Confidence Intervals for Data-driven Inventory Policies with Demand Censoring

Gah-Yi Ban
Management Science & Operations, London Business School, Regent's Park, London, NW1 4SA, United Kingdom.
gban@london.edu

We revisit the classical dynamic inventory management problem of Scarf (1959b) from the perspective of a decision-maker who has n historical selling seasons of data and must make ordering decisions for the upcoming season. We develop a nonparametric estimation procedure for the (S, s) policy that is consistent, then characterize the finite-sample properties of the estimated (S, s) levels by deriving their asymptotic confidence intervals. We also consider having at least some of the past selling seasons of data censored from the absence of backlogging, and show that the intuitive procedure of first correcting for censoring in the demand data yields inconsistent estimates. We then show how to correctly use the censored data to obtain consistent estimates and derive asymptotic confidence intervals for this policy using Stein's method. We further show the confidence intervals can be used to effectively bound the difference between the expected total cost of an estimated policy and that of the optimal policy. We validate our results with extensive computations on simulated data. Our results extend to the repeated newsvendor problem and the base-stock policy problem by appropriate parameter choices.

Key words: inventory management; nonparametric estimation; demand censoring; confidence intervals;
History: This version: April 20, 2019. Forthcoming in *Operations Research*.

1. Introduction

The stochastic dynamic inventory problem constitutes an important class of decision problems in operations management. In this paper, we revisit the classical problem of Scarf (1959b) from the perspective of a decision-maker (DM) who does not know the demand distribution but has historical data to base her decision on.

Specifically, the DM has observations of n independent, identically distributed (iid) selling seasons of data, where the selling horizon is of length T. We consider three cases for the available data. In Scenario (a), the demands for all n selling seasons are available, in Scenario (b), only sales data are available for some (n_1, $1 \leq n_1 \leq n-1$) of the past selling seasons, and in Scenario (c), only sales data are available for all n selling seasons.

For these cases, we ask the following questions: (Q1) given the n seasons of data, how should the DM order for the upcoming selling season? and (Q2) what finite-sample performance bounds can she assign for the estimated ordering policy, and the total cost? In other words, what is the precision of an inventory policy estimated from historical data? Note the classical dynamic inventory problem we consider simplifies to the repeated newsvendor problem and the base-stock policy problem by appropriate parameter choices, so answers to the aforementioned questions apply to these problems as well.

While the first question has been addressed to varying degrees in the literature, the second question, which is of significant practical value, has not yet been addressed. The main aim of this paper is thus to characterize confidence intervals around sensible estimates of the optimal $(S_t, s_t)_{t=1}^T$ policy of the classical dynamic inventory problem.

To address (Q1), we propose nonparametric estimation procedures that yield consistent policies for all three scenarios of available data, in Sec. 3. For Scenario (a), an intuitive use of the data to estimate the stochastic dynamic program yields consistent decisions. For Scenario (b), we show that the DM must be careful in using the censored demand data. In particular, we show that the intuitive procedure of correcting for censoring in the demand data itself, then using the corrected data to solve the estimated dynamic programming (DP) equations necessarily yields inconsistent estimates of the optimal inventory policy. We then show how the DM can correctly use the censored data to yield consistent decisions. Finally, we show that the estimation problem under Scenario (c) can be broken down into Scenario (a) or Scenario (b) depending on the relative positions of the S_t's to the censoring levels.

To address (Q2), we derive asymptotic confidence intervals for the estimates of $(S_t, s_t)_{t=1}^T$ using the asymptotic normality property of M-estimators (Van der Vaart 2000). This can be found in Sec. 4. For Scenario (a), we can use the classical results with an inductive argument, but for Scenario (b), we need an extension of the classical results using Stein's method (Stein 1972) as our proposed estimation procedure introduces correlations in the estimation objective. The confidence intervals under under Scenario (c) equal to those under Scenario (a) or Scenario (b) depending on the relative positions of the S_t's to the censoring levels.

In Sec. 5, we investigate how the confidence intervals of an estimated $(S_t, s_t)_{t=1}^T$ policy can be used to bound its worst-case expected total cost. We then validate the theoretical results via extensive computations on simulated data in Sec. 6.

1.1. Summary of main contributions

We make three main contributions. First, we establish estimation procedures for finding consistent estimators of the optimal $(S_t, s_t)_{t=1}^T$ policy, under both uncensored and censored data scenarios. For the uncensored data Scenario (a), the intuitive Sample Average Approximation (SAA; see Shapiro et al. 2009) procedure works. For the censored data Scenario (b), however, we show that an intuitive approach of correcting for censoring in the demand data first yields inconsistent estimates. We propose an alternative procedure which corrects for the censoring by re-weighting the demand data indirectly through the estimating equations, and show this yields consistent estimators. This is a significant departure from much of the demand censoring literature, which has focused on the correction of the estimation of the demand distribution (Conrad 1976, Wecker 1978, Nahmias 1994, Agrawal and Smith 1996, Anupindi et al. 1998, Vulcano et al. 2012). While understanding the full demand distribution is of inherent value, our results show that if the end goal is to estimate the optimal reorder points and order-up-to levels, the correction for censoring must be done for the objective function, rather than for the demand distribution.

Second, we analytically derive asymptotic confidence intervals of estimated $(S_t, s_t)_{t=1}^T$ policies, which, although asymptotic formulas, are accurate enough to be used in practice. While the use of confidence intervals has been a key component of decision-making in other arenas (e.g., evaluations of economic policies and drug trials), it has thus far been overlooked by the operations literature. This work purports to fill the gap between what is now available (data and statistical theory) and the classical operational problem. Furthermore, confidence intervals for the censored data Scenario (b) adds new perspective to the literature on data-driven inventory management with demand censoring, which we discuss in the literature review below.

Third, we provide upper bounds on the estimated total cost. One implication of the theoretical bound is that the expected total cost of an estimated policy can be bounded by a linear combination of the confidence intervals of the estimated order-up-to levels, plus a small error term. This means that, if the confidence intervals around the estimated order-up-to levels are reduced by 10%, this would translate directly to a 10% reduction in upper bounding the expected total cost as well. Thus, direct improvements to the precision of estimating optimal order-up-to levels reduce uncertainty in the estimation of the total cost.

1.2. Literature Review

The earliest papers on stochastic dynamic inventory management establish the structure of the optimal inventory policy. The seminal work of Scarf (1959b) showed that the optimal policy is of (S,s) type when the ordering cost consists of a fixed setup cost and a linear per-unit cost. That is, at the beginning of period t, the DM observes the current inventory level and orders up to S if this level is below the critical level s. Works showing the optimality of the (S,s) structure for other settings followed (e.g. infinite horizon problem Iglehart 1963, generalized cost structures Porteus 1971 and Markovian demand Sethi and Cheng 1997); as well as efforts in efficient computations (see Federgruen and Zipkin 1984, Zheng and Federgruen 1991 and references therein). We refer the reader to Zipkin (2000) and Porteus (2002) for an overview.

In practice, however, the full distributional information of the demand is not available, thus how one ought to make inventory decisions under uncertainty has formed a significant line of inquiry. Two main approaches exist in the literature: the Bayesian approach, whereby unknown parameters of the demand distribution are dynamically learned (Scarf 1959a, Azoury 1985, Lovejoy 1990) and the nonparametric approach, whereby the DM has access to samples of demand data from an unknown distribution. Our distribution-free, data-driven setting thus falls under the nonparametric category.

One focus of nonparametric inventory management papers has been on the efficient computation of data-driven (or, equivalently, sampling-based) policies. Burnetas and Smith (2000), Huh and Rusmevichientong (2009) and Kunnumkal and Topaloglu (2008) consider stochastic gradient algorithms, Godfrey and Powell (2001) and Powell et al. (2004) consider the adaptive value estimation method, Levi et al. (2007) provide a customized method based on weaving the actual DP through a shadow DP, Levi and Shi (2013) propose algorithms based on randomized decision rules, and Ban and Rudin (2019) provide machine learning algorithms for the repeated newsvendor problem. We make clear that this is not the focus of the current paper; our focus is to derive finite-sample properties of data-driven policies, which is distinct from existing works.

While there is no precedent for characterizations of confidence intervals for estimated inventory policies, Levi et al. (2007) and Levi et al. (2015), which provide data-driven inventory policies with probabilistic guarantees, could be considered the most similar to this work. Levi et al. (2007) studies the single-period and dynamic inventory problems with zero setup cost from a nonparametric perspective and provides algorithms inspired by approximating the SAA problem with convex counterparts to solve them. The paper then provides probabilistic bounds on the minimal number of iid demand observations that are needed for the algorithms to be near-optimal. Levi et al. (2015) improves upon this bound for the single-period case.

However, the results in Levi et al. (2007) and Levi et al. (2015) are based on the Hoeffding and Bernstein inequalities respectively, both of which are very loose. Hence one cannot obtain confidence interval estimates from their probabilistic bounds. To illustrate, for the examples considered in Sec. 6, one needs over 56,000–240,000 iid observations of the demand to give bounds on the base stock policy within 50% accuracy and at 5% significance level using the bounds of Levi et al. (2015) and Levi et al. (2007). In contrast, our asymptotic confidence interval bounds can be used to quote confidence bounds for n as small as 50.

Recent works on the issue of demand censoring for the data-driven inventory problem include Heese and Swaminathan (2010), Huh et al. (2011), Dai and Jerath (2013), Besbes and Muharremoglu (2013), Jain et al. (2014), Chen and Mersereau (2015) (which gives a comprehensive review), Shi et al. (2016), Chen et al. (2017) and Zhang et al. (2018). Of these, the closest to our work are Huh et al. (2011) and Besbes and Muharremoglu (2013), both of which study the simpler repeated newsvendor problem from a nonparametric perspective. We discuss these two papers in more detail below.

Huh et al. (2011) proposes a consistent policy based on the Kaplan-Meier estimator for general discrete demands; in contrast our work focuses on consistent estimation of the dynamic inventory policy for continuous demands. Besbes and Muharremoglu (2013) shows that there is no marked difference between demand censoring and full-information cases when the underlying demand is continuous, in that the minimum worst-case regret in both cases grow logarithmically in the number of periods (which in our setting is n, the number of observations). In the current paper, we support this result through a different type of analysis. Instead of the worst-case regret, we consider confidence bounds on the total cost, and show that this shrinks at the same rate in the number of observations for both censored and uncensored data.

Finally, we mention a recent work, Arlotto and Steele (2016), that derives Central Limit Theorem (CLT) results for temporally non-homogenous Markov chains, with implications on the infinite-horizon inventory problem. While the inventory example considered in Arlotto and Steele (2016) is very different (infinite horizon, no demand censoring, no setup cost, Markovian demands with identical distributions over time, and not data-driven), as far as we are aware, it is the only other existing work to analyze an inventory problem via CLT.

1.3. Preliminaries

We denote convergence in probability by $\xrightarrow{P}$ and convergence in distribution by $\Rightarrow$. As convention, we denote random variables in capital letters and their realizations in lower-case letters. All proofs can be found in the supplementary online Appendix.

2. The Model

There are two key ingredients to our problem — the structure of the stochastic dynamic program and the nature of the data available to the DM. Let us describe the problem structure first, then detail the nature of the available data.

2.1. Dynamic Inventory Management Problem

The inventory problem we consider is the classical problem of Scarf (1959b), where unmet demands are backlogged. The time-horizon is finite, with T planning periods. We use t to denote the actual time period, starting at 1. At each period, the backordering cost is denoted b_t and the holding cost h_t. There is also a fixed setup cost $K_t \geq 0$ if an order is made at time t; thus the total cost of ordering a quantity q at time t is denoted by:

$$O_t(q) = \begin{cases} K_t + c_t q & \text{if } q > 0 \\ 0 & \text{otherwise,} \end{cases}$$

where c_t is the per unit ordering cost. Denote the inventory level at the beginning of period t by I_t; note this can be negative due to backlogging. The initial inventory level, I_1, is known to the DM. Lead time is zero, so any orders placed in period t arrive within the same period. The random demand in period t, D_t, is realized after any orders made in period t arrive. The value of future cash flows is discounted at the rate $\alpha_t \in [0,1)$. Any remaining inventory at the end of the selling horizon T has zero salvage value.

Let us introduce $C_t(\cdot,\cdot) : \mathbb{R}_+ \times [\underline{D}, \bar{D}] \to \mathbb{R}$, the single-period newsvendor cost function:

$$C_t(y;d) := b_t(d-y)^+ + h_t(y-d)^+,$$

where the variable y denotes the stock level after deliveries in period t. Then the Bellman equations for the dynamic inventory problem can be stated as follows:

$$\begin{aligned} V_{T+1}(x) &= 0 \quad \forall\, x \geq 0 \\ V_t(I_t) &= \min_{y \geq I_t} O_t(y - I_t) + \mathbb{E}[C_t(y, D_t) + \alpha_t V_{t+1}(y - D_t)], \quad 1 \leq t \leq T. \end{aligned} \tag{1}$$

Scarf (1959b) showed that the solution to (1) is of $(S_t, s_t)_{t=1}^T$-type— that is, if at period t, $I_t < s_t$, it is optimal to order up to S_t. Using this insight, Scarf (1959b) shows that for each $t = 1, \dots, T$, the optimal order-up-to level S_t is the global minimum of the function

$$G_t(y) := c_t y + \mathbb{E}_t[C_t(y, D_t)] + \alpha_t \mathbb{E}_t[V_{t+1}(y - D_t)], \tag{2}$$

over the domain $[\underline{D}, \bar{D}]$, where $\alpha_T = 0$ and the cost-to-go function is given by

$$\begin{aligned} &V_{T+1}(x) := 0 \;\; \forall\, x \geq 0 \\ &V_{t+1}(x) := \begin{cases} G_{t+1}(S_{t+1}) + K - c_t x & \text{if } x < s_{t+1} \\ G_{t+1}(x) - c_t x & \text{if } x \geq s_{t+1}, \end{cases} \quad 1 \leq t \leq T-1. \end{aligned} \tag{3}$$

The optimal reorder point s_t is the smallest s less than S_t such that:

$$G_t(s) = G_t(S_t) + K.$$

With (3) we can rewrite (2) as

$$\begin{aligned} G_T(y) &= c_T y + \mathbb{E}_T[C_T(y, D_T)], \\ G_t(y) &= (1 - \alpha_t) c_t y + \mathbb{E}_t[C_t(y, D_t)] + \alpha_t c_t \mathbb{E}_t[D_t] + \alpha_t G_{t+1}(s_{t+1}) \mathbb{E}_t[\mathbb{I}_t(y - s_{t+1})] \\ &\quad + \alpha_t \mathbb{E}_t[G_{t+1}(y - D_t) \mathbb{I}_t^c(y - s_{t+1})], \quad t = 1, \dots, T-1, \end{aligned}$$

where $\mathbb{I}_t(x) := \mathbb{I}(x < D_t)$ and $\mathbb{I}_t^c(x)$ is its complement.

Note if the setup cost K is zero, then the problem reduces to one for which the optimal policy is the base-stock policy $R_1, \dots, R_T$, where at each time period t it is optimal to order $R_t - I_t$ if the inventory level is below the critical level R_t. In this case, the optimal reorder points and order-up-to levels coincide, and equal $R_1, \dots, R_T$. While most recent works in inventory management have focused on this case, we consider $K \geq 0$ for completeness. If for all $t = 1, \dots, T$, $\alpha_t = 0$ and the D_t's have the same distribution, the problem reduces to the repeated newsvendor problem; if $T = 1$, the problem is the single-period newsvendor problem.

2.2. Problem Assumptions

We make the following assumptions throughout this paper.

1. *Assumption on the stochastic demand*: For each $t = 1, \dots, T$, the random demand at time t, D_t, is a continuous random variable bounded on the interval $[\underline{D}, \bar{D}]$, which

is known to the DM. The demands are independent but not necessarily identically distributed across time.

2. *Identifiability*: For each $t = 1, \ldots, T$,

$$\forall \varepsilon > 0, \inf_{y \in [\underline{D}, \bar{D}]} \{G_t(y) : ||y - S_t||_2 \geq \varepsilon\} > G_t(S_t). \tag{ID}$$

Assumption 1 pertains to the nature of the stochastic demand. The assumption on the continuity of the demand random variable is needed for the differentiability of the objective functions $G_t(\cdot)$, $t = 1, \ldots, T$. The assumption that the DM knows a lower and upper bound on the demand is realistic in practice because 0 is a universal lower bound on the demand for any product, and an upper bound can be obtained from estimating the firm's total customer base. Finally, the assumption that the demands are independent across time respects the original problem structure of Scarf (1959b). Relaxing this assumption would require additional assumptions on the temporal evolution of the demand process, which we leave for future work.

Assumption 2 is essential for statistical inference. One sufficient condition for (ID) to be satisfied is if $G_t(\cdot)$ has a unique minimizer over $[\underline{D}, \bar{D}]$. Note we do not need a similar assumption on s_t, $t = 1, \ldots, T$, because by definition they are unique.

2.3. Description of the Available Data

The firm has collected n selling seasons of data, which are iid across the seasons. That is, the T-dimensional demand vector $[d_1^i, \ldots, d_T^i]$ from season i is iid to the demand vector $[d_1^j, \ldots, d_T^j]$ from season j.

We consider three different cases. In Scenario (a), the DM has access to the actual demand data, $\mathbf{D} = \{[d_1^i, \ldots, d_T^i]_{i=1}^n\}$ for all n seasons. In Scenario (b), the DM has access to n_0, $1 \leq n_0 \leq n-1$ seasons of demand data and $n_1 := n - n_0$ seasons of censored demand data as well as the past stocking levels for the censored seasons, $\{\mathbf{x}^i\}_{i \in \mathcal{J}} = \{[x_1^i, \ldots, x_T^i]\}_{i \in \mathcal{J}}$, where $\mathcal{J} \subseteq \{1, \ldots, n\}$ is the index set denoting the censored selling seasons, with $|\mathcal{J}| = n_1$. Finally, in Scenario (c) the DM has access to only the sales data $\mathbf{Z} = \{[z_1^i, \ldots, z_T^i]_{i=1}^n\}$, where z_t^i is the realization of the censored random variable $Z_t^i := \min(D_t, x_t^i)$ for season i, and the past stocking levels $\{\mathbf{x}^i\}_{i=1}^n = \{[x_1^i, \ldots, x_T^i]\}_{i=1}^n$.

The three scenarios are illustrated in Fig. 1.

$$
\begin{array}{r l}
 & \overbrace{\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad}^{\text{Selling horizon: T periods}} \\
\begin{array}{c} n_0 = (1-r)n \\ \text{uncensored seasons} \end{array} & \left\{ \begin{array}{cccc} d_1^1 & d_2^1 & \dots & d_T^1 \\ \vdots & \vdots & \vdots & \vdots \\ d_1^{n_0} & d_2^{n_0} & \dots & d_T^{n_0} \end{array} \right. \\
\begin{array}{c} n_1 = rn \\ \text{censored seasons} \end{array} & \left\{ \begin{array}{cccc} d_1^{n_0+1} & d_2^{n_0+1} & \dots & d_T^{n_0+1} \\ \vdots & \vdots & \vdots & \vdots \\ d_1^{n} & d_2^{n} & \dots & d_T^{n} \end{array} \right.
\end{array}
$$

Figure 1: A schematic for the available data. Under Scenario (a), the full information case, $n_0 = n$, under Scenario (b), the partially censored case, $1 \le n_0 \le n-1$ and in the fully censored case Scenario (c), $n_0 = 0$. Note the uncensored and censored seasons do not have to be in separable blocks as shown; what matters is the total number of uncensored versus censored seasons.

3. Asymptotically consistent estimation of $(S_t, s_t)_{t=1}^T$

In this section, we propose estimators for the optimal policy $(S_t, s_t)_{t=1}^T$ under the three Scenarios (a)–(c) and show that they are consistent.

