Management Science

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

On Implications of Demand Censoring in the Newsvendor Problem

Omar Besbes, Alp Muharremoglu,

To cite this article:
Omar Besbes, Alp Muharremoglu, (2013) On Implications of Demand Censoring in the Newsvendor Problem. Management Science 59(6):1407-1424. http://dx.doi.org/10.1287/mnsc.1120.1654

MANAGEMENT SCIENCE
Vol. 59, No. 6, June 2013, pp. 1407–1424
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.1120.1654

On Implications of Demand Censoring in the Newsvendor Problem

Omar Besbes
Graduate School of Business, Columbia University, New York, New York 10027, ob2105@columbia.edu

Alp Muharremoglu
Naveen Jindal School of Management, University of Texas at Dallas, Richardson, Texas 75080, alp@utdallas.edu

We consider a repeated newsvendor problem in which the decision maker (DM) does *not* have access to the underlying demand distribution. The goal of this paper is to characterize the implications of demand censoring on performance. To that end, we compare the benchmark setting in which the DM has access to *demand* observations to a setting in which the DM may only rely on *sales* data. We measure performance in terms of regret: the difference between the cumulative costs of a policy and the optimal cumulative costs with knowledge of the demand distribution. Through upper and lower bounds, we characterize the optimal magnitude of the worst-case regret for the two settings, enabling one to isolate the implications of demand censoring. In particular, the results imply that the exploration–exploitation trade-off introduced by demand censoring is fundamentally different in the continuous and discrete demand cases, and that active exploration plays a much stronger role in the latter case. We further establish that in the discrete demand case, the need for active exploration almost disappears as soon as a lost sales indicator (that records whether demand was censored or not) becomes available, in addition to the censored demand samples.

Key words: demand censoring; inventory management; newsvendor; estimation; nonparametric
History: Received January 29, 2010; accepted April 23, 2012, by Gérard P. Cachon, stochastic models and simulation. Published online in *Articles in Advance* January 28, 2013.

1. Introduction

An important factor driving firms to hold inventories is uncertainty in demand. Consequently, a key input in any stochastic inventory model is a description of the demand process, most often given in the form of a probability distribution. In practice, such a demand distribution would need to be estimated from historical data, and in many instances these data might be influenced by past inventory decisions. For example in retail, it is common for the demand that exceeds the available inventory to be lost, and in most cases this excess demand is *not* observable by the firm. Therefore, the firm may only have records of past *sales*, as opposed to actual *demand*. This restriction, commonly referred to as demand censoring, inevitably comes at a cost to the firm, and the purpose of this paper is to further one's understanding of the implications of censoring.

Inventory control with an unknown demand distribution was first studied by Scarf (1959) in a Bayesian framework with observable demand, with many studies following to include censoring, in which case significant tractability issues may arise (Braden and Freimer 1991). In parallel, an increasing number of studies have focused on nonparametric representations of the unknown demand, and our study falls in the latter category.

At a high level, the main contributions of the present paper are twofold. First, we show that the impact of censoring differs at a fundamental level as a function of demand being continuous or discrete. In particular, the tension between exploration and exploitation introduced by censoring is more significant for discrete distributions and impacts the type of experimentation that should be conducted. From a practical perspective, the distinction between discrete versus continuous distributions is a matter of granularity, i.e., how large is a "demand unit" compared to typical demand. Second, we establish that collecting even minimal information about lost sales can yield significant value. In discrete demand settings with a high level of granularity, even knowing whether or not any sales were lost can greatly mitigate the impact of demand censoring.

In more detail, we study a classical inventory system, a repeated newsvendor problem. We assume that the demand distribution is *unknown* to the decision-maker, and he or she has access only to data he or she collects over time. Demand censoring introduces a fundamental tension. One may select an action to minimize the current period cost (exploitation) given the current information, or explore by ordering more than this level, to reduce the extent of censoring and obtain more information for future periods. To isolate

the effects of censoring, we analyze different informational settings. We first consider the benchmark case in which demand is observable, and the decision maker may use past demand observations to determine her or his inventory decisions. We then focus on the case in which only sales are observable, as opposed to demand; this is the censored case. For example, if the inventory level in a given period is 15 and demand turns out to be 20, then sales are equal to 15. In the observable case, the decision maker would observe the value of demand, 20 and in the censored case, the decision maker would only know that demand was *greater than or equal to* 15. The difference between the performance in both settings can be interpreted as the price in performance that one pays due to censoring.

We measure performance of a policy through regret, the difference between the expected cost of the policy and that of an oracle with access to the true demand distribution acting optimally, and we set the objective of the decision maker to be one of minimizing the worst-case regret over a given class of demand distributions. Such a notion of minimax regret enables one to determine the performance level one should *aim at* given the specific information available. Although the framework is fairly general, we narrow down the analysis to a subclass of demand distributions to isolate cases in which censoring is a key driver of performance.[1] We aim at understanding how the minimax regret varies given the information available and the implications on the exploration–exploitation trade-off faced by the decision maker.

1.1. Summary of Main Results

When demand is continuous, we establish that the minimax regret grows logarithmically with the number of periods in both the observable and censored demand cases. This result is based on a lower bound for the minimax regret in the observable demand case (presented in Theorem 1) and the fact that some existing policies achieve the lower bound up to a multiplicative constant (see, e.g., Huh and Rusmevichientong 2009, §3.5). To achieve the best rate of growth of the minimax regret, one does not need to actively explore, because a stochastic direction of cost improvement is available even when the demand samples are censored, and as a result, one may apply stochastic gradient type algorithms such as the above mentioned policy.

When demand is discrete, we establish that for the observable case, the decision maker can achieve a worst-case regret that is bounded, i.e., regret will not grow beyond a certain value as the number of periods increases, regardless of the underlying demand distribution in the class under consideration (Theorem 2). On the other hand, when demand samples are censored, the minimax regret now grows logarithmically with the number of periods (Theorems 3 and 4). In contrast with the continuous distributions analysis, the minimax regret has a different order of magnitude depending on the demand samples being censored or not, and the exploration–exploitation trade-off comes to the foreground explicitly. Now, the decision maker no longer has access to a local (stochastic) direction of cost improvement. In the example above, if the ordering level is kept at 15, even an infinite number of sales observations would not allow the decision maker to systematically determine whether the optimal ordering quantity is strictly higher than 15 or not. To achieve the growth rate of the minimax regret, a policy has to occasionally order more than the current best estimate, and our results imply that the policy would have to experiment over a number of periods that grows logarithmically with the total number of periods.

The theoretical distinction between continuous and discrete cases should be interpreted carefully. The insights derived in our discrete demand setting should be valid in the presence of a high level of demand granularity (e.g., expected demand is low or demands are multiples of a large minimum order size), whereas situations with a low level of granularity should be closer to the continuous demand setting. As the granularity of demand decreases, the requirement of active exploration becomes less costly. We illustrate this progression with a numerical example in §5.

Various nonparametric approaches for multiperiod inventory management have been proposed in the literature. Stochastic gradient algorithms were studied by Burnetas and Smith (2000), Huh and Rusmevichientong (2009), and Kunnumkal and Topaloglu (2008), and by van Ryzin and McGill (2000) and Kunnumkal and Topaloglu (2009) in the related setting of repeated capacity booking problems. An adaptive value estimation method was studied by Godfrey and Powell (2001) and Powell et al. (2004), a maximum-entropy approach was analyzed by Eren and Maglaras (2013), and an algorithm based on the Kaplan–Meier estimator was presented by Huh et al. (2011). The nonparametric studies above focus on providing prescriptions for potentially censored demand settings, and sometimes analyze the performance of these policies via upper bounds or convergence of the prescribed decisions. In contrast, we analyze each level of information from a fundamental perspective and compare and contrast informational settings, as opposed to policies. The resulting profile of minimax regrets highlights the implications of demand censoring in a nonparametric setting, bringing to the foreground the role of discreteness and its relationship to active exploration.

[1] This subclass is detailed in §2.

To elucidate the impact of partial observations of lost sales, we study in the discrete case an intermediate informational setting, "partial censoring," in which the decision maker, in addition to observing sales, also observes whether demand exceeded (strictly) sales or not, i.e., observes whether any sales were lost.[2] In the example above, this corresponds to a decision maker observing sales of 15 and, in addition, knowing that demand was strictly greater than 15. We establish that in this setting, the availability of the lost sales indicator enables a decision maker to recover bounded regret as in the case of observable demand (see Theorem 5). The availability of the lost sales indicator alters the minimax regret growth significantly, highlighting its value. Intuitively, this result stems from the fact that the lost sales indicator provides the decision maker with "free experimentation," and the need for active exploration essentially disappears. Viewed differently, the availability of the indicator enables the decision maker to obtain a noisy signal about the potential need for an upward correction, removing the necessity of active exploration. We further illustrate numerically in §5 that observing a small percentage of lost sales (as opposed to only the first one) can eliminate most of the impact of censoring.

The availability of the lost sales indicator has been assumed, to the best of our knowledge, in all nonparametric studies that have appeared in the literature and that analyze newsvendor problems with an unknown discrete demand distribution (and inventory levels). The analysis of this setting also enables one to understand the important implications of such an a priori innocuous assumption.

1.2. Connection to Bayesian Formulations and General Sequential Decision Problems

It is worthwhile to compare the results above with those typically obtained in the context of Bayesian formulations. We refer the reader to Chen (2010), Bensoussan et al. (2009b), Akcay et al. (2009), and Chen and Plambeck (2008) for some recent studies in this context. The focus of these studies is mainly on structural properties of the optimal ordering policy and its relationship to one in the absence of censoring. The "stock more" result, namely, that it is optimal to order a higher level than the myopic optimal, and variants of it have been recurring; see Harpaz et al. (1982), Lariviere and Porteus (1999), and Ding et al. (2002) (and the related notes by Lu et al. 2008 and Bensoussan et al. 2009a). Such a stock-more property was absent from all nonparametric studies. By bringing to the foreground how discreteness drives the need to systematically stock more (active exploration), the present study closes a disconnect between the Bayesian approaches to demand censoring and the nonparametric ones. In addition, through the frequentist approach taken, one obtains a characterization of the optimal frequency for such deviations.

In the case of discrete distributions, the problem studied is one of finding the best possible inventory level among a finite number of possibilities and, as such, resembles at first sight a multiarmed bandit problem (see, e.g., Lai and Robbins 1985). However, the rewards of a given arm provide feedback about those of other arms, resulting in more information for the decision maker, and our analysis exploits this fact to obtain tight bounds as a function of the informational setting. The use of minimax regret objectives has also appeared in various streams of the economics and computer science literatures to analyze dynamic adversarial environments. Early references include Blackwell (1956) and Hannan (1957), and a review of this line of work appears in Foster and Vohra (1999) and Cesa-Bianchi and Lugosi (2006).

1.3. Measuring Lost Sales

Conrad (1976) observed that treating sales data as if they are equivalent to demand may lead to poor decisions. This is an example of model misspecification, and in such situations, a host of phenomena can arise; see also Cooper et al. (2006) and Cachon and Kök (2007). Although the importance of distinguishing between sales and demand data is well recognized by many practitioners, it is in many cases costly or simply impossible to keep track of the exact number of lost sales. The present study highlights the potential significant value associated with recording only a portion of those, and this can in general be done by means of procedures and/or information technology. For example, brick and mortar retailers that do not have all their stock on the shop floor may often post signs prompting customers to ask for a sales associate to check if the item is available. Whereas this ensures that demand is satisfied when possible, it also provides a means to measure stockouts.[3] Sales people can be instructed to make a record of cases in which the customer accepted to substitute an out-of-stock item with another similar product. Another possibility is for firms to record when an item runs out of stock between ordering periods, which may be an indication of the existence of lost sales, even if it is not possible to accurately measure those. In the mail order catalog setting, Schleifer (1992) highlighted how LL Bean has gone through the process of modifying their

[2] Note that in the continuous demand setting, there is no distinction between censored versus partially censored settings.

[3] Such a process may only ensure that one observes stockouts with some likelihood, because some customers might still decide not to communicate their needs if a product is not readily visible.

information technology system to record lost sales, and Anderson et al. (2006) studied the effectiveness of various innovative responses to mitigate the impact of stockouts, including the possibility of offering discounts or free shipping to a subset of customers to convince them to wait. Besides having the potential of convincing a subset of customers to not cancel their order, such approaches could give the firm a possible means for recording some stockouts. In B2B settings, tracking lost sales may be somewhat easier, as established relationships with some of the customers make it possible to know with relatively high confidence that one could have sold more. This can once again be supported with appropriate information technology; for example, add-on modules have been developed for SAP Business One to ensure that lost sales are documented. In online settings, it may be possible to ascertain with high likelihood that lost sales occurred, based on customer activity after the item is stocked out (e.g., a high number of clicks).

