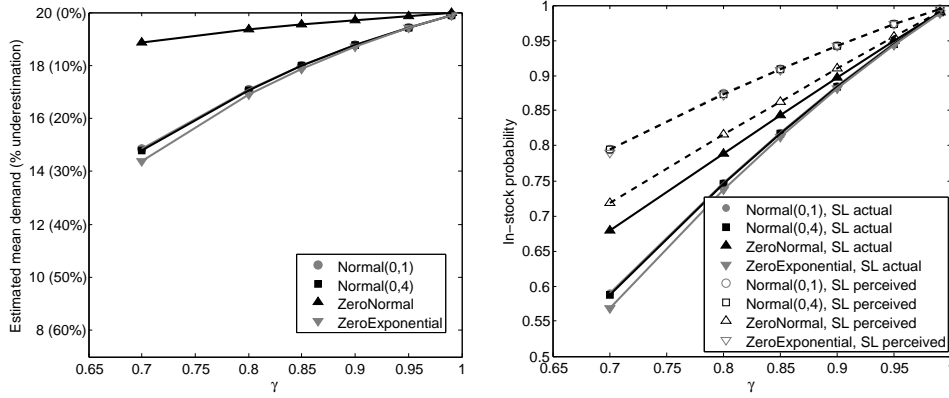


Figure 2: Plot of solutions to equation (12) for $1/\theta = 20$.

Equation (12) depends on the discrepancy distribution in two ways: through the probability $1 - H(0) = \Pr\{\epsilon_t > 0\}$ of positive discrepancies and through the moment-generating function $M_\epsilon(\theta)$. For all four discrepancy distributions $M_\epsilon(1/20)$ is close to one, so the dependence of the estimation biases on the discrepancy distributions is primarily through the statistic $H(0)$. Indeed, the Normal(0,1), Normal(0,4), and ZeroExponential distributions all have $H(0) = 0.5$ and display similar biases in Figure 2, whereas the ZeroNormal distribution has larger $H(0)$ (i.e., smaller $\Pr\{\epsilon_t > 0\}$) and results in less bias.

In the right-hand panel of Figure 2, we plot both the “perceived” service level $1 - \mathbb{E}[\tilde{C}^n]$ and the “actual” service level, calculated analytically as the (type 1) service level achieved by the physical stock $\tilde{j} - \epsilon_t$ with respect to the true exponential distribution of demand. We observe in all instances not only that the service level seen by customers is lower than the target, but also that the naive seller’s perceived service level exceeds both the target and actual service levels. These differences are again largest for low γ and high $\Pr\{\epsilon_t > 0\}$. We conclude that when a seller neglects inventory record inaccuracy when estimating demand, customers may experience lower than intended service levels even as the seller measures high service levels and therefore may fail to detect the problem.

Figure 3 presents solutions $\tilde{\theta}$ to equation (12) alongside simulated long-run demand estimates for instances with true demand parameter $\theta = 1/2$. (Details are available in Appendix B.) We again use discrepancy distributions from Table 2, meaning that errors are larger relative to demand compared with the $\theta = 1/20$ case, and we do not necessarily have $R_t \approx \epsilon_t$ as assumed in the derivation of equations (8) and (12). We are not able to find a solution to equation (12) for several instances in which $M_\epsilon(\theta)$ is relatively large and the critical fractile parameter γ is relatively low, but we obtain simulated estimates for all instances. Figure 3 shows that the equation, when it has a solution, underestimates the simulated estimates by a margin that decreases in the target service

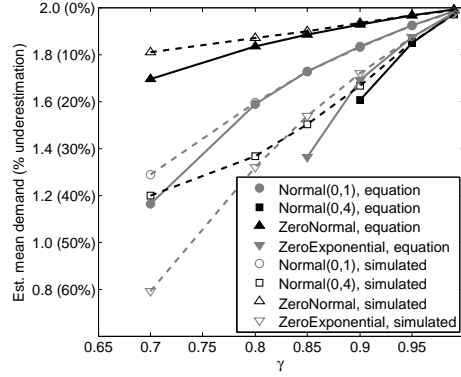


Figure 3: Demand estimates from equation (12) and long-run simulation for $\theta = 1/2$.

level. Looking at both the equation solutions and simulated estimates in Figure 3, we see that many of the same insights found for the $\theta = 1/20$ instances carry over to the $\theta = 1/2$ instances. The naive seller underestimates demand in all cases, and this underestimation is largest when γ is small. Furthermore, the ZeroNormal discrepancy distribution, which has the smallest $\Pr\{\epsilon_t > 0\}$, also yields the smallest biases. However, as predicted by equations (11) and (12), the relationship between discrepancy distribution and bias is more complicated when errors are large relative to demand. We observe that the biases are most significant for the Normal(0,4) and ZeroExponential discrepancy distributions, which have the largest values for $M_\epsilon(1/2)$.