3.1. Asymptotically consistent estimation with fully uncensored data ($n_0 = n$)

In the full information case of Scenario (a), we estimate $(S_t, s_t)_{t=1}^T$ with $(\hat{S}_t, \hat{s}_t)_{t=1}^T$, where for each $t = T, \dots, 1$,

$$\hat{S}_t := \underset{y \in [\underline{D}, \bar{D}]}{\operatorname{argmin}} \hat{G}_t(y), \tag{4}$$

and

$$\hat{s}_t := \min_s \left\{ \underline{D} \le s \le \hat{S}_t \middle| \ \hat{G}_t(s) - \hat{G}_t(\hat{S}_t) - K = 0 \right\}, \tag{5}$$

where $\hat{G}_t(y) = \frac{1}{n}\sum_{i=1}^n g(y, d_t^i)$, with

$$
\begin{aligned}
g_T(y,d) &= c_T y + C_T(y,d), \\
g_t(y,d) &= (1-\alpha_t)c_t y + C_t(y,d) + \alpha_t c_t d + \alpha_t \hat{G}_{t+1}(\hat{s}_{t+1}) \mathbb{I}(y - \hat{s}_{t+1} \le d) \\
&\qquad + \alpha_t \hat{G}_{t+1}(y-d)\mathbb{I}(y - \hat{s}_{t+1} > d), \quad t = 1, \dots, T-1.
\end{aligned} \tag{6}
$$

The asymptotic consistency of estimated quantities $\hat{s}_t$ and $\hat{S}_t$ to their true respective quantities s_t and S_t requires proving convergence of solutions of estimated optimization problems. For this purpose, we utilize the theory of M-estimation from statistics (Van der

Vaart 2000). M-estimation refers to estimation through optimizing an objective function (the "M" stands for maximization or minimization); for example Maximum Likelihood Estimation is a well-known example. This literature thus provides a basis for analyzing estimators obtained through optimization.

Three conditions are needed to ensure the convergence of an estimated optimal solution to the true value. They are: (i) near-optimality of the estimator for the estimated problem, (ii) the true optimal solution is well-defined, and (iii) the estimated objective function converges uniformly to the true objective function over the domain of the problem.

In our setting, the first condition is satisfied because we can compute the in-sample critical values to an arbitrary accuracy because they are defined by a continuous function on a finite domain $[\underline{D}, \bar{D}]$. The second condition is a condition on the true problem, and is satisfied by the identifiability condition (ID). The third and final condition is the uniform convergence of the estimated cost function $\hat{G}_t(\cdot)$ to its true value $G_t(\cdot)$.

To get to the main result of asymptotic consistency, we piece together the key results in an inductive argument, starting at $t = T$ and working backwards in time.

THEOREM 1 (**$(\hat{S}_t, \hat{s}_t)_{t=1}^T$ is consistent**). *The estimated reorder points and optimal order-up-to levels $(\hat{S}_t, \hat{s}_t)_{t=1}^T$ are consistent, i.e., for each $t = 1, \ldots, T$, as $n \to \infty$, $\hat{S}_t \xrightarrow{P} S_t$ and $\hat{s}_t \xrightarrow{P} s_t$.*

3.2. Asymptotically consistent estimation with censored data ($1 \leq n_0 \leq n-1$)

In the classical problem setting, there is no demand censoring as back-ordering is allowed. However, this may not not hold in practice if, for instance, customers refuse backlogging their demand when there is no stock, or there are human or machine errors in recording back-orders. Alternatively, a firm could initially have stocked a large amount of its product for the first few selling seasons to learn about the demand, then subsequently introduce back-ordering to reduce costs at a later season. Let us thus assume that for n_0, $1 \leq n_0 \leq n-1$ selling seasons in the past, backlogging was allowed, but not so for $n_1 = 1 - n_0$ selling seasons. Let $r := n_1/n$ denote the proportion of censored selling seasons to the total number of selling seasons. We show in Sec. 4 that the analysis of this scenario informs decision-making for the full demand censoring Scenario (c) as well ($n_0 = 0$).

To simplify the analysis, we assume that in the selling seasons with no backlogging, the DM had set the stocking levels at $(x_1, \ldots, x_T)$ — i.e., the stocking levels during the n_0 selling seasons were the same across seasons. Note that the numbers $(x_1, \ldots, x_T)$ are arbitrary and nonrandom, reflecting the stocking decisions that were made in the past, be they a result of algorithmic or human decisions. Importantly, we note that this information is readily available to the DM in practice. Generalizing this to allow for different stocking levels from one selling season to the next (i.e., the stocking levels can differ from one season to the next) is a straight-forward extension of the simpler case, but with much more notational complexity, hence we forego the analysis in full generality.

Given the partially censored data, the DM may wish to discard observations that correspond to selling seasons with no backlogging, i.e., discard all $(d_1^i, \ldots, d_T^i)$'s where $i \in \mathcal{J}$. In such a case, the effective size of the data set reduces to $(1-r)n$. The DM can still obtain consistent policies by solving (4)–(5) on the reduced data set — however, this is not ideal as this reduces the sample size. A natural question that follows is whether the DM could use the censored data more effectively, rather than discard them.

A large part of the demand censoring literature has focused on the correction of the estimation of the demand distribution (Conrad 1976, Wecker 1978, Nahmias 1994, Agrawal and Smith 1996, Anupindi et al. 1998, Vulcano et al. 2012). However, we show in Sec. 3.3 below that in the context of an inventory problem, correcting for censoring in the demand data directly can lead to inconsistent estimates of $(S_t, s_t)_{t=1}^T$. In Sec. 3.4 we show that censoring needs to be corrected through estimates of the objective functions $G_t(\cdot)$ in order to yield consistent estimates of $(S_t, s_t)_{t=1}^T$.

3.3. Inconsistent use of censored demand data

Consider correcting for the demand censoring in the following way. For $i \in \mathcal{J}$, let

$$\breve{d}_t^i = \begin{cases} d_t^i & \text{if } d_t^i < x_t \\ \bar{d}_t := \dfrac{\sum_{i \in \mathcal{J}^c} d_t^i \mathbb{I}(d_t^i \geq x_t)}{\sum_{i \in \mathcal{J}^c} \mathbb{I}(d_t^i \geq x_t)} & \text{otherwise,} \end{cases}$$

for all $t = 1, \ldots, T$.

In other words, estimate the demand at points of possible lost sales by the conditional average of the demands above the stock level from the uncensored data set $\mathcal{J}^c$. Thus $\breve{d}_t^i$ is, approximately, a realization of the following random variable:

$$\breve{D}_t = \begin{cases} D_t & \text{if } D_t < x_t \\ \mathbb{E}[D_t | D_t \geq x_t] & \text{otherwise} \end{cases}$$

The DM can now use the transformed data set $\breve{\mathcal{D}} = \{(\breve{d}_1^i, \ldots, \breve{d}_T^i)_{i=1}^n\}$, where $\breve{d}_t^i = d_t^i$ for $i \in \mathcal{J}^c$, to compute the optimal ordering policy. In other words, the DM estimates $(S_t, s_t)_{t=1}^T$ by $(\breve{s}_t, \breve{S}_t)_{t=1}^T$, where for each $t = T, \ldots, 1$,

$$\breve{S}_t := \underset{y \in [\underline{D}, \bar{D}]}{\operatorname{argmin}} \ \breve{G}_t(y),$$

and

$$\breve{s}_t := \min_s \left\{ \underline{D} \le s \le \breve{S}_t | \ \breve{G}_t(s) - \breve{G}_t(\breve{S}_t) - K = 0 \right\},$$

where $\breve{G}_t(y) = \frac{1}{n} \sum_{i=1}^n g_t(y, \breve{d}_t^i)$, and $g_t(\cdot, \cdot)$ are as defined before, in (6).

We state below that $(\breve{s}_t, \breve{S}_t)_{t=1}^T$ is not a consistent estimator of $(S_t, s_t)_{t=1}^T$.

THEOREM 2 **($(\breve{s}_t, \breve{S}_t)_{t=1}^T$ is not consistent)**. *The estimated reorder points and optimal order-up-to levels $(\breve{s}_t, \breve{S}_t)_{t=1}^T$ are not consistent, i.e., for each $t = 1, \ldots, T$, as $n \to \infty$, $\breve{s}_t \overset{P}{\not\to} s_t$ and $\breve{S}_t \overset{P}{\not\to} S_t$.*

3.4. A consistent use of censored demand data

We now show a consistent way to incorporate the censored demand data by adjusting for the censoring in the objective function estimation. Instead of rescaling the demand data directly, estimate $G_t(y)$ by

$$\tilde{G}_t(y) = \frac{1}{n} \sum_{i=1}^n \tilde{g}_t(y, d_t^i),$$

where

$$\tilde{g}_t(y, d_t^i) := \begin{cases} \bar{g}_t(y) \mathbb{I}(d_t^i \ge x_t) + g_t(y, d_t^i) \mathbb{I}(d_t^i < x_t) & \text{for } i \in \mathcal{J} \\ g_t(y, d_t^i) & \text{for } i \in \mathcal{J}^c \end{cases}, \tag{7}$$

where $g_t(\cdot, \cdot)$ are as in (6) as before and

$$\bar{g}_t(y) := \frac{\sum_{j \in \mathcal{J}^c} g_t(y, d_t^j) \mathbb{I}(d_t^j \ge x_t)}{\sum_{j \in \mathcal{J}^c} \mathbb{I}(d_t^j \ge x_t)},$$

and solve the dynamic program with $\tilde{G}_t(\cdot)$ instead of $\hat{G}_t(\cdot)$. In other words, the DM can estimate $(S_t, s_t)_{t=1}^T$ by $(\tilde{S}_t, \tilde{s}_t)_{t=1}^T$, where for each $t = T, \ldots, 1$,

$$\tilde{S}_t := \underset{y \in [\underline{D}, \bar{D}]}{\operatorname{argmin}} \ \tilde{G}_t(y),$$

and

$$\tilde{s}_t := \min_s \left\{ \underline{D} \leq s \leq \tilde{S}_t | \ \tilde{G}_t(s) - \tilde{G}_t(\tilde{S}_t) - K = 0 \right\}.$$

The proposed estimation procedure above maintains the empirical estimation of the $G_t(\cdot)$ function if the data comes from an uncensored season ($i \in \mathcal{J}^c$), or if the data comes from a censored season ($i \in \mathcal{J}$) but is the actual demand data ($d_t^i < x_t$). If the data comes from a censored season *and* is censored ($d_t^i > x_t$), then we estimate $g_t(y, d_t^i)$ by a sample average estimator using the data from the uncensored seasons, which is what $\bar{g}_t(y)$ stands for. That is, when there is censoring this method uses $\bar{g}_t(y)$, the sample average estimator of $\mathbb{E}[g_t(y, D_t)|D_t \geq x_t]$. Note this is different from the inconsistent method discussed earlier, where $g_t(y, D_t)\mathbb{I}(D_t \geq x_t)$ is estimated by the sample average estimate of $g_t(y, \mathbb{E}[D_t|D_t \geq x_t])$.

We state below $(\tilde{S}_t, \tilde{s}_t)_{t=1}^T$ is consistent.

THEOREM 3 (**$(\tilde{S}_t, \tilde{s}_t)_{t=1}^T$ is consistent**). *The reorder points and optimal order-up-to levels $(\tilde{S}_t, \tilde{s}_t)_{t=1}^T$ estimated under Scenario (b) are consistent, i.e., for each $t = 1, \ldots, T$, as $n \to \infty$, $\tilde{s}_t \xrightarrow{P} s_t$ and $\tilde{S}_t \xrightarrow{P} S_t$.*

3.5. Asymptotically consistent estimation with fully censored data ($n_0 = 0$)

In this section, we consider the most realistic case for the past demand data available to the decision-maker, Scenario (c).

If there was no backlogging for all past selling seasons, then we cannot identify $\hat{G}_t(y)$ for all $y \in [\underline{D}, \bar{D}]$. However, as we shall show, the DM can still compute the consistent, full information estimate $(\hat{S}_t, \hat{s}_t)$ of the optimal inventory policy as long as the functions $\hat{G}_t(\cdot)$ are identifiable up to at least S_t. This is a minimal requirement for any meaningful estimation of the optimal policies, and can be satisfied by the following condition.

Condition (c1): For all $t = 1, \ldots, T$, the past stocking levels x_t were greater than or equal to S_t.

Under Condition (c1), one can still compute the subgradients of $\hat{G}_t(\cdot)$ on the domain $[\underline{D}, S_t]$ because $\hat{G}_t(\cdot)$ consists of a sum of multiple check-loss functions. Thus its gradient is a sum of multiple indicator functions at points of differentiability and an interval on the remaining non-differentiable points. To illustrate, for $y \in [\underline{D}, S_t] \backslash \{d_T^1, \ldots, d_T^n\}$, the gradient of $\hat{G}_T(\cdot)$ is given by:

$$\dot{\hat{G}}_T(y) = c_T + \frac{1}{n} \sum_{i=1}^n [-b\mathbb{I}(d_T^i \geq y) + h\mathbb{I}(y \leq d_T^i)],$$

and for $y \in \{d_T^1, \dots, d_T^n\}$, the set of subgradients of $\hat{G}_t(\cdot)$ is the interval $[c_T - b, c_T + h]$. Clearly, the subgradients of $\hat{G}_T(\cdot)$ are computable for all $y \in [\underline{D}, S_T]$ with censored data $z_T^i := \min(x_T, d_T^i)$, $i = 1, \dots, n$, by Condition (c1). We can recursively show the same for $t = T-1, \dots, 1$. Thus $\hat{S}_t$, for all $t = 1, \dots, T$ can still be found through subgradient methods, as in Kunnumkal and Topaloglu (2008), who find estimates of the optimal base-stock policy with censored data via the same argument. We note that Huh and Rusmevichientong (2009), Huh et al. (2011) and Besbes and Muharremoglu (2013) make identifiability assumptions similar to Condition (c1) for the repeated newsvendor problem with censored data (specifically, they assume the DM has knowledge of an upper bound to the critical quantile, which is similar to assuming $\hat{G}_t(\cdot)$ is observable on $[\underline{D}, \bar{S}_t]$ for some upper bound $\bar{S}_t$ on S_t, $t = 1, \dots, T$).

Condition (c1) also ensures that s_t, which is strictly less than S_t, is identifiable with the censored data. One can find $\hat{s}_t$ even with censored data because the critical equation $|\hat{G}_t(s) - \hat{G}_t(\hat{S}_t) - K|$ is still computable, because the difference between two check-loss functions at different locations only depend on indicators at either extremes of the domain. To illustrate, observe that

$$
\begin{aligned}
&|\hat{G}_T(s) - \hat{G}_T(\hat{S}_T) - K| \\
&= \left| c_T(s - \hat{S}_T) + \frac{1}{n}\sum_{i=1}^{n} -b(s - \hat{S}_T)\mathbb{I}(d_T^i > s) + h(s - \hat{S}_T)\mathbb{I}(d_T^i < s) + (b+h)(\hat{S}_T - d_T^i)\mathbb{I}(s \le d_T^i \le \hat{S}_T) \right|,
\end{aligned}
$$

which is clearly computable for $s \in [0, S_t)$ even with censored data $\{z_T^i\}_{i=1}^n$. We can recursively show the same for $t = T-1, \dots, 1$. Thus, one can still compute the consistent estimates $(\hat{S}_t, \hat{s}_t)_{t=1}^T$ by searching over the restricted domains $y \in [0, x_t)$, $t = 1, \dots, T$.

We can relax Condition (c1) by allowing for some stocking decisions that were not necessarily larger than S_t. At the minimum, however, we require the stocking levels to have been greater than or equal to S_t for at least some of the time in the past, to allow for identifiability. We state this minimal assumption formally below.

Condition (c2): For each $t \in [1, \dots, T]$, let $\bar{x}_t$ denote the maximal stocking level over all censored seasons. For all $t \in [1, \dots, T]$, the maximal stocking levels $\bar{x}_t$ were greater than or equal to S_t.

Under Condition (c2), we can use the n seasons of data as follows. The seasons during which the stocking level was at the maximal level can be treated as uncensored seasons of data in the study of partially censored data Scenario (b). The remaining seasons of data can be treated as the censored seasons in the same scenario. Then the DM can compute $(\tilde{S}_t, \tilde{s}_t)_{t=1}^T$, which was proven to be consistent in Sec. 3.2.

A limitation to operating under either Condition (c1) or (c2) is that the DM cannot know a priori if either condition is satisfied or not, because both conditions require prior knowledge of the relative position of past stocking levels to the optimal order-up-to levels, $\{S_t\}_{t=1}^T$. However, one can ensure either conditions are met by having the stocks set at the maximum demand levels for at least some selling seasons in the past. Reasonable values for the maximum demand levels can be found by market research and expert insight.

Alternatively, if this had not been the case, then the DM needs to first perform an identifiability test to see if the data at hand are sufficient for estimation of the $(S_t, s_t)_{t=1}^T$ policy. One way to do this is to plot empirical estimates of the $G_t(\cdot)$ function on the observable domain, and see if a pair of numbers satisfying optimality conditions for (S_t, s_t) can be found on this domain. By K-convexity of the $G_t(\cdot)$ function, it is known that there can be just one pair of numbers satisfying optimality (Scarf (1959b)). Thus if such a pair of numbers cannot be found on the observable domain, then the DM can conclude S_t is not identifiable with the available data.

4. Confidence intervals of the estimated policies

In this section, we derive formulas for the asymptotic confidence intervals of the estimated policies introduced in Sec. 3. The formulas are necessarily asymptotic because the complexity of the estimated policies is such that the confidence intervals cannot be characterized analytically in finite-sample. The specific result we prove is asymptotic normality of the estimated policies. (Note: this is distributional convergence, as opposed to probabilistic convergence analysis of Sec. 3.) In other words, for reasonably large n, we show that the estimated $(\hat{S}_t, \hat{s}_t)_{t=1}^T$ and $(\tilde{S}_t, \tilde{s}_t)_{t=1}^T$ values are approximately normally distributed with centers at the respective true values $(s_t,\, S_t)$ with variances $(\rho_t^2,\, \sigma_t^2)$ and $(\tilde{\rho}_t^2,\, \tilde{\sigma}_t^2)$ respectively.

Confidence intervals are useful because they provide statistically meaningful error bars for estimators. For instance, by taking the square root of the asymptotic variance formulas, the DM can then quote approximate 90%, 95% or 99% confidence intervals for the estimates of the $(S_t, s_t)_{t=1}^T$ policy.

4.1. Confidence intervals with fully uncensored data ($n_0 = n$)

THEOREM 4 **(Asymptotic normality of $(\hat{S}_t, \hat{s}_t)_{t=1}^T$).** *Assume Scenario (a). Then for all $t = 1, \ldots, T$, $\sqrt{n}(\hat{S}_t - S_t) \Rightarrow \mathcal{N}(0, \sigma_t^2)$ and $\sqrt{n}(\hat{s}_t - s_t) \Rightarrow \mathcal{N}(0, \rho_t^2)$ as $n \to \infty$, where*

$$\sigma_t^2 = \frac{\mathbb{E}_t[\dot{g}_t(S_t, D_t)]^2}{[\mathbb{E}\ddot{g}_t(S_t, D_t)]^2}, \tag{8}$$

and

$$\rho_t^2 = \frac{\mathbb{E}[g_t(s_t, D_t) - g_t(S_t, D_t) - K]^2}{[\mathbb{E}_t \dot{g}_t(s_t, D_t)]^2}, \tag{9}$$

where $g_t(\cdot,\cdot)$ is defined in (6) and $\dot{g}_t(\cdot,\cdot)$, $\ddot{g}_t(\cdot,\cdot)$ are its first and second derivatives in the first argument.

Remark. Theorem 4 also applies to Scenario (c), the fully censored data case under Condition (c1).

Observe that σ_t^2 depends on the second moment of $\dot{g}_t(\cdot)$, i.e., depends on the variability of the critical equation for S_t. Likewise, ρ_t^2 depends directly on the variability of $[g_t(s_t, D_t) - g_t(S_t, D_t) - K]$. This is intuitive — the more variable the critical equation for S_t or s_t, the more variable the estimates $\hat{S}_t$ and $\hat{s}_t$. Further, both σ_t^2 and ρ_t^2 are normalized by the squared expected gradients of the respective critical equations. This is also intuitive because zeros of equations are easier to find the larger the gradients of the equations at zero, hence the inverse relationship between the squared expected gradients of the critical equations and the asymptotic variances.

Theorem 4 is a statement about the asymptotic normality of estimated quantities $\hat{s}_t$ and $\hat{S}_t$ around their true respective quantities s_t and S_t. As with Theorem 1, we use the theory of M-estimators to the dynamic inventory problem through an inductive argument to prove Theorem 4.