2. Problem Formulation

We consider a multiperiod newsvendor problem, in which unmet demand is lost and leftover inventory perishes at the end of each period. Let D_t denote the demand in period t. In what follows, we assume that $\{D_t\colon t=1,2,\ldots\}$ are independent and identically distributed random variables with support $\mathcal{S}\subseteq\mathbb{R}_+$, finite expectation, and cumulative distribution F. Let x_t denote the inventory decision in period t. The cost for period t is assumed to be given by

$$C(F,x_t)=h\,\mathbb{E}[(x_t-D_t)^+]+b\,\mathbb{E}[(D_t-x_t)^+], \qquad (1)$$

where b is the per unit underage cost, and h is the per unit overage cost; both are assumed to be positive. Throughout this paper, we assume zero leadtime. If one lets

$$\beta=\frac{b}{h+b}, \qquad (2)$$

a standard derivation yields that an optimal ordering quantity is given by

$$x_F^*=\min\{x\in\mathcal{S}\colon F(x)\geq\beta\}.$$

We will use the convention of calling x_F^* the β-quantile of the distribution F. We let $C(F,x_F^*)$ denote the optimal per period newsvendor cost, and let $\mathcal{C}^*(F,T)=T\cdot C(F,x_F^*)$ denote the optimal cumulative cost over T periods, noting that both of these quantities can only be computed with knowledge of the distribution F. In this paper, we do not assume that the demand distribution F is known to the decision maker. We will only assume that F belongs to a class $\mathcal{F}$, to be specified later.

2.1. Informational Levels and Admissible Policies

Although the decision maker does not know the demand distribution F, he or she may use information collected over time to infer information about F and refine her/his decisions. We consider three settings corresponding to different data being available to the decision maker when ordering a quantity at time t. In the *observable demand* case, the decision maker has access to all past decisions x_s and demands D_s, for $s\leq t-1$. The second setting we consider is one in which the firm has only access to past decisions x_s and sales given by $\min\{D_s,x_s\}$ for $s\leq t-1$. We will refer to this setting as the *censored demand* case. In addition, when demand is discrete, we will also consider an intermediate setting referred to as *partially censored demand* case. In this setting, the decision maker, in addition to having access to the information available in the censored demand case, also observes whether demand strictly exceeded the ordering quantity in all past periods, $\mathbf{1}\{D_s>x_s\}$ for $s\leq t-1$ (where $\mathbf{1}\{\cdot\}$ is the indicator function). In other words, the decision maker observes whether any sales were lost.

We will denote quantities with superscript $a=u$, c, pc to refer to the uncensored, censored, and partially censored, settings, respectively. For each of the settings, a policy will be said to be nonanticipating if the quantity ordered in the tth period, x_t, is determined by the available history.[4] We will restrict attention to the set of nonanticipating policies denoted by $\mathcal{P}^a$, $a=u$, c, pc, and for any policy $\pi\in\mathcal{P}^a$, we denote the quantity ordered in the tth period by x_t.

2.2. Objective

We focus on the performance over a finite horizon T. Let π be an admissible policy for informational setting $a=u$, c or pc, and let

$$\mathcal{C}^\pi(F,T)=\mathbb{E}^\pi\left[\sum_{t=1}^{T}C(F,x_t)\right]$$

denote the expected cost over the first T periods when using policy π. In the three informational settings above, the decision maker does not know initially the distribution of the demand F, and as a result cannot compute $\mathcal{C}^\pi(F,T)$ and a fortiori cannot set as an objective to maximize $\mathcal{C}^\pi(F,T)$. We adopt the following minimax regret objective: select an admissible policy to minimize the worst-case difference between the cost[5] incurred and the optimal cost one could have

[4] A formal description of the histories in each setting is presented in the preliminaries of Appendix A.

[5] If one uses a standard profit maximization model with a constant selling price p, unit purchase cost c, and salvage cost s, one can show that the regret in terms of profit is equal to the regret in costs, if one uses an overage cost of $c-s$ and an underage cost of $p-c$.

incurred with knowledge of F, $\sup_{F\in\mathcal{F}}\{\mathcal{C}^\pi(F,T) - \mathcal{C}^*(F,T)\}$. We let

$$\mathcal{R}^a(\mathcal{F},T) = \inf_{\pi\in\mathcal{P}^a}\sup_{F\in\mathcal{F}}\left\{\mathcal{C}^\pi(F,T) - \mathcal{C}^*(F,T)\right\}, \quad (3)$$

where $a = u$, p, or pc, depending on the informational setting studied. This objective is well posed and can be seen as a game between the decision maker and "nature." The decision maker initially selects an admissible policy π, and subsequently nature can choose any distribution in $\mathcal{F}$ that would maximize the regret $\mathcal{C}^\pi(F,T) - \mathcal{C}^*(F,T)$. Given that the sets of admissible policies satisfy $\mathcal{P}^c \subseteq \mathcal{P}^{pc} \subseteq \mathcal{P}^u$, the minimax regret performances for the three informational settings are ordered as $\mathcal{R}^u(\mathcal{F},T) \le \mathcal{R}^{pc}(\mathcal{F},T) \le \mathcal{R}^c(\mathcal{F},T)$. Because characterizing, in an exact manner, each of the quantities $\mathcal{R}^u(\mathcal{F},T)$, $\mathcal{R}^{pc}(\mathcal{F},T)$, and $\mathcal{R}^c(\mathcal{F},T)$ is likely to be a highly intractable problem, in the rest of this paper, we will focus on quantifying these quantities through lower and upper bounds when the class $\mathcal{F}$ is selected to appropriately isolate the demand censoring effects.

2.3. Assumptions

The objective of the present study is to further one's understanding of the implications of demand censoring on performance and to provide insights on the exploration–exploitation trade-off resulting from censoring. To crisply understand those, it is important to separate the implications of censoring from other drivers of performance. In general, when the demand distribution is unknown, there are various drivers of the minimax regret. In particular, it might be fundamentally difficult to estimate the optimal decision, and this might lead to a high regret, independently of the demand being censored or not. To avoid this, we will assume that the admissible cost function is not flat around the optimal quantity. In particular, we consider the class $\mathcal{F}$ of demand distributions with finite expectation that satisfy the following for some $M > 0$ and ε in $(0,1)$:

(i) $x_F^* \le M$ (bounded optimal order quantity);

(ii) if demand is continuous, $F(\cdot)$ is differentiable, and $F'(x) \ge \varepsilon$ for all $x \ge 0$ (strictly convex cost function);

(iii) if demand is discrete, $|F(x) - \beta| \ge \varepsilon$ for $x = x_F^* - 1$, x_F^* (minimal separation around optimal quantity).

The first condition is mild; in settings where demand is bounded, one could take M to be a bound on the maximal value that demand may take. The second condition ensures that the objective function is strictly convex in the case of continuous demands. Noting that $F(x_F^* - 1) < \beta \le F(x_F^*)$ by definition, the third condition implies that $C(F,x) - C(F,x^*) > (h+b)\varepsilon$ for all $x \ne x^*$, which precludes too much flatness of the objective function around the optimal quantity. To highlight some of the issues one faces when the assumptions above are not satisfied, consider the following example with discrete demand.

Example 1. Fix the time horizon T such that $1/\sqrt{T} < \min\{\beta, 1-\beta\}$ and suppose that the class $\mathcal{F}$ consists of only two distributions, F_a and F_b, defined as follows:

$$F_a(k) = \begin{cases} \beta + \delta_T & \text{if } k = 0, \\ 1 & \text{if } k \ge 1; \end{cases} \quad F_b(k) = \begin{cases} \beta - \delta_T & \text{if } k = 0, \\ 1 & \text{if } k \ge 1. \end{cases}$$

The optimal ordering quantity is 0 for F_a and 1 for F_b.

If $\delta_T = 1/\sqrt{T}$, it is possible to show that, in the case of *observable* demand, no policy can achieve a better worst-case regret than $O(\sqrt{T})$ (i.e., $\mathcal{R}^u(\mathcal{F},T) \ge C\sqrt{T}$),[6] whereas if $\delta_T = \delta_0$ is a positive constant, independent of the horizon, then the regret will be bounded (see §4.1). Intuitively, the difference in performance stems from the fact that when $\delta_T = 1/\sqrt{T}$, it is not possible to distinguish reliably between the two distributions F_a and F_b in T periods because the estimation noise will be of the same order as the separation between the demand distributions. This implies in turn that it is not possible to determine the optimal ordering quantity reliably, and the order of magnitude of regret under both censored and uncensored demands would be dominated by the $\sqrt{T}$ term without the minimum separation assumption.

In the example above, the $\sqrt{T}$ loss in performance does not stem from demand censoring (demand samples are observable) but stems exclusively from the difficulty associated with identifying which of the distributions generates demand and the resulting challenge of estimating an optimal ordering quantity x^*. This situation arises due to the presence of a very small probability mass around x^* and the fact that the cost function can be very flat around this value.

3. The Case of Continuous Distributions

We discuss here the case in which there is a continuum of decisions and the demand distribution is continuous. The next result establishes a fundamental lower bound on achievable performance in the onservable demand setting.

Theorem 1. *Suppose $T \ge 2$. Consider the setting in which demand is observable. For any policy π in $\mathcal{P}^u$,*

$$\sup_{F\in\mathcal{F}}\{\mathcal{C}^\pi(F,T) - \mathcal{C}^*(F,T)\} \ge \underline{K}_u[M + \log T], \quad (4)$$

[6] A close inspection of the proof of Lemma 4 in the online companion provides the lower bound; in particular, replacing ε by $1/\sqrt{T}$ in (C-8) yields the result.

where $\underline{K}_u$ is a positive constant that only depends on ε and on the cost parameters b and h.

This result states that, independently of the policy one selects, when demand samples are observable, it will never be possible to achieve a smaller regret than $O(\log T)$ uniformly over all admissible distributions. Whereas the above result is stated for the observable demand case, it also of course provides a lower bound on the performance for the case in which demand samples are censored. However, policies that can achieve a minimax regret of order $\log T$ in the censored case have been developed by, e.g., Huh and Rusmevichientong (2009, §3.5). Hence, when the demand distribution and decisions are continuous, the minimax regret is of the same order of magnitude for both the censored and observable cases.

The policy of Huh and Rusmevichientong (2009) is based on constructing a stochastic gradient of the cost function, $G_t = h\mathbf{1}\{x_t > D_t\} - b\mathbf{1}\{x_t \le D_t\}$, which has expectation $\mathbb{E}[G_t] = C'(F, x_t)$, and the policy prescribes to follow the resulting direction of improvement with carefully designed steps. In other words, despite the fact that demand is censored, local information is available to move in the "right" direction, and this enables one to achieve the same growth in the rate of regret as in the observable demand case. As we will see in the next section, the picture differs significantly in the case of discrete distributions.

4. The Case of Discrete Distributions

We now assume that D has a discrete distribution and, without loss of generality, that it has support in the set of nonnegative integers.

4.1. Observable Demand

In the case in which demand samples are observable, the decision maker's ordering decisions do not have any impact on the observations that are collected. We first analyze the performance of a natural candidate policy, the policy π^u, defined below. At every time $t \ge 2$, it orders x_t (the minimum of the empirical sample quantile) and M (the bound on the optimal ordering quantity), and orders an arbitrary quantity in the first period. Such empirical quantile policies have been used as a benchmark in numerical experiments in the literature.

Algorithm 1 (π^u)

Step 1. *Initialization*: Select $x_1 \in \{0, \ldots, M\}$ arbitrarily

Step 2. *Ordering*: For $t \ge 2$

$$\text{Set} \quad q_t = \inf\left\{k\colon (t-1)^{-1}\sum_{i=1}^{t-1}\mathbf{1}\{D_i \le k\} \ge \beta\right\} \qquad \text{[sample } \beta\text{-quantile]}$$

$$x_t = \min\{q_t, M\}$$

Theorem 2. *The sample quantile policy π^u described in Algorithm* 1 *achieves a worst-case regret that satisfies*

$$\sup_{F\in\mathcal{F}}\{\mathcal{C}^{\pi^u}(F,T) - \mathcal{C}^*(F,T)\} \le \bar{K}_u M, \tag{5}$$

where $\bar{K}_u$ is a positive constant that only depends on ε and the cost parameters b and h.

This result states that the difference between the expected cost under π^u and under an optimal policy that has access to the demand distribution is actually bounded by a number that does not depend on the length of the planning horizon T and thus will not grow indefinitely as T grows. We refer to this property as bounded regret. This result contrasts with the limit in performance derived in the continuous setting, highlighting the differing nature of the two problems, even in the observable demand case. When demand is observable, there is no interaction between observations and decisions, and the difference above stems from the fact that for a discrete support, one may identify the β-quantile exactly with very high probability as we detail below, whereas for a continuous support, it is only possible to do so up to some correction factor.

Proof Sketch. The analysis of the performance of π^u relies on the following key large deviations result, which is also a building block for Theorems 4 and 5 in the coming sections.

