informs
https://pubsonline.informs.org/journal/opre

OPERATIONS RESEARCH
Vol. 71, No. 6, November–December 2023, pp. 2092–2110
ISSN 0030-364X (print), ISSN 1526-5463 (online)

Contextual Areas

Inventory Control and Learning for One-Warehouse Multistore System with Censored Demand

Recep Yusuf Bekci,[a,]* Mehmet Gümüş,[b] Sentao Miao[c]

[a] Department of Management Sciences, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada; [b] Desautels Faculty of Management, McGill University, Montréal, Québec H3A 0G4, Canada; [c] Leeds School of Business, University of Colorado Boulder, Boulder, Colorado 80309

*Corresponding author

Contact: recep.bekci@uwaterloo.ca, https://orcid.org/0009-0000-5315-6955 (RYB); mehmet.gumus@mcgill.ca, https://orcid.org/0000-0003-3814-896X (MG); sentao.miao@colorado.edu, https://orcid.org/0000-0002-0380-0797 (SM)

Received: October 29, 2021
Revised: August 29, 2022; February 4, 2023; May 4, 2023
Accepted: June 9, 2023
Published Online in Articles in Advance: August 2, 2023

Area of Review: Operations and Supply Chain

https://doi.org/10.1287/opre.2021.0694

Abstract. Motivated by our collaboration with one of the largest fast-fashion retailers in Europe, we study a two-echelon inventory control problem called the one-warehouse multistore (OWMS) problem when the demand distribution is unknown. This system has a central warehouse that receives an initial replenishment and distributes its inventory to multiple stores in each time period during a finite horizon. The goal is to minimize the total expected cost, which consists of shipment, holding, lost-sales, and end-of-horizon disposal costs. The OWMS system is ubiquitous in supply chain management, yet its optimal policy is notoriously difficult to calculate even under the complete demand distribution case. In this work, we consider the OWMS problem when the demand is censored and its distribution is unknown a priori. The main challenge under the censored demand case is the difficulty in generating unbiased demand estimation. In order to address this, we propose a primal-dual algorithm in which we continuously learn the demand and make inventory control decisions on the fly. Results show that our approach has great theoretical and empirical performances.

Funding: The work of M. Gümüş was supported in part by research grants from the Natural Sciences and Engineering Research Council of Canada [Grant 217601 NSERC RGPIN-2019-06091]; and the Institute for Data Valorization [IVADO G254088]. The work of S. Miao was supported by the Strategic Management Society Strategy Research Foundation [Grant SRF-2015DP-0063]; the Social Science and Humanities Research Council of Canada [Grant 752-2014-0378]; and the Natural Sciences and Engineering Research Council of Canada [Grant G259160 NSERC RGPIN-2022-03247].

Supplemental Material: The online appendix is available at https://doi.org/10.1287/opre.2021.0694.

Keywords: inventory control • demand learning • one-warehouse multistore system • inventory allocation • censoring • heuristics

1. Introduction

Rapidly changing market conditions, product proliferation, and ever-shifting buying habits of consumers make the retail sector more unpredictable than ever. As if that were not enough, many retailers are now faced with supply-side glitches because of rising globalization, which leave many struggling to find ways to fulfill demand as quickly and cost-efficiently as possible (Petro 2021). Along these lines, the last decade witnessed increasing use of analytics and data-driven technologies that enabled supply chain planners to connect the voice of customers into their inventory management decisions (Nicasio 2021). Indeed, a recent survey conducted by Deloitte Consulting (Davenport et al. 2019) among 1,048 executives who work at large companies (501 and more employees) suggests that executives who incorporate data-driven insights into their approach are 24% more likely to exceed their business goals. That said, the same study also reports that, whereas everyone is aware of the benefits of data-driven approaches, the majority of them either lack the technological or human resource infrastructure or fail to integrate real-time data into their decision-making process.

Inventory planners in retail companies are no strangers to these challenges. According to the 2021 Inventory Optimization Survey Report (enVista 2021) conducted among 100 inventory management professionals, whereas 35% of the respondents consider inventory optimization a strategic priority and have projects in place to deliver the required capabilities, the majority report that the top three challenges are the lack of an effective process to incorporate real-time data into inventory planning, high demand volatility, and reduced forecast accuracy. Aimed at addressing these challenges, in this paper, we focus on developing effective real-time, data-driven inventory control strategies whose objective is not only to match demand with supply, but also learn it on the fly.