In Figure 4, we explore the impact of the correction heuristic proposed in Section 5.2. For $\theta = 1/20$, we show solutions to equation (13) using $M_\epsilon(\theta)$ derived from the Normal(0,4) and ZeroExponential discrepancy distributions described in Section 5.4. Simulations (omitted to conserve space) show that these fixed points are representative of long-run behavior. For $\theta = 1/2$, where the approximation $R_t \approx \epsilon_t$ is less accurate, the plot shows simulated estimates of long-run behavior.

We observe that the heuristic sharply reduces the bias for all critical fractiles γ and for both discrepancy distributions, although the heuristic is relatively less effective for the case $\theta = 1/2$. Recall that the fixed point equation depends on the discrepancy distribution through two statistics, $\Pr\{\epsilon_t > 0\}$ and $M_{R_t}(\theta, \tilde{j})$, the latter of which is inconsequential when demand is large relative to errors. Therefore it makes sense that our heuristic correction, which is based on $\Pr\{\epsilon_t > 0\}$, is most effective when demand is large relative to errors.

5.5. Beyond the Exponential Demand Distribution

The fixed point equations and our numerical results so far have all assumed exponential demand distributions. A heretofore unanswered question is the impact of the demand distribution on the estimation bias we identify. The exponential demand distribution has a constant coefficient of

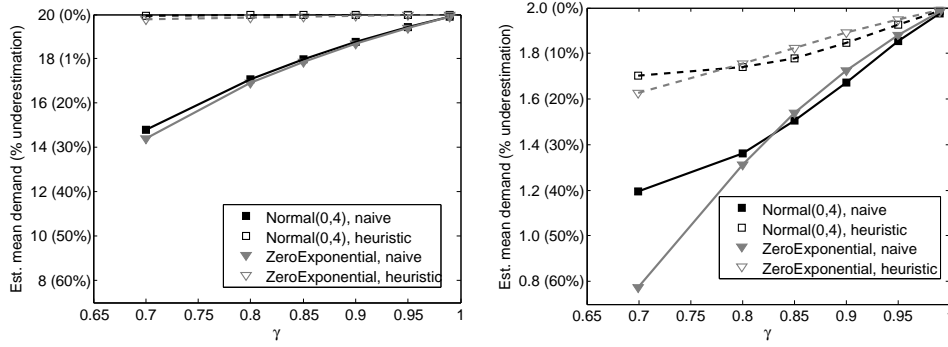


Figure 4: Mean demand estimates under heuristic prescription for $\theta = 1/20$ (left) and $\theta = 1/2$ (right).

variation (CV) equal to one. We explore the sensitivity of our results to the shape of the demand distribution.

Here we use a Weibull demand distribution $F(d|\theta, k) = 1 - \exp\{-\theta d^k\}$, a newsvendor distribution that subsumes the exponential distribution. In particular, we assume true demand distributions with CV 2.0 (“high variance”) and 0.5 (“low variance”). For mean demand 20, this translates to Weibull distributions with shape and rate parameters $k = 0.5427$, $\theta = 0.0014$ for “high variance” and $k = 2.1013$, $\theta = 0.0111$ for “low variance.” We assume the shape parameters k to be known and fixed, and the rate parameters θ to be unknown and subject to estimation by the seller. The “base case” demand distribution is the exponential one considered previously, which is equivalent to a Weibull distribution with $k = 1$.