Lemma 1. *Let y be an integer, and let $Z_i = \min\{y, D_i\}$ for $i \ge 1$. Let $\tilde{x}_t = \inf\{k\colon (t-1)^{-1}\sum_{i=1}^{t-1}\mathbf{1}\{Z_i \le k\} \ge \beta\}$ be the sample β-quantile based on the first $t-1$ observations of Z_i. Then $\tilde{x}_t$ satisfies*

$$\mathbb{P}\{\tilde{x}_t \le j\} \le \alpha_j^{t-1} \quad \text{for all } j \le \min\{x_F^*, y\} - 1, \tag{6}$$

$$\mathbb{P}\{\tilde{x}_t \ge j\} \le \alpha_{j-1}^{t-1} \quad \text{for all } j \ge x_F^* + 1, \tag{7}$$

for some nonnegative constant α_j such that $\alpha_j \le 1 - m_\beta(F(j)-\beta)^2 \le 1 - m_\beta\varepsilon^2$ for all $j \le y$, where m_β (defined in (A3)*) is a positive constant that only depends on β.*

The result above is true for $y = +\infty$. In such a case, Lemma 1 bounds the probability of the sample β-quantile being away from the optimal newsvendor quantity x_F^*, and shows that this probability converges to zero exponentially fast as the number of observations grows. This fast convergence is the key factor leading to the performance guarantee of the sample quantile policy π^u.

4.2. Censored Demand

4.2.1. A Lower Bound on Achievable Performance. The next result provides a fundamental limit on the performance for the censored demand case.

THEOREM 3. *Suppose that* $\varepsilon \le \min\{\beta, 1-\beta\}/2$. *Then for any policy* π *in* $\mathcal{P}^c$,

$$\sup_{F\in\mathcal{F}}\{\mathcal{C}^\pi(F,T) - \mathcal{C}^*(F,T)\} \ge \underline{K}_c[M + \log T], \quad \text{for all } T \ge T_0, \quad (8)$$

where $\underline{K}_c$ *is a positive constant that only depends on* ε *and the cost parameters* b *and* h, *and* T_0 *is an integer that only depends on* ε.

This result, in conjunction with the upper bound provided in Theorem 2 for the observable demand case, highlights that the impact of having access to only censored demand observations leads to at least a degradation of order $\log T$ with respect to the worst-case regret criterion. In particular, it is impossible to design policies with bounded worst-case regret when demand is censored. At an intuitive level, the lower bound (8) can be interpreted as follows: the factor of M is the loss that one must incur to "zoom in" on the neighborhood of the optimal ordering quantity x^* and the factor of $\log T$ represents the loss that one must incur to refine one's estimate of x^* once one operates around x^*.

Proof Sketch and Intuition. The necessary regret growth with the time horizon T stems from the fact that, when having access to sales only, one will not be able to refine one's confidence about the current "best estimate" of the optimal ordering level while using it. Indeed, suppose a policy orders $\hat{x}$, its current estimate of x_F^*, for some periods. For those periods, because of censoring, the inventory manager observes whether demand was greater than or equal to $\hat{x}$ (which is equivalent to observing $\mathbf{1}\{D_t \le \hat{x}-1\}$), but not whether demand was equal to $\hat{x}$ (i.e., $\mathbf{1}\{D_t = \hat{x}\}$), and hence cannot use these observations to reliably refine his or her estimate of $\mathbb{P}(D_t \le \hat{x})$. However, the optimal inventory level x_F^* is the first value x for which $\mathbb{P}\{D_t \le x\} \ge \beta$, and hence having access to a better estimate of $\mathbb{P}(D_t \le \hat{x})$ is in general needed to improve the confidence in the estimate of x_F^*. As a result, when the decision maker uses the current best estimate of the optimal level, he or she cannot decrease the "risk" of having used the wrong level while using its best estimate. In turn, the decision maker needs to periodically order quantities that are strictly above $\hat{x}$ to refine his or her estimate as time progresses. The proof of Theorem 3 formalizes the above intuition through information theoretical arguments. In particular, we consider the case in which there are only two possible distributions, F_a and F_b, such that it is optimal not to order any units under F_a and to order one unit under F_b. In such a case, if one does not order any unit (the optimal decision under F_a), then one cannot infer any additional information about the true distribution because demand is censored and one will only observe sales (which are equal to zero). To learn about the true distribution, one has to order a positive quantity. After T periods, if the decision maker wants to settle on ordering zero, they will need to have experimented with one for at least $O(\log T)$ periods to ensure a small probability of error (which, roughly speaking, decreases exponentially with the number of periods when one orders one unit). It is this minimal amount of required experimentation that drives the fact that the worst-case regret has to grow at least at a rate of $\log T$.

It is interesting to contrast the result above with that of continuous demand. In the latter case, local information regarding a (stochastic) direction of improvement was available. In contrast, in the discrete case, when one orders x_t while one still observes $G_t = h\mathbf{1}\{x_t > D_t\} - b\mathbf{1}\{x_t \le D_t\}$, G_t only provides one-sided information, $\mathbb{E}[G_t] = C(F,x_t) - C(F,x_t-1)$, and only indicates the potential for a downward correction but not for an upward one. In other words, through sales observations while ordering x_t, one may not obtain a noisy signal about $C(F,x_t+1) - C(F,x_t)$. This unavailability of local information plays a crucial role in the added regret and drives the need for the systematic experimentation highlighted above. In particular, in §4.3, we will see that as soon as one recovers such local information for both downward and upward corrections, then one may obtain a bounded regret, as in the uncensored case, with no active experimentation required.

4.2.2. A Near-Optimal Policy. We now construct a policy for the censored demand setting. It operates in stages and maintains an estimate of the optimal ordering quantity whose "accuracy" improves from stage to stage. Each stage j starts with an exploitation phase of length Δ_j^e, during which the current estimate is applied. Then, the algorithm computes the empirical β-quantile of the observations in this phase. If this quantity is strictly less than the ordering quantity, this suggests that the current estimate was too high, and the empirical β-quantile becomes the new estimate, initiating a new phase. If, on the other hand, the β-quantile is equal to the experimentation level, which suggests that the current estimate is *greater than or equal to* the optimal ordering quantity, the algorithm enters into an exploration phase of length Δ_j^e with an ordering quantity that is increased beyond the current estimate by a given factor. After the first exploration phase, if the empirical β-quantile is now below the experimentation level, the estimate for the optimal ordering quantity is updated and the algorithm enters into a new stage. If, however, the new β-quantile is still equal to the experimentation level, then one increases the experimentation level once again by a given factor and starts another exploration phase. The

policy proceeds in this fashion with successive contingent exploration phases until the latest empirical β-quantile is strictly below the corresponding experimentation level, at which point it proceeds to a new stage. A formal description of the policy is provided below.

Algorithm 2 (Input parameters: $\{(\Delta_j^c, \Delta_j^e, m_j^n)\colon j \geq 1, n \geq 1\}$)

Step 1. *Initialization*:
Set $t_s = 0$, $y_1^{(0)} \in \{1, \ldots, M\}$, $n = 0$, $j = 1$
Step 2. *Ordering*:
Set $x_t = y_j^{(0)}$ for $t = t_s + 1, \ldots, t_s + \Delta_j^c$ [exploitation]
Estimate β-quantile of $\min\{D_t, x_t\}$ over the exploitation phase:

Compute

$$q_j^{(0)} = \inf\left\{k\colon |\Delta_j^c|^{-1} \sum_{t=t_s+1}^{t_s+\Delta_j^c} \mathbf{1}\{\min\{D_t, x_t\} \leq k\} \geq \beta\right\}$$

Set $n = 0$ and $t_s = t_s + \Delta_j^c$
While $q_j^{(n)} = y_j^{(n)}$ and $y_j^{(n)} \leq M - 1$
Set $y_j^{(n+1)} = \min\{y_j^{(n)} + \max\{\lceil m_j^{n+1} y_j^{(n)} \rceil, 1\}, M\}$
Set $x_t = y_j^{(n+1)}$ for $t = t_s + 1, \ldots, t_s + \Delta_j^e$ [exploration]
Estimate β-quantile of $\min\{D_t, x_t\}$ over the exploration phase:

Compute

$$q_j^{(n+1)} = \inf\left\{k\colon |\Delta_j^e|^{-1} \sum_{t=t_s+1}^{t_s+\Delta_j^e} \mathbf{1}\{\min\{D_t, x_t\} \leq k\} \geq \beta\right\}$$

Set $n = n + 1$ and $t_s = t_s + \Delta_j^e$
End
Set $y_{j+1}^{(0)} = q_j^{(n)}$
Step 3. *Loop*: Set $j = j + 1$ and go to Step 2.

The policy outlined above may take an arbitrary initial ordering level as an input. In particular, it is possible to start at a level based on the input of experts or some sort of estimate based on observations from related products. The decision x_t prescribed by the policy at time t only depends on the sales observations up until time $t - 1$, and hence it is admissible in the censored demand setting.

The role of the sequence m_j^n in the algorithm is to modulate the steps taken as time progresses. Suppose that $m_j^n = 0.5$ for all n in cycle j, and the algorithm is at a low ordering level compared to the optimal. With high likelihood, there will be successive contingent exploration phases, and each new exploration level will be approximately 1.5 times the previous one. This enables the algorithm to reach the neighborhood of the optimal quantity rather quickly. However, when the algorithm has had some time to learn about the demand distribution and is likely to be closer to the optimal quantity, one should select a smaller quantity m_j^n, to take smaller steps while performing successive contingent exploration phases.

We next specify a particular sequence of exploration and exploitations phases that will lead to near-optimal performance. The intuition for the selection is based on the insights gleaned from the proof of the lower bound of Theorem 3 that indicated that experimentation should take place at logarithmic frequency. Hence, we will take sequences such that $\Delta_j^e \approx \log \Delta_j^c$. In particular, fix $a, z \in (1, 2)$, and suppose one selects

$$\gamma_1 = \gamma_e = 2\lceil -[\log(1 - m_\beta \varepsilon^2)]^{-1} \log(\max\{M, 2\})(\log a)^{-1} \rceil, \quad (9)$$

$$\Delta_j^c = \gamma_1 \lceil a^{z^{j-1}} \rceil, \quad j \geq 1 \quad [\textit{exploitation} \text{ phase lengths}], \quad (10)$$

$$\Delta_j^e = \gamma_e z^{j-1}, \quad j \geq 1 \quad [\textit{exploration} \text{ phase lengths}], \quad (11)$$

where m_β was defined in Lemma 1 (see (A3)). Suppose also that $m_j^n > 0$ for all $j \geq 1$ and $n \geq 2$, and that for some given j', $m_j^1 = 0$ for $j \geq j'$.[7] Let π^c denote the resulting policy. The next result provides a characterization of the performance of π^c.

Theorem 4. *π^c satisfies*

$$\sup_{F \in \mathcal{F}} \{\mathcal{C}^{\pi^c}(F, T) - \mathcal{C}^*(F, T)\} \leq \bar{K}_c \log M[M \log M + \log T], \quad (12)$$

where $\bar{K}_c$ is a positive constant.

This result, in conjunction with the result of Theorem 3, yields that the minimax regret grows logarithmically with time in the censored setting, and that the policy π^c is near optimal.

Proof Sketch and Intuition. The intuition underlying the performance of the policy above is based on the fact that the policy π^c ensures that sufficient experimentation is performed to refine one's estimate of the β-quantile as time progresses, and such experimentation is performed essentially at the "optimal" frequency of $(\log T)/T$. More precisely, the proof establishes by induction that the estimate of the β-quantile at the beginning of stage j, $y_j^{(0)}$, satisfies

$$\mathbb{P}\{y_j^{(0)} \leq i\} \leq a_j \alpha_i^{\Delta_j^e}, \quad \text{for } i \leq x_F^* - 1,$$

$$\mathbb{P}\{y_j^{(0)} \geq i\} \leq a_j \alpha_{i-1}^{\Delta_j^e}, \quad \text{for } i \geq x_F^* + 1,$$

[7] The rationale for the latter is to ensure that only local experimentation is conducted once the algorithm has zoomed in on the correct region.

where a_j is an appropriately bounded sequence, and α_i is the same constant as that defined in Lemma 1. Hence, the probability of ordering a quantity that differs from the optimal one during an exploitation phase shrinks exponentially fast as the number of stages increases. We use this result to account for the losses over the exploitation and exploration phases, compared to the optimal cost with knowledge of the demand distribution. In addition to losses of order M due to the initial exploration phase of the algorithm, one will incur bounded losses for the exploration phases; these are of length $\Delta_j^e \approx \log \Delta_j^c$. We establish that in each stage, the algorithm will—with very high likelihood—actively explore, by ordering more than the current estimate of the optimal quantity, and that the policy will most often enter a single exploration phase. The total losses over those phases can be shown to be of order $\log T$. The probability bounds above imply that the number of stages in which more than one contingent exploration is performed has a finite expectation and that losses in each can be bounded by a factor of M. Similarly, one can establish that losses over all exploitation phases are bounded by a factor of M. This is what drives the bound in (12).