Our setting in this paper is motivated by our collaboration with one of the largest fashion retailers in Europe. Similar to many fast-fashion retail chains characterized by quick response capabilities (Caro and Gallien 2010), our industry partner first receives new products from its suppliers into its main warehouse and then distributes the new products to its network of stores. Whereas inbound shipments are executed in very large quantities and less frequent fashion, outbound shipments are executed in small quantities and a more frequent fashion. This is very typical in a fast-fashion supply chain with tens of thousands of new stock-keeping units (SKUs) introduced to hundreds of stores in a relatively short period of time and the last-mile shipments to the stores are spread out in a frequent fashion across the selling season because of logistical considerations. In fact, our industry partner operates more than 70 stores in a densely populated area (overall, more than 400 stores in its network); as such, the shipments of new products from its central warehouse to its stores are conducted on an almost daily basis. Furthermore, because the majority of the SKUs designed for a new season have never been sold before, retailers in this industry almost always start every season with very limited data and improve their forecast on the fly based on actual sales during the season. Indeed, before each selling season, our industry partner conducts focus group analyses to gather opinions from a group that consists of 10–20 potential customers about the future demand of a selected SKU. Because the sample size is extremely low, the resulting range for the potential demand turns out to be inevitably very wide, which leaves the inventory planner with almost no clue about the true underlying distribution. Using this as our main setting, in this paper, we aim to develop several data-driven strategies for the in-season inventory planning problem between the central warehouse and the multiple stores that face demand uncertainty with unknown distribution. In practice, the majority of the SKUs managed by the fast-fashion retailers (e.g., approximately more than 75% of the SKUs in our partner retailer) fall into this category in the sense that once the initial inventory is shipped to the central warehouse, there is no inventory replenishment decision (i.e., in-season inventory planning) between the suppliers and central warehouse throughout the selling season. Therefore, once the season starts, the focus of an inventory planner in a typical fast-fashion retailer shifts from deciding on initial preseason order quantity to optimizing the in-season inventory shipments from the central warehouse to the stores based on realized sales.

In order to capture this setting, we develop a two-tier supply chain model, in which the upstream tier consists of one central warehouse and the downstream tier consists of multiple stores. The central warehouse is replenished at the beginning of the horizon, whereas sales are observed at the store level, and the inventory is distributed periodically from the central warehouse to different stores during the planning horizon. To model the business environment described, we assume that the initial inventory in the central warehouse is fixed and not replenished during the planning horizon. Aligned with this setting, the main decisions are periodic shipments to the stores, which face stochastic demand with unknown distribution. Unsatisfied demand is lost, and remaining inventories in the stores are carried over to the next period. The goal of the firm is to minimize the total expected cost, which includes shipment cost (of transferring items from the central warehouse to stores), holding cost, lost-sales cost (of unsatisfied demands in the stores), and end-of-horizon disposal cost (of remaining inventory in the central warehouse).

The aforementioned setting, also referred to as the one-warehouse multistore (OWMS) model in operations management literature, has received a lot of attention from academia for its prevalence not only in fast fashion but also in many other real-life applications, such as grocery chains, cash management at ATMs, gas stations, and production planning. As shown in a seminal paper by Clark and Scarf (1960), it is a well-known fact that the OWMS model has complicated, state-dependent optimal policies, which suffer from the curse of dimensionality. Therefore, preceding studies on the OWMS model (see, e.g., Marklund and Rosling 2012, Lei et al. 2020, Nambiar et al. 2020, Chao et al. 2021, Miao et al. 2022) restrict their attention to developing computationally efficient algorithms with provable worst case bounds. That said, all these studies assume that the demand distribution is known in advance. This paper is the first work focusing on the setting of unknown demand information, which is prevalent in the fast-fashion industry because of the majority of SKUs being new seasonal products. In this work, we analyze the OWMS model with demand learning in two different settings, namely, with and without demand censoring. When the firm is able to observe the uncensored demand, we show that an efficient inventory control policy can be developed using sample average approximation by constructing empirical demand distributions. However, when the demand is censored so that only sales (instead of actual demand) can be observed, this approach no longer works as it requires gathering uncensored samples to draw inferences from the data (see, e.g., Kaplan and Meier 1958, Huh et al. 2011). This complication can be rendered as an exploration–exploitation trade-off. Exploration, in this manner, can be defined as gathering uncensored samples, which requires keeping excessive inventory on hand to prevent censorship. On the other hand, exploitation is keeping optimal (or near-optimal) inventory levels that are identified from the data. It is clear that efficient exploitation requires sufficient exploration; on the other hand, too much exploration leads to excessive costs from overordering, which may even hurt our decisions later.