4.2. Confidence intervals with censored data ($1 \le n_0 \le n$)

THEOREM 5 **(Asymptotic normality of $(\tilde{S}_t, \tilde{s}_t)_{t=1}^T$).** *Assume Scenario (b). Then for all $t = 1, \dots, T$, $\sqrt{n}(\tilde{S}_t - S_t) \Rightarrow \mathcal{N}(0, \tilde{\sigma}_t^2)$, and $\sqrt{n}(\tilde{s}_t - s_t) \Rightarrow \mathcal{N}(0, \tilde{\rho}_t^2)$, where*

$$\tilde{\sigma}_t^2 = \sigma_t^2 + r\sigma_t^c \quad \textit{and} \quad \tilde{\rho}_t^2 = \rho_t^2 + r\rho_t^c,$$

where σ_t^2 and ρ_t^2 are as in (8)–(9), the variances of $\hat{S}_t$ and $\hat{s}_t$ that correspond to Scenario (a),

$$\sigma_t^c = \frac{\left\{ \frac{1+p_t}{p_t} [\mathbb{E}\dot{g}_t(S_t, D_t)\mathbb{I}(D_t \ge x_t)]^2 + \frac{r}{1-r} Var\left(\dot{g}_t(S_t, D_t)\mathbb{I}(D_t \ge x_t)\right) \right\}}{[\mathbb{E}\ddot{g}_t(S_t, D_t)]^2} \quad \textit{and} \tag{10}$$

$$\begin{aligned}\rho_t^c = \Bigg\{ &\frac{1+p_t}{p_t} \left[\mathbb{E}(g_t(s_t, D_t) - g_t(S_t, D_t) - K)\mathbb{I}(D_t \ge x_t)\right]^2 + \\ &\frac{r}{1-r} Var\left((g_t(s_t, D_t) - g_t(S_t, D_t) - K)\mathbb{I}(D_t \ge x_t)\right) \Bigg\} \Big/ [\mathbb{E}\dot{g}_t(s_t, D_t)]^2, \end{aligned} \tag{11}$$

where $p_t = \mathbb{P}(D_t \ge x_t)$ and $g_t(\cdot,\cdot)$ is defined in (6).

Remark. Theorem 5 also applies to Scenario (c), the fully censored data case under Condition (c2) by setting $x_t = \bar{x}_t$.

Thus, the asymptotic variances of $(\tilde{S}_t, \tilde{s}_t)_{t=1}^T$ for the censored demand Scenario (b) are equal to the variances of $(\hat{S}_t, \hat{s}_t)_{t=1}^T$ plus $r\sigma_t^c/n$ and $r\rho_t^c/n$, respectively. As a sanity check, observe that we retrieve the uncensored results when r, the proportion of the data that is censored, is set to zero.

We also retrieve the uncensored case results when $x_t = \bar{D}$, the upper bound on the random demand. This can be seen by taking the limit $x_t \to \bar{D}$ in Eqs. (10)–(11); the first terms are zero by taking the limit using L'Hôpital's rule and the second terms of the equations are trivially zero because $\mathbb{I}(D_t \geq \bar{D}) = 0$ (recall we assume D_t is continuous). Theorem 5 also shows that the confidence intervals on the estimates with censoring scale at the same rate $n^{-1/2}$ as in the uncensored case.

4.3. Estimating the Asymptotic Confidence Interval Formulas

In practice, the quantities $\sigma_t^2, \rho_t^2, \sigma_t^c$ and ρ_t^c need to be estimated with data. One consistent method of estimation is to estimate them via sample averages. To illustrate, one can estimate σ_t^2 and ρ_t^2, $t = 1, \ldots, T$, with:

$$\hat{\sigma}_t^2 = \frac{n^{-1}\sum_{i=1}^n [\dot{\hat{g}}_t(\hat{S}_t, d_t^i)]^2}{[n^{-1}\sum_{i=1}^n \ddot{\hat{g}}_t(\hat{S}_t, d_t^i)]^2}, \tag{12}$$

$$\hat{\rho}_t^2 = \frac{n^{-1}\sum_{i=1}^n [\hat{g}_t(\hat{s}_t, d_t^i) - \hat{g}_t(\hat{S}_t, d_t^i) - K]^2}{[n^{-1}\sum_{i=1}^n \dot{\hat{g}}_t(\hat{s}_t, d_t^i)]^2}, \tag{13}$$

where $\hat{g}_T(\cdot) = g_T(\cdot, \cdot)$, and

$$\begin{aligned}\hat{g}_t(y,d) &= (1-\alpha_t)cy + b_t(d-y)^+ + h_t(y-d)^+ + \alpha_t cd \\ &\quad + \alpha_t \hat{G}_{t+1}(s_{t+1})\mathbb{I}(y - s_{t+1} \leq d) + \alpha_t \hat{G}_{t+1}(y-d)\mathbb{I}(y - s_{t+1} > d),\end{aligned}$$

where $\hat{G}_{t+1}(\cdot) = n^{-1}\sum_{i=1}^n \hat{g}_{t+1}(\cdot, d_{t+1}^i)$. The estimators $\hat{\rho}_t^2$ and $\hat{\sigma}_t^2$ are consistent by the Strong Law of Large Numbers on the respective numerator and denominator, combined with Slutsky's lemma. Similar consistent, sample average estimators for σ_t^c and ρ_t^c can be written down.

Note the above estimators are computable even under the fully censored Scenario (c) by arguments similar to those outlined in Sec. 3.5. Under the Conditions (c1) and (c2), both the numerators and denominators of $\hat{\rho}_t^2$ and $\hat{\sigma}_t^2$ are identifiable, with the estimates remaining unchanged when d_t^i are replaced by the censored counterpart z_t^i, for all $i=1,\ldots,n$ and $t=1,\ldots,T$. This is also the case for the sample average estimators for σ_t^c and ρ_t^c.

Finally, we remark that in practice, when the DM quotes confidence intervals around the estimated optimal inventory policy, the confidence intervals are themselves subject to estimation errors. We quantify such errors numerically in Sec. 6.

5. Implications on the Total Cost

In this section, we show how the errors associated with estimating the optimal policy translate to errors in estimating the total cost.

Given a vector $\mathbf{q}=[q_1,\ldots,q_T]$ of ordering quantities, the corresponding expected total cost is given by

$$V_1(I_1;\mathbf{q})=\sum_{t=1}^{T}(\Pi_{\tau=1}^{t-1}\alpha_\tau)\mathbb{E}[K\mathbb{I}(I_t<s_t)+c_tq_t+C_t(q_t+I_t,D_t)],$$

where I_1 is the available inventory at the beginning of period 1 and $I_{t+1}=I_t+q_t-D_t$, $1\le t\le T-1$. Denote the optimal ordering quantities by $\mathbf{q}^*=[q_1^*,\ldots,q_T^*]$ and the corresponding inventory levels by $I_1^*,\ldots,I_T^*$. Then $q_t^*=(S_t-I_t^*)\mathbb{I}(I_t^*<s_t)$ for $1\le t\le T$ and $I_{t+1}^*=I_t^*+q_t^*-D_t$, $1\le t\le T-1$, with $I_1^*=I_1$.

We now show that estimated ordering quantities are consistent if the corresponding order-up-to levels are consistent.

Proposition 1. *Suppose $(S_t',s_t')_{t=1}^T$ is a consistent estimate of the optimal $(S_t,s_t)_{t=1}^T$ policy. Denote the corresponding ordering quantities by $\mathbf{q}'=[q_1',\ldots,q_T']$, where $q_t'=(S_t'-I_t')\mathbb{I}(I_t'<s_t')$, and where $I_{t+1}'=I_t'+q_t'-D_t$, $1\le t\le T$, with $I_1'=I_1$ a known constant. Then for each $t=1,\ldots,T$, $q_t'\xrightarrow{P}q_t^*$ as $n\to\infty$.*

We can also conclude that the expected total cost of a consistent estimated policy is also consistent, which follows from applying the Continuous Mapping Theorem (CMT) to Proposition 1:

Corollary 1. *Let $\mathbf{q}'$ and I_1' be as in Proposition 1. Then $V_1(I_1';\mathbf{q}')\xrightarrow{P}V_1(I_1;\mathbf{q}^*)$ as $n\to\infty$.*

Proposition 1 and Corollary 1 ensure that as the DM collects more data, she is able to order quantities that are closer to the optimal ordering quantities (in a probabilistic sense), and that the corresponding expected total cost of the estimated ordering quantities also converge to that of the optimal ordering quantities.

In practice, however, the DM only has access to a finite sample of data. As such, her estimated inventory policy would not coincide exactly with the optimal policy. Suppose that the DM estimates the optimal order-up-to-level S_t by S'_t, and estimates the optimal reorder point s_t by s'_t. The corresponding ordering quantities are given by $\mathbf{q}' = [q'_1, \dots, q'_T]$, where $q'_t = (S'_t - I'_t)\mathbb{I}(I'_t < s'_t)$, and the corresponding inventories are $I'_{t+1} = I'_t + q'_t - D_t$ for $1 \le t \le T$, with $I'_1 = I_1$.

Let us investigate the expected total cost of such an estimated inventory policy. Observe that an error with the critical level s_t affects the inventory policy only if the error is large enough that the DM either orders when it is optimal not to, or not orders when it is optimal to do so. Let $\mathcal{I} \subset \{1, \dots, T\}$ denote the time periods at which the DM makes such mistakes.

Consider first the event $\mathcal{A}_t := \{t \notin \mathcal{I}\} = \{\mathbb{I}(I^*_t < s_t) = \mathbb{I}(I'_t < s'_t)\}$. If we also have $\mathcal{B}_t = \{\mathbb{I}(I^*_t < s_t) = 0\}$, the DM does not order so $q'_t = q^*_t = 0$, and the contribution to the expected total cost difference at time t is given by

$$\mathbb{E}\left\{[C_t(I'_t, D_t) - C_t(I^*_t, D_t)]\mathbb{I}(\mathcal{A}_t \cap \mathcal{B}_t)\right\}.$$

If on the other hand we are in the event $\mathcal{A}_t \cap \mathcal{B}^c_t = \{\mathbb{I}(I^*_t < s_t) = \mathbb{I}(I'_t < s'_t) = 1\}$, the DM does order and the inaccuracy in the order quantity results in an inaccuracy in the expected total cost by the amount

$$\begin{aligned}
&\mathbb{E}\left\{[c_t(q'_t - q^*_t) + C_t(I'_t + q'_t, D_t) - C_t(I^*_t + q^*_t, D_t)]\mathbb{I}\left(\mathcal{A}_t \cap \mathcal{B}^c_t\right)\right\} \\
&\quad = \mathbb{E}\left\{[c_t(S'_t - S_t) - c_t(I'_t - I^*_t) + C_t(S'_t, D_t) - C_t(S_t, D_t)]\mathbb{I}\left(\mathcal{A}_t \cap \mathcal{B}^c_t\right)\right\}.
\end{aligned}$$

Now consider the event $\mathcal{A}^c_t = \{t \in \mathcal{I}\} = \{\mathbb{I}(I^*_t < s_t) \neq \mathbb{I}(I'_t < s'_t)\}$. The DM either orders when it is optimal not to, or does not order when it is optimal to do so. In the former case, the expected cost difference from the optimal policy at time t is

$$\mathbb{E}\left\{[K + c_t(S'_t - I'_t) + C_t(S'_t, D_t) - C_t(I^*_t, D_t)]\mathbb{I}(\mathcal{A}^c_t \cap \mathcal{B}_t)\right\},$$

and in the latter case, the cost difference is

$$\mathbb{E}\left\{[-K - c_t(S_t - I^*_t) + C_t(I'_t, D_t) - C_t(S_t, D_t)]\mathbb{I}(\mathcal{A}^c_t \cap \mathcal{B}^c_t)\right\}.$$

Combining both $t \notin \mathcal{I}$ and $t \in \mathcal{I}$ cases, we arrive at the following result, which holds for any estimated policy $(S'_t, s'_t)_{t=1}^T$.

THEOREM 6. *The difference between the expected total cost of an ordering policy $(S'_t, s'_t)_{t=1}^T$ and the expected total cost of the optimal policy $(S_t, s_t)_{t=1}^T$ can be bounded according to:*

$$
\begin{aligned}
&|V_1(I_1;\mathbf{q}') - V_1(I_1;\mathbf{q}^*)| \\
&\le \sum_{t=1}^{T} (\Pi_{\tau=1}^{t-1}\alpha_\tau) \Bigg\{ \mathbb{E}\Bigg[(c_t + (b_t \vee h_t))\Lambda_t^1 |S'_t - S_t| + \sum_{k=1}^{t-1} [c_t\Lambda_t^1 + (b_t\vee h_t)(1-\Lambda_t^1)]\Gamma_{k,t-1}|S'_k - S_k| \Bigg] \\
&+\mathbb{E}\Bigg[[K + (c_t + (b_t\vee h_t))|S_t - I_t^*|]\Lambda_t^2 + \sum_{k=1}^{t-1}(1+\Gamma_{k,t-2})[c_t\Lambda_t^1 + (b_t\vee h_t)(1-\Lambda_t^1)]\Lambda_k^2|S_k - I_k^*| \Bigg] \Bigg\},
\end{aligned}
\tag{14}
$$

where

$$
\begin{aligned}
\Lambda_t^1 &= \mathbb{I}(\mathcal{A}_t \cap \mathcal{B}_t^c) + \mathbb{I}(\mathcal{A}_t^c \cap \mathcal{B}_t), \textit{ and} \\
\Lambda_t^2 &= \mathbb{I}(\mathcal{A}_t^c)(\mathbb{I}(\mathcal{B}_t) - \mathbb{I}(\mathcal{B}_t^c)).
\end{aligned}
$$

The only simplifications used to derive the upper bound in (14) are the triangle inequality and the Lipschitz property of the single-period cost function $C_t(y,\cdot)$. Inspecting inside the large parentheses in (14), we see that the first expectation contains a linear combination of $|S'_\tau - S_\tau|$, $\tau = 1,\dots,t$, which we can directly relate to a confidence interval on S'_τ. The second expectation contains terms that can be bounded by a constant times $\mathbb{P}(\mathcal{A}_t^c)$, which is the chance that the estimated policy makes mistakes in the ordering epochs. For a sensible policy such as $(\hat{S}_t, \hat{s}_t)_{t=1}^T$ or $(\tilde{S}_t, \tilde{s}_t)_{t=1}^T)$ which converges to the truth asymptotically normally, we intuit that the chance of making such mistakes would be negligible compared to the first term. We investigate these insights numerically in Sec. 6.

6. Computational results

In this section, we validate the consistency and asymptotic normality results of Secs. 3–4 on simulated data, and investigate the tightness the bound on the expected total cost stated in Theorem 6.

We estimate the optimal order-up-to levels, $\hat{S}_t$ and $\tilde{S}_t$, $t = 1,\dots,T$, which are minimizers of continuous functions over a bounded domain, by a grid-search, and the reorder points $\hat{s}_t$ and $\tilde{s}_t$, $t = 1,\dots,T$, which are zeros of given equations, by the secant method. While this approach can be computationally expensive, we can compute $(\hat{S}_t, \hat{s}_t)_{t=1}^T$ and $(\tilde{S}_t, \tilde{s}_t)_{t=1}^T$ to an arbitrary accuracy this way.

Faster algorithms are known for special cases of our problem, e.g. if the discount factors α_t are zero, the problem decouples into separate newsvendor problems, which are single-period convex optimization problems; or if the setup cost K is zero, the problem is a convex stochastic dynamic program, which can be solved efficiently using the stochastic approximation method, as shown by Kunnumkal and Topaloglu (2008). For brevity and focus, we leave open the question of finding efficient algorithms for the general estimation problem, noting that randomized decision rules of Levi and Shi (2013), heuristic-based approaches for similar problems (Bollapragada and Morton 1999, Cheung and Simchi-Levi 2019) and fast algorithms for the infinite horizon case (Federgruen and Zipkin 1984, Zheng and Federgruen 1991) could provide useful starting points.

For the following computational results, we consider a 3-period inventory problem ($T = 3$), where the demand in each period is independent from each other and have truncated normal distributions. Specifically, $[\underline{D}, \bar{D}] = [30, 100]$, $D_1 \sim \mathcal{N}(75, 20^2) \cap [30, 100]$, $D_1 \sim \mathcal{N}(70, 30^2) \cap [30, 100]$, $D_2 \sim \mathcal{N}(55, 20^2) \cap [30, 100]$. For simplicity, the discount factor, unit ordering, backordering and holding costs are the same for all periods; i.e., $\alpha_t = 0.1$, $c_t = 0.2$, $b_t = 1$ and $h_t = 0.5$ for $t = 1, 2, 3$. The setup cost is $K = 2$.

To test the asymptotic theory, we simulate 100 iid demand data sets $\{d_1^i, d_2^i, d_3^i\}_{i=1}^n$ of size n, where we consider $n = 50, 100, 200$. For each data set, we compute the estimates $\{(\hat{S}_1, \hat{s}_1), (\hat{S}_2, \hat{s}_2), (\hat{S}_3, \hat{s}_3)\}$ and $\{(\tilde{S}_1, \tilde{s}_1), (\tilde{S}_2, \tilde{s}_2), (\tilde{S}_3, \tilde{s}_3)\}$ according to Secs. 3.1-3.2. We take the average of 50 simulations for a dataset with $n = 500$ as the numerically converged optimal values $\{(S_1, s_1), (S_2, s_2), (S_3, s_3)\}$.

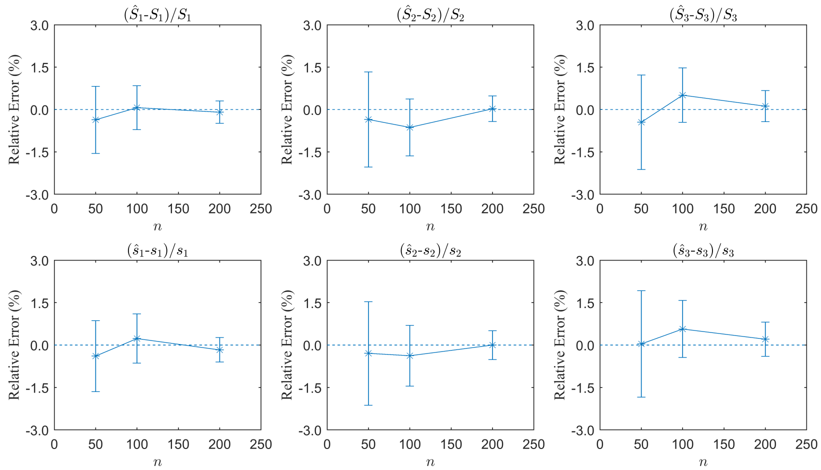
6.1. Validation of Consistency

Figure 2a shows the convergence of the estimated levels $(\hat{S}_t, \hat{s}_t)$ to (S_t, s_t), $t = 1, 2, 3$ as the number of observations n grows. Displayed are normalized values, i.e., $(\hat{S}_t - S_t)/S_t$ and $(\hat{s}_t - s_t)/s_t$, $t = 1, 2, 3$, with error bars indicating 95% confidence intervals. We observe that the estimates $(\hat{S}_t, \hat{s}_t)$, $t = 1, 2, 3$ are accurate even for $n = 50$, with the 95% error bars falling within $\pm 3\%$ of the converged optimal levels (S_t, s_t).

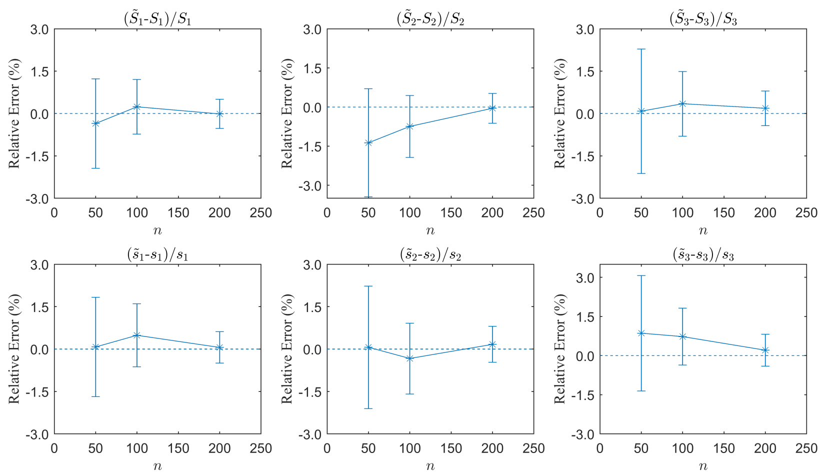
Figure 2b shows the convergence of the estimated levels $(\tilde{S}_t, \tilde{s}_t)$ to (S_t, s_t), $t = 1, 2, 3$ for the censored data case, where $r = 0.5$ and $\mathbf{x} = [50, 50, 50]$, as the number of observations n grows. That is, we consider the case where 50% of the time in the past, the stock level was capped at 50 units, and the other 50% of the time there was no demand censoring. Here we also observe that the accuracy of $(\tilde{S}_t, \tilde{s}_t)$ is high even for $n = 50$.