The policy π^c bases its decisions on the information gathered solely in the preceding phase. A natural alternative is to use information from all periods with an order level of at least as much as the current one,[8] and base decisions on observations in this set. When data are not aggregated from stage to stage, the decision in stage $j+1$ is only based on the decision at the start of stage j and demands in stage j that are independent of the decision. When data are aggregated from stage to stage, the decision in stage $j+1$ is now based on the decision at the start of stage j and past observations, which may be correlated with the decision. For example, the fact that the decision at the start of stage j is (erroneously) low is an indication of past demand realizations that are also low. This lack of decoupling between decisions and observations introduces significant technical complications. Although we have restricted the proof to the case in which no aggregation is performed for technical considerations, we briefly explore numerically the impact of aggregation in §5.

4.3. Partially Censored Demand

We now turn to the intermediate setting in which, in addition to sales, the decision maker observes a lost sales indicator, i.e., is able to see whether any demand was not satisfied by the available inventory. In this setting, consider a slightly modified version of Algorithm 2, with the following two changes. (i) In the line where the "While" statement appears, replace $q_j^{(n)} = y_j^{(n)}$ with $q_j^{(n)} = y_j^{(n)} + 1$, and (ii) replace $\min\{D_t, x_t\}$ with $\min\{D_t, x_t\} + \mathbf{1}\{D_t > x_t\}$ throughout the algorithm. For completeness, a formal description is provided in Appendix D in the online companion (available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1983270). Let π^{pc} denote the policy with the same parameters as π^c with the modifications described above. In other words, π^{pc} is the version of the policy π^c that uses the lost sales indicator ($\mathbf{1}\{D > y_j^{(n)}\}$) to determine whether demand strictly exceeded inventory at various points, which was not possible in the censored case. Observe that at these steps in the policy, the lost sales indicator is added to the sales observation, and the sample quantile of the sum is computed. The sum is equal to the inventory level whenever demand is exactly equal to the inventory level (i.e., there were no lost sales), and the sum is equal to the inventory level plus one, whenever there are lost sales. Consequently, one can refine the estimate of the probability that demand is less than or equal to the inventory level (which allows to estimate whether the current level $\hat{x}$ is a good estimate of the optimal ordering quantity) without deviating to a strictly higher inventory level. In fact, observing sales plus the lost sales indicator at ordering level $\hat{x}$ is exactly equal to observing sales at ordering level $\hat{x}+1$. In a sense, the lost sales indicator allows the decision maker to perform "free experimentation" at $\hat{x}+1$, while still incurring costs for $\hat{x}$.

The performance of π^{pc} is provided in the next result.

THEOREM 5. *π^{pc} satisfies*

$$\sup_{F\in\mathcal{F}}\{\mathcal{C}^{\pi^{pc}}(F,T) - \mathcal{C}^*(F,T)\} \le \bar{K}_{pc}(\log M)^2 M, \quad (13)$$

where $\bar{K}_{pc}$ is a positive constant.

This result establishes that it is possible to recover bounded regret with the availability of the lost sales indicator. As opposed to the policy π^c, which uses sales observations only, and which performed frequent active exploration to refine its estimate of the optimal ordering quantity, exploration phases will now be avoided with high likelihood when the current estimate is correct. It is now possible to use the data from the exploitation phase to refine one's estimate of the optimal ordering quantity. The policy π^{pc} still possesses the ability to counter the milder censoring that is present through the contingent exploration phases.

[8] If one uses observations from periods with an order level smaller than the current one, any sales observations during such periods may be censored. In this case, the fact that such a sales observation is below the current estimate may not necessarily mean that the associated demand was below the current estimate. The same is not true for the other direction.

A key insight that arises from the result above is that for discrete demand distributions, the availability of the lost sales indicator, which could be interpreted as the minimal information one could collect beyond sales, enables one to mitigate the performance degradation stemming from censoring. Although the policies in the uncensored and partially censored settings both admit bounded minimax regret, it is worth noting that their performance will in general be different because one still operates with much less information in the partially censored setting. However, the two settings share a key characteristic, the absence of the need for systematic exploration at suboptimal levels to achieve the best possible order of magnitude of regret, that was key in the censored setting.

We now revisit the discussion on the availability of local information with respect to improvement directions and the contrast between the continuous and discrete settings for the case of censored demand. Whereas in the discrete setting with censored observations the only information available was with respect to downward corrections, the availability of the lost sales indicator enables the decision maker to now obtain information regarding upward corrections. Indeed, letting $G_t^+ = G_t + (h+b)[\mathbf{1}\{D_t \geq x_t\} - \mathbf{1}\{D_t > x_t\}]$, one has that $\mathbb{E}[G_t^+] = C(F, x_t+1) - C(F, x_t)$. As a result, in the partially censored setting, one recovers the local information for improvement directions that was available in the continuous setting. Hence, in both the discrete and the continuous settings, we conclude that when such local information is available, the minimax regret growth under censoring is identical to that of the observable demand case.

5. Numerical Experiments

In the previous sections, we characterized the minimax regret under different informational settings. This order of magnitude characterization provides useful insights with respect to the value of different types of information and the implications on the exploration–exploitation trade-off. To complement these theoretical results, we explore numerically three questions pertaining to the impact of data aggregation, demand granularity, and the marginal value of observing additional lost sales.

For this, we will focus on the following five policies: the policy π^u described in Algorithm 1, which uses demand observations; the policy π^c described in Algorithm 2, which uses sales observations only (we select the tuning parameters $\Delta_j^c = \lceil \gamma_1 a^{z^{j-1}} \rceil$ and $\Delta_j^e = \gamma_e z^{j-1}$, $j \geq 1$, with $a = 2$, $z = 1.25$, $\gamma_1 = 10$, $\gamma_e = 10$, and $m_j^n = 1/j^2$ for all j and n); the policy π^{pc} described in §4.3 that uses sales observations as well as a lost sales indicator, which uses the same tuning parameters as π^c; and the policies π_a^c and π_a^{pc}, which are identical to π^c and π^{pc}, except they aggregate past data as described below. The values of M and ε are not used by the policies we test. In the algorithms, we replace M by ∞. Although specific conditions on the multiplicative constants were required for the proofs of theoretical performance (see (9)), it is in general difficult to provide "optimal" values for those. In the present case, the numerical performance of the policies seemed to be robust across a broad set of parameters. For example, we chose to start the phases with 10 periods, but as long as reasonable values are selected to ensure that meaningful inference and learning can take place, without those overtaking exploitation periods, the overall qualitative behavior between the different information settings is preserved. Although we did not do so, in general, to select tuning constants, it might be desirable to test different sets of parameters on a large set of test instances and select the ones that lead to the most robust performance.

5.1. Data Aggregation

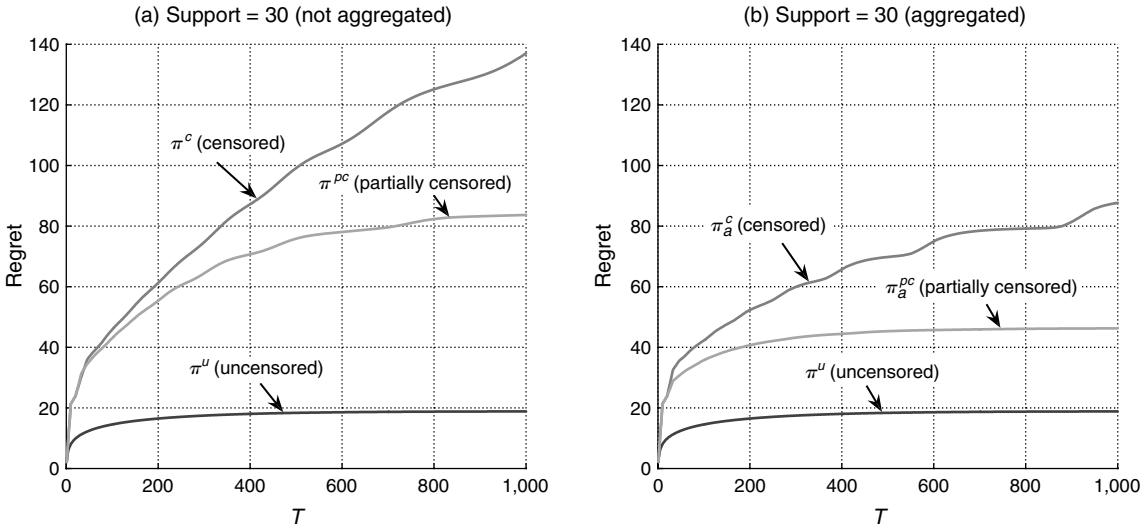
The policies π^c and π^{pc} base their decisions on the information gathered solely in the preceding phase. A natural alternative is to use information from all periods with an order level of at least as much as the current one, and base decisions on observations in this set. Intuitively, aggregation should improve performance, because each decision is based on a larger number of observations. Numerically, we observed that this is indeed the case. The uncensored policy also aggregates past observations, and studying the performance of the aggregated versions of our algorithms in a sense levels the playing field. Let $I_t(y) = \{i < t \mid x_i \geq y\}$ be the set of all periods before t with an ordering quantity of at least y. Consider policy π^c. Recall that the exploration level in cycle j on the nth exploration phase is $y_j^{(n)}$. Let t be the first period of this exploration phase. When computing the sample quantile $q_j^{(n)}$ after this exploration phase, π^c uses the observations in this final exploration phase. Consider the alternative policy π_a^c, where one uses all the observations in the set $I_t(y_j^{(n)})$ by setting

$$q_j^{(n)} = \inf\left\{k: \frac{1}{|I_t(y_j^{(n)})|} \cdot \sum_{i \in I_t(y_j^{(n)})} \mathbf{1}\{\min\{D_i, y_j^{(n)}\} \leq k\} \geq \beta\right\} \quad (14)$$

in the corresponding step of the algorithm. The resulting algorithm operates in a similar fashion. A similar modification is made when defining the policy π_a^{pc} in the partially censored setting.

Figure 1 depicts the regret $\mathcal{C}^\pi(F, T) - \mathcal{C}^*(F, T)$ as a function of the time horizon T for the policies

Figure 1 Performance



Notes. The figure depicts the regret $\mathcal{C}^{\pi}(F,T) - \mathcal{C}^{*}(F,T)$ as a function of time. The regret was estimated through simulation, using 10^4 replications.

above. For the censored and partially censored settings, Figure 1(a) contains the policies π^c and π^{pc}, which do not aggregate past observations, whereas part (b) contains their counterparts π_a^c and π_a^{pc}, which aggregate past observations. Both parts of the figure also include the policy π^u, which uses uncensored demand observations. The case under consideration is one where the underlying demand distribution is Binomial with 30 trials and a probability of success of 0.5. In addition, the underage and overage costs are fixed at $b = 2$ and $h = 1$. The optimal order level is $x^* = 16$ in this case, and all the algorithms start with an initial ordering level of 20.

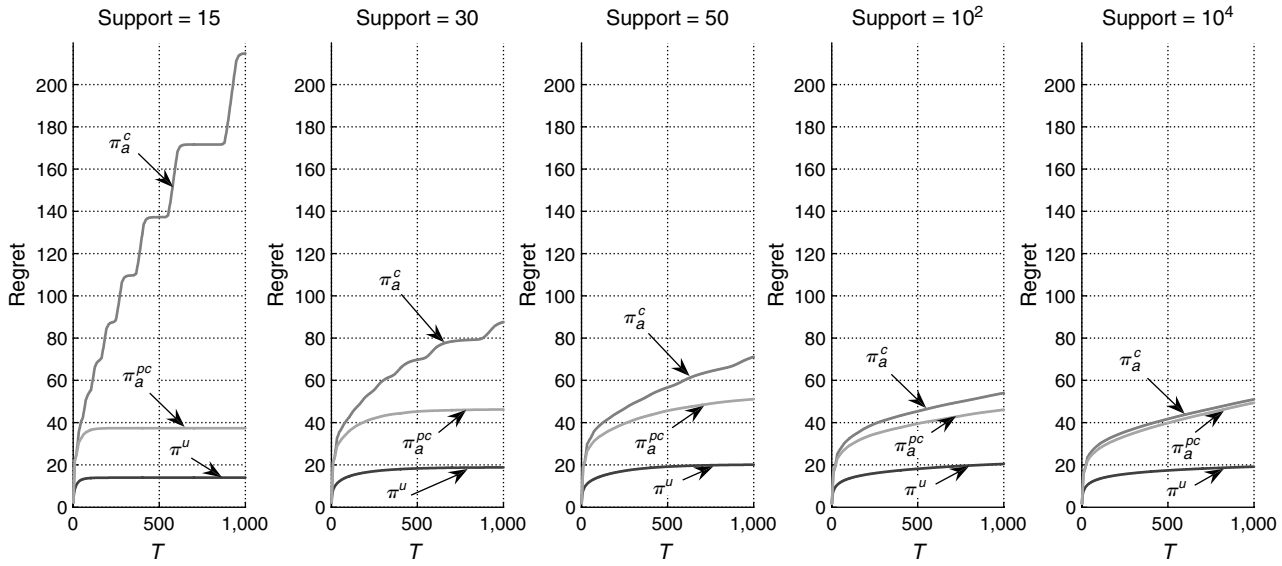
In Figure 1(a), we observe that the policy π^u that has access to the greatest level of information (uncensored demand) has the lowest regret, and the policy π^c, which has access to the least information (sales), has the highest regret, as expected. The policy π^{pc}, which has access to partially censored demand, achieves a performance between those two. The regret associated with policy π^c increases with time, at a logarithmic rate. This is due to the "jump" in regret during the frequent experimentation phases, which have length $O(\log T)$ and are performed at each stage. On the other hand, π^{pc} does not incur such systematic losses, because once the policy has already zoomed in on the right region, it only conducts experimentation with small probability. Initially, the performance of π^{pc} is closer to that of π^c than π^u; as time progresses, the regret under π^c grows without bound, whereas it eventually stabilizes under both π^u and π^{pc}. Using aggregation, policies that operate under censoring do better, but their qualitative comparison does not change.