Throughout the paper, we denote “defined as” with “:=”. x^+ represents $\max(x, 0)$. $O(\cdot)$ is the classic big-O notation that is common in algorithm analysis. In $\tilde{O}(\cdot)$ notation, the logarithmic terms are hidden.

Our main results and contributions can be summarized as follows.

1. Double binary search (DBS) algorithm with demand censoring: When the demand is censored, the main challenge stems from our inability to observe uncensored demand data, which might hurt the performance of the algorithm by deliberately ordering too many inventories to the stores. To tackle this challenge, we propose an algorithm named DBS based on the Lagrangian relaxation heuristic for continuous and bounded demand distributions with censored demand. This algorithm is based on a primal-dual learning and optimization framework and involves two intertwined loops in which we actively search for the optimal dual variable in the outer loop and the optimal base-stock levels for a given dual variable in the inner loop. Theoretical results show that the algorithm DBS achieves a regret of $\tilde{O}(N\sqrt{T})$, where N is the number of stores and T is the length of the horizon.

2. Modifications for uncensored, discrete, and unbounded demand cases: We developed the DBS algorithm under the assumptions that the demand information is censored and underlying demand follows a continuous distribution with bounded support. In the case of uncensored demand, we can eliminate both inner and outer loops of the DBS algorithm and directly generate unbiased estimates for dual variable and order-up-to levels by constructing empirical distributions. In the case of discrete demand, we need to significantly modify the DBS algorithm in order to accommodate the irregularities caused by multiple optimal solutions. Finally, in the case of unbounded demand, we can revise the regret analysis of the DBS algorithm by establishing bounds on order-up-to levels. To summarize, we are able to prove that the similar primal-dual learning and optimization framework developed for the DBS achieves the same regret bound $\tilde{O}(N\sqrt{T})$ for uncensored, discrete, and unbounded demand cases.

3. Thorough numerical experiments based on synthetic and real data: We present two sets of experiments. First, using synthetic data, we conduct experiments in order to perform a sensitivity analysis with respect to input parameters such as horizon length, cost parameters, and demand variability. In the second set of experiments, we utilize a real data set obtained from our partner fast-fashion company to infer parameters and demand distributions. In this set of experiments, we used a representative subset of the stores and products of our partner to add robustness to our results. In both experiments, the numerical results demonstrate that our DBS algorithm significantly outperforms all benchmarks.

The rest of the paper is organized as follows. In Section 2, we review the relevant literature. Section 3 contains the description of the OWMS model and its formulations. In Section 4, we present the details about the DBS algorithm for the censored demand case together with its theoretical performance. In Section 5, we analyze three modifications based on uncensored, discrete, and unbounded demand cases. We carry out numerical experiments in Section 6 using synthetic and real data. Finally, we conclude in Section 7.

2. Literature Review

Our research falls within the general theme of inventory control in the presence of unknown demand distribution. However, to the best of our knowledge, ours is the first paper in the literature that studies this problem in the context of the OWMS model. As such, there are two streams of research directly related to our paper. In what follows, we review each stream separately.

2.1. Inventory Control with Demand Learning

Inventory control has been an important research area since decades ago (see, e.g., Zipkin 2000, Porteus 2002 for detailed analysis and survey for some classical work). Whereas earlier work mostly focuses on inventory control and optimization with complete demand information, recently there is a growing body of literature investigating the case when demand is unknown and has to be learned either from an off-line data set or on the fly. For the demand model, the literature either assumes that it is a parametric model with unknown parameters or it is a nonparametric model, which is the one we consider in this paper.