6.2. Validation of Asymptotic Normality

Figure 3 shows the histograms for the uncensored data estimates $(\hat{S}_t, \hat{s}_t)$, $t = 1, 2, 3$ for $n = 50, 100, 200$. Superimposed on the histograms are the theoretical asymptotic normal



(a)



(b)

Figure 2: Convergence of estimated (S_t, s_t), $t = 1,2,3$ levels for (a) the uncensored data case and (b) the censored data case where $r = 0.5, \mathbf{x} = [50,50,50]$. The estimates are de-meaned and normalized by the respective converged values, so the vertical axes correspond to relative errors. The error bars correspond to 95% confidence intervals of the estimations.

distributions from Theorem 4. Visually, the histograms fit the theoretical normal distributions very well, with increasing accuracy as n increases. This observation is also supported by the one-sample Kolmogorov-Smirnov test, which do not reject the null hypothesis that the histograms are indistinguishable from the asymptotic normal distributions at the 1% significance level.

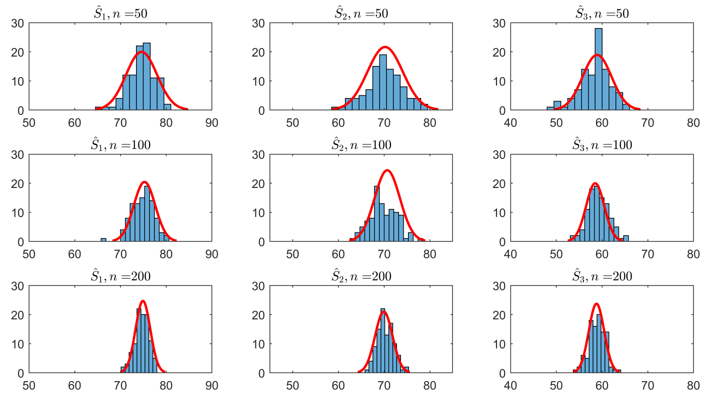
Figure 4 shows the histograms for the censored estimates $(\tilde{S}_t, \tilde{s}_t)$, $t = 1,2,3$ for $n = 50,100,200$, $r = 0.5$ and $\mathbf{x} = [50,50,50]$. Superimposed on the histograms are the theoretical asymptotic normal distributions from Theorem 5. Again, the histograms fit the theoretical normal distributions very well, with increasing accuracy as n increases, and the one-sample Kolmogorov-Smirnov test does not reject the null hypothesis that the histograms are indistinguishable from the asymptotic normal distributions at the 1% significance level.

In practice, the DM would not know $\sigma_t, \rho_t, \tilde{\sigma}_t$ and $\tilde{\rho}_t$, $t=1,2,3$ so would need to estimate these values with the data she has, for instance by using the estimators in Eqs. (12)–(13). To explore the additional errors associated with estimating the asymptotic variances, we plot in Figure 5 the relative estimation errors for $n=50,100,200$ with error bars to indicate 95% confidence intervals. We observe that the error bars for σ_t and $\tilde{\sigma}_t$ are within $\pm 3\%$ of the converged value, and for ρ_t and $\tilde{\rho}_t$ are within $\pm 8\%$ of the converged value for $n=50,100,200$. Thus, estimating σ_t is associated with smaller errors than estimating ρ_t; this we believe is due to the fact that estimating σ_t has one less source of error. Scanning Eqs. (8)–(9), one can see that ρ_t depends on both S_t and s_t, which are sources of estimation errors, whereas σ_t depends on S_t but not on s_t.

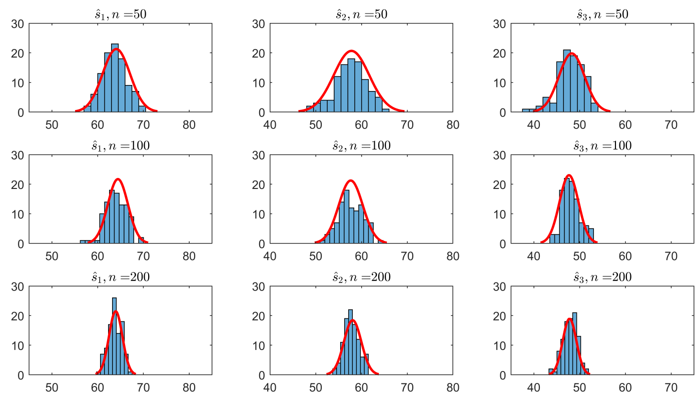
To gauge the relative importance of the estimation errors, we perform the one-sample Kolmogorov-Smirnov test for the equality of the theoretical normal distributions of Theorems 4–5 with estimated asymptotic variances and the empirical histograms of Figs. 3–4. We find that all tests still do not reject the null hypothesis that the empirical histograms are indistinguishable from the asymptotic normal distributions with estimated variances at the 1% significance level. This means that while there are errors in estimating $\sigma_t, \rho_t, \tilde{\sigma}_t$ and $\tilde{\rho}_t$, $t=1,2,3$, the errors are not so large as to invalidate the normal approximation of $(\hat{S}_t, \hat{s}_t)_{t=1}^T$ and $(\tilde{S}_t, \tilde{s}_t)_{t=1}^T$ using estimated values of $\sigma_t, \rho_t, \tilde{\sigma}_t$ and $\tilde{\rho}_t$, $t=1,2,3$.

6.3. Validation of Total Cost Analysis

In this subsection, we investigate the out-of-sample total cost of the estimated policies. In Table 1, we display (i) the expected total cost of the estimated policies versus the expected total cost of the optimal policy and (ii) 95th percentile of the differences between the total costs of 100 iid estimated policies from the converged optimal, versus the theoretical upper bound on the expected total cost difference from Theorem 6. The table displays results for (a) uncensored data case and (b) a censored data case where $r=0.5$ and $\mathbf{x}=[50,50,50]$. For the theoretical upper bounds, we set $S_t' - S_t = \Delta_t$, where $\Delta_t = 1.96\sigma_t/\sqrt{n}$ for the uncensored case and $\Delta_t = 1.96\tilde{\sigma}_t/\sqrt{n}$ for the censored case, so that 95% of the estimates $\hat{S}_t$ will fall within this range. As intuited in Sec. 5, we find that the second term in (14) is negligible



(a)

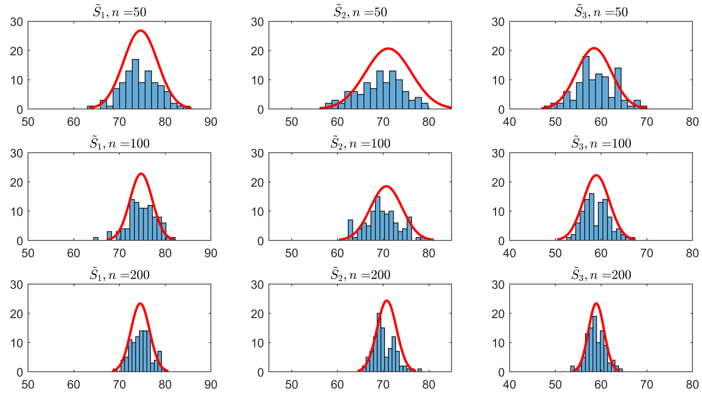


(b)

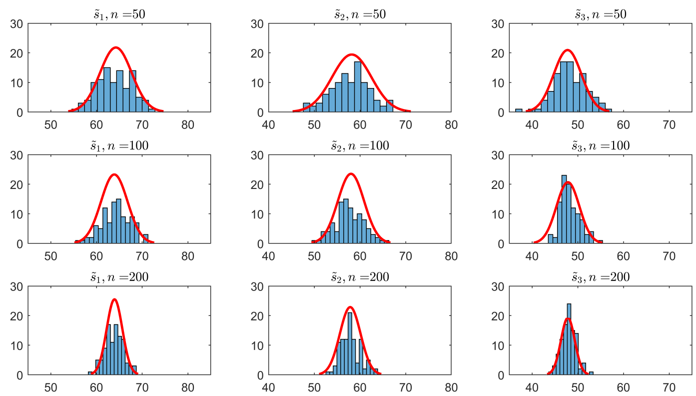
Figure 3: Histograms of (a) $\hat{S}_t$ and (b) $\hat{s}_t$, $t=1,2,3$ for $n=50,100,200$ from 100 simulations. The superimposed normal curves correspond to the asymptotic normal distribution from Theorem 4.

compared to the first term, as it can be bounded by $\mathbb{P}(\mathcal{A}_t^c)$ which is zero to two decimal places.

We make two important observations. The first point of interest is that the expected total costs of the estimated policies, in both uncensored and censored data scenarios, are remarkably close to those of the true optimal policy. The differences in the expected total cost are just 0.6% and 1.1% respectively for $(\hat{S}_t, \hat{s}_t)_{t=1}^T$ and $(\tilde{S}_t, \tilde{s}_t)_{t=1}^T$ for $n=50$, and even less for $n=100$ and 200. This is very promising, as it shows that the estimated policies are near-optimal in terms of the expected total cost, which is arguably the most important metric.



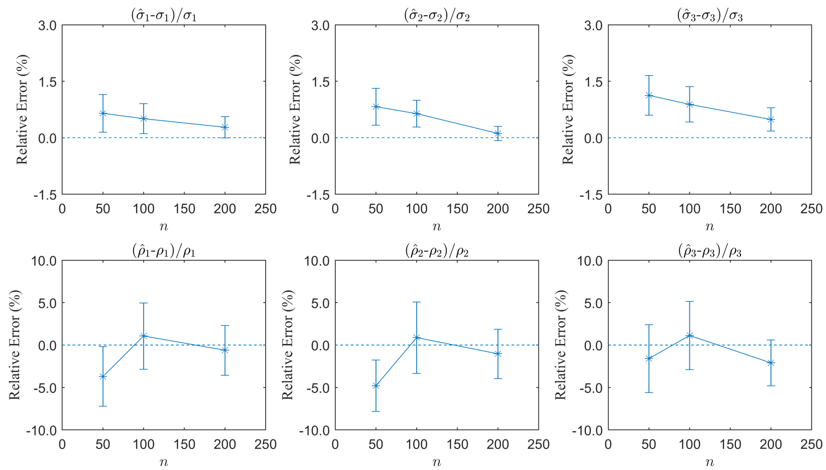
(a)



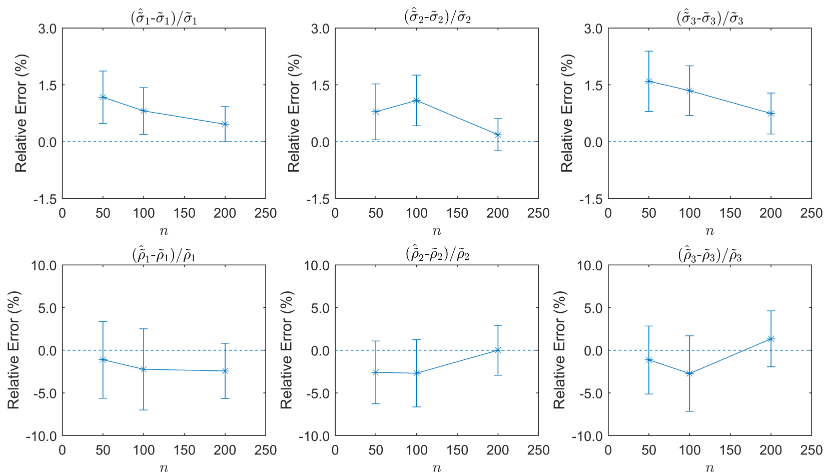
(b)

Figure 4: Histograms of (a) $\tilde{S}_t$ and (b) $\tilde{s}_t$, $t = 1, 2, 3$ where $r = 0.5, \mathbf{x} = [50, 50, 50]$ for $n = 50, 100, 200$ from 100 simulations. The superimposed normal curves correspond to the asymptotic normal distribution from Theorem 5.

The second point of interest is that theoretical upper bounds from (14) are effective upper bounds on the the empirical 95th percentile of the total cost differences. The theoretical upper bounds are conservative, with the empirical percentiles being 46–58% of the theoretical upper bounds, but this is expected as setting $S_t' - S_t = \Delta_t$ for all $t = 1, 2, 3$ in (14) is equivalent to taking the supremum over $S_t' \in [S_t - \Delta_t, S_t + \Delta_t]$, $t = 1, 2, 3$. The numerical result here suggests that taking Δ_t to just one asymptotic standard deviation, not two, is sufficient to capture the 95th percentile of the total cost differences. From a practical perspective, we can conclude that the theoretical upper bounds on the expected total costs provide effective upper bounds and provide complimentary information to just the expected value.



(a)



(b)

Figure 5: Convergence of (a) estimated (σ_t, ρ_t), $t = 1, 2, 3$ and (b) estimated $(\tilde{\sigma}_t, \tilde{\rho}_t)$, $t = 1, 2, 3$ where $r = 0.5, \mathbf{x} = [50, 50, 50]$. The estimates are demeaned and normalized by the respective converged values, so the vertical axes correspond to relative errors. The error bars correspond to 95% confidence intervals of the estimations.

In Figure 6, we plot histograms of the normalized absolute differences in the expected total cost of the 100 estimated policies and the converged optimal policy for (a) uncensored and (b) censored data cases where $r = 0.5$, $\mathbf{x} = [50, 50, 50]$. (Note: The normalization is by the total cost of the converged optimal policy, and the expectation is computed by taking the average of the total costs over 500 new, out-of-sample iid demand paths.) We observe that estimated total costs are very accurate, as can be seen by the heavy weights at zero for all histograms in Figure 6, as well as the limited total range. Even for $n = 50$, most total costs of the estimated policies fall within 10% of the total cost of the converged optimal policy, with no cases off by more than 25%. Nevertheless, there is a non-negligible distribution past zero, and this justifies the need for upper bounds on the expected total cost (recall that in practice, the DM would not know the distributions in Figure 6, but can compute the theoretical upper bounds using in-sample data).

(a) Uncensored data results

n	Exp. total cost (est. policy)	Exp. total cost (opt. policy)	% diff.	95th perc. total cost diff. (Diff)	Theoretical Upper Bound (UB)	Diff/UB
50	31.04	30.86	0.6%	3.91	8.52	45.9%
100	30.98	30.86	0.4%	2.98	6.02	49.5%
200	30.91	30.86	0.2%	2.05	4.26	48.1%

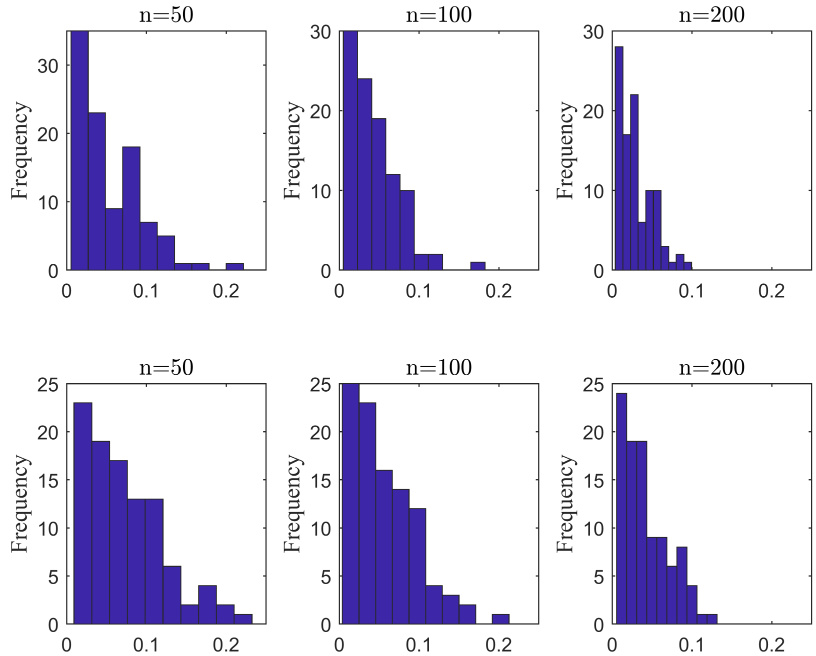
(d) Censored data results ($r = 0.5, \mathbf{x} = [50, 50, 50]$)

n	Exp. total cost (est. policy)	Exp. total cost (opt. policy)	% diff.	95th perc. total cost diff. (Diff)	Theoretical Upper Bound (UB)	Diff/UB
50	31.20	30.86	1.1%	5.46	10.30	53.0%
100	31.05	30.86	0.6%	4.29	7.28	58.9%
200	30.95	30.86	0.3%	2.97	5.15	57.7%

Table 1: Left: Expected total cost of the estimated policies versus the expected total cost of the optimal policy. Right: 95th percentile of the total cost difference of 100 independent estimated policies versus the theoretical upper bound on the expected total cost from Theorem 6, using 95th percentile on estimated S_t, $t = 1, \ldots, T$. See Sec. 6.3 for details on the computations.

7. Conclusion

Confidence intervals provide important information for decision-makers, but has been largely ignored in the operations management literature. In this paper, we address this gap by investigating both finite-sample and asymptotic behaviors of data-driven dynamic inventory policies, for both uncensored and censored demand data settings. We first show that appropriate estimation procedures yield consistent estimators of the optimal $(S_t, s_t)_{t=1}^T$ policy. We then explore the finite-sample precision of the estimated policies by using CLT analysis for M-estimators. We further derived an upper bound on the expected total cost



(a)

(b)

Figure 6: Histograms of the 100 absolute relative differences in the total cost of the estimated policy from the optimal policy, computed by averaging over 500 iid out-of-sample demand paths for (a) uncensored data case and (b) censored data case, where $r = 0.5, \mathbf{x} = [50, 50, 50]$.

of an estimated policy, which can use the asymptotic confidence intervals as inputs. All our theoretical results are numerically validated on simulated data.

A key direction for follow-up work is to relax the assumption that the demands are independently distributed across time. This is clearly not the case in real-life as past sales influence future sales through network effects. To analyze this setting, one would need to combine CLT results for non-iid data with M-estimation.

Another direction for follow-up work is in developing efficient algorithms to solve the estimation problems in full generality. As mentioned in the main text, we believe the algorithms in Levi and Shi (2013), Bollapragada and Morton (1999), Cheung and Simchi-Levi (2019), Federgruen and Zipkin (1984), Zheng and Federgruen (1991) could provide useful starting points.

Finally, we mention that implementing decision-support tools that show confidence intervals in practice, and evaluating the value of this extra information in real-life decison-making would form an important study that bridges the gap between theory and practice.

Acknowledgments

The author is grateful to the Associate Editor and two anonymous referees who have given much constructive feedback.

References

Agrawal N, Smith SA (1996) Estimating negative binomial demand for retail inventory management with unobservable lost sales. *Naval Research Logistics (NRL)* 43(6):839–861.

Anupindi R, Dada M, Gupta S (1998) Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science* 17(4):406–423.

Arlotto A, Steele JM (2016) A central limit theorem for temporally nonhomogenous markov chains with applications to dynamic programming. *Mathematics of Operations Research* 41(4):1448–1468.

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

Ban GY, Rudin C (2019) The big data newsvendor: Practical insights from machine learning. *Operations Research* 67(1):90–108.

Bartlett P (2013) Lecture notes on theoretical statistics URL www.stat.berkeley.edu/~bartlett/courses/2013spring-stat210b/.

Bertsekas DP (1995) *Dynamic programming and optimal control*, volume 1 (Athena Scientific Belmont, MA).

Besbes O, Muharremoglu A (2013) On implications of demand censoring in the newsvendor problem. *Management Science* 59(6):1407–1424.

Bollapragada S, Morton TE (1999) A simple heuristic for computing nonstationary (s, s) policies. *Operations Research* 47(4):576–584.

Burnetas AN, Smith CE (2000) Adaptive ordering and pricing for perishable products. *Operations Research* 48(3):436–443.

Chatterjee S (2014) A short survey of stein's method. *Proceedings of ICM* IV:1–24.