The example focused on a problem with $b = 2$. In problems with larger underage costs, a higher quantile of the distribution needs to be estimated, which is in a sense more difficult, because there are fewer observations above higher quantiles. We observed that although learning happens more slowly in such settings, the qualitative observations are preserved, and an example with $b = 15$ is provided in Appendix E in the online companion.

5.2. Role of Granularity

As discussed in the introduction, one would expect the value of the lost sales indicator to decrease, as demand becomes less and less granular. We next investigate numerically the transition from a discrete to a continuous setting. In Figure 2, the binomial distribution that was used in Figure 1 is—in a sense—scaled to have less granularity. Consider two demand models: one where the discrete points correspond to, say, "kilograms" of the good, and a scaled version, where the unit of measurement is "grams." The scaled demand distribution is constructed in a way such that the cumulative distribution function (cdf) of the original demand distribution evaluated at x kilograms has the same value as the cdf of the scaled demand distribution evaluated at $10^3 \cdot x$ grams, for any integer x. However, the scaled demand distribution has positive mass also between multiples of 10^3. The cdf for these values are chosen as a linear interpolation of the neighboring multiples of 10^3. For example, the cdf of the scaled demand distribution at 1,500 grams is the average of the cdf of the original demand distribution at 1 kilogram and 2 kilograms. The backorder and holding cost rates are also scaled accordingly. Recall that the demand distribution used in Figure 1 is binomial with a support of 30. The demand distributions underlying Figure 2 are scaled versions of

Figure 2 Granularity



Notes. The figure depicts the regret as a function of time, for different levels of granularity for policies π_a^c, π_a^{pc}, and π^u. The regret was estimated through simulation, using 10^4 replications.

this binomial distribution, with support values of 15, 30, 50, 100, and 10^4 respectively.

As granularity decreases, the value of the lost sales indicator decreases, as observed by the diminishing gap between the regret curves corresponding to the censored and partially censored settings. In addition, when the support is large, all the three policies (π_a^c, π_a^{pc} and π^u)[9] now appear to have a regret that grows logarithmically with time, which is what one would expect if demand was continuous (based on the analysis of §3).

5.3. Impact of Information Beyond the Lost Sales Indicator

We have so far focused exclusively on three information levels: uncensored, censored, and partially censored. There may be settings in which the firm has more information than the partially censored setting, but less information than the observable demand setting. To better understand the value of such additional information, and just for illustration purposes, we implemented policies that can observe the first i lost sales, for increasing values of i. The case $i = 0$ corresponds to censored demand, $i = 1$ corresponds to partially censored demand, and $i = \infty$ corresponds to observable demand. We consider policies with the same structure as π_a^{pc}, except that as i increases, the size of the exploration and exploitation phases decreases in the following fashion: $\gamma_e = \gamma_1 = \lceil 10/(i-1) \rceil$ and $a = \max\{2/(i-1), 1\}$. This is a heuristic meant to capture the fact that as demand is less censored, one requires less exploration.

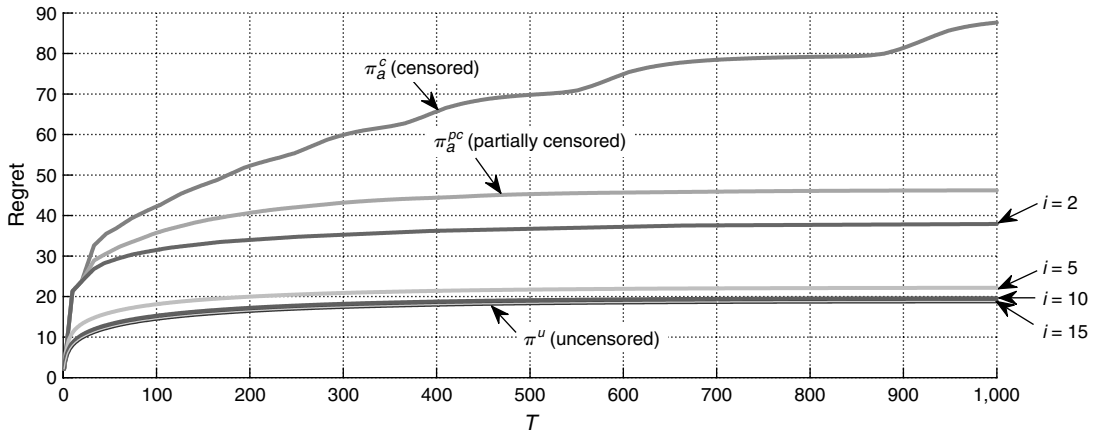
We illustrate this in Figure 3. The lines that are between those of the partially censored and uncensored cases correspond to policies that have progressively higher levels of information, in particular, the first i lost sales are observed, for increasing values of i. The levels of intermediate settings shown are $i = 2, 5, 10$, and 15. As expected, each level of additional information improves performance. The policies span the gap between the partially censored setting and the uncensored setting, and converge to the uncensored setting for i large enough. Although it is the first lost sales indicator that removes the need for fine-tuning, additional information enables one to converge faster to the neighborhood of the optimal quantity. The example suggests that observing even a small percentage of lost sales can enable firms to mitigate the effects of censoring to a substantial degree.

6. Conclusions

The present paper assumed that inventory is perishable, and unused inventory is lost at the end of every period. If inventory is not perishable, in addition to being coupled through information, periods also become coupled through inventory carryover. Under an assumption on salvage value (see Veinott 1965), the optimal policy with the knowledge of the demand distribution would be a stationary order-up-to policy, and the framework developed here is likely to be applicable to such cases, under additional assumptions on the form of demand. The key would now be to carefully account for both the impact of information and inventory position. To control for the

[9] We illustrate the aggregated version of the policies here, but all qualitative comparisons remain the same when data are not aggregated. A similar figure with nonaggregated policies is provided in Appendix E in the online companion.

Figure 3 Partial Censoring



Notes. The figure depicts the regret $\mathcal{C}^\pi(F,T) - \mathcal{C}^*(F,T)$ as a function of time for the three policies π^u, π_a^c, and π_a^{pc} and policies that have access to the first i lost sales, for $i = 2, 5, 10$, and 15. The regret was estimated through simulation, using 10^4 replications.

impact of overordering, demand would now need to be bounded away from zero, at least in expectation. For example, with such assumptions, for the case of continuous distributions and a given stochastic gradient algorithm, Huh and Rusmevichientong (2009) proposed an approach to account for losses due to potentially having a wrong estimate of the optimal order-up-to level, as well as losses associated with the inability to bring the inventory level down to the current estimate.

The characterization of the minimax regret obtained under various informational settings and demand types in the present study enhances the understanding of the implications of demand censoring. Although the present work focused on the repeated newsvendor problem, the approach and tools developed herein will be useful in other settings where one is interested in comparing different informational settings with respect to the cost stemming from censored information, as well as the fundamental changes in the exploration–exploitation trade-off that such censoring drives. Information censoring is prevalent in a variety of applications beyond inventory management, ranging from capacity planning in airlines with multiple fare classes to bidding in auctions with uncertain valuations.

From a more practical perspective, demand censoring is an important phenomenon faced by many businesses. Firms may choose to invest in information technology in an attempt to collect data on lost sales, and a key input for deciding to do so is associated with the cost of such censoring. Although the present paper has focused on a theoretical characterization of demand censoring through the lens of minimax regret, we believe that working to refine our understanding of the cost of censoring, and developing associated practical frameworks, is an important direction for further research.

Acknowledgments
The authors are grateful to Gérard Cachon, the associate editor, and three anonymous reviewers for their suggestions that helped improve the paper.

Appendix A. Selected Proofs
In the present section, we provide proofs of selected results. All other proofs are provided in the online companion, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1983270.

Preliminaries
Formally, the history up to (and including) time t available to the decision maker in each of the three informational settings is given as follows:

$$\mathcal{H}_t^u = \{(D_s, x_s)\colon 1 \le s \le t\} \quad \text{for } t \ge 1 \text{ and } \mathcal{H}_0^u = \varnothing,$$
$$\mathcal{H}_t^c = \{(\min\{D_s, x_s\}, x_s)\colon 1 \le s \le t\} \quad \text{for } t \ge 1 \text{ and } \mathcal{H}_0^c = \varnothing,$$
$$\mathcal{H}_t^{pc} = \{(\min\{D_s, x_s\}, x_s, \mathbf{1}\{D_s > x_s\})\colon 1 \le s \le t\} \quad \text{for } t \ge 1 \text{ and } \mathcal{H}_0^{pc} = \varnothing.$$

For each of the settings above, $a = u, c, pc$, a policy is nonanticipating if the quantity ordered in the tth period is determined by $\mathcal{H}_{t-1}^a$ (i.e., is $\mathcal{H}_{t-1}^a$-measurable). For any $\pi \in \mathcal{P}^a$, we denote the quantity ordered in the tth period, x_t, by $\psi_t(\mathcal{H}_{t-1}^a)$.

For any $a = u$, c or pc, the expected cost associated with an admissible policy π over T periods when the demand distribution is F is given by

$$\begin{aligned}\mathcal{C}^\pi(F,T) &= \mathbb{E}_F^\pi\Bigg[\sum_{t=1}^T [h(\psi_t(\mathcal{H}_{t-1}^a) - D_t)^+ + b(D_t - \psi_t(\mathcal{H}_{t-1}^a))^+]\Bigg] \\ &= \mathbb{E}_F^\pi\Bigg[\sum_{t=1}^T \mathbb{E}_F^\pi[h(\psi_t(\mathcal{H}_{t-1}^a) - D_t)^+ + b(D_t - \psi_t(\mathcal{H}_{t-1}^a))^+ \mid \psi_t(\mathcal{H}_{t-1}^a)]\Bigg] \\ &= \mathbb{E}_F^\pi\Bigg[\sum_{t=1}^T C(F, \psi_t(\mathcal{H}_{t-1}^a))\Bigg],\end{aligned}$$

where the last equality follows from the fact that D_t is independent of $\psi_t(\mathcal{H}^a_{t-1})$. The equality above will be used extensively in all the proofs below. We will also use the fact that, for discrete distributions, for any integer $x \geq 0$,

$$C(F,x) - C(F,x^*) = (h+b) \sum_{i=\min\{x^*,x\}}^{\max\{x^*,x\}-1} |\beta - F(i)|. \tag{A1}$$

Proof of Lemma 1. The proof is based on a large deviation argument. Let $W_{i,j} = \mathbf{1}\{Z_i \leq j\}$ for $i \geq 1$ and $j = 0, \ldots, y$, and note that, for $j \in \{0, \ldots, y-1\}$, $W_{i,j}$ is a Bernoulli random variable with expectation $F(j)$.

Suppose first that $j \leq x_F^* - 1$ and $F(j) \in (0,1)$. Then for $t \geq 2$,

$$\mathbb{P}\{\tilde{x}_t \leq j\} = \mathbb{P}\left\{\sum_{i=1}^{t-1} \mathbf{1}\{Z_i \leq j\} \geq (t-1)\beta\right\} = \mathbb{P}\left\{\sum_{i=1}^{t-1} W_{i,j} \geq (t-1)\beta\right\},$$

and a Chernoff bound yields that for any nonnegative parameter θ,

$$\begin{aligned}\mathbb{P}\{\tilde{x}_t \leq j\} &\leq \mathbb{E}\left[\exp\left\{\theta \sum_{i=1}^{t-1} W_{i,j}\right\}\right] \exp\{-\theta(t-1)\beta\} \\ &= ((\exp\{\theta\} - 1)F(j) + 1)^{t-1} \exp\{-\theta(t-1)\beta\}.\end{aligned}$$

Selecting the value of θ that minimizes the right-hand side above, $\theta^* = \log[\beta(1-\beta)^{-1}(1-F(j))(F(j))^{-1}]$, which is positive because $j \leq x_F^* - 1$ implies that $F(j) < \beta$, one obtains

$$\mathbb{P}\left\{\sum_{i=1}^{t-1} \mathbf{1}\{Z_i \leq j\} \geq (t-1)\beta\right\} \leq \alpha_j^{t-1},$$

with the value of α_j given by

$$\alpha_j = \left(\frac{1-F(j)}{1-\beta}\right)^{1-\beta} \left(\frac{F(j)}{\beta}\right)^{\beta}. \tag{A2}$$

If $F(j) = 0$, then we have $\mathbb{P}\{\tilde{x}_t \leq j\} = 0 = \alpha_j$.