Since we do not have closed-form fixed point expressions for the Weibull demand case, the results presented here represent results of long-run simulations starting from gamma prior distributions. In Figure 5, we plot the results of these simulations for the Normal(0,4) stationary discrepancy distribution for both mean 2 and mean 20 demand instances. (We have generated similar results for the ZeroExponential discrepancy distribution, with similar conclusions. We omit these results to conserve space.) We observe that the shape of the demand distribution indeed impacts the observed estimation bias. The estimation bias is larger for larger demand variance. Nevertheless, conditional on the demand distribution, the insights observed previously appear to carry over. For example, the proposed heuristic nearly eliminates the estimation bias for examples with mean demand 20 (where demand is large relative to errors) and reduces it significantly for examples with mean demand 2 (where demand is small relative to errors), consistent with previous results.

6. Discrete Demand Distributions

We look to further generalize our findings to other demand forms. For example, retail demand is often modeled as discrete, following a Poisson or negative binomial probability distribution. As

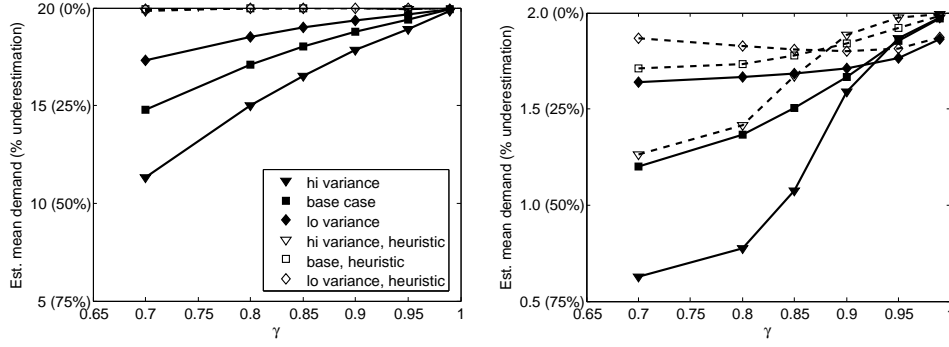


Figure 5: Simulated mean demand estimates for Normal(0,4) discrepancy distribution and Weibull demand distributions with means 20 (left) and 2 (right).

Braden and Freimer (1991) suggest, there does not seem to be a meaningful data reduction for these demand distributions as for the newsvendor distributions described in Section 3.1. This makes computation of the Bayes update of equation (1) considerably more complicated, and impossible to meaningfully express in closed form. In this section we present a framework for numerically evaluating fixed points of the estimate-stock-estimate process when inventory and demand are discrete. We demonstrate the framework on instances assuming Poisson demand. As in Sections 5.4 and 5.5, we assume stationary discrepancies for simplicity.

6.1. Procedure

Recall that the fixed point equation (11) can be derived by setting the demand estimate $\tilde{\theta}$ equal to the limiting maximum likelihood demand parameter estimate given the stationary stocking quantity \tilde{j} . We take the maximum likelihood perspective here.

We use the following procedure to calculate fixed points of the naive and heuristic sellers' estimate-stock-estimate processes. For each potential integer stocking quantity j (from a range of potential stocking quantities), we calculate the joint pmf $g(s, c|\theta)$ of sales and stockout indicators (S_t, C_t) the seller would see, assuming the true demand and error processes. We then let $\tilde{\theta}(j)$ be the maximizer (computed using MATLAB's *fminbnd* function) of the seller's expected log-likelihood function (i.e., ignoring errors). Letting $f(\cdot|\theta)$ indicate the pmf of demand, we can write the expected log-likelihood functions for the naive and heuristic sellers, respectively, as

$$\begin{aligned} \ell^n(\tilde{\theta}) &= \sum_{s=0}^{\infty} \left[g(s, 1|\theta) \ln \left(1 - F(s-1|\tilde{\theta}) \right) + g(s, 0|\theta) \ln \left(f(s|\tilde{\theta}) \right) \right], \\ \ell^h(\tilde{\theta}) &= \sum_{s=0}^{\infty} \left[g(s, 1|\theta) (1/\Pr\{\epsilon \leq 0\}) \ln \left(1 - F(s-1|\tilde{\theta}) \right) + g(s, 1|\theta) (1 - 1/\Pr\{\epsilon \leq 0\}) \ln \left(f(s|\tilde{\theta}) \right) \right. \\ &\quad \left. + g(s, 0|\theta) \ln \left(f(s|\tilde{\theta}) \right) \right]. \end{aligned}$$

Having generated the vector $\tilde{\theta}(j)$ for a range of j 's, we search for the \tilde{j} such that the critical fractile stocking quantity for $\tilde{\theta}(\tilde{j})$ equals \tilde{j} . We then declare that $(\tilde{\theta}(\tilde{j}), \tilde{j})$ is a fixed point of the estimate-stock-estimate process.