Chen L, Mersereau AJ (2015) Analytics for operational visibility in the retail store: The cases of censored demand and inventory record inaccuracy. *Retail Supply Chain Management*, 79–112 (Springer).

Chen L, Mersereau AJ, Wang Z (2017) Optimal merchandise testing with limited inventory. *Operations Research* 65(4):968–991.

Cheung WC, Simchi-Levi D (2019) Sampling-based approximation schemes for capacitated stochastic inventory control models. *Mathematics of Operations Research* Articles in Advance.

Conrad S (1976) Sales data and the estimation of demand. *Journal of the Operational Research Society* 27(1):123–127.

Dai T, Jerath K (2013) Salesforce compensation with inventory considerations. *Management Science* 59(11):2490–2501.

Federgruen A, Zipkin P (1984) An efficient algorithm for computing optimal (s, s) policies. *Operations research* 32(6):1268–1285.

Godfrey GA, Powell WB (2001) An adaptive, distribution-free algorithm for the newsvendor problem with censored demands, with applications to inventory and distribution. *Management Science* 47(8):1101–1112.

Heese HS, Swaminathan JM (2010) Inventory and sales effort management under unobservable lost sales. *European Journal of Operational Research* 207(3):1263–1268.

Huh WT, Levi R, Rusmevichientong P, Orlin JB (2011) Adaptive data-driven inventory control with censored demand based on kaplan-meier estimator. *Operations Research* 59(4):929–941.

Huh WT, Rusmevichientong P (2009) A nonparametric asymptotic analysis of inventory planning with censored demand. *Mathematics of Operations Research* 34(1):103–123.

Iglehart DL (1963) Optimality of (s, s) policies in the infinite horizon dynamic inventory problem. *Management Science* 9(2):259–267.

Jain A, Rudi N, Wang T (2014) Demand estimation and ordering under censoring: Stock-out timing is (almost) all you need. *Operations Research* 63(1):134–150.

Kunnumkal S, Topaloglu H (2008) Using stochastic approximation methods to compute optimal base-stock levels in inventory control problems. *Operations Research* 56(3):646–664.

Levi R, Perakis G, Uichanco J (2015) The data-driven newsvendor problem: new bounds and insights. *Operations Research* 63(6):1294–1306.

Levi R, Roundy RO, Shmoys DB (2007) Provably near-optimal sampling-based policies for stochastic inventory control models. *Mathematics of Operations Research* 32(4):821–839.

Levi R, Shi C (2013) Approximation algorithms for the stochastic lot-sizing problem with order lead times. *Operations Research* 61(3):593–602.

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

Nahmias S (1994) Demand estimation in lost sales inventory systems. *Naval Research Logistics* 41(6):739–758.

Porteus EL (1971) On the optimality of generalized $(s,\ s)$ policies. *Management Science* 17(7):411–426.

Porteus EL (2002) *Foundations of stochastic inventory theory* (Stanford University Press).

Powell W, Ruszczyński A, Topaloglu H (2004) Learning algorithms for separable approximations of discrete stochastic optimization problems. *Mathematics of Operations Research* 29(4):814–836.

Scarf H (1959a) Bayes solutions of the statistical inventory problem. *The Annals of Mathematical Statistics* 490–508.

Scarf H (1959b) *The optimality of (s, S) policies in the dynamic inventory problem. Mathematical Methods in the Social Science, KJ Arrow, S. Karlin, P. Suppes, eds.* (Stanford University Press, Stanford).

Sethi SP, Cheng F (1997) Optimality of (s, s) policies in inventory models with markovian demand. *Operations Research* 45(6):931–939.

Shapiro A, Dentcheva D, Ruszczyński AP (2009) *Lectures on stochastic programming: modeling and theory*, volume 9 (SIAM).

Shi C, Chen W, Duenyas I (2016) Nonparametric data-driven algorithms for multiproduct inventory systems with censored demand. *Operations Research* 64(2):362–370.

Stein C (1972) A bound for the error in the normal approximation to the distribution of a sum of dependent random variables. *Proceedings of the Fifth Berkeley symposium on mathematical statistics and probability*, volume 2, 583–602.

Van der Vaart AW (2000) *Asymptotic statistics*, volume 3 (Cambridge University press).

Vulcano G, Van Ryzin G, Ratliff R (2012) Estimating primary demand for substitutable products from sales transaction data. *Operations Research* 60(2):313–334.

Wecker WE (1978) Predicting demand from sales data in the presence of stockouts. *Management Science* 24(10):1043–1054.

Zhang H, Chao X, Shi C (2018) Perishable inventory systems: Convexity results for base-stock policies and learning algorithms under censored demand. *Operations Research* 66(5):1276–1286.

Zheng YS, Federgruen A (1991) Finding optimal (s, s) policies is about as simple as evaluating a single policy. *Operations Research* 39(4):654–665.

Zipkin PH (2000) *Foundations of inventory management* (McGraw-Hill New York).

Electronic Companion to "Confidence Intervals for Data-driven Inventory Policies with Demand Censoring"

Appendix A: Proofs of results in Section 3

A.1. Lemma EC.1

First, we need the following technical lemma on the uniform convergence of the $\hat{G}_t(\cdot)$ function.

LEMMA EC.1 **(Uniform convergence of $\hat{G}_t(\cdot)$).** *As $n \to \infty$, $\sup_{y \in [I_t, \bar{D}]} |\hat{G}_t(y) - G_t(y)| \xrightarrow{P} 0$ for all $t = 1, \ldots, T$.*

Proof. We prove the statement inductively, starting at $t = T$ and working backwards.

Step 1. For $t = T$, this is true because for all $y_1, y_2 \in [\underline{D}, \bar{D}]$ we have

$$|g_T(d; y_1) - g_T(d; y_2)| \le (c_t + (b \vee h))|y_1 - y_2|, \tag{EC.1}$$

i.e., the function $g_T(d; \cdot)$ is Lipschitz with an integral coefficient. One can then show that the class of functions $\mathcal{G}_T = \{g_T(\cdot; y) | y \in [\underline{D}, \bar{D}]\}$ has a finite bracketing number $N_{[\,]}(\varepsilon, \mathcal{G}_T, L_1(P))$ for every $\varepsilon > 0$. Then by Theorem 19.4 of Van der Vaart (2000), $\mathcal{G}_T$ is P_T-Glivenko-Cantelli, i.e., $\sup_{y \in [\underline{D}, \bar{D}]} |n^{-1} \sum_{i=1}^n g_T(d_T^i; y) - \mathbb{E}_T g_T(D_T; y)| \to 0$ almost surely.

Step 2. Assume $\sup_{y \in \mathcal{Y}} |\hat{G}_\tau(y) - G_\tau(y)| \xrightarrow{P} 0$ for some $1 \le \tau \le T$.

Step 3. For $1 \le t \le \tau - 1$, let us first define

$$\begin{aligned}\tilde{G}_t(y) &= (1 - \alpha_t) c_t y + \hat{\mathbb{E}}_t[C_t(y, D_t)] + \alpha_t c_t \hat{\mathbb{E}}_t[D_t] \\ &\quad + \alpha_t G_{t+1}(s_{t+1}) \hat{\mathbb{E}}_t[\mathbb{I}_t(y - s_{t+1})] + \alpha_t \hat{\mathbb{E}}_t[G_{t+1}(y - D_t) \mathbb{I}_t^c(y - s_{t+1})]\end{aligned} \tag{EC.2}$$

In other words, $\tilde{G}_t$ is the same as $\hat{G}_t$, except that any portion that depends on data in the future is replaced by the corresponding true value.

We have $|\hat{G}_t(y) - G_t(y)| \le \underbrace{|\hat{G}_t(y) - \tilde{G}_t(y)|}_{(A)} + \underbrace{|\tilde{G}_t(y) - G_t(y)|}_{(B)}$, and we proceed to bound the two terms on the RHS separately.

For term (A) we have the following:

$$\begin{aligned}|\hat{G}_t(y) - \tilde{G}_t(y)| &\le \alpha_t |\hat{G}_{t+1}(\hat{S}_{t+1}) \hat{\mathbb{E}}_t[\hat{\mathbb{I}}_t(y - \hat{s}_{t+1})] - G_{t+1}(S_{t+1}) \hat{\mathbb{E}}_t[\mathbb{I}_t(y - s_{t+1})]| && \text{(Aa)} \\ &\quad + \alpha_t K |\hat{\mathbb{E}}_t[\hat{\mathbb{I}}_t(y - \hat{s}_{t+1})] - \hat{\mathbb{E}}_t[\mathbb{I}_t(y - s_{t+1})]| && \text{(Ab)} \\ &\quad + \alpha_t |\hat{\mathbb{E}}_t[\hat{G}_{t+1}(y - D_t) \hat{\mathbb{I}}_t^c(y - \hat{s}_{t+1})] - \hat{\mathbb{E}}_t[G_{t+1}(y - D_t) \mathbb{I}_t^c(y - s_{t+1})]| && \text{(Ac)}\end{aligned}$$

We can bound (Aa) by:

$$|\hat{G}_{t+1}(\hat{S}_{t+1}) \hat{\mathbb{E}}_t[\hat{\mathbb{I}}_t(y - \hat{s}_{t+1})] - G_{t+1}(S_{t+1}) \hat{\mathbb{E}}_t[\mathbb{I}_t(y - s_{t+1})]|$$

$$\leq |\hat{G}_{t+1}(\hat{S}_{t+1})\hat{\mathbb{E}}_t[\hat{\mathbb{I}}_t(y-\hat{s}_{t+1})] - \hat{G}_{t+1}(\hat{S}_{t+1})\mathbb{E}_t[\hat{\mathbb{I}}_t(y-\hat{s}_{t+1})]| \tag{Aa-1}$$

$$+ |\hat{G}_{t+1}(\hat{S}_{t+1})\mathbb{E}_t[\hat{\mathbb{I}}_t(y-\hat{s}_{t+1})] - G_{t+1}(S_{t+1})\mathbb{E}_t[\hat{\mathbb{I}}_t(y-\hat{s}_{t+1})]| \tag{Aa-2}$$

$$+ |G_{t+1}(S_{t+1})\mathbb{E}_t[\hat{\mathbb{I}}_t(y-\hat{s}_{t+1})] - G_{t+1}(S_{t+1})\hat{\mathbb{E}}_t[\hat{\mathbb{I}}_t(y-\hat{s}_{t+1})]| \tag{Aa-3}$$

$$+ |G_{t+1}(S_{t+1})\hat{\mathbb{E}}_t[\hat{\mathbb{I}}_t(y-\hat{s}_{t+1})] - G_{t+1}(S_{t+1})\hat{\mathbb{E}}_t[\mathbb{I}_t(y-s_{t+1})]|. \tag{Aa-4}$$

The terms (Aa-1) and (Aa-3) converge uniformly on $y \in [\underline{D}, \bar{D}]$ by the Glivenko-Cantelli Theorem; the term (Aa-2) by CMT, the induction hypothesis that $\hat{G}_{t+1} \to G_{t+1}$ uniformly, and since $\hat{\mathbb{E}}_t[\hat{\mathbb{I}}_t(y - \hat{s}_{t+1})]$ can be bounded by 1; and the term (Aa-4) converges uniformly on $y \in [\underline{D}, \bar{D}]$ because

$$\begin{aligned} |\hat{\mathbb{E}}_t[\hat{\mathbb{I}}_t(y-\hat{s}_{t+1})] - \hat{\mathbb{E}}_t[\mathbb{I}_t(y-s_{t+1})]| &= \frac{1}{n}\sum_{i=1}^{n} |\mathbb{I}(y \leq d_t^i + \hat{s}_{t+1}] - \mathbb{I}(y \leq d_t^i + s_{t+1}]| \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{I}(y \in [d_t^i + \min(\hat{s}_{t+1}, s_{t+1}), d_t^i + \max(\hat{s}_{t+1}, s_{t+1})]); \end{aligned}$$

since $\hat{s}_{t+1} \to s_{t+1}$ as $n_{t+1} \to \infty$ by the induction hypothesis, $\hat{s}_{t+1}$ and s_{t+1} can be made arbitrarily close; now assume n_{t+1} is large enough $(n_{t+1} > N_{t+1} = N_{t+1}(n))$ such that the intervals $[d_t^i + \min(\hat{s}_{t+1}, s_{t+1}), d_t^i + \max(\hat{s}_{t+1}, s_{t+1})]$, $i = 1, \ldots, n$ do not overlap; then, the above expression is bounded by n^{-1}, which converges to zero as $n \to \infty$.

We can bound (Ab) by the same argument as for (Aa-4), and for (Ac) we have:

$$|\hat{\mathbb{E}}_t[\hat{G}_{t+1}(y-D_t)\hat{\mathbb{I}}_t^c(y-\hat{s}_{t+1})] - \hat{\mathbb{E}}_t[G_{t+1}(y-D_t)\mathbb{I}_t^c(y-s_{t+1})]| \tag{EC.4}$$

$$\leq |\hat{\mathbb{E}}_t[\hat{G}_{t+1}(y-D_t)\hat{\mathbb{I}}_t^c(y-\hat{s}_{t+1})] - \hat{\mathbb{E}}_t[G_{t+1}(y-D_t)\hat{\mathbb{I}}_t^c(y-\hat{s}_{t+1})]| \tag{EC.5}$$

$$+ |\hat{\mathbb{E}}_t[G_{t+1}(y-D_t)\hat{\mathbb{I}}_t^c(y-\hat{s}_{t+1})] - \hat{\mathbb{E}}_t[G_{t+1}(y-D_t)\mathbb{I}_t^c(y-s_{t+1})]|, \tag{EC.6}$$

where the first term converges to zero uniformly on $y \in [\underline{D}, \bar{D}]$ by the induction hypothesis (and by bounding the term $\hat{\mathbb{I}}_t^c(y - s_{t+1})$ by 1) and the second term by an argument similar to (Aa-4), by first bounding the common term $G_{t+1}(y - D_t)$ by an upper bound $\bar{G}_{t+1}$ on $G_{t+1}(\cdot)$, which exists because G_{t+1} is continuous with a bounded domain.

For term (B) we have:

$$|\tilde{G}_t(y) - G_t(y)| \leq |\hat{\mathbb{E}}_t[C_t(y, D_t)] - \mathbb{E}_t[C_t(y, D_t)]| \tag{Ba}$$

$$+ \alpha_t c_t |\hat{\mathbb{E}}_t[D_t] - \mathbb{E}_t[D_t]| + \alpha_t (G_{t+1}(S_{t+1} + K)|\hat{\mathbb{E}}_t[\mathbb{I}_t(y-s_{t+1})] - \mathbb{E}_t[\mathbb{I}_t(y-s_{t+1})]| \tag{Bb}$$

$$+ \alpha_t |\hat{\mathbb{E}}_t[G_{t+1}(y-D_t)\mathbb{I}_t^c(y-s_{t+1})] - \mathbb{E}_t[G_{t+1}(y-D_t)\mathbb{I}_t^c(y-s_{t+1})])|. \tag{Bc}$$

The term (Ba) converges uniformly on $y \in [\underline{D}, \bar{D}]$ by the same argument as showing uniform convergence of $\hat{G}_T(\cdot)$ in Step 1 of the induction; and for (Bb) and (Bc) we can apply the Glivenko-Cantelli Theorem (for term (Bc) we bound the common term $G_{t+1}(y - D_t)$ by $\bar{G}_{t+1}$). $\square$

A.2. Proof of Theorem 1

The theory of M-estimation concerns the following scenario. Consider the parametric function $m_\theta : \mathcal{X} \mapsto \bar{\mathbb{R}}$, where θ is a parameter chosen from Θ, and $\mathcal{X}$ is a subset of the Euclidean space. We are interested in finding the parameter θ^* that maximizes (for minimization, we can use $-m_\theta$ instead) the expected value of this function $M(\theta) = \mathbb{E} m_\theta(X)$, where X is drawn from the probability space $(\Omega, \mathcal{F}, P)$. In the absence of the true distributional knowledge, but in the presence of iid observations $X_1, \ldots, X_n$, one can estimate θ^* by minimizing instead the empirical function $M_n(\theta) = n^{-1}\sum_{i=1}^n m_\theta(X_i)$. Of central importance is whether the solution (or, a near-optimal solution) to the empirical problem is consistent, i.e., whether it converges to the true optimal as the number of observations tend to infinity. Theorem 5.7 in Van der Vaart (2000) provides sufficient conditions for asymptotic consistency, which we show are satisfied by our problem.

To prove the consistency of $(\hat{s}_t)_{t=1}^T$, we make use of Theorem 2 from Lecture 15 of Bartlett (2013) for Z-estimators ("Z" meaning that the quantity of estimation interest is the zero of a function).

We now prove Theorem 1.

Step 1. **Consistency of $(\hat{s}_T, \hat{S}_T)$:**

1. $\hat{S}_T \xrightarrow{P} S_T$: $\hat{S}_T$ is the empirical quantile of $\hat{P}_n$ at level $(b-c_t)/(b+h)$ so $\hat{S}_T \xrightarrow{P} S_T$ follows from the Glivenko-Cantelli Theorem (which is in fact a stronger, almost-sure consistency result).
2. $\hat{s}_T \xrightarrow{P} s_T$: Let $\Psi(y) = G_T(y) - G_T(S_T) - K$ and $\Psi_n(y) = \hat{G}_T(y) - \hat{G}_T(\hat{S}_T) - K$. By definition, the solution to $\Psi(y) = 0$, s_T, is unique, so the second condition of Theorem 2 of Bartlett (2013) Lecture 15 is satisfied. Also, the Algorithm solves for $\hat{s}_T$ to an arbitrary accuracy, so the near-zero assumption is also satisfied. For the first condition we have:

$$
\begin{aligned}
|\Psi_n(y) - \Psi(y)| &= |\hat{G}_T(y) - G_T(y) - (\hat{G}_T(\hat{S}_T) - G_T(S_T))| \\
&\leq |\hat{G}_T(y) - G_T(y)| + |\hat{G}_T(\hat{S}_T) - G_T(\hat{S}_T)| + |G_T(\hat{S}_T) - G_T(S_T)|,
\end{aligned}
$$

   by adding and subtracting the term $\hat{G}_T(\hat{S}_T)$ and employing the triangle inequality. Now we have already shown that the first two terms converge uniformly to zero in y. The last term converges to zero (uniformly on $\mathcal{Y}$) by CMT, since $\hat{S}_T \xrightarrow{P} S_T$ and $G_T(\cdot)$ is continuous. Hence we have $\hat{s}_t \to s_T$ by Theorem 2 of Bartlett (2013) Lecture 15.

Step 2. **Induction hypothesis:** assume $\sup_{y\in\mathcal{Y}} |\hat{G}_\tau(y) - G_\tau(y)| \xrightarrow{P} 0$ and $(\hat{s}_\tau, \hat{S}_\tau) \xrightarrow{P} (s_\tau, S_\tau)$ as $n \to \infty$.