Now, similarly, if $x_F^* + 1 \leq j \leq y$ and $F(j-1) \in (0,1)$, we have that, for $t \geq 2$,

$$\begin{aligned}\mathbb{P}\{\tilde{x}_t \geq j\} &= \mathbb{P}\left\{\sum_{i=1}^{t-1} \mathbf{1}\{Z_i \leq j-1\} < (t-1)\beta\right\} \\ &= \mathbb{P}\left\{\sum_{i=1}^{t-1} W_{i,j-1} < (t-1)\beta\right\},\end{aligned}$$

and through a Chernoff bound that, for any nonnegative parameter θ,

$$\mathbb{P}\{\tilde{x}_t \geq j\} \leq ((\exp\{-\theta\} - 1)F(j-1) + 1)^{t-1} \exp\{\theta(t-1)\beta\}.$$

Selecting the value of θ that minimizes the right-hand side above, $\theta^* = -\log[\beta(1-\beta)^{-1}(1-F(j-1))(F(j-1))^{-1}]$, which is nonnegative because $j \geq x_F^* + 1$ and hence $F(j-1) > \beta$, one obtains

$$\mathbb{P}\left\{\sum_{i=1}^{t-1} \mathbf{1}\{Z_i \leq j-1\} \leq (t-1)\beta\right\} \leq \alpha_{j-1}^{t-1},$$

with α_j given in (A2).

If $F(j-1) = 1$, then we have $\mathbb{P}\{\tilde{x}_t \geq j\} = 0 = \alpha_{j-1}$. If $j \geq y+1$, then the result also follows trivially.

We now turn to establish the bound for α_j. For all $x \in (0,1)$, let

$$g(x) = \left(\frac{1-x}{1-\beta}\right)^{1-\beta} \left(\frac{x}{\beta}\right)^{\beta}.$$

Clearly, $g(\cdot)$ is infinitely differentiable on $(0,1)$ with second derivative given by

$$g^{(2)}(x) = -\frac{(1-x)^{-\beta-1}x^{\beta}}{(1-\beta)^{-\beta}\beta^{\beta-1}} - 2\frac{(1-x)^{\beta}x^{\beta-1}}{(1-\beta)^{\beta}\beta^{\beta-1}} - \frac{(1-x)^{1-\beta}x^{\beta-2}}{(1-\beta)^{-\beta}\beta^{\beta-1}},$$

which is negative for all $x \in (0,1)$; hence $g(\cdot)$ is strictly concave on $(0,1)$. In addition, one can establish that $g(\cdot)$ is maximized at $x = \beta$ with value 1. The fact that $g^{(2)}(\cdot)$ tends to $-\infty$ as $x \to 0$ and as $x \to 1$, coupled with the fact that $g^{(2)}(\cdot)$ is continuous, implies that $g^{(2)}(\cdot)$ admits a maximizer in $(0,1)$. Let x_g be this maximizer, $m_g = g^{(2)}(x_g)$, and let

$$m_\beta = (1/2)|m_g|. \tag{A3}$$

We have $g^{(2)}(x) \leq m_g < 0$ for all $x \in (0,1)$. We deduce, through a Taylor expansion that for all $x \in (0,1)$, there exists $\tilde{x} \in (0,1)$ such that $g(x) - g(\beta) = (1/2)g^{(2)}(\tilde{x})(x-\beta)^2$, and hence $1 - g(x) \geq (1/2)|m_g|(x-\beta)^2$. By continuity of $g(\cdot)$ on $[0,1]$, the latter inequality is also valid at the boundaries 0 and 1. Noting that for all $j \geq 0$, $\alpha_j = g(F(j))$, we hence have that

$$\alpha_j \leq 1 - m_\beta(F(j) - \beta)^2. \quad \square$$

Proof of Theorem 2. Let F be an arbitrary distribution in $\mathcal{F}$. Recall that x_t denotes the ordering quantity of policy π^u at time t. Consider the regret over T periods:

$$\begin{aligned}&\mathcal{C}^{\pi^u}(F,T) - \mathcal{C}^*(F,T) \\ &= \mathbb{E}\left[\sum_{t=1}^{T} [C(F,x_t) - C(F,x^*)]\right] \\ &= \sum_{t=1}^{T} \sum_{j \in \{0,\ldots,M\} \setminus x^*} \mathbb{E}[C(F,x_t) - C(F,x^*) \mid x_t = j] \mathbb{P}\{x_t = j\} \\ &\overset{(a)}{=} (b+h) \sum_{t=1}^{T} \left[\sum_{j=0}^{x^*-1} \left[\sum_{i=j}^{x^*-1} (\beta - F(i))\right] \mathbb{P}\{x_t = j\} \right. \\ &\qquad \left. + \sum_{j=x^*+1}^{M} \left[\sum_{i=x^*}^{j-1} (F(i) - \beta)\right] \mathbb{P}\{x_t = j\}\right] \\ &= (b+h) \sum_{t=1}^{T} \left[\sum_{i=0}^{x^*-1} (\beta - F(i)) \sum_{j=0}^{i} \mathbb{P}\{x_t = j\} \right. \\ &\qquad \left. + \sum_{i=x^*}^{M} (F(i) - \beta) \sum_{j=i+1}^{M} \mathbb{P}\{x_t = j\}\right] \\ &= (b+h) \sum_{t=1}^{T} \left[\sum_{i=0}^{x^*-1} (\beta - F(i)) \mathbb{P}\{x_t \leq i\} \right. \\ &\qquad \left. + \sum_{i=x^*}^{M-1} (F(i) - \beta) \mathbb{P}\{x_t \geq i+1\}\right],\end{aligned}$$

where (a) follows from (A1). Now, applying Lemma 1(a) with $y = M+1$ (in which case $\tilde{x}_t = x_t$ for all $t \geq 2$) and

bounding the terms $\mathbb{P}\{x_1 \le i\}$ and $\mathbb{P}\{x_1 \ge i+1\}$ by 1 yields that

$$\begin{aligned}
&\mathcal{C}^{\pi^u}(F,T) - \mathcal{C}^*(F,T)\\
&\quad \le (b+h)\sum_{t=1}^{T}\left[\sum_{i=0}^{x^*-1}(\beta - F(i))\alpha_i^{t-1} + \sum_{i=x^*}^{M-1}(F(i)-\beta)\alpha_i^{t-1}\right]\\
&\quad = (b+h)\sum_{i\in\{0,\dots,M-1\}}|\beta - F(i)|\sum_{t=1}^{T}\alpha_i^{t-1}\\
&\quad \overset{(a)}{\le} (b+h)\sum_{i\in\{0,\dots,M-1\}}\min\left\{|\beta - F(i)|T, |\beta - F(i)|\frac{1}{1-\alpha_i}\right\},
\end{aligned}$$

where (a) follows from upper bounding $\sum_{t=1}^{T}\alpha_i^{t-1}$ by either T or $\sum_{t=1}^{\infty}\alpha_i^t$. An application of Lemma 1 implies that $\alpha_i \le 1 - m_\beta(F(i)-\beta)^2$, and we deduce that $|\beta - F(i)|(1-\alpha_i)^{-1} \le (m_\beta|F(i)-\beta|)^{-1}$. Returning to the regret, we have

$$\begin{aligned}
&\mathcal{C}^{\pi^u}(F,T) - \mathcal{C}^*(F,T)\\
&\quad \le (b+h)\sum_{i=0}^{M-1}\min\left\{|\beta - F(i)|T, \frac{1}{m_\beta|F(i)-\beta|}\right\}\\
&\quad \le \frac{(b+h)M}{m_\beta\varepsilon}. \qquad \text{(A4)}
\end{aligned}$$

This completes the proof. □

Proof of Theorem 3. We analyze the worst-case performance of any policy when nature can only select between two demand functions, which provides a lower bound on the minimax regret, which scales logarithmically with T. This part uses ideas that have appeared in the context of multiarmed bandit problems (see, e.g., Lai and Robbins 1985).

Dependence on T. We will develop a lower bound on performance that scales with the time horizon T. To that end, we will again analyze the worst-case performance of any policy $\pi \in \mathcal{P}^c$ when nature is restricted to select between the following two judiciously selected distribution functions in $\mathcal{F}$:

$$F_a(k) = \begin{cases}\beta+\varepsilon & \text{if } k=0,\\ 1 & \text{if } k\ge 1,\end{cases} \qquad F_b(k) = \begin{cases}\beta-\varepsilon & \text{if } k=0,\\ 1 & \text{if } k\ge 1,\end{cases}$$

which admit positive mass at both 0 and 1 because we assume that $\varepsilon \in (0, \min\{\beta, 1-\beta\}/2)$. For $\ell = a, b$, and for any event $\mathcal{A}$ and random variable X, we let $\mathbb{P}_\ell(\mathcal{A})$ and $\mathbb{E}_\ell[X]$ denote the probability of $\mathcal{A}$ and the expectation of X (when it is well defined), respectively, when $F = F_\ell$.

For the rest of the analysis in Step 2, we fix an arbitrary policy π in $\mathcal{P}^c$, and we will establish by contradiction that the worst-case performance regret associated with this policy is necessarily bounded below by $K\varepsilon^{-1}\log T$ for some appropriate positive constant K.

Fix an arbitrary $\xi \in (0,1)$, and define the following constants that depend only on ξ, b, and h:

$$K_1 = \xi 4^{-1}[2\beta^{-1} + (1-\beta)^{-1} + \beta^{-2}]^{-1}, \qquad \text{(A5)}$$

$$K_2 = (8\beta^{-1} + 12(1-\beta)^{-1})^2, \qquad \text{(A6)}$$

$$K_3 = 2^{-1}K_1\exp\{-4\xi K_2^{-1}\}(h+b), \qquad \text{(A7)}$$

$$\begin{aligned}
K_4 = \max\{&2K_1, 2\exp\{1\}\\
&\cdot(1 - \min\{1/2, 2\exp\{-4\xi K_2^{-1}\}\})^{-1}K_3(h+b)^{-1}\}. \qquad \text{(A8)}
\end{aligned}$$

Note that $T^{1-\xi}(\log T)^{-1}$ tends to infinity as T tends to infinity, and hence there exists T_0 such that

$$T^{1-\xi}(\log T)^{-1} \ge K_4\varepsilon^{-2}, \quad \text{for all } T \ge T_0. \qquad \text{(A9)}$$

For the rest of the proof, we assume that $T \ge T_0$. Suppose for a moment that

$$\sup_{\ell=a,b}\{\mathcal{C}^\pi(F_\ell, T) - \mathcal{C}^*(F_\ell, T)\} < K_3\varepsilon^{-1}\log T. \qquad \text{(A10)}$$

Let τ denote the number of periods until T when the policy orders a positive quantity, i.e., $\tau = \sum_{t=1}^{T}\mathbf{1}\{\psi_t(\mathcal{H}_{t-1}^c) \ne 0\}$; τ is also the number of periods when the policy orders a number of units that differs from the optimal number of units when the demand distribution is F_a.

The analysis in the remainder of this step will proceed as follows. We first establish that given the performance guarantee (A10), the probability that $\tau \le K_1\varepsilon^{-2}\log T$ is appropriately "small" (for K_1 defined in (A5)) when demand is generated according to F_b. This will imply through a change of measure argument that the probability that $\tau \le K_1\varepsilon^{-2}\log T$ is also small when demand is generated according to F_a. The latter, in turn, will yield a contradiction with (A10).