For the parametric demand model, usually, the decision maker has an idea or prior belief about the demand distribution. Afterward, the decision maker updates the belief as the decision maker discovers new information. Bayesian inference is one of the methods to update the parameter (see Scarf 1959, Azoury 1985), and this method is also applied in censored demands (see Chen 2010).

Compared with the parametric demand model, many other works focus on the nonparametric demand model, which is more flexible and requires fewer assumptions. Nonparametric demand models do not assume any functional form of demand distribution. If the decision maker has access to a historical data set, the learning and optimization can be done with this on-hand information. Because there is no active interaction with the environment (i.e., sales operation and the data collection process), this type of learning is called off-line learning. On the other hand, in contrast with the off-line setting, online learning starts with zero data and requires direct interaction with the environment.

The off-line learning problem for inventory control is extensively studied. Levi et al. (2007) apply an empirical distribution function–based solution to the off-line newsvendor problem, and they report the number of samples required for achieving an approximately optimal solution. Later, Levi et al. (2015) improve this bound for fewer samples. Similar problems are studied in other literature as well with different settings: Cheung and Simchi-Levi (2019) study this problem with capacity constraint, Qin et al. (2019) address the joint pricing and inventory problem using the empirical distribution method, Ban and Rudin (2019) extend the solution using contextual information, Bu et al. (2020) study the joint pricing and inventory control problem with censored data, and Ban (2020) considers fixed cost of ordering with censored demand.

Our paper belongs to the online learning stream. For the newsvendor model, Besbes and Muharremoglu (2013) study the implication of demand censoring, and Godfrey and Powell (2001) use sample gradients to estimate piecewise-linear concave functions to approximate the cost function. Compared with the newsvendor model, more research focuses on the dynamic programming model in which inventories are carried over to the next period. For instance, Huh and Rusmevichientong (2009) solve a multiperiod inventory control problem with demand censoring by applying some stochastic gradient descent method (this method is used in other work as well; see, e.g., Shi et al. 2016, Zhang et al. 2018, Yuan et al. 2021), and Huh et al. (2011) use a Kaplan–Meier estimator to tackle the censored data. Besides pure inventory control, some recent work also studies the joint pricing and inventory problem: Chen et al. (2019) use linear function approximation for the demand estimation, Chen et al. (2020c) consider fixed costs, Chen et al. (2020b) apply binary and trisection search, Chen et al. (2021) use spline approximation to tackle demand censoring, and Chen et al. (2020a) study limited price changes.

2.2. One-Warehouse Multistore System

The research interest for the OWMS system dates back to the seminal paper by Clark and Scarf (1960) on multiechelon inventory systems. Because of the computational difficulty pointed out in Clark and Scarf (1960), researchers focus on developing efficient heuristics for this problem. One stream of the OWMS system is when the inventory in the central warehouse is given at the beginning without replenishment, which is first studied by Jackson (1988), who proposes a constant base-stock heuristic without theoretical performance. Jackson et al. (2019) is another paper on this topic that investigates a two-period model with backordering and develops a heuristic based on an approximation in which the number of stores goes to infinity (see McGavin et al. 1993, 1997 for work on a similar topic).

Recently, some work on OWMS develops heuristics with theoretical performance (Marklund and Rosling 2012, Lei et al. 2020, Nambiar et al. 2020, Chao et al. 2021, Miao et al. 2022). In particular, Marklund and Rosling (2012) and Nambiar et al. (2020) propose some heuristics based on Lagrangian relaxation for backorder and lost-sales OWMS systems, respectively, and they show that their heuristics are asymptotically optimal in the number of stores N but not in the length of horizon T. Later, Miao et al. (2022) generalize their result to a multiwarehouse multistore system and prove that Lagrangian relaxation heuristics are asymptotically optimal in both N and T. Similar results are obtained in Chao et al. (2021) when there is a fixed ordering cost, and the authors show that the performance of the heuristics can be further improved using appropriate readjustment of policy. Besides the Lagrangian relaxation method, asymptotically optimal heuristics based on certainty-equivalent optimization are used in Lei et al. (2020), who study a joint pricing and inventory optimization problem in the OWMS system.

It is noted that the aforementioned work on this subject assumes that the demand distribution is known, whereas, in our study, it is learned along the way.