Step 3. **Consistency of $(\hat{S}_t, \hat{s}_t)_{t=1}^T$ for $1 \leq t < \tau$:**

1. $\hat{S}_t \xrightarrow{P} S_t$: We can compute $\hat{S}_t$ to an arbitrary accuracy so it is a near-optimal minimizer of $\hat{G}_t$. Also, we have $\sup_{y\in[\underline{D},\bar{D}]} |\hat{G}_t(y) - G_t(y)| \xrightarrow{P} 0$ by Lemma EC.1, and by assumption (ID), Theorem 5.7 in Van der Vaart (2000) applies and we can conclude $\hat{S}_t \xrightarrow{P} S_t$.
2. $\hat{s}_t \xrightarrow{P} s_t$: Let $\Psi(y) = G_t(y) - G_t(S_t) - K$ and $\Psi_n(y) = \hat{G}_t(y) - \hat{G}_t(\hat{S}_t) - K$. By definition, the solution to $\Psi(y) = 0$, s_t, is unique, so the second condition of Theorem 2 from Lecture 15

of Bartlett (2013) is satisfied. Also, since we can solve for $\hat{s}_t$ to an arbitrary accuracy, the near-zero assumption is also satisfied. For the first condition we have:

$$|\Psi_n(y) - \Psi(y)| \le |\hat{G}_t(y) - G_t(y)| + |\hat{G}_t(\hat{S}_t) - G_t(\hat{S}_t)| + |G_t(\hat{S}_t) - G_t(S_t)|,$$

by adding and subtracting the term $\hat{G}_T(\hat{S}_T)$ and employing the triangle inequality. Now by Lemma EC.1, we have already shown that the first two terms converge uniformly to zero in y. The last term converges to zero (uniformly on $\mathcal{Y}_t$) by CMT, since $\hat{S}_T \xrightarrow{P} S_T$ and $G_t(\cdot)$ is continuous. Hence we have $\hat{s}_t \to s_t$ by Theorem 2 from Lecture 15 of Bartlett (2013). $\square$

A.3. Proof of Theorem 2

We have

$$\begin{aligned}\frac{1}{n}\sum_{i\in\mathcal{J}} g_t(\breve{d}_t^i;y) &= \frac{1}{n}\sum_{i\in\mathcal{J}} g_t(\bar{d}_t^1;y)\mathbb{I}(d_t^i \ge x_t) + \frac{1}{n}\sum_{i\in\mathcal{J}} g_t(d_t^i;y)\mathbb{I}(d_t^i < x_t)\\ &\xrightarrow{P} r g_t(\mathbb{E}[D_t|D_t \ge x_t];y)\mathbb{P}(D_t \ge x_t) + r\mathbb{E}[g_t(D_t;y)\mathbb{I}(D_t < x_t)] \ne r\mathbb{E}[g_t(D_t;y)]\end{aligned}$$

because in the first term of the third line, $\bar{d}_t^1 \xrightarrow{P} \mathbb{E}[D_t|D_t \ge x_t]$ by the Weak Law of Large Numbers (WLLN), which means $g_t(\bar{d}_t^1;y) \xrightarrow{P} g_t(\mathbb{E}[D_t|D_t \ge x_t];y)$ by CMT, and this together with Portmanteau lemma and WLLN on the average of the indicators $n^{-1}\sum_{i\in\mathcal{J}}\mathbb{I}(D_t \ge x_t)$ gives the first limit, and the second term of the third line tends to $r\mathbb{E}[g_t(D_t;y)\mathbb{I}(D_t \ge x_t)]$ in probability by WLLN.

Hence the objective function for $\breve{S}_t$ and the governing equation for $\breve{s}_t$ do not converge to the objective function for S_t and the governing equation for s_t. Inconsistency then follows. $\square$

A.4. Proof of Theorem 3

It suffices to show that Condition 1 of Theorem 5.7 in Van der Vaart (2000) is satisfied by $\tilde{G}_t(y)$, i.e., $\sup_{y\in\mathcal{Y}_t} |\tilde{G}_t(y) - G_t(y)| \xrightarrow{P} 0$. We have $|\tilde{G}_t(y) - G_t(y)| \le |\tilde{G}_t(y) - \hat{G}_t(y)| + |\hat{G}_t(y) - G_t(y)|$, where the second term goes to zero in probability by Lemma EC.1. The first term is

$$\begin{aligned}|\tilde{G}_t(y) - \hat{G}_t(y)| \le &\left|\frac{1}{n}\sum_{i\in\mathcal{J}} \tilde{g}_t(y)\mathbb{I}(d_t^i \ge x_t) - |\mathcal{J}|\mathbb{E}[g_t(y,D_t)\mathbb{I}(D_t \ge x_t)]\right|\\ &+ \left|\frac{1}{n}\sum_{i\in\mathcal{J}} g_t(y,d_t^i) - |\mathcal{J}|\mathbb{E}[g_t(y,D_t)\mathbb{I}(D_t \ge x_t)]\mathbb{I}(d_t^i \ge x_t)\right|.\end{aligned}$$

Now $\mathcal{Y}_t$ is bounded, both $\Gamma_1(y,d) = \tilde{g}_t(y)\mathbb{I}(d \ge x_t)$ and $\Gamma_2(y,d) = g_t(y,d)$ are continuous at each y for almost all $d \in [\underline{D}, \bar{D}]$, and are measurable functions of d at each y. Furthermore, both $\Gamma_1(\cdot,\cdot)$ and $\Gamma_2(\cdot,\cdot)$ are bounded, continuous functions over a bounded domain for both arguments, so can be upper bounded by a constant. In all, this means that the Uniform Law of Large numbers applies to both $n^{-1}\sum_{i\in\mathcal{J}} \tilde{g}_t(y)\mathbb{I}(d_t^i \ge x_t)$ and $n^{-1}\sum_{i\in\mathcal{J}} g_t(y,d_t^i)$, hence $\sup_{y\in\mathcal{Y}_t} |\tilde{G}_t(y) - \hat{G}_t(y)| \xrightarrow{P} 0$, and we have the desired result. $\square$

Appendix B: Proofs of results in Section 4

B.1. Proof of Theorem 4

We make use of Theorem 5.23 of Van der Vaart (2000) and Theorem 2 from Lecture 17 of Bartlett (2013). To apply Theorem 5.23 of Van der Vaart (2000) for the asymptotics of $\hat{S}_t$, we need the following technical lemma.

LEMMA EC.2. *For every y_1, y_2 in a neighborhood of S_t, there exists a measurable function $\dot{h}_t : [\underline{D}, \bar{D}] \mapsto \mathbb{R}$ with $\mathbb{E}_t \dot{h}_t^2(D_t) < \infty$ such that*

$$|g_t(y_1, d) - g_t(y_2, d)| \le \dot{h}_t(d)|y_1 - y_2| \tag{EC.8}$$

for all $t = 1, \dots, T$, where $g_t(\cdot,\cdot)$ is defined (6) for $1 \le t \le T$.

Proof of Lemma EC.2:

Step 1. $\mathbf{t = T}$**:** For $t = T$, it is straight-forward to show that

$$|g_T(y_1, d) - g_T(y_2, d)| \le c_t|y_1 - y_2| + (b_T \vee h_T)|y_1 - y_2|,$$

hence the Lipschitz property (EC.8) holds with $\dot{h}_T(d) = c_t + (b_T \vee h_T)$.

Step 2. **Induction hypothesis:** assume the Lipschitz property (EC.8) holds with $\dot{h}_\tau(d) = (1 - \alpha_t)c_t + (b_\tau \vee h_\tau) + \alpha_t|\mathbb{E}_{\tau+1}\dot{h}_{\tau+1}(D_{\tau+1})|$, for some $1 \le \tau \le T - 1$.

Step 3. $\mathbf{t = \tau - 1}$**:** Observe

$$\begin{aligned}|g_t(y_1, d) - g_t(y_2, d)| \le\ & (1 - \alpha_t)c_t|y_1 - y_2| + (b_t \vee h_t)|y_1 - y_2| \\ & + \alpha_t G_{t+1}(s_{t+1})|\mathbb{I}(y_1 - s_{t+1} \le d) - \mathbb{I}(y_2 - s_{t+1} \le d)| \\ & + \alpha_t|G_{t+1}(y_1 - d)\mathbb{I}(y_1 - s_{t+1} > d) - G_{t+1}(y_2 - d)\mathbb{I}(y_2 - s_{t+1} > d)|.\end{aligned}$$

Without loss of generality, assume $y_1 < y_2$. Then

$$|\mathbb{I}(y_1 - s_{t+1} \le d) - \mathbb{I}(y_2 - s_{t+1} \le d)| = \mathbb{I}(y_1 \le d + s_{t+1} \le y_2).$$

We consider three cases: $S_t < \underline{D} + s_{t+1}$, $\underline{D} + s_{t+1} \le S_t \le \bar{D} + s_{t+1}$, and $S_t \ge \bar{D} + s_{t+1}$.

Case I: $S_t < \underline{D} + s_{t+1}$. In this case, there exists a neighbourhood $\mathcal{N}_t$ of S_t such that $y < \underline{D} + s_{t+1}$ for all $y \in \mathcal{N}_t$. If y_1 and y_2 are in this neighborhood, we have $\mathbb{I}(y_1 - s_{t+1} \le d) = \mathbb{I}(y_2 - s_{t+1} \le d) = 1$ and $\mathbb{I}(y_1 - s_{t+1} > d) = \mathbb{I}(y_2 - s_{t+1} > d) = 0$.

Case II: $\underline{D} + s_{t+1} \le S_t \le \bar{D} + s_{t+1}$. In this case, for any $y_1, y_2 \in [\underline{D} + s_{t+1}, \bar{D} + s_{t+1}]$,

$$\sup_{D_t \in [\underline{D}, \bar{D}]} \mathbb{I}(y_1 \le d + s_{t+1} \le y_2) = 1.$$

and

$$\sup_{D_t \in [\underline{D}, \bar{D}]} |G_{t+1}(y_1 - d)\mathbb{I}(y_1 > d) - G_{t+1}(y_2 - d)\mathbb{I}(y_2 > d)| = G_{t+1}(y_2 - D^*),$$

where D^* is some number such that $y_1 - s_{t+1} < D^* < y_2 - s_{t+1}$. Thus for any neighborhood of S_t in the same range,

$$\begin{aligned}&\alpha_t|G_{t+1}(s_{t+1})(\mathbb{I}(y_1 - s_{t+1} \le d) - \mathbb{I}(y_2 - s_{t+1} \le d))\\&+ G_{t+1}(y_1 - d)\mathbb{I}(y_1 - s_{t+1} > d) - G_{t+1}(y_2 - d)\mathbb{I}(y_2 - s_{t+1} > d)| \le \alpha_t|\mathbb{E}_{t+1}\dot{h}_{t+1}(D_{t+1})||y_1 - y_2|\end{aligned}$$

since $y_1 - D^* < s_{t+1}$ in the neighborhood of consideration and because, by the K-convexity of the G_{t+1} function, any point to the left of the cutoff level s_{t+1} is necessarily larger than $G_{t+1}(s_{t+1})$ by Lemma 4.2.1 of Bertsekas (1995).

Case III: $S_t \ge \bar{D} + s_{t+1}$. In this case, there exists a neighborhood $\mathcal{N}_t$ of S_t such that $y > \bar{D} + s_{t+1}$ for all $y \in \mathcal{N}_t$. In this neighborhood $\mathcal{N}_t$, $\mathbb{I}(y_1 - s_{t+1} \le d) = \mathbb{I}(y_2 - s_{t+1} \le d) = 0$ for all $y_1, y_2 \in \mathcal{N}_t$ and we also have $\mathbb{I}(y_1 - s_{t+1} > d) = \mathbb{I}(y_2 - s_{t+1} > d) = 1$. Furthermore, noting that $G_t(y) = \mathbb{E}_t[g_t(y)]$, we have

$$|G_{t+1}(S_1 - d)\mathbb{I}(y_1 - s_{t+1} > d) - G_{t+1}(y_2 - d)\mathbb{I}(y_2 - s_{t+1} > d)| \quad \le |\mathbb{E}_{t+1}\dot{h}_{t+1}(D_{t+1})||y_1 - y_2|. \quad \square$$

We now prove the main result by induction.

Step 1. **Asymptotic normality of** $(\hat{s}_T, \hat{S}_T)$**:**

1. Asymptotic normality of $\hat{S}_T$: It is clear that $d \mapsto g_T(S, d)$ is measurable for each $S \in \mathcal{Y}_T$ and the map $S \mapsto g_T(S, D_T)$ differentiable at S_T for P-a.s. D_T (it is differentiable everywhere except at $S_T = D_T$, which has mass zero since we assume the demand is continuous). The derivative with respect to the first argument is given by:

$$\frac{d}{dy}g_T(y,d) = \dot{g}_T(y,d) = c_t - b_T\mathbb{I}(d - y > 0) + h_T\mathbb{I}(d - y < 0), \tag{EC.9}$$

where $\mathrm{dom}(\dot{g}_T) = \mathcal{Y}_T \backslash \{D_T\}$. Condition 2 of Theorem 4 is satisfied due to (EC.1), Condition 3 by Theorem 1 and since S_T is the exact minimizer of $\hat{G}_T(\cdot)$. Finally, let the pdf of the demand at time T be f_T. Then

$$\begin{aligned}&\frac{d}{dy}\mathbb{E}_T g_T(y, D_T) = c_t\mathbb{E}_T D_T - b_T\int_y^{\bar{D}_T} f_T(x)dx + h_T\int_{\underline{D}_T}^{y} f_T(x)dx\\ \Longrightarrow\ &\frac{d^2}{dy^2}\mathbb{E}_T g_T(y, D_T) = (b_T + h_T)f_T(y).\end{aligned}$$

We thus have $\sqrt{n}(\hat{S}_T - S_T) \Rightarrow \mathcal{N}(0, \sigma_T^2)$, where

$$\sigma_T^2 = \frac{\mathbb{E}_T[c_t - b_T\mathbb{I}(D_T > S_T) + h_T\mathbb{I}(D_T < S_T)]^2}{[(b_T + h_T)f_T(S_T)]^2}. \tag{EC.10}$$

2. Asymptotic normality of $\hat{s}_T$: Let $\Psi(y) = G_T(y) - G_T(S_T) - K$ and $\Psi_n(y) = \hat{G}_T(y) - \hat{G}_T(\hat{S}_T) - K$ as before, and $\psi(y,d) = g_T(y,d) - g_T(S_T,d) - K$ and $\psi_n(y,d) = \hat{g}_T(y,d) - \hat{g}_T(\hat{S}_T,d) - K$. We have that $\hat{s}_t$ is a near zero of $\Psi_n(y) = n^{-1}\sum_{i=1}^n \hat{\psi}_n(y, d_T^i)$, $\mathbb{E}\psi(s_T, D_T)^2$ exists and $\mathbb{E}\dot{\psi}(s_T, D_T)$ exists and is equal to

$$\mathbb{E}\dot{\psi}(s_T, D_T) = \mathbb{E}\dot{g}_T(s_T, D_T) = c_t - b_T\mathbb{P}(D_T > s_T) + h_T\mathbb{P}(D_T < s_T),$$

which is non-zero by assumption. Finally, we have $\ddot{\psi}(y,d)=\ddot{g}_T(y,d)=(b_T+h_T)f_T(y)$, and this is bounded (in probability) for all $y\in\mathcal{Y}_T$. Thus all conditions of Theorem 2 from Lecture 17 of Bartlett (2013) are satisfied and together with the consistency result Theorem 1, we have $\sqrt{n}(\hat{s}_T-s_T)\Rightarrow\mathcal{N}(0,\rho_T^2)$ where

$$\rho_T^2=\frac{\mathbb{E}[g_T(s_T,D_T)-g_T(S_T,D_T)-K]^2}{[c_t-b_T\mathbb{P}(D_T>s_T)+h_T\mathbb{P}(D_T<s_T)]^2}. \tag{EC.11}$$

Step 2. **Induction hypothesis:** assume $\sqrt{n}(\hat{S}_\tau-S_\tau)\Rightarrow\mathcal{N}(0,\sigma_\tau^2)$ and $\sqrt{n}(\hat{s}_\tau-s_t)\Rightarrow\mathcal{N}(0,\rho_\tau^2)$ as $n\to\infty$ for some $1\le\tau\le T-1$.

Step 3. **Asymptotic normality of $(\hat{S}_t,\hat{s}_t)_{t=1}^T$, $t\le T-1$:** Recall

$$\begin{aligned}g_t(y,d)&=(1-\alpha_t)cy+b_t(d-y)^++h_t(y-d)^++\alpha_t cd+\alpha_t G_{t+1}(s_{t+1})\mathbb{I}_t(y-s_{t+1})\\&\quad+\alpha_t G_{t+1}(y-d)\mathbb{I}_t^c(y-s_{t+1}).\end{aligned}$$

1. Asymptotic normality of $\hat{S}_t$: It is clear that for each $y\in\mathcal{Y}$, $D_t\mapsto g_t(D_t;y)$ is measurable such that $y\mapsto g_t(D_t;y)$ is differentiable at S_t for P_t-a.s. D_t because the number of non-differentiable points are finite. The derivative with respect to the first argument is given by

$$\begin{aligned}\frac{d}{dy}g_t(y,d)&=(1-\alpha_t)c_t-b_t\mathbb{I}(d-y>0)+h_t\mathbb{I}(y-d>0)-\alpha_t G_{t+1}(s_{t+1})\delta(y-s_{t+1}-d)\\&\quad+\alpha_t G_{t+1}(y-d)\delta(y-s_{t+1}-d)+\alpha_t\dot{G}_{t+1}(y-d)\mathbb{I}_t^c(y-s_{t+1}),\end{aligned}\tag{EC.12}$$

where $\delta(\cdot)$ is the Dirac delta function, defined by

$$\delta(z)=\begin{cases}+\infty & \text{if } y=0\\ 0 & \text{otherwise.}\end{cases}$$

Condition 2 of Theorem 5.23 of Van der Vaart (2000) is satisfied due to Proposition EC.2, and Condition 3 is satisfied trivially. Hence

$$\begin{aligned}&\frac{d}{dy}\mathbb{E}_t g_t(y,D_t)=(1-\alpha_t)c_t-b_t\int_y^{\bar{D}_t}f_t(x)dx+h_t\int_{\underline{D}_t}^y f_t(x)dx+\alpha_t\mathbb{E}_t[\dot{G}_{t+1}(y-D_t)\mathbb{I}_t^c(y-s_{t+1})]\\&\implies\frac{d^2}{dy^2}\mathbb{E}_t g_t(y,D_t)=(b_t+h_t)f_t(y)+\alpha_t\dot{G}_{t+1}(s_{t+1})+\alpha_t\mathbb{E}_t[\ddot{G}_{t+1}(y-D_t)\mathbb{I}_t^c(y-s_{t+1})],\end{aligned}$$

and we conclude $\sqrt{n}(\hat{S}_t-S_t)\Rightarrow\mathcal{N}(0,\sigma_t^2)$, where σ_t^2 is given by (8).

2. Asymptotic normality of $\hat{s}_t$: Let $\Psi(y)=G_t(y)-G_t(S_t)-K$ and $\Psi_n(y)=\hat{G}_t(y)-\hat{G}_t(\hat{S}_t)-K$ as before, and $\psi(y)=g_t(y)-g_t(S_t)-K$ and $\psi_n(y)=\hat{g}_t(y)-\hat{g}_t(\hat{S}_t)-K$. Then by Theorem 2 from Lecture 17 of Bartlett (2013) and Theorem 1 and arguments similar to the $t=T$ case, $\sqrt{n}(\hat{s}_t-s_t)\Rightarrow\mathcal{N}(0,\rho_t^2)$, where ρ_t^2 is given by (9). □

B.2. Proof of Theorem 5

For the censored demand data case, we cannot use asymptotic results of M- and Z-estimators as we did for Theorem 4 because M- and Z- estimators assume iid data. Specifically, the asymptotic normality result of M-estimators (stated as Theorem 5.23 of Van der Vaart (2000)), breaks down for $\tilde{S}_t$ because the normalized sum in the statement

$$\sqrt{n}(\tilde{S}_t - S_t) = -V_1(S_t)^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_t^1(S_t, d_t^i) + o_P(1) \tag{EC.13}$$

where

$$\eta_t^1(y,d) = \frac{d}{dy}\tilde{g}_t(y,d) = \dot{\tilde{g}}_t(y,d), \tag{EC.14}$$

with $\tilde{g}_t(y,d)$ defined in (7), and

$$V_1(y) = \frac{d^2}{dy^2}\mathbb{E}[g_t(y, D_t)], \tag{EC.15}$$

consists of terms $\eta_t^1(S_t, d_t^i)$ that are correlated with each other, and Theorem 5.23 of Van der Vaart (2000) relies on the standard CLT for normalized average of iid terms. Likewise, the asymptotic normality result of Z-estimators (see Lecture 17 of Bartlett (2013)), breaks down for $\tilde{s}_t$ because the normalized sum in the statement

$$\sqrt{n}(\tilde{s}_t - s_t) = \frac{-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\eta_t^2(s_t, d_t^i)}{\frac{1}{n}\sum_{i=1}^{n}\dot{\eta}_t^2(s_t, d_t^i) + o_P(1)}, \tag{EC.16}$$

where $\eta_t^2(y,d) = \tilde{g}_t(y,d) - \tilde{g}_t(S_t,d) - K$ consists of terms $\eta_t^2(S_t, d_t^i)$ that are correlated with each other.

Thus the key to proving Theorem 5 is in extending M- and Z-estimator results to handle correlated data. In what follows, we rely on Stein's method (Stein 1972) for establishing Guassian limit results. In particular, we use one of the main theorems from Chatterjee (2014) to prove technical lemmas Lemmas EC.3 and EC.4, which state asymptotic normality of normalized sums of dependent random variables that arise in Eqs. (EC.13)-(EC.16). The proofs of Lemmas EC.3 and EC.4 can be found in the following subsections.