6.2. Illustrations for Poisson Demand

We illustrate the use of this procedure for a set of repeated newsvendor instances in which the true demand distribution is Poisson with mean 5. Inventory discrepancies in each period are stationary and given by differences of Poisson random variables with means 0.718 and 0.359, respectively. (Such a random variable is said to follow the Skellam distribution.) This implies an inaccuracy probability $\Pr\{\epsilon \neq 0\} = 0.57$ and a probability $\Pr\{\epsilon > 0\} = 0.4$ of positive discrepancies, which are in line with empirical studies of inventory inaccuracy in retail. Kang and Gershwin (2005) report 51% of SKUs having inaccurate inventory records. DeHoratius and Raman (2008) report an inaccuracy rate (*after* correction of misplacements) of 65%, of which 59% of the discrepancies were positive. Figure 18 of Gruen and Corsten (2008) shows an inaccuracy rate of 68%, with the distribution skewed towards positive discrepancies.

Results appear in Figure 6. We see that the insights are consistent with those for exponential distributions in Section 5.3: the naive seller underestimates demand over all instances, the bias is largest for relatively low service levels, and the heuristic correction significantly alleviates the bias.

Using the same discrepancy distribution as in Figure 6 and fixing $\gamma = 0.85$, we calculate steady state demand estimates for both the naive and heuristic sellers for a range of true demand means: 1.00, 1.25, 1.50, ..., 10.00. The results appear in Figure 7, expressed relative to the true mean demand. We see that the relative bias is largest when the demand mean is small relative to the fixed discrepancy distribution, consistent with what we found in Section 5.3. Given that the CV of a Poisson distribution is decreasing in its mean, these results are also consistent with results in Section 5.5 showing the estimation bias to be increasing in the CV of demand.

Demand underestimation generally leads to suboptimal decisions. We translate the demand underestimation results in Figure 7 into expected sales and profit terms in order to establish the correlation between estimation bias and performance and to show that the impact can be significant. Estimating this impact is complicated by the fact that inventory record inaccuracy is known to have a negative impact on matching supply with demand even when the seller has accurate demand distribution estimates. Therefore, we measure sales shortfall in two ways: Method A and Method B. In both methods, we assume the naive and heuristic sellers choose stocking quantities by targeting service level 0.85 in their estimated Poisson distributions. In Method A, we

compare results against a tracking seller who knows the discrepancy distribution and accounts for the presence of inventory errors. (Specifically, we assume a unit cost 1 charged on units I_t , a selling price 1.5, and a unit salvage value 0.912 for remaining units at the end of each period, which imply a target service level of 0.85. We compute the tracking seller’s optimal policy by enumeration. See Mersereau 2013 for analysis of optimal policies under record inaccuracy.) In Method B, we assume the tracking seller does not account for the presence of errors. Method A is a more reasonable model of the tracking seller, but Method B better isolates the performance impact attributable to demand underestimation.

Because decisions are discrete in the Poisson case, in many instances in Figure 7 the estimation bias does not lead to a difference in stocking quantities and therefore has no impact on performance. The differences that do arise track the estimation bias and are largest when demand is small relative to the discrepancy distribution. We believe that a 1.2% average sales loss due to underestimation (the average of the Method B sales shortfall percentages in the figure) is significant, and it is reduced significantly to 0.3% by the heuristic prescription, which requires little investment to implement. Comparing the results for Method A with those for Method B, we see that estimation bias is a substantial component of the overall sales impact due to inventory record inaccuracy.

We have also translated these sales shortfalls into profit shortfalls, estimated using the revenue and cost framework used to choose the tracking seller’s policy in Method A. We find that the relative profit impact is correlated with the relative sales impact but is smaller in magnitude. Averaged over the figure instances, we find profit shortfall to be 0.65% for the naive seller relative to the tracking seller using Method B and less than 0.1% for the heuristic seller. We note that any percentage gains in this profit are essentially bottom-line additions to gross margins, which tend to be slim in retail.