3. Model Description and Formulations

3.1. Model Primitives

We consider the OWMS inventory problem in which there is a central warehouse with an initial capacity W that distributes its inventory to N stores, denoted by $i = 1, \dots, N$, during a finite horizon $t = 1, \dots, T$. There is no external replenishment during the horizon to the central warehouse, and the inventory is nonperishable. Transshipment between stores is not allowed. Demands $D_{i,t}$, which are unknown in our problem, are independent and identically distributed (i.i.d.) over time t and independent across stores i (i.e., the demand distributions for different stores can be different).

Before providing a detailed description of our model, we present notations in Table 1.

3.2. The Decision-Making Process and the Optimization Problem

The decision-making process, together with the cost structure, of our problem is summarized as follows:

- At the beginning of each period $t \in [T]$, the decision maker observes the initial on-hand inventory levels $I_{i,t}$ at stores $i \in [N]$ and the remaining inventory at the central warehouse.
- The decision maker then decides the amount of inventory to allocate to each store from the central warehouse. There is an exogenous delivery cost per item c_i, which is specific to the store i. Deliveries are instantaneous, and the corresponding inventory level after the delivery is denoted by $y_{i,t}$ for each store i. As a result, each store i incurs a delivery cost $c_i(y_{i,t} - I_{i,t})$.
- Demand $D_{i,t}$ is realized at each store i. Any unsatisfied demand is lost and incurs lost-sales cost $b_i(D_{i,t} - y_{i,t})^+$. If there is positive inventory left at store i, the holding cost $h_i(y_{i,t} - D_{i,t})^+$ is incurred, and the

Table 1. List of Notations

$D_{i,t}$	Random demand for store $i \in [N]$ at period $t \in [T]$
$I_{i,t}$	Initial on-hand inventory at store $i \in [N]$ and period $t \in [T]$
$y_{i,t}$	Order-up-to level for store $i \in [N]$ and period $t \in [T]$
π	Inventory control policy
V^π	Total expected cost of the policy π in the original system
V^*	Total expected cost of the optimal policy
$\tilde{V}^\lambda$	Total expected cost of the Lagrangian relaxation system with dual variable λ
$C_i(y_i;\lambda)$	Single-period cost function for store $i \in [N]$
λ	Lagrangian multiplier
$\kappa_i(\lambda)$	Critical fractile function
Parameters	
$h_i > 0$	Holding cost per unit for store $i \in [N]$
$b_i > 0$	Lost-sales cost per unit for store $i \in [N]$
$c_i \geq 0$	Shipment cost from the warehouse to store $i \in [N]$
$w \in \mathbb{R}$	End-of-horizon disposal cost
W	Central warehouse capacity

leftover inventory is carried over to the next time period.

- At the end of the time horizon, the central warehouse incurs an end-of-horizon disposal cost/value $w \in \mathbb{R}$ for each unit of the leftover inventory (i.e., w is positive if it is a cost and negative if it is a value). For the sake of brevity, we just call it the end-of-horizon disposal cost.

According to this description, the decision maker's goal is to minimize the total expected cost, which can be modeled as a multiperiod stochastic control problem formulation as follows:

$$
\begin{aligned}
V^* := \min_{\pi} \sum_{t=1}^{T}\sum_{i=1}^{N} & (\mathbb{E}[b_i(D_{i,t} - y^\pi_{i,t})^+ + h_i(y^\pi_{i,t} - D_{i,t})^+] \\
& + c_i'\mathbb{E}[y^\pi_{i,t} - I^\pi_{i,t}]) + wW \\
\text{st} \quad & \sum_{t=1}^{T}\sum_{i=1}^{N}(y^\pi_{i,t} - I^\pi_{i,t}) \leq W \\
& I^\pi_{i,t+1} = (y^\pi_{i,t} - D_{i,t})^+ \quad t \in [T],\, i \in [N] \\
& y^\pi_{i,t} \geq I^\pi_{i,t} \quad t \in [T],\, i \in [N],
\end{aligned}
\tag{P}
$$

where $c_i' = c_i - w$. Note that we denote the policy with π, and in the later context, it may be dropped sometimes for brevity whenever there is no confusion. The objective value is the total expected cost. The first constraint is the inventory constraint in the central warehouse; the second constraint is the system dynamics of the inventory carryover at the stores; the last constraint means that the delivery from the warehouse to stores cannot be negative. Note that, in our multiperiod stochastic control problem formulation, the constraints are satisfied almost surely.