Lemma EC.3. *Let*

$$W_t^1(y) = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n}\eta_t^1(y, D_t^i) - \tilde{\mu}_{t,1}(y)}{\tilde{\sigma}_{t,1}(y)}, \tag{EC.17}$$

where $\eta_t^1(\cdot,\cdot)$ is as defined in (EC.14), $\tilde{\mu}_{t,1}(y) = \mathbb{E}\dot{g}_t(y, D_t)$, and

$$\begin{aligned}\tilde{\sigma}_{t,1}^2(y) = Var(\dot{g}_t(y,D_t)) + \frac{r(1+p_t)}{p_t}[\mathbb{E}\dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t)]^2 + \frac{r^2}{(1-r)}Var\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t)\right) \\ - 2r[\mathbb{E}\dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t)]\mathbb{E}\dot{g}_t(y,D_t)\end{aligned}$$

Then $W_t^1(y)$ converges in distribution to the standard normal as n tends to infinity.

Lemma EC.4. *Let*

$$W_t^2(y) = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^{n}\eta_t^2(y, D_t^i) - \tilde{\mu}_{t,2}(y)}{\tilde{\rho}_{t,1}(y)}, \tag{EC.18}$$

where $\eta_t^2(y,d) = \tilde{g}_t(y,d) - \tilde{g}_t(S_t,d) - K$, $\tilde{\mu}_{t,2}(y) = \mathbb{E}[g_t(y,D_t) - g_t(S_t,D_t) - K]$, *and*

$$\tilde{\rho}_{t,1}^2(y) = Var\left(g_t(y,D_t) - g_t(S_t,D_t) - K\right) + \frac{r(1+p_t)}{p_t}\left[\mathbb{E}(g_t(y,D_t) - g_t(S_t,D_t) - K)\mathbb{I}(D_t \geq x_t)\right]^2$$
$$+ \frac{r^2}{1-r} Var\left((g_t(y,D_t) - g_t(S_t,D_t) - K)\mathbb{I}(D_t \geq x_t)\right)$$
$$- 2r\mathbb{E}\left[(g_t(y,D_t) - g_t(S_t,D_t) - K)\mathbb{I}(D_t \geq x_t)\right]\mathbb{E}\left[g_t(y,D_t) - g_t(S_t,D_t) - K\right].$$

Then $W_t^2(y)$ *converges in distribution to the standard normal as* n *tends to infinity.*

The statement of Theorem 5 follows from combining the results of Lemmas EC.3 and EC.4 with the proof of Theorem 4 (asymptotic normality in the uncensored data case), and using the fact that $\mathbb{E}[g_t(s_t,D_t) - g_t(S_t,D_t) - K]$ equals zero by definition of s_t. $\square$

B.3. Proof of Lemma EC.3

Let

$$\tilde{W}_t^1(y) = \sqrt{n}\frac{\frac{1}{n}\sum_{i=1}^n \eta_t^1(y,D_t^i) - \tilde{\mu}_{t,n}(y)}{\tilde{\sigma}_{t,n}(y)}, \tag{EC.19}$$

where $\tilde{\mu}_{t,1}$ and $\tilde{\sigma}_{t,1}$ in W_t^1 have been replaced by finite-sample versions $\tilde{\mu}_{t,n}$ and $\tilde{\sigma}_{t,n}$.

Fix $y \in \mathcal{Y}$ and let $\mathbf{D}_t = \{D_t^1,\ldots,D_t^n\}$ and $f : [\underline{D},\bar{D}]^n \to \mathbb{R}$ be a measurable function such that $\tilde{W}_t^1(y) = f(\mathbf{D}_t)$. Note that in our setup, $\tilde{W}_t^1(y)$ has been standardized: $\mathbb{E}\tilde{W}_t^1(y) = 0$ and $Var(\tilde{W}_t^1(y)) = 1$. Also let $\mathbf{D}'_t = \{D_t'^1,\ldots,D_t'^n\}$ be an independent copy of $\mathbf{D}_t$, $[n] = \{1,\ldots,n\}$ and for each $A \subset [n]$, define the random vector $\mathbf{D}_t^A$ as

$$D_{t,i}^A = \begin{cases} D_t'^i & \text{if } i \in A \\ D_t^i & \text{if } i \notin A. \end{cases}$$

For simplicity, if A is a singleton such as $\{i\}$, then we write $\mathbf{D}_t^i$. Similarly, write $A \cup i$ instead of $A \cup \{i\}$. Define $\Delta_i f := f(\mathbf{D}) - f(\mathbf{D}^i)$ and for each $A \subset [n]$ and $i \notin A$, let $\Delta_i f^A := f(\mathbf{D}^A) - f(\mathbf{D}^{A\cup i})$. Finally, let

$$T := \frac{1}{2}\sum_{i=1}^n \sum_{A \subset [n]\setminus\{i\}} \frac{1}{n\binom{n-1}{|A|}} \Delta_i f \Delta_i f^A.$$

Then Theorem 3.1 of Chatterjee (2014) states that

$$\sup_{t\in\mathbb{R}} |\mathbb{P}(\tilde{W}_t^1(y) \leq t) - \mathbb{P}(Z \leq t)| \leq 2\left(\sqrt{Var(\mathbb{E}(T|W_t^1(y)))} + \frac{1}{4}\sum_{i=1}^n \mathbb{E}|\Delta_i f|^3\right)^{1/2}.$$

In our setup, we have

$$\frac{\tilde{\sigma}_{t,n}(y)}{\sqrt{n}}\Delta_i f = \frac{1}{n}(\dot{\tilde{g}}_t(y,D_t^i) - \dot{\tilde{g}}_t(y,(D_t')^i)) \implies \Delta_i f = \frac{1}{\tilde{\sigma}_{t,n}(y)\sqrt{n}}(\dot{\tilde{g}}_t(y,D_t^i) - \dot{\tilde{g}}_t(y,(D_t')^i))$$

and for all subset $A \subset [1,\ldots,n]\setminus\{i\}$, we can show $\Delta_A f = \Delta_i f$. Hence

$$T = \frac{1}{2n}\sum_{i=1}^n (\Delta_i f)^2 = \frac{1}{2\tilde{\sigma}_{t,1}^2(y)n^2}\sum_{i=1}^n [\dot{\tilde{g}}_t(y,D_t^i) - \dot{\tilde{g}}_t(y,D(_t')^i)]^2$$

which gives

$$
\begin{aligned}
Var[\mathbb{E}(T|\tilde{W}_t^1(y))] &\leq Var[\mathbb{E}(T|\mathbf{D}_t)] \\
&\leq Var\left(\frac{1}{2\tilde{\sigma}^2_{t,n}(y)n^2}\sum_{i=1}^n[\dot{\bar{g}}_t(y,D_t^i)^2 - 2\dot{\bar{g}}_t(y,D_t^i)\mathbb{E}\dot{\bar{g}}_t(y,(D_t')^i) + \mathbb{E}[\dot{\bar{g}}_t(y,(D_t')^i)^2]]\right) \\
&= \frac{1}{4\tilde{\sigma}_{t,n}(y)^4n^4}Var\left(\sum_{i=1}^n[\dot{\bar{g}}_t(y,D_t^i)^2 - 2\dot{\bar{g}}_t(y,D_t^i)\mathbb{E}\dot{\bar{g}}_t(y,(D_t')^i)]\right) \\
&= \frac{1}{4\tilde{\sigma}_{t,n}(y)^4n^4}Var\left(\sum_{i=1}^n[\dot{\bar{g}}_t(y,D_t^i)^2 - 2\dot{\bar{g}}_t(y,D_t^i)\mathbb{E}\dot{\bar{g}}_t(y,(D_t')^i)]\right) \\
&= \frac{1}{4\tilde{\sigma}_{t,n}(y)^4n^4}Var\left(\sum_{i\in\mathcal{J}}\Gamma_i\mathbb{I}(D_t^i\geq x_t) + \sum_{i\in\mathcal{J}}\Gamma_i\mathbb{I}(D_t^i< x_t) + \sum_{i\in\mathcal{J}^c}\Gamma_i\right) \\
&= \frac{1}{4\tilde{\sigma}_{t,n}(y)^4n^4}\left[\sum_{i\in\mathcal{J}}Var\left(\Gamma_i\mathbb{I}(D_t^i\geq x_t)\right) + \sum_{i\in\mathcal{J}}Var\left(\Gamma_i\mathbb{I}(D_t^i< x_t)\right) + \sum_{i\in\mathcal{J}^c}Var\left(\Gamma_i\right)\right. \\
&\left.+2Cov\left(\sum_{i\in\mathcal{J}}Var\left(\Gamma_i\mathbb{I}(D_t^i\geq x_t)\right), \sum_{i\in\mathcal{J}^c}Var\left(\Gamma_i\right)\right)\right] \\
&= \frac{1}{n^2}\left(\frac{C_{1,1}}{n} + C_{1,2}\right),
\end{aligned}
$$

where $\Gamma_i = \dot{\bar{g}}_t(y,D_t^i)^2 - 2\dot{\bar{g}}_t(y,D_t^i)\mathbb{E}\dot{\bar{g}}_t(y,(D_t')^i)$, and $C_{1,1}$ and $C_{1,2}$ are appropriately matched constants. We also have

$$
\frac{1}{4}\sum_{i=1}^n\mathbb{E}|\Delta_i f|^3 = \frac{1}{4\tilde{\sigma}_{t,n}(y)^3\sqrt{n}}\mathbb{E}|\dot{\bar{g}}_t(y,d_t^i) - \dot{\bar{g}}_t(y,(d_t^i)')|^3 = \frac{C_2}{\sqrt{n}},
$$

thus

$$
\sup_{x\in\mathbb{R}}|\mathbb{P}_t(\tilde{W}_t^1\leq x) - \mathbb{P}(Z\leq x)| \leq \left(\frac{1}{n}\sqrt{\frac{C_{1,1}}{n} + C_{1,2}} + \frac{C_2}{\sqrt{n}}\right)^{1/2}.
$$

Furthermore, $|W_t^1 - \tilde{W}_t^1| = o_P(1)$ by construction, so the conclusion follows.

Variance computation:

$$
\begin{aligned}
\tilde{\sigma}^2_{t,n}(y) &= Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n\dot{\bar{g}}_t(y,D_t^i)\right) \\
&= \frac{1}{n}Var\left(\sum_{i\in\mathcal{J}}\dot{\bar{g}}_t(y)\mathbb{I}(D_t^i\geq x_t) + \sum_{i\in\mathcal{J}}\dot{g}_t(y,D_t^i)\mathbb{I}(D_t^i< x_t) + \sum_{i\in\mathcal{J}^c}\dot{g}_t(y,D_t^i)\right) \\
&= \frac{1}{n}Var\left(\sum_{i\in\mathcal{J}}[\dot{\bar{g}}_t(y) - \dot{g}_t(y,D_t^i)]\mathbb{I}(D_t^i\geq x_t) + \sum_{i=1}^n\dot{g}_t(y,D_t^i)\right) \\
&= Var(\dot{g}_t(y,D_t)) + \frac{1}{n}Var\left(\sum_{i\in\mathcal{J}}[\dot{\bar{g}}_t(y) - \dot{g}_t(y,D_t^i)]\mathbb{I}(D_t^i\geq x_t)\right) \\
&+ \frac{2}{n}Cov\left(\sum_{i\in\mathcal{J}}[\dot{\bar{g}}_t(y) - \dot{g}_t(y,D_t^i)]\mathbb{I}(D_t^i\geq x_t), \sum_{i=1}^n\dot{g}_t(y,D_t^i)\right)
\end{aligned}
$$

$$
\begin{aligned}
&= Var(\dot{g}_t(y,D_t)) + \frac{1}{n}Var\left(\sum_{i\in\mathcal{J}}\dot{\bar{g}}_t(y)\mathbb{I}(D_t^i \geq x_t)\right) + \frac{1}{n}Var\left(\sum_{i\in\mathcal{J}}\dot{g}_t(y,D_t^i)\mathbb{I}(D_t^i \geq x_t)\right)\\
&\quad - \frac{2}{n}Cov\left(\sum_{i\in\mathcal{J}}\dot{\bar{g}}_t(y)\mathbb{I}(D_t^i \geq x_t), \sum_{i\in\mathcal{J}}\dot{g}_t(y,D_t^i)\mathbb{I}(D_t^i \geq x_t)\right)\\
&+ \frac{2}{n}Cov\left(\sum_{i\in\mathcal{J}}\dot{\bar{g}}_t(y)\mathbb{I}(D_t^i \geq x_t), \sum_{i=1}^{n}\dot{g}_t(y,D_t^i)\right) - \frac{2}{n}Cov\left(\sum_{i\in\mathcal{J}}\dot{g}_t(y,D_t^i)\mathbb{I}(D_t^i \geq x_t), \sum_{i=1}^{n}\dot{g}_t(y,D_t^i)\right)\\
&= Var(\dot{g}_t(y,D_t)) + \frac{1}{n}Var\left(\sum_{i\in\mathcal{J}}\dot{\bar{g}}_t(y)\mathbb{I}(D_t^i \geq x_t)\right) + \frac{1}{n}Var\left(\sum_{i\in\mathcal{J}}\dot{g}_t(y,D_t^i)\mathbb{I}(D_t^i \geq x_t)\right)\\
&\quad - \frac{2|\mathcal{J}|}{n}\mathbb{E}[\dot{\bar{g}}_t(y)]Cov\left(\mathbb{I}(D_t \geq x_t), \dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t)\right)\\
&+ \frac{2}{n}Cov\left(\sum_{i\in\mathcal{J}}\dot{\bar{g}}_t(y)\mathbb{I}(D_t^i \geq x_t), \sum_{i\in\mathcal{J}}\dot{g}_t(y,D_t^i)\right) + \frac{2}{n}Cov\left(\sum_{i\in\mathcal{J}}\dot{\bar{g}}_t(y)\mathbb{I}(D_t^i \geq x_t), \sum_{i\in\mathcal{J}^c}\dot{g}_t(y,D_t^i)\right)\\
&- \frac{2}{n}Cov\left(\sum_{i\in\mathcal{J}}\dot{g}_t(y,D_t^i)\mathbb{I}(D_t^i \geq x_t), \sum_{i\in\mathcal{J}}\dot{g}_t(y,D_t^i)\right)\\
&= Var(\dot{g}_t(y,D_t)) + \frac{|\mathcal{J}|p_t(1-p_t)}{n}[\mathbb{E}\dot{\bar{g}}_t(y)]^2 + \frac{|\mathcal{J}|^2p_t^2}{n}Var\left(\dot{\bar{g}}_t(y)\right) + \frac{|\mathcal{J}|p_t(1-p_t)}{n}Var\left(\dot{\bar{g}}_t(y)\right)\\
&\quad + \frac{|\mathcal{J}|}{n}Var\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t)\right)\\
&\quad - \frac{2|\mathcal{J}|}{n}\mathbb{E}[\dot{\bar{g}}_t(y)]\mathbb{E}[\dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t)] + \frac{2|\mathcal{J}|p_t}{n}\mathbb{E}[\dot{\bar{g}}_t(y)]\mathbb{E}[\dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t)]\\
&\quad + \frac{2|\mathcal{J}|}{n}\mathbb{E}[\dot{\bar{g}}_t(y)]Cov\left(\mathbb{I}(D_t \geq x_t), \dot{g}_t(y,D_t)\right) + \frac{2|\mathcal{J}|p_t}{n}Cov\left(\dot{\bar{g}}_t(y), \sum_{i\in\mathcal{J}^c}\dot{g}_t(y,D_t^i)\right)\\
&- \frac{2|\mathcal{J}|}{n}Cov\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t), \dot{g}_t(y,D_t)\right)\\
&= Var(\dot{g}_t(y,D_t)) + rp_t(1-p_t)[\mathbb{E}\dot{\bar{g}}_t(y)]^2 + rp_t\left(rp_tn + (1-p_t)\right)Var\left(\dot{\bar{g}}_t(y)\right) + rVar\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t)\right)\\
&\quad + 2rp_t\mathbb{E}[\dot{\bar{g}}_t(y)]\mathbb{E}[\dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t)] - 2rp_t\mathbb{E}\dot{\bar{g}}_t(y)\mathbb{E}\dot{g}_t(y,D_t) + 2rp_tCov\left(\dot{\bar{g}}_t(y), \sum_{i\in\mathcal{J}^c}\dot{g}_t(y,D_t^i)\right)\\
&- 2r\mathbb{E}[\dot{g}_t(y,D_t)^2\mathbb{I}(D_t \geq x_t)] + 2r\mathbb{E}\dot{g}_t(y,D_t)\mathbb{E}[\dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t)].
\end{aligned}
$$

Now

$$
\dot{\bar{g}}_t(y) = \frac{\sum_{j\in\mathcal{J}^c}\dot{g}_t(y,D_t^j)\mathbb{I}(D_t^j \geq x_t)}{\sum_{j\in\mathcal{J}^c}\mathbb{I}(D_t^j \geq x_t)} = \frac{R_n}{p_t}\frac{\sum_{j\in\mathcal{J}^c}\dot{g}_t(y,D_t^j)\mathbb{I}(D_t^j \geq x_t)}{(1-r)n},
$$

where

$$
R_n := \frac{p_t(1-r)n}{\sum_{j\in\mathcal{J}^c}\mathbb{I}(D_t^j \geq x_t)}.
$$

The random variable R_n tends to 1 as n tends to infinity, so we have

$$
\lim_{n\to\infty}\dot{\bar{g}}_t(y) = \frac{1}{p_t}\mathbb{E}[\dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t)],
$$

and

$$\lim_{n\to\infty} nVar(\dot{\bar{g}}_t(y)) = \lim_{n\to\infty} \frac{1}{(1-r)^2 p_t^2 n}\mathbb{E}\left(\sum_{j\in\mathcal{J}^c} \dot{g}_t(y,D_t^j)\mathbb{I}(D_t^j \geq x_t) - \mathbb{E}[\dot{g}_t(y,D_t^j)\mathbb{I}(D_t^j \geq x_t)]\right)^2$$
$$= \frac{1}{(1-r)p_t^2} Var(\dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t)).$$

We also have

$$\lim_{n\to\infty} Cov\left(\dot{\bar{g}}_t(y), \sum_{i\in\mathcal{J}^c} \dot{g}_t(y,D_t^i)\right) = \lim_{n\to\infty} Cov\left(\frac{R_n}{p_t}\frac{\sum_{j\in\mathcal{J}^c}\dot{g}_t(y,D_t^j)\mathbb{I}(D_t^j\geq x_t)}{(1-r)n}, \sum_{i\in\mathcal{J}^c}\dot{g}_t(y,D_t^i)\right)$$
$$= \frac{1}{p_t} Cov\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t \geq x_t), \dot{g}_t(y,D_t)\right).$$

Thus

$$\begin{aligned}
\tilde{\sigma}_{t,1}^2(y) &:= \lim_{n\to\infty}\tilde{\sigma}_{t,n}^2(y)\\
&= \lim_{n\to\infty} Var(\dot{g}_t(y,D_t)) + rp_t(1-p_t)[\mathbb{E}\dot{\bar{g}}_t(y)]^2 + rp_t(rp_t n + (1-p_t))Var\left(\dot{\bar{g}}_t(y)\right)\\
&\quad - r\mathbb{E}\left(\dot{g}_t(y,D_t)^2\mathbb{I}(D_t\geq x_t)\right) - r[\mathbb{E}\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)\right)]^2\\
&\quad + 2r\mathbb{E}\dot{g}_t(y,D_t)\mathbb{E}\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)\right) - 2rp_t\mathbb{E}[\dot{\bar{g}}_t(y)]\mathbb{E}[\dot{g}_t(y,D_t)]\\
&\quad + 2rp_t\mathbb{E}[\dot{\bar{g}}_t(y)]\mathbb{E}[\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)] + 2rp_t Cov\left(\dot{\bar{g}}_t(y), \sum_{i\in\mathcal{J}^c}\dot{g}_t(y,D_t^i)\right)\\
&= Var(\dot{g}_t(y,D_t)) + r\frac{(1-p_t)}{p_t}[\mathbb{E}\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)]^2 + \frac{r^2}{(1-r)}Var(\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t))\\
&\quad - r\mathbb{E}\left(\dot{g}_t(y,D_t)^2\mathbb{I}(D_t\geq x_t)\right) + r\left[\mathbb{E}\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)\right)\right]^2\\
&\quad + 2rCov\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t), \dot{g}_t(y,D_t)\right)\\
&= Var(\dot{g}_t(y,D_t)) + \frac{r}{p_t}[\mathbb{E}\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)]^2 + \frac{r^2}{(1-r)}Var\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)\right)\\
&\quad + r\mathbb{E}\left(\dot{g}_t(y,D_t)^2\mathbb{I}(D_t\geq x_t)\right) - 2r\mathbb{E}\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)\right)\mathbb{E}\dot{g}_t(y,D_t)\\
&= Var(\dot{g}_t(y,D_t)) + \frac{r}{p_t}[\mathbb{E}\dot{g}_t(y,D_t^i)\mathbb{I}(D_t\geq x_t)]^2 + \frac{r^2}{(1-r)}Var\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)\right)\\
&\quad + rVar\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)\right) + r[\mathbb{E}\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)]^2\\
&\quad - 2r\mathbb{E}\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)\right)\mathbb{E}\dot{g}_t(y,D_t)\\
&= Var(\dot{g}_t(y,D_t)) + \frac{r(1+p_t)}{p_t}[\mathbb{E}\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)]^2 + \frac{r}{(1-r)}Var\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)\right)\\
&\quad - 2r\mathbb{E}\left(\dot{g}_t(y,D_t)\mathbb{I}(D_t\geq x_t)\right)\mathbb{E}\dot{g}_t(y,D_t). \quad \square
\end{aligned}$$

B.4. Proof of Lemma EC.4

The proof parallels the proof of Lemma EC.3. Below we show the variance computation explicitly.