Figure 7 is admittedly a rough evaluation. First, even though we assume that demand is learned and possibly mis-estimated by the seller, we assume that the form of demand is correctly assumed to be Poisson and that the cost and error parameters are known. This may not be the case in some real applications, where the real impact could differ from our estimates. Second, we assume the same error process regardless of demand volume. There is empirical evidence that inventory inaccuracy is more pronounced for high-volume items (e.g., DeHoratius and Raman 2008). Because of this, we might expect that the results in Figure 7 overestimate the bias for low-volume items and underestimate the bias for high-volume items. A more accurate analysis for a particular situation would require detailed information on demand volumes and associated discrepancy distributions, not to mention cost parameters and inventory policies. We expect that the bias we identify may be

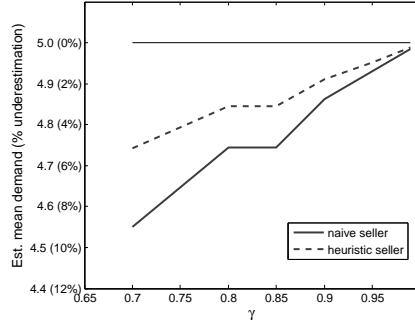


Figure 6: Fixed point demand estimates of the naive and heuristic sellers under Poisson demand.

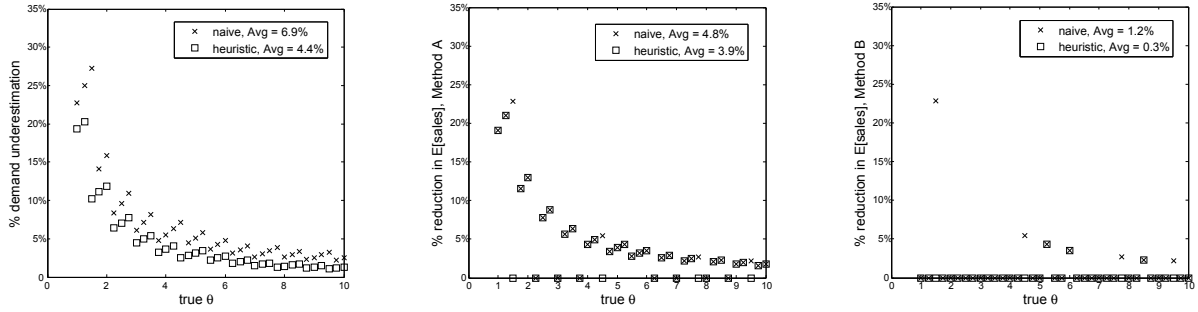


Figure 7: (a) underestimation of demand at fixed point, (b) expected sales shortfall under Method A, and (c) expected sales shortfall under Method B for naive and heuristic sellers facing Poisson demand.

small for some typical situations (e.g., with high target service levels and relatively small errors), but we believe that understanding when the estimation bias is relatively large or small is part of our contribution.

7. Concluding Remarks

We conclude that ignoring inventory record inaccuracy when trying to learn demand from censored observations can have a negative impact on a firm's operations, in most realistic cases leading to systematic underestimation of demand that in turn reduces the true service levels experienced by the firm's customers. Even as this happens, the service levels may appear to the firm to exceed targets. We find that the firm's demand underestimation is largest when service levels are relatively low and when there is a significant probability of positive (i.e., stock-reducing) errors. We suggest a simple heuristic correction that adjusts stockout observations based on the probability of positive errors. Our work suggests a new component of the value of inventory tracking technologies. In addition, we complement existing literature on inventory record inaccuracy by revealing another incentive for firms to systematically understand their inventory errors. We contribute to the literature on demand censoring by modeling the practically relevant case in which stockout observations

are unreliable.

Our work leaves several theoretical questions unanswered. While the fixed points we have identified in our analyses seem to represent steady-state behavior in our numerical studies, we have side-stepped a theoretical study of convergence. We also leave open the study of optimal stocking policies in the presence of demand learning and inventory errors. In this vein, a reasonable question is whether the “stock more” results of Lariviere and Porteus (1999), Ding et al. (2002), and others carry over to the setting with inventory errors.