3.3. Demand Learning and Regret

In this paper, we discuss how to solve the optimization problem (P) when the distribution of $D_{i,t}$ is unknown and censored. Similar to other online learning problems, we face the challenge of the trade-off between demand learning and cost minimization (also known as the exploration–exploitation trade-off). As a result, the common performance metric of online learning algorithms, named regret, is used throughout this paper. In particular, we have the following definition.

Definition 1. For a policy π, the (expected) regret is defined as

$$
\mathcal{R}(\pi) = V^\pi - V^*,
$$

which is the difference between the expected cost of policy π and the optimal value of the clairvoyant.

We note that, according to this definition, minimizing V^π is equivalent to minimizing the regret $\mathcal{R}(\pi)$. Before providing our main algorithm and its performance under demand censoring, we first present a Lagrangian relaxation–based formulation, which is used for bounding the regret as well as developing the algorithm.

3.4. Lagrangian Relaxation–Based Benchmark

It is known that the inventory control in an OWMS system has a complicated optimal structure, and hence, heuristics are needed to find an approximately optimal solution of (P). The heuristics with full demand information are studied in Miao et al. (2022), who propose two heuristics to solve the OWMS inventory control based on Lagrangian relaxation. In this paper, we use the heuristic named LaBS in Miao et al. (2022) as our benchmark for both developing the learning algorithm and bounding the regret. This heuristic solves a modified Lagrangian relaxation of (P) as follows:

$$
\begin{aligned}
\tilde{V}^\lambda := \min_{\pi} \sum_{t=1}^{T}\sum_{i=1}^{N} & (\mathbb{E}[b_i(D_{i,t} - y^\pi_{i,t})^+ + h_i(y^\pi_{i,t} - D_{i,t})^+] \\
& + c_i'\mathbb{E}[y^\pi_{i,t} - I^\pi_{i,t}]) + wW - \sum_{i=1}^{N} c_i'\mathbb{E}[(y^\pi_{i,T} - D_{i,T})^+] \\
& + \lambda\left(\sum_{t=1}^{T}\sum_{i=1}^{N}\mathbb{E}[(y^\pi_{i,t} - I^\pi_{i,t})] - W - \sum_{i=1}^{N}\mathbb{E}[(y^\pi_{i,T} - D_{i,T})^+]\right) \\
\text{st} \quad & I^\pi_{i,t+1} = (y^\pi_{i,t} - D_{i,t})^+ \quad t \in [T]\ i \in [N] \\
& y^\pi_{i,t} \geq I^\pi_{i,t} \quad t \in [T]\, i \in [N].
\end{aligned}
\tag{PS}
$$

The idea of LaBS is to apply Lagrangian relaxation for the inventory constraint in the central warehouse and allow a hypothetical sell-back option from all stores back to the warehouse at the end of the horizon (to see how we derive (PS) and the purpose of the sell-back option, we refer to Section EC.1 in the online appendix).

By this relaxation, (PS) can be decomposed into a store-level inventory control problem as

$$\tilde{V}^{\lambda} = (w - \lambda)W + \sum_{i=1}^{N} \tilde{V}_i^{\lambda}, \tag{1}$$

where each $\tilde{V}_i^{\lambda}$ can be written as

$$\begin{aligned} \tilde{V}_i^{\lambda} := \min_{\pi} \sum_{t=1}^{T} (\mathbb{E}[b_i(D_{i,t} - y_{i,t}^{\pi})^+ + (h_i - c_i(\lambda)) \\ (y_{i,t}^{\pi} - D_{i,t})^+] + c_i(\lambda)\mathbb{E}[y_{i,t}^{\pi}]) \qquad \text{(NV)} \\ \text{st } I_{i,t+1}^{\pi} = (y_{i,t}^{\pi} - D_{i,t})^+ \quad t \in [T] \\ y_{i,t}^{\pi} \geq I_{i,t}^{\pi} \quad t \in [T], \end{aligned}$$