Variance computation:

$$
\begin{aligned}
&\tilde{\rho}_{t,n}^2(y) = Var\left(\frac{1}{\sqrt{n}}\sum_{i=1}^n \eta_t^2(y, D_t^i)\right)\\
&= \frac{1}{n}Var\left(\sum_{i=1}^n \tilde{g}_t(y, D_t^i) - \tilde{g}_t(S_t, D_t^i) - K\right)\\
&= \frac{1}{n}Var\left(\sum_{i\in\mathcal{J}}(\bar{g}_t(y) - \bar{g}_t(S_t) - K)\mathbb{I}(D_t^i \geq x_t) + \sum_{i\in\mathcal{J}}(g_t(y, D_t^i) - g_t(S_t, D_t^i) - K)\mathbb{I}(D_t^i < x_t)\right.\\
&\qquad\qquad \left. + \sum_{i\in\mathcal{J}^c} g_t(y, D_t^i) - g_t(S_t, D_t^i) - K\right)\\
&= \frac{1}{n}Var\left(\sum_{i\in\mathcal{J}}\left[\bar{g}_t(y) - \bar{g}_t(S_t) - K - (g_t(y, D_t^i) - g_t(S_t, D_t^i) - K)\right]\mathbb{I}(D_t^i \geq x_t)\right.\\
&\qquad\qquad \left. + \sum_{i=1}^n g_t(y, D_t^i) - g_t(S_t, D_t^i) - K\right)\\
&= \frac{1}{n}Var\left(\sum_{i\in\mathcal{J}}\left[\bar{g}_t(y) - \bar{g}_t(S_t) - K - (g_t(y, D_t^i) - g_t(S_t, D_t^i) - K)\right]\mathbb{I}(D_t^i \geq x_t)\right)\\
&+ \frac{1}{n}Var\left(\sum_{i=1}^n g_t(y, D_t^i) - g_t(S_t, D_t^i) - K\right)\\
&+ \frac{2}{n}Cov\left(\sum_{i\in\mathcal{J}}\left[\bar{g}_t(y) - \bar{g}_t(S_t) - K - (g_t(y, D_t^i) - g_t(S_t, D_t^i) - K)\right]\mathbb{I}(D_t^i \geq x_t),\right.\\
&\qquad\qquad \left. \sum_{i=1}^n g_t(y, D_t^i) - g_t(S_t, D_t^i) - K\right)\\
&= Var\left(g_t(y, D_t) - g_t(S_t, D_t) - K\right) + \frac{1}{n}Var\left(\sum_{i\in\mathcal{J}}[\bar{g}_t(y) - \bar{g}_t(S_t) - K]\,\mathbb{I}(D_t^i \geq x_t)\right)\\
&+ \frac{1}{n}Var\left(\sum_{i\in\mathcal{J}}\left[g_t(y, D_t^i) - g_t(S_t, D_t^i) - K\right]\mathbb{I}(D_t^i \geq x_t)\right)\\
&\quad - \frac{2}{n}Cov\left(\sum_{i\in\mathcal{J}}[\bar{g}_t(y) - \bar{g}_t(S_t) - K]\,\mathbb{I}(D_t^i \geq x_t), \sum_{i\in\mathcal{J}}\left[(g_t(y, D_t^i) - g_t(S_t, D_t^i) - K)\right]\mathbb{I}(d_t^i \geq x_t)\right)\\
&\quad + \frac{2}{n}Cov\left(\sum_{i\in\mathcal{J}}[\bar{g}_t(y) - \bar{g}_t(S_t) - K]\,\mathbb{I}(D_t^i \geq x_t), \sum_{i\in\mathcal{J}} g_t(y, D_t^i) - g_t(S_t, D_t^i) - K\right)\\
&\quad + \frac{2}{n}Cov\left(\sum_{i\in\mathcal{J}}[\bar{g}_t(y) - \bar{g}_t(S_t) - K)]\,\mathbb{I}(D_t^i \geq x_t), \sum_{i\in\mathcal{J}^c} g_t(y, D_t^i) - g_t(S_t, D_t^i) - K\right)\\
&\quad - \frac{2}{n}Cov\left(\sum_{i\in\mathcal{J}}\left[g_t(y, D_t^i) - g_t(S_t, D_t^i) - K\right]\mathbb{I}(D_t^i \geq x_t), \sum_{i\in\mathcal{J}} g_t(y, D_t^i) - g_t(S_t, D_t^i) - K\right)\\
&= Var\left(g_t(y, D_t) - g_t(S_t, D_t) - K\right) + \frac{|\mathcal{J}|p_t(1-p_t)}{n}\left[\mathbb{E}\bar{g}_t(y) - \bar{g}_t(S_t) - K\right]^2\\
&\quad + \frac{|\mathcal{J}|^2 p_t^2}{n}Var\left(\bar{g}_t(y) - \bar{g}_t(S_t) - K\right) + \frac{|\mathcal{J}|p_t(1-p_t)}{n}Var\left(\bar{g}_t(y) - \bar{g}_t(S_t) - K\right)
\end{aligned}
$$

$$
\begin{aligned}
&+\frac{|\mathcal{J}|}{n}Var\left(\left[g_t(y,D_t^i)-g_t(S_t,D_t)-K\right]\mathbb{I}(D_t\geq x_t)\right)\\
&-\frac{2|\mathcal{J}|}{n}\mathbb{E}\left[\bar{g}_t(y)-\bar{g}_t(S_t)-K\right]Cov\left(\mathbb{I}(D_t\geq x_t),\left[g_t(y,D_t)-g_t(S_t,D_t)-K\right]\mathbb{I}(D_t^i\geq x_t)\right)\\
&+\frac{2|\mathcal{J}|}{n}\mathbb{E}\left[\bar{g}_t(y)-\bar{g}_t(S_t)-K\right]Cov\left(\mathbb{I}(D_t\geq x_t),g_t(y,D_t)-g_t(S_t,D_t)-K\right)\\
&+\frac{2|\mathcal{J}|p_t}{n}Cov\left(\bar{g}_t(y)-\bar{g}_t(S_t)-K),\sum_{i\in\mathcal{J}^c}g_t(y,D_t)-g_t(S_t,D_t)-K\right)\\
&-\frac{2|\mathcal{J}|}{n}Cov\left(\left[g_t(y,D_t)-g_t(S_t,D_t)-K\right]\mathbb{I}(D_t\geq x_t),g_t(y,D_t)-g_t(S_t,D_t)-K\right)\\
=&\,Var\left(g_t(y,D_t)-g_t(S_t,D_t)-K\right)+rp_t(1-p_t)\left[\mathbb{E}\bar{g}_t(y)-\bar{g}_t(S_t)-K\right]^2\\
&+r^2p_t^2nVar\left(\bar{g}_t(y)-\bar{g}_t(S_t)-K\right)+rp_t(1-p_t)Var\left(\bar{g}_t(y)-\bar{g}_t(S_t)-K\right)\\
&-r\mathbb{E}\left[\left(g_t(y,D_t)-g_t(S_t,D_t)-K\right)^2\mathbb{I}(D_t\geq x_t)\right]\\
&+2r\mathbb{E}\left[\left(g_t(y,D_t)-g_t(S_t,D_t)-K\right)\mathbb{I}(D_t\geq x_t)\right]\mathbb{E}[g_t(y,D_t)-g_t(S_t,D_t)-K]\\
&+2rp_t\mathbb{E}\left[\bar{g}_t(y)-\bar{g}_t(S_t)-K\right]\mathbb{E}\left[\left(g_t(y,D_t)-g_t(S_t,D_t)-K\right)\mathbb{I}(D_t\geq x_t)\right]\\
&-2rp_t\mathbb{E}\left[\bar{g}_t(y)-\bar{g}_t(S_t)-K\right]\mathbb{E}\left[g_t(y,D_t)-g_t(S_t,D_t)-K\right]\\
&+2rp_tCov\left(\bar{g}_t(y)-\bar{g}_t(S_t)-K),\sum_{i\in\mathcal{J}^c}g_t(y,D_t^i)-g_t(S_t,D_t^i)-K\right)
\end{aligned}
$$

By similar arguments as for $\tilde{\sigma}^2_{t,n}(y)$ and $\tilde{\sigma}^2_{t,1}(y)$, we can show

$$
\begin{aligned}
\tilde{\rho}^2_{t,1}(y)=&\lim_{n\to\infty}\tilde{\rho}^2_{t,n}(y)\\
&=Var\left(g_t(y,D_t)-g_t(S_t,D_t)-K\right)+\frac{r(1+p_t)}{p_t}\left[\mathbb{E}(g_t(y,D_t)-g_t(S_t,D_t)-K)\mathbb{I}(D_t\geq x_t)\right]^2\\
&\quad+\frac{r}{1-r}Var\left((g_t(y,D_t)-g_t(S_t,D_t)-K)\mathbb{I}(D_t\geq x_t)\right)\\
&\quad-2r\left[\mathbb{E}(g_t(y,D_t)-g_t(S_t,D_t)-K)\mathbb{I}(D_t\geq x_t)\right]\left[\mathbb{E}(g_t(y,D_t)-g_t(S_t,D_t)-K)\right],
\end{aligned}
$$

and in particular,

$$
\begin{aligned}
\tilde{\rho}^2_{t,1}(s_t)=&\,Var\left(g_t(s_t,D_t)-g_t(S_t,D_t)-K\right)+\frac{r(1+p_t)}{p_t}\left[\mathbb{E}(g_t(s_t,D_t)-g_t(S_t,D_t)-K)\mathbb{I}(D_t\geq x_t)\right]^2\\
&+\frac{r}{1-r}Var\left((g_t(s_t,D_t)-g_t(S_t,D_t)-K)\mathbb{I}(D_t\geq x_t)\right).\quad\square
\end{aligned}
$$

Appendix C: Proofs of results in Section 5

C.1. Proof of Proposition 1.

We prove by forward induction.

- $t=1$: $\hat{q}_1=(\hat{S}_1-I_1)\mathbb{I}(I_1<\hat{s}_1)$ converges to $q_1=(S_1-I_1)\mathbb{I}(I_1<s_1)$ in probability since $(\hat{s}_1,\hat{S}_1)$ are consistent by Theorem 1.
- $t=\tau$: assume $\hat{q}_\tau\xrightarrow{P}q_\tau$.
- $t=\tau+1$: Note we can rewrite I_{t+1} as $I_{t+1}=I_1+\sum_{k=1}^t(q_k^*-D_k)$. Hence it is straight-forward to see that $\hat{I}_{\tau+1}\xrightarrow{P}I_{\tau+1}$, by the induction hypothesis. Together with Theorem 1 we then have

$\hat{I}_{\tau+1} - \hat{s}_{\tau+1} \xrightarrow{P} I_{\tau+1} - s_{\tau+1}$, and by the Continuous Mapping Theorem (CMT) we can conclude $\mathbb{I}(\hat{I}_{\tau+1} - \hat{s}_{\tau+1} < 0) \xrightarrow{P} \mathbb{I}(I_{\tau+1} - s_{\tau+1} < 0)$, since $\mathbb{P}(I_{\tau+1} - s_{\tau+1} = 0) = 0$ by assumption that the underlying demand is continuous. Since we also have $\hat{S}_{\tau+1} - \hat{I}_{\tau+1} \xrightarrow{P} S_{\tau+1} - I_{\tau+1}$, we can conclude $\hat{q}_{\tau+1} = (\hat{S}_{\tau+1} - \hat{I}_{\tau+1})\mathbb{I}(\hat{I}_{\tau+1} < \hat{s}_{\tau+1})$ converges to $q_{\tau+1}$ in probability. $\square$

C.2. Proof of Theorem 6.

We have

$$\begin{aligned}
&|V_1(I_1;\mathbf{q}') - V_1(I_1;\mathbf{q}^*)| \\
&= \Bigg| \sum_{t=1}^{T} (\Pi_{\tau=1}^{t-1}\alpha_\tau)\mathbb{E}\{[C_t(I_t', D_t) - C_t(I_t^*, D_t)]\mathbb{I}(\mathcal{A}_t \cap \mathcal{B}_t) \\
&+ [c_t(S_t' - S_t) - c_t(I_t' - I_t^*) + C_t(S_t', D_t) - C_t(S_t, D_t)]\mathbb{I}(\mathcal{A}_t \cap \mathcal{B}_t^c) \\
&+ [K + c_t(S_t' - I_t') + C_t(S_t', D_t) - C_t(I_t^*, D_t)]\mathbb{I}(\mathcal{A}_t^c \cap \mathcal{B}_t) \\
&+ [-K - c_t(S_t - I_t^*) + C_t(I_t', D_t) - C_t(S_t, D_t)]\mathbb{I}(\mathcal{A}_t^c \cap \mathcal{B}_t^c) \Bigg| \\
&= \Bigg| \sum_{t=1}^{T} (\Pi_{\tau=1}^{t-1}\alpha_\tau)\mathbb{E}\{[C_t(I_t', D_t) - C_t(I_t^*, D_t)]\mathbb{I}(\mathcal{A}_t \cap \mathcal{B}_t) \\
&+ [c_t(S_t' - S_t) - c_t(I_t' - I_t^*) + C_t(S_t', D_t) - C_t(S_t, D_t)]\mathbb{I}(\mathcal{A}_t \cap \mathcal{B}_t^c) \\
&+ [K(\mathbb{I}(\mathcal{B}_t) - \mathbb{I}(\mathcal{B}_t^c)) + c_t(S_t'\mathbb{I}(\mathcal{B}_t) - S_t\mathbb{I}(\mathcal{B}_t^c)) - c_t(I_t'\mathbb{I}(\mathcal{B}_t) - I_t^*\mathbb{I}(\mathcal{B}_t^c)) \\
&+ C_t(S_t', D_t)\mathbb{I}(\mathcal{B}_t) - C_t(S_t, D_t)\mathbb{I}(\mathcal{B}_t^c) - C_t(I_t^*, D_t)\mathbb{I}(\mathcal{B}_t) + C_t(I_t', D_t)\mathbb{I}(\mathcal{B}_t^c)]\mathbb{I}(\mathcal{A}_t^c)\} \Bigg| \\
&= \Bigg| \sum_{t=1}^{T} (\Pi_{\tau=1}^{t-1}\alpha_\tau)\mathbb{E}\{(C_t(I_t', D_t) - C_t(I_t^*, D_t))(\mathbb{I}(\mathcal{A}_t \cap \mathcal{B}_t) + \mathbb{I}(\mathcal{A}_t^c \cap \mathcal{B}_t^c)) \\
&+ [c_t(S_t' - S_t) - c_t(I_t' - I_t^*) + C_t(S_t', D_t) - C_t(S_t, D_t)](\mathbb{I}(\mathcal{A}_t \cap \mathcal{B}_t^c) + \mathbb{I}(\mathcal{A}_t^c \cap \mathcal{B}_t)) \\
&+ [K + c_t(S_t - I_t^*) + C_t(S_t, D_t) - C_t(I_t^*, D_t)]\,\mathbb{I}(\mathcal{A}_t^c)(\mathbb{I}(\mathcal{B}_t) - \mathbb{I}(\mathcal{B}_t^c)) \Bigg| \\
&\le \sum_{t=1}^{T} (\Pi_{\tau=1}^{t-1}\alpha_\tau) \Bigg\{ \mathbb{E}[|I_t' - I_t^*|[c_t - (c_t - (b_t \vee h_t))(\mathbb{I}(\mathcal{A}_t \cap \mathcal{B}_t) + \mathbb{I}(\mathcal{A}_t^c \cap \mathcal{B}_t^c))]] \\
&+ (c_t + (b_t \vee h_t))\mathbb{E}[|S_t' - S_t|(\mathbb{I}(\mathcal{A}_t \cap \mathcal{B}_t^c) + \mathbb{I}(\mathcal{A}_t^c \cap \mathcal{B}_t))] \\
&+ \mathbb{E}[[K + (c_t + (b_t \vee h_t))|S_t - I_t^*|]\mathbb{I}(\mathcal{A}_t^c)(\mathbb{I}(\mathcal{B}_t) - \mathbb{I}(\mathcal{B}_t^c))] \Bigg\},
\end{aligned} \tag{EC.20}$$

where the inequality is due to the Lipschitz property of $C_t(y, \cdot)$,

$$\sup_{D_t \in [\underline{D}, \bar{D}]} |C_t(y_1, D_t) - C_t(y_2, D_t)| \le (b_t \vee h_t)|y_1 - y_2|.$$

We can make further simplifications to (EC.20) by observing that:

$$(I_t' - I_t^*) = \sum_{k=1}^{t-1}(q_k' - q_k^*) = \sum_{k=1}^{t-1}(S_k' - I_k')\mathbb{I}(I_k' < s_k') - (S_k - I_k^*)\mathbb{I}(I_k^* < s_k^*),$$

thus

$$
\begin{aligned}
(I'_t - I^*_t) &= \sum_{k=1}^{t-1} [(S'_k - I'_k)\mathbb{I}(I'_k < s'_k) - (S_k - I^*_k)\mathbb{I}(I^*_k < s^*_k)]\mathbb{I}(\mathcal{A}_k) \\
&\quad + [(S'_k - I'_k)\mathbb{I}(I'_k < s'_k) - (S_k - I^*_k)\mathbb{I}(I^*_k < s^*_k)]\mathbb{I}(\mathcal{A}^c_k) \\
&= \sum_{k=1}^{t-1} [(S'_k - S_k) - (I'_k - I^*_k)]\mathbb{I}(\mathcal{A}_k) + (S'_k - I'_k)\mathbb{I}(\mathcal{A}^c_k \cap \mathcal{B}_k) - (S_k - I^*_k)\mathbb{I}(\mathcal{A}^c_k \cap \mathcal{B}^c_k) \\
&= \sum_{k=1}^{t-1} [(S'_k - S_k) - (I'_k - I^*_k)][\mathbb{I}(\mathcal{A}_k) + \mathbb{I}(\mathcal{A}^c_k \cap \mathcal{B}_k)] + (S_k - I^*_k)\mathbb{I}(\mathcal{A}^c_k)(\mathbb{I}(\mathcal{B}_k) - \mathbb{I}(\mathcal{B}^c_k)),
\end{aligned}
$$

and by induction we can show

$$
I'_t - I^*_t = \sum_{k=1}^{t-1} \Gamma_{k,t-1}(S'_k - S_k) + (1 + \Gamma_{k,t-2})(S_k - I^*_k)\mathbb{I}(\mathcal{A}^c_k)(\mathbb{I}(\mathcal{B}_k) - \mathbb{I}(\mathcal{B}^c_k)), \tag{EC.21}
$$

where

$$
\Gamma_{k,\tau} = \sum_{m=k}^{\tau} \prod_{\ell=k}^{m} [\mathbb{I}(\mathcal{A}_k) + \mathbb{I}(\mathcal{A}^c_k \cap \mathcal{B}_k)].
$$

Combining (EC.20) and (EC.21), the difference in the expected total cost of the estimated policy from the optimal policy can be bounded by (14). $\square$