Finally, some work would be required to apply our ideas to practical settings. While our parametric model is rigorous and convenient for analysis, it would be interesting to investigate to what extent our results extend to non-parametric settings or to estimation paradigms used in practice. We are aware of retailers who account for censoring via extrapolation schemes. If in a seven-day week the firm believes that the customer-available inventory was zero for two of the days, for example, then the firm multiplies observed sales by $7/5$ to approximate the week’s demand. Preliminary experiments with an estimation paradigm modeled on this idea reveal that inventory inaccuracy indeed leads to underestimation biases as in our model. We leave it for future work to thoroughly understand the impact of inventory inaccuracy and to invent effective prescriptions in such settings.

Acknowledgments

The author thanks Steve Graves, an anonymous associate editor, and three anonymous referees for constructive reviews that resulted in substantial improvements to the paper. He thanks Vinayak Deshpande and Linus Schrage for comments on a draft version of this manuscript, and he thanks seminar participants at the University of Florida and UCLA for helpful feedback. He thanks the University of North Carolina Kenan-Flagler Business School and the Sarah Graham Kenan Foundation for financial support.

References

- Agrawal, N., S. Smith. 1996. Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Nav. Res. Log.* **43** 839–861.
- Atali, A., H. Lee, Ö. Özer. 2011. If the inventory manager knew: Value of visibility and RFID under imperfect inventory information. Stanford University. Working paper.
- Bensoussan, A., M. Cakanyildirim, A. Royal, S. Sethi. 2009a. Bayesian and adaptive controls for a newsvendor facing exponential demand. *Risk and Decision Analysis* **1** 197–210.
- Bensoussan, A., M. Cakanyildirim, S. Sethi. 2007. Partially observed inventory systems: The case of zero balance walk. *SIAM J. Contr. Opt.* **46**(1) 176–209.
- Bensoussan, A., M. Cakanyildirim, S. Sethi. 2009b. A note on ‘the censored newsvendor and the optimal acquisition of information’. *Oper. Res.* **57**(3) 791–794.

-
- Besbes, O., A. Muharremoglu. 2013. On implications of demand censoring in the newsvendor problem. *Management Sci.* **59**(6) 1407–1424.
- Bisi, A., M. Dada, S. Tokdar. 2011. A censored-data multiperiod inventory problem with newsvendor demand distributions. *Manufacturing Service Oper. Management* **37**(11) 1390–1405.
- Braden, D. J., M. Freimer. 1991. Informational dynamics of censored observations. *Management Sci.* **37**(11) 1390–1405.
- Burnetas, A., C. Smith. 2000. Adaptive ordering and pricing for perishable products. *Oper. Res.* **48**(3) 436–443.
- Cachon, G. P., A. G. Kök. 2007. How to (and how not to) estimate the salvage value in the newsvendor model. *Manufacturing Service Oper. Management*. **9**(3) 276–290.
- Camdereli, A. Z., J. M. Swaminathan. 2010. Misplaced inventory and radio-frequency identification (RFID) technology: Information and coordination. *Production Oper. Management* **19**(1) 1–18.
- Chen, L. 2010. Bounds and heuristics for optimal Bayesian inventory control with unobserved lost sales. *Oper. Res.* **58**(2) 396–413.
- Chen, L. 2013. Fixing phantom stockouts: Optimal data-driven shelf inspection policies. Duke University. Working Paper.
- Chen, L., A. J. Mersereau. 2014. Analytics for operational visibility in the retail store: The cases of censored demand and inventory record inaccuracy. Duke University. Working Paper.
- Chen, L., E. L. Plambeck. 2008. Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing Service Oper. Management*. **10**(2) 236–256.
- Conrad, S. A. 1976. Sales data and the estimation of demand. *Operational Res. Quart.* **27**(1) 123–127.
- Cooper, W. L., T. Homem-de-Mello, A. J. Kleywegt. 2006. Models of the spiral down effect in revenue management. *Oper. Res.* **54**(5) 968–987.
- Cooper, W. L., T. Homem-de-Mello, A. J. Kleywegt. 2009. Learning and pricing with models that do not explicitly incorporate competition. Working paper, University of Minnesota.
- DeHoratius, N., A. Mersereau, L. Schrage. 2008. Retail inventory management when records are inaccurate. *Manufacturing Service Oper. Management*. **10**(2) 257–277.
- DeHoratius, N., A. Raman. 2008. Inventory record inaccuracy: An empirical analysis. *Management Sci.* **54**(4) 627–641.
- DeHoratius, N., Z. Ton. 2009. The role of execution in managing product availability. N. Agrawal, S. A. Smith, eds., *Retail Supply Chain Management*. Springer Science+Business Media, LLC, 53–77.
- Ding, X., M. Puterman, A. Bisi. 2002. The censored newsvendor and the optimal acquisition of information. *Oper. Res.* **50**(3) 517–527.