where $c_i(\lambda) = c_i' + \lambda$. Without loss of generality, we omit c_i' in our analysis and use c_i in subsequent sections for brevity. Note that $\tilde{V}_i^{\lambda}$ is simply a single-store inventory control problem with stationary demand and end-of-horizon sell-back, and it has been studied very well (see, e.g., Zipkin 2000). An important structure of the optimal policy of $\tilde{V}_i^{\lambda}$ is that it is a base-stock policy with a constant base-stock level, which can be solved by minimizing the following newsvendor problem of the myopic cost function:

$$\begin{aligned} C_i(y_i;\lambda) := \mathbb{E}[c_i(\lambda)y_i + (h_i - c_i(\lambda))(y_i - D_{i,t})^+ \\ + b_i(D_{i,t} - y_i)^+], \end{aligned}$$

and its optimal solution $y_i^*(\lambda)$ satisfies that $F^i(y_i^*(\lambda)) = \kappa_i(\lambda)$, where

$$\kappa_i(\lambda) := \frac{b_i - c_i - \lambda}{b_i + h_i - c_i - \lambda},$$

and we have that $\tilde{V}_i^{\lambda} = T \cdot C_i(y_i^*(\lambda); \lambda)$.

Knowing the optimal value of $\tilde{V}^{\lambda}$, we are able to find the optimal dual variable λ^* by solving the convex optimization problem

$$\lambda^* = \arg\max_{\lambda \geq 0} \tilde{V}^{\lambda}, \tag{2}$$

and its corresponding optimal base-stock levels.

Above all, it is clear that the main purpose we choose LaBS as our benchmark to develop the learning algorithm is for its efficient computation as well as the simple optimal policy structure. Because Lagrangian relaxation always acts as a lower bound for the optimal value function, it is shown that $\tilde{V}^{\lambda^*} \leq V^*$ in Miao et al. (2022). As a result, we are able to bound the regret of any policy π by $\mathcal{R}(\pi) \leq V^{\pi} - \tilde{V}^{\lambda^*}$.

3.5. Technical Assumptions

In inventory learning literature, the standard technical assumptions are made to bound the optimal primal and dual variables so that we ensure convergence in finite time. The primal variable in our context corresponds to order-up-to levels y_i for each store i. In order to ensure that the optimal order-up-to level for each store is finite, we make the following assumption.

Assumption 1. *Demand $D_{i,t} \in [\underline{D}, \overline{D}]$ almost surely, where $0 < \underline{D} < \overline{D} < \infty$. Its distribution $F^i : [\underline{D}, \overline{D}] \to [0,1]$ is continuous, and it has a probability density function f^i with $0 < \underline{f} \leq f_i(d) \leq \overline{f} < \infty$ for all $i \in [N]$ and $d \in [\underline{D}, \overline{D}]$. All the parameters $\underline{D}, \overline{D}, \underline{f}, \overline{f}$ are known.*

Assumption 1 is a standard assumption regarding the regularity of the demand distribution, which is also made in other inventory learning literature (see, e.g., Chen et al. 2020b). We discuss relaxing the boundedness assumption on $D_{i,t} \in [\underline{D}, \overline{D}]$ in Section 5.3.

Assumption 1 is satisfied by many popular demand distributions, such as uniform distribution and truncated normal distribution. Under this assumption, optimal base-stock level $y_i^*(\lambda)$ can be shown to be finite and equal to $(F^i)^{-1}(\kappa_i(\lambda))$ as long as $\lambda \in [0, b_i - c_i]$ (otherwise, $y_i^*(\lambda) = 0$ if $\lambda > b_i - c_i$). This also leads to our second assumption.

Assumption 2. *For the optimal Lagrangian dual variable $\lambda^* \geq 0$, we have $\lambda^* \leq \min_{i \in [N]}(b_i - c_i)$.*