Given (A9), $K_1\varepsilon^{-2}\log T \le T/2$, and one can bound $\mathbb{P}_b^\pi\{\tau \le K_1\varepsilon^{-2}\log T\}$ as follows:

$$\begin{aligned}
\mathbb{P}_b^\pi\{\tau \le K_1\varepsilon^{-2}\log T\} &= \mathbb{P}_b^\pi\{T - \tau \ge T - K_1\varepsilon^{-2}\log T\}\\
&\overset{(a)}{\le} (T - K_1\varepsilon^{-2}\log T)^{-1}\mathbb{E}_b^\pi[T-\tau]\\
&\overset{(b)}{\le} (T - K_1\varepsilon^{-2}\log T)^{-1}\\
&\quad \cdot K_3\varepsilon^{-2}(b+h)^{-1}\log T\\
&\le 2T^{-1}K_3\varepsilon^{-2}(b+h)^{-1}\log T, \qquad \text{(A11)}
\end{aligned}$$

where (a) follows from Markov's inequality, and (b) follows from (A10) in conjunction with the following sequence of inequalities that links $\mathbb{E}_b^\pi[T-\tau]$ and $\mathcal{C}^\pi(F_b,T) - \mathcal{C}^*(F_b,T)$:

$$\begin{aligned}
\mathcal{C}^\pi(F_b,T) - \mathcal{C}^*(F_b,T) &= \mathbb{E}_b^\pi\left[\sum_{t=1}^{T}[C(F_b, \psi_t(\mathcal{H}_{t-1}^c)) - C(F_b,1)]\right]\\
&\overset{(a)}{\ge} \sum_{t=1}^{T}\mathbb{E}_b^\pi[[C(F_b, \psi_t(\mathcal{H}_{t-1}^c)) - C(F_b,1)]\\
&\qquad \cdot \mathbf{1}\{\psi_t(\mathcal{H}_{t-1}^c) = 0\}]\\
&= (C(F_b,0) - C(F_b,1))\,\mathbb{E}_b^\pi[T-\tau]\\
&\overset{(b)}{=} (b+h)\varepsilon\,\mathbb{E}_b^\pi[T-\tau].
\end{aligned}$$

In the latter, (a) is a consequence of the optimality of ordering one unit when the distribution is F_b, and (b) follows from (A1).

Based on (A11), we develop an upper bound on $\mathbb{P}_a^\pi\{\tau \le K_1\varepsilon^{-2}\log T\}$, whose proof appears in Appendix C in the online companion.

Lemma 2. *The following inequality holds*:

$$\mathbb{P}_a^\pi\{\tau > K_1\varepsilon^{-2}\log T\} \ge \exp\{-4\xi K_2^{-1}\}. \qquad \text{(A12)}$$

We conclude Step 2 with the analysis of the regret associated with π when $F = F_a$:

$$\begin{aligned}
&\mathcal{C}^\pi(F_a, T) - \mathcal{C}^*(F_a, T)\\
&= \mathbb{E}_a^\pi\left[\sum_{t=1}^{T}[C(F_a, \psi_t(\mathcal{H}_{t-1}^c)) - C(F_a, 0)]\right]\\
&\geq \mathbb{E}_a^\pi\left[\sum_{t=1}^{T}[C(F_a, \psi_t(\mathcal{H}_{t-1}^c)) - C(F_a, 0)]\,\middle|\,\tau > K_1\varepsilon^{-2}\log T\right]\\
&\qquad \cdot \mathbb{P}_a^\pi\{\tau > K_1\varepsilon^{-2}\log T\}\\
&\overset{(a)}{\geq} \exp\{-4\xi K_2^{-1}\}\,\mathbb{E}_a^\pi\\
&\quad \cdot\left[\sum_{t=1}^{T}[C(F_a, \psi_t(\mathcal{H}_{t-1}^c)) - C(F_a, 0)]\,\middle|\,\tau > K_1\varepsilon^{-1}\log T\right]\\
&\overset{(b)}{\geq} \exp\{-4\xi K_2^{-1}\}(C(F_a, 1) - C(F_a, 0))\\
&\quad \cdot \mathbb{E}_a^\pi\left[\sum_{t=1}^{T}\mathbf{1}\{\psi_t(\mathcal{H}_{t-1}^c) \neq 0\}\,\middle|\,\tau > K_1\varepsilon^{-2}\log T\right]\\
&> \exp\{-4\xi K_2^{-1}\}(C(F_a, 1) - C(F_a, 0))K_1\varepsilon^{-2}\log T,
\end{aligned}$$

where (a) follows from (A12) and (b) follows from the fact that $C(F_a, x) - C(F_a, 0) \geq C(F_a, 1) - C(F_a, 0)$ whenever $x \neq 0$. Equation (A1) implies that $C(F_a, 1) - C(F_a, 0) \geq (h+b)\varepsilon$, and hence

$$\mathcal{C}^\pi(F_a, T) - \mathcal{C}^*(F_a, T) \geq \exp\{-4\xi K_2^{-1}\}K_1(h+b)\varepsilon^{-1}\log T. \quad \text{(A13)}$$

We have therefore established that if (A10) holds, then (A13) necessarily holds. This is a contradiction because $\exp\{-4\xi K_2^{-1}\}(h+b)K_1 = 2K_3 > K_3$ (see (A5) and (A7)). Hence (A10) cannot hold, and it must be the case that

$$\sup_{\ell=a,\,b}\{\mathcal{C}^\pi(F_\ell, T) - \mathcal{C}^*(F_\ell, T)\} \geq K_3\varepsilon^{-1}\log T. \quad \text{(A14)}$$

General lower bound. We note that the worst-case regret may always be of order M given the limited initial information (see Lemma 4 in Appendix C in the online companion). We deduce that for some $K > 0$, $\inf_{\pi\in\mathcal{P}^c}\sup_{F\in\mathcal{F}}\{\mathcal{C}^\pi(F, T) - \mathcal{C}^*(F, T)\} \geq K\max\{M, \log T\}$, which in turn implies that

$$\inf_{\pi\in\mathcal{P}^c}\sup_{F\in\mathcal{F}}\{\mathcal{C}^\pi(F, T) - \mathcal{C}^*(F, T)\} \geq \tfrac{1}{2}K[M + \log T].$$

This concludes the proof. □

Proof of Theorem 4. Let F be an arbitrary distribution in $\mathcal{F}$, and let x^* denote the optimal newsvendor solution when one knows F ($x^* = \inf\{k \geq 0\colon F(k) \geq \beta\}$); the latter is necessarily bounded by M because $F \in \mathcal{F}$. The proof proceeds in two main steps. First, we establish that, from stage to stage, one will refine the estimate of x^* in a manner made precise in Lemma 3. We then conduct a performance analysis, evaluating separately the expected regret over the exploitation and exploration phases. For simplicity, we fix $z = 2$ throughout the proof.

Step 1: *Refining the estimate of* x^*. The policy π^c consists of consecutive stages of increasing length. Each stage starts with an exploitation phase, followed by possibly one or more contingent additional exploration phases. The ordering level in the exploitation phase of stage $j \geq 2$ is $y_j^{(0)} = q_{j-1}^{(\tilde{n}_j)}$, where $\tilde{n}_j$ is the number of contingent exploration phases in stage j, i.e., $\tilde{n}_j = \inf\{n \geq 0\colon \{q_{j+1}^{(n)} < y_j^{(n)}\} \cup \{y_j^{(n)} = M\}\}$. Let $z_j^n(k)$ be the ordering level in the nth exploration phase (if necessary) of cycle j, given that the cycle starts with exploitation level $y_j^{(0)} = k$:

$$\begin{aligned}
z_j^1(k) &= \min\{k + \max\{\lceil m_j^1 \cdot k\rceil, 1\}, M\},\\
z_j^n(k) &= \min\{z_j^{n-1} + \max\{\lceil m_j^n \cdot z_j^{n-1}(k)\rceil, 1\}, M\}, \quad \forall\, n \geq 2.
\end{aligned}$$

For all $j \geq 1$, let $n_j(k, x) = \min\{\ell\colon z_j^\ell(k) \geq x\}$. The following lemma (proved in Appendix C in the online companion) bounds the probability of using an exploitation level other than x^* for each cycle j.

Lemma 3. *For* $j \geq 2$, *the exploitation level at stage* j, $y_j^{(0)}$, *satisfies*

$$\mathbb{P}\{y_j^{(0)} \leq i\} \leq a_j\alpha_i^{\Delta_{j-1}^e} \quad \textit{for all } i \leq x^* - 1, \quad \text{(A15)}$$

$$\mathbb{P}\{y_j^{(0)} \geq i\} \leq a_j\alpha_{i-1}^{\Delta_{j-1}^e} \quad \textit{for all } i \geq x^* + 1, \quad \text{(A16)}$$

where $a_2 = 3 + n_1(1, M)$ *and, for* $j \geq 2$,

$$a_{j+1} = \frac{1}{1 - \alpha_{x^*}^{\Delta_j^e}} + 1 + n_{j+1}(1, M)a_j\alpha_{x^*-1}^{\Delta_{j-1}^e}. \quad \text{(A17)}$$

Step 2: *Performance analysis*. We now turn to analyze the regret of the proposed policy. Let t_j be the first time period in cycle j, and let $t_{j,n}$ be the last period of (contingent) exploration phase n in cycle j (with $t_{j,0} := t_j + \Delta_j^c - 1$). Consider any time T and $\kappa = \inf\{j \geq 1\colon t_{j+1} \geq T\}$. We first note that

$$\begin{aligned}
\mathcal{C}^\pi(F, T) - \mathcal{C}^*(F, T) &\leq \mathcal{C}^\pi(F, t_{\kappa+1}) - \mathcal{C}^*(F, t_{\kappa+1})\\
&= \mathbb{E}\left[\sum_{j=1}^{\kappa}[\mathcal{D}_j^c + \mathcal{D}_j^e]\right], \quad \text{where for } j \geq 1,
\end{aligned}$$

$$\mathcal{D}_j^c = \sum_{t=t_j}^{t_j+\Delta_j^c-1}[C(F, x_t) - C(F, x^*)]$$

[regret over the exploitation phase of cycle j],

$$\mathcal{D}_j^e = \sum_{t=t_{j,0}+1}^{t_{j,\tilde{n}_j}}[C(F, x_t) - C(F, x^*)]$$

[regret over the $\tilde{n}_j \geq 0$ exploitation phases of cycle j].

Let $p_j(k) = \mathbb{P}\{y_j^{(0)} = k\}$ for all $j \geq 2$ and $k \geq 1$.

Analysis of $\mathcal{D}_j^c$. For $j = 1$, one obtains trivially $\mathbb{E}[\mathcal{D}_j^c] \leq (h + b)M2\gamma_1$. For $j \geq 2$, the expected regret during the exploitation phase of can be bounded as follows:

$$\begin{aligned}
\mathbb{E}[\mathcal{D}_j^c] &= \Delta_j^c\sum_{k=0}^{M}\mathbb{E}[C(F, y_j^{(0)}) - C(F, x^*) \mid y_j^{(0)} = k]p_j(k)\\
&= (h+b)\Delta_j^c\sum_{k=0}^{M}\sum_{i=\min\{x^*,k\}}^{\max\{x^*,k\}-1}|\beta - F(i)|p_j(k)\\
&= (h+b)\Delta_j^c\left[\sum_{i=0}^{x^*-1}|\beta - F(i)|\,\mathbb{P}\{y_j^{(0)} \leq i\}\right.\\
&\qquad\left. + \sum_{i=x^*}^{M-1}|\beta - F(i)|\,\mathbb{P}\{y_j^{(0)} \geq i+1\}\right]\\
&\overset{(a)}{\leq} (h+b)\Delta_j^c a_j\sum_{i=0}^{M-1}|\beta - F(i)|\alpha_i^{\Delta_{j-1}^e},
\end{aligned}$$

where (a) follows from Lemma 3. Noting that $\alpha_i^{\gamma_e} \le 1/4$, we have

$$\mathbb{E}[\mathscr{D}_j^c] \le (h+b)\gamma_1 a^{2^{j-1}} a_j \sum_{i=0}^{M-1} |\beta - F(i)| \alpha_i^{\Delta_{j-1}^e}$$
$$\le (h+b)\gamma_1 a_j (a/4)^{4^{j-2}} M. \tag{A18}$$

Analysis of $\mathscr{D}_j^e$. Suppose first that $j = 1$. Then, noting that there are at most $n_1(1, M)$ contingent explorations, $\mathbb{E}[\mathscr{D}_1^e] \le n_1(1, M)(h+b)M\gamma_e \le (h+b)M[(\log M)/\log(1+m_1)]\gamma_e$, where the last inequality follows from the fact that $n_1(1, M) \le (\log M)/\log(1+m_j)$. Suppose now that $j \ge 2$. We have

$$\mathbb{E}[\mathscr{D}_j^e] = \mathbb{E}\left[\sum_{n=1}^{\tilde{n}_j} \sum_{t=t_{j,n-1}+1}^{t_{j,n}} [C(F, y_j^{(n)}) - C(F, x^*)]\right]$$
$$= \sum_{k=1}^{M} \mathbb{E}\left[\sum_{n=1}^{n_j(k,M)} \sum_{t=t_{j,n-1}+1}^{t_{j,n}} [C(F, z_j^n(k)) - C(F, x^*)] \cdot \mathbf{1}\{\tilde{n}_j \ge n\} \mid y_j^{(0)} = k\right] p_j(k).$$