-
- Gaukler, G. M., R. W. Seifert, W. H. Hausman. 2007. Item-level RFID in the retail supply chain. *Production Oper. Management* **16**(1) 65–76.
- Gruen, T. W., D. Corsten. 2008. A comprehensive guide to retail out-of-stock reduction in the fast-moving consumer goods industry. Arlington, VA: Food Marketing Institute.
- Harpaz, G., W. Lee, R. Winkler. 1982. Learning, experimentation, and the optimal output decisions of a competitive firm. *Management Sci.* **28** 589–603.
- Heese, H. S. 2007. Inventory record inaccuracy, double marginalization, and RFID adoption. *Production Oper. Management* **16**(5) 542–553.
- Huh, W. T., R. Levi, P. Rusmevichientong, J. B. Orlin. 2011. Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. *Oper. Res.* **59**(4) 929–941.
- Huh, W. T., P. Rusmevichientong. 2009. A non-parametric asymptotic analysis of inventory planning with censored demand. *Math. Oper. Res.* **34**(1) 103–123.
- Iglehart, D. L., R. C. Morey. 1972. Inventory systems with imperfect asset information. *Management Sci.* **18**(8) B338–B394.
- Kang, Y., S. Gershwin. 2005. Information inaccuracy in inventory systems: Stock loss and stockout. *IIE Trans.* **37** 843–859.
- Kök, A. G., K. Shang. 2007. Inspection and replenishment policies for systems with inventory record inaccuracy. *Manufacturing Service Oper. Management* **9**(2) 185–205.
- Kök, A. G., K. Shang. 2014. Evaluation of cycle-count policies for supply chains with inventory inaccuracy and implications on RFID investments. *Europ. J. Oper. Res.* **237**(1) 91–105.
- Kumar, S., M. J. Anselmo, K. J. Berndt. 2009. Transforming the retail industry: Potential and challenges with RFID technology. *Transportation Journal* **48**(4) 61–71.
- Lariviere, M., E. Porteus. 1999. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Sci.* **45**(3) 346–363.
- Le Cam, L. 1986. *Asymptotic Methods in Statistical Decision Theory*. Springer-Verlag.
- Lee, H., Ö. Özer. 2007. Unlocking the value of RFID. *Prod. Oper. Management* **16**(1) 40–64.
- Lee, S., T. Homem-de-Mello, A. J. Kleywegt. 2012. Newsvendor-type models with decision-dependent uncertainty. *Math. Methods Oper. Res.* **76**(2) 189–221.
- Lu, X., J.-S. Song, K. Zhu. 2005. On ‘the censored newsvendor and the optimal acquisition of information’. *Oper. Res.* **53**(6) 1024–1027.
- Lu, X., J.-S. Song, K. Zhu. 2007. Inventory control with unobservable lost sales and Bayesian updates. Duke University. Working paper.
- Lu, X., J.-S. Song, K. Zhu. 2008. Analysis of perishable-inventory systems with censored demand data. *Oper. Res.* **56**(4) 1034–1038.

-
- Mersereau, A. J. 2013. Information-sensitive replenishment when inventory records are inaccurate. *Prod. Oper. Management* **22**(4) 792–810.
- Nahmias, S. 1994. Demand estimation in lost sales inventory systems. *Naval Res. Log.* **41** 739–757.
- Rekik, Y., E. Sahin, Y. Dallery. 2008. Analysis of the impact of RFID technology on reducing product misplacement errors at retail stores. *Int. J. Prod. Econom.* **112**(1) 264–278.
- Ton, Z., A. Raman. 2006. Cross sectional analysis of phantom products at retail stores. Harvard University. Working paper.
- Wecker, W. E. 1978. Predicting demand from sales data in the presence of stockouts. *Management Sci.* **24**(10) 1043–1054.