Conditional on $y_j^{(0)} = k$, the event $\{\tilde{n}_j \ge n\}$ only depends on $\mathscr{H}^c_{t_{j,n-1}}$ and is hence independent of the demand observations $\{D_t\colon t \ge t_{j,n-1}+1\}$. We have

$$\mathbb{E}[\mathscr{D}_j^e] = \sum_{k=1}^{M}\left[\sum_{n=1}^{n_j(k,M)} \sum_{t=t_{j,n-1}+1}^{t_{j,n}} \mathbb{E}[C(F, z_j^n(k)) - C(F, x^*) \mid y_j^{(0)} = k] \cdot \mathbb{P}\{\tilde{n}_j \ge n \mid y_j^{(0)} = k\}\right] p_j(k)$$
$$\le \Delta_j^e \sum_{k=1}^{M}\left(\sum_{n=1}^{n_j(k,M)} \mathbb{E}[C(F, z_j^n(k)) - C(F, x^*) \mid y_j^{(0)} = k] \cdot \mathbb{P}\{\tilde{n}_j \ge n \mid y_j^{(0)} = k\}\right) p_j(k)$$
$$\le \Delta_j^e [A_1 + A_2 + A_3 + A_4],$$

where

$$A_1 = \sum_{k=1}^{x^*-1}\left(\sum_{n=1}^{n_j(k,M)} \mathbb{E}[C(F, z_j^n(k)) - C(F, x^*) \mid y_j^{(0)} = k] \cdot \mathbb{P}\{\tilde{n}_j \ge n \mid y_j^{(0)} = k\}\right) p_j(k),$$
$$A_2 = \mathbb{E}[C(F, z_j^1(x^*)) - C(F, x^*) \mid y_j^{(0)} = x^*] \cdot \mathbb{P}\{\tilde{n}_j \ge 1 \mid y_j^{(0)} = x^*\} p_j(x^*), \tag{A19}$$
$$A_3 = \sum_{n=2}^{n_j(x^*,M)} \mathbb{E}[C(F, z_j^n(x^*)) - C(F, x^*) \mid y_j^{(0)} = x^*] \cdot \mathbb{P}\{\tilde{n}_j \ge n \mid y_j^{(0)} = x^*\} p_j(x^*),$$
$$A_4 = \sum_{k=x^*+1}^{M}\left(\sum_{n=1}^{n_j(k,M)} \mathbb{E}[C(F, z_j^n(k)) - C(F, x^*) \mid y_j^{(0)} = k] \cdot \mathbb{P}\{\tilde{n}_j \ge n \mid y_j^{(0)} = k\}\right) p_j(k)$$

$$A_1 \le \sum_{k=1}^{x^*-1} n_j(0, M)(h+b)Mp_j(k)$$
$$\le n_j(0, M)(h+b)M\,\mathbb{P}\{y_j^{(0)} \le x^* - 1\}$$
$$\le n_j(0, M)(h+b)Ma_{j-1}\alpha_{x^*-1}^{\Delta_{j-1}^e},$$

where the last inequality follows from Lemma 3. We also have

$$A_2 \le \mathbb{E}[C(F, z_j^1(x^*)) - C(F, x^*)] = (h+b)\sum_{i=x^*}^{z_j^1(x^*)-1} |\beta - F(i)| \quad \text{and}$$
$$A_3 \le n_j(x^*, M)(h+b)M\,\mathbb{P}\{\tilde{n} \ge 2 \mid y_j^{(0)} = x^*\}$$
$$\le n_j(x^*, M)(h+b)Ma_{j-1}\alpha_{x^*}^{\Delta_{j-1}^e},$$

where the last inequality follows from noting that $\mathbb{P}\{\tilde{n}_j \ge 2 \mid y_j^{(0)} = x^*\} = \mathbb{P}\{q_j^{(0)} = x^*, q_j^{(1)} = z_j^1(x^*) \mid y_j^{(0)} x^*\}$, and hence one may upper bound this probability by $\mathbb{P}\{q_j^{(1)} \ge x^* + 1 \mid y_j^{(0)} = x^*, q_j^{(0)} = x^*\} \le \alpha_{x^*}^{\Delta_{j-1}^e}$. Finally,

$$A_4 \le n_j(x^*+1, M)(h+b)M\,\mathbb{P}\{y_j^{(0)} \ge x^* + 1\}$$
$$\le n_j(x^*+1, M)(h+b)Ma_{j-1}\alpha_{x^*}^{\Delta_{j-1}^e},$$

where the last inequality follows from Lemma 3.

Note that for $i \in \{0, \ldots, M\}$,

$$\Delta_j^e \alpha_i^{\Delta_{j-1}^e} = 4\Delta_{j-2}^e \alpha_i^{\Delta_{j-2}^e} \alpha_i^{\Delta_{j-2}^e} \overset{(a)}{\le} \frac{4\exp\{-1\}}{-(\log \alpha_i)} \alpha_i^{\Delta_{j-2}^e}$$
$$\overset{(b)}{\le} 4\exp\{-1\}(\min\{1/M^2, 1/2^2\})^{2^{j-3}} m_\beta^{-1}\varepsilon^{-2},$$

where (a) follows from the fact as long as $\alpha_i \in (0, 1)$, for any $i = 0, \ldots, M$ and any z,

$$z\alpha_i^z \le -(\log \alpha_i)^{-1}\exp\{-1\}, \tag{A20}$$

and (b) follows from the fact that $\alpha_i^{\gamma_e} \le \min\{1/M^2, 1/2^2\}$ for all $i \ge 0$ given the choice of γ_e in (9). We deduce that

$$\mathbb{E}[\mathscr{D}_j^e] \le (h+b)\sum_{i=x^*}^{z_j-1} |\beta - F(i)|\Delta_j^e + n_j(0, M)(h+b)M\Delta_j^e a_{j-1}\left[\alpha_{x^*-1}^{\Delta_{j-1}^e} + 2\alpha_{x^*}^{\Delta_{j-1}^e}\right]$$
$$\le (h+b)\left[\sum_{i=x^*}^{z_j-1} |\beta - F(i)|\Delta_j^e + 12n_j(0, M)Ma_{j-1}e^{-1} \cdot (\min\{1/M^2, 1/2^2\})^{2^{j-3}} m_\beta^{-1}\varepsilon^{-2}\right]. \tag{A21}$$

Conclusion. Now, noting that $\alpha_{x^*}^{\gamma_e} \le \min\{M^{-2}, 2^{-2}\}$, $n_j(1, M) \le M$ and using (A17), we have $a_{j+1} \le 3 + Ma_j \cdot \min\{M^{-2^j}, 2^{-2^j}\}$. Recalling that $a_2 = 3 + n_1(1, M) \le 3 + M$, one can establish that $a_j \le 6$ for all $j \ge 3$. Since $t_{j+1} - t_j \ge \gamma_1 a^{2^j}$, $\kappa \le (\log\log(T/\gamma_1) - \log\log a)/\log 2 \le (\log\log T - \log\log a)/\log 2$. Recalling the definition of γ_e in (9) and

putting together (A18), and (A21) with the facts stated in above, one obtains

$$\begin{aligned}\mathcal{C}^{\pi}(F,T)-\mathcal{C}^{*}(F,T) \le{}& (h+b)M2\gamma_1+6(h+b)M\gamma_1\sum_{j=2}^{\kappa}(a/4)^{4^{j-2}}\\ &+(h+b)M[(\log M)/\log(1+m_1)]\gamma_e\\ &+(h+b)\sum_{j=2}^{\kappa}\sum_{i=x^*}^{z_j^1(x^*)-1}|\beta-F(i)|\Delta_j^e\\ &+72(h+b)M\exp\{-1\}m_\beta^{-1}\varepsilon^{-2}\\ &\cdot\sum_{j=2}^{\kappa}n_j(0,M)(\min\{1/M^2,1/2^2\})^{2^{j-3}}.\end{aligned}$$

Note that $z_j^1(x^*)=x^*+1$ for $j\ge j'$, because $m_j^1=0$ for all $j\ge j'$. Thus,

$$\sum_{j=j'}^{\kappa}\sum_{i=x^*}^{z_j^1(x^*)-1}|\beta-F(i)|\Delta_j^e=|\beta-F(x^*)|\gamma_e\sum_{j=j'}^{\kappa}2^{j-1}\le\gamma_e 2^{\kappa}\le\gamma_e\frac{\log T}{\log a}.$$

One deduces that

$$\mathcal{C}^{\pi}(F,T)-\mathcal{C}^{*}(F,T)\le K\log M\varepsilon^{-2}[M\log M+\log T],$$

where K is a suitable positive constant that depends only on b and h. This completes the proof. □

References

Akcay A, Biller B, Tayur S (2009) Setting inventory targets in the presence of finite historical demand data. Working paper, Carnegie Mellon University, Pittsburgh.

Anderson ET, Fitzsimons GJ, Simester D (2006) Measuring and mitigating the costs of stockouts. *Management Sci.* 52:1751–1763.

Bensoussan A, Cakanyildirim M, Sethi SP (2009a) Technical note: The censored newsvendor and the optimal acquisition of information. *Oper. Res.* 57:791–794.

Bensoussan A, Cakanyildirim M, Feng Q, Sethi SP (2009b) Optimal ordering policy and value of information under partially observed lost sale. Working paper, University of Texas at Dallas, Richardson.

Blackwell D (1956) An analog of the minimax theorem for vector payoffs. *Pacific J. Math.* 6:1–8.

Braden DJ, Freimer M (1991) Informational dynamics of censored observations. *Management Sci.* 37:1390–1404.

Burnetas AN, Smith CE (2000) Adaptive ordering and pricing for perishable products. *Oper. Res.* 48:436–443.

Cachon GP, Kök AG (2007) How to (and how not to) estimate the salvage value in the newsvendor model. *Manufacturing Service Oper. Management* 9:276–290.

Cesa-Bianchi N, Lugosi G (2006) *Prediction, Learning, and Games* (Cambridge University Press, New York).

Chen L (2010) Bounds and heuristics for optimal Bayesian inventory control with unobserved lost sales. *Oper. Res.* 58:396–413.

Chen L, Plambeck EL (2008) Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing Service Oper. Management* 10:236–256.

Conrad SA (1976) Sales data and the estimation of demand. *Oper. Res. Quart.* 27:121–127.

Cooper WL, Homem-de-Mello T, Kleywegt AJ (2006) Models of the spiral-down effect in revenue management. *Oper. Res.* 54:968–987.

Ding X, Puterman ML, Bisi A (2002) The censored newsvendor and the optimal acquisition of information. *Oper. Res.* 50:517–527.

Eren S, Maglaras C (2013) A maximum entropy joint demand estimation and capacity control policy for revenue management. *Production Oper. Management*. Forthcoming.

Foster DP, Vohra R (1999) Regret in the on-line decision problem. *Games Econom. Behav.* 29:7–35.

Godfrey GA, Powell WB (2001) An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. *Management Sci.* 47:1101–1112.

Hannan J (1957) Approximation to Bayes risk in repeated play. *Contributions to the Theory of Games*, Vol. 3 (Princeton University Press, Princeton, NJ), 97–139.

Harpaz G, Lee WY, Winkler RL (1982) Learning, experimentation, and the optimal output decisions of a competitive firm. *Management Sci.* 28:589–603.

Huh WT, Rusmevichientong P (2009) A non-parametric asymptotic analysis of inventory planning with censored demand. *Math. Oper. Res.* 34:103–123.

Huh WT, Levi R, Rusmevichientong P, Orlin JB (2011) Adaptive data-driven inventory control with censored demand based on Kaplan-Meier estimator. *Oper. Res.* 59:929–941.

Kunnumkal S, Topaloglu H (2008) Using stochastic approximation methods to compute optimal base-stock levels in inventory control problems. *Oper. Res.* 56:646–664.

Kunnumkal S, Topaloglu H (2009) A stochastic approximation method for the single-leg revenue management problem with discrete demand distributions. *Math. Methods Oper. Res.* 70:477–504.

Lai TL, Robbins H (1985) Asymptotically efficient adaptive allocation rules. *Adv. Appl. Math.* 6:4–22.

Lariviere MA, Porteus EL (1999) Stalking information: Bayesian inventory management with unobserved lost sales. *Management Sci.* 45:346–363.

Lu X, Song J-S, Zhu K (2008) Technical note: Analysis of perishable-inventory systems with censored demand data. *Oper. Res.* 56:1034–1038.

Powell W, Ruszczyński A, Topaloglu H (2004) Learning algorithms for separable approximations of discrete stochastic optimization problems. *Math. Oper. Res.* 29:814–836.

Scarf H (1959) Bayes solutions of the statistical inventory problem. *Ann. Math. Statist.* 30:490–508.

Schleifer A (1992) L. L. Bean Inc.: Item forecasting and inventory management. Case teaching note, Harvard Business School, Boston.

van Ryzin G, McGill J (2000) Revenue management without forecasting or optimization: An adaptive algorithm for determining airline seat protection levels. *Management Sci.* 46:760–775.

Veinott A (1965) Optimal policy for a multi-product, dynamic, nonstationary inventory problem. *Management Sci.* 12:206–222.