In essence, we make Assumption 2 because, if there exists a store i whose underage cost $b_i - c_i$ is less than optimal λ^* (i.e., the marginal cost of inventory), then we can completely remove that store i from consideration and never replenish its inventory throughout the planning period in the full information setting. We give the details regarding sufficient conditions that imply Assumption 2 in Online Section EC.3. Note that we have $\lambda^* \leq \max_{i \in [N]}(b_i - c_i)$. Thus, one sufficient condition is that all the stores have symmetric cost structure (i.e., $b_i = b_{i'}, c_i = c_{i'}$ for all $i \neq i' \in [N]$). This situation is not uncommon because one central warehouse typically serves local stores that are similar to each other. The second sufficient condition is when the warehouse inventory is at least as much as the expected demand of all stores over the time horizon. This condition is due to the complementary slackness condition (see lemma 5 of Miao et al. 2022), which can be expressed as $\lambda^* \cdot (\sum_{t=1}^{T} \sum_{i=1}^{N} \mathbb{E}[\min(y_i^*, D_{i,t})] - W) = 0$. This is again likely because retailers typically tend to overstock because of the high underage cost. Indeed, in most fast-fashion retail applications, the underage cost (i.e., lost sales because of shortages) tends to be the dominating factor in the cost equation. In our data set collected from our industrial partner, the underage cost (as commonly measured by the difference between the unit selling price and marginal cost) accounts for, on average, more than 50% of the unit price for all stores.

In case Assumption 2 does not hold, we can analyze the implications of relaxing Assumption 2 on our algorithm and show that it still works under relaxed Assumption 2, but its regret bound needs to be modified accordingly. We present the detailed discussion after Theorem 1 in Section 4.

4. Proposed Algorithm for Censored Demand

This section presents the proposed algorithm for censored demand, its theoretical performance, and the main steps of the proof for the theoretical result. In Section 4.1, we give the detailed algorithm and its theoretical performance together with some high-level ideas. Then, the theoretical performance is proved in Section 4.2 with some technical details provided in the online appendix.

4.1. The Algorithm and Its Performance

In order to jointly learn the demand and optimize the inventory decisions, we need to accurately estimate the demand distribution. However, when the demand is censored, the decision maker can only observe the sales instead of the actual demand. This suggests that, in order to have an unbiased estimation, active exploration of the demand distribution is required. Exploration for inventory control can generally be done by ordering high amounts of inventory to cope with the censorship. However, this deliberate extra ordering hurts the overall expected cost, which may be particularly suboptimal if the central warehouse inventory is scarce. Consequently, in this paper, we propose a smarter way that iteratively improves the base-stock levels at the store level and the dual variable at the warehouse level.

The key idea of our algorithm is that, if we can find λ^*, lemma 10 of Miao et al. (2022) shows that the gradient of $\tilde{V}^\lambda$ (divided by T; for the sake of brevity, we just call it the gradient of $\tilde{V}^\lambda$) can be derived as

$$\sum_i \mathbb{E}[\min(y_i^*(\lambda), D_{i,t})] - \frac{W}{T}. \tag{3}$$

Thus, we can estimate the marginal value of warehouse inventory, which then helps us to decide whether we replenish each store now or later. Note that the gradient of $\tilde{V}^\lambda$ can possibly be estimated without observing the full demand. More specifically, suppose we know $y_i^*(\lambda)$; if we apply $y_i^*(\lambda)$ for multiple times, we are able to observe samples of $\min(y_i^*(\lambda), D_{i,t})$, and by taking an average of these samples, we get an unbiased estimation of $\mathbb{E}[\min(y_i^*(\lambda), D_{i,t})]$. We can then use this to obtain an unbiased estimation of the gradient of $\tilde{V}^\lambda$. This nice property is another important reason that we use LaBS in Miao et al. (2022) as our benchmark instead of other heuristics such as SLaBS in Miao et al. (2022) and the algorithm in Nambiar et al. (2020). Of course, there are still two questions left: First, we do not know $y_i^*(\lambda)$ obviously. Second, how should we divide the time horizon smartly to have samples to estimate the gradient of $\tilde{V}^\lambda$? Our algorithm, referred to as the DBS algorithm, is developed to address these two challenges.

4.1.1. Algorithm Overview.

The overall idea of the algorithm is taking turns to find λ^* (i.e., update λ) and $y_i^*(\lambda)$ (for a fixed λ) via a binary search method. More specifically, the algorithm has an outer loop of searching for λ^* (with carefully designed exponentially increasing loop length) using binary search. In each outer loop, with λ fixed, there are inner loops of estimating $y_i^*(\lambda)$ for each $i \in [N]$ using binary search again. More details are explained later, but we briefly present the key steps in outer and inner loops.

Note that setting (3) to zero and solving it for T yields the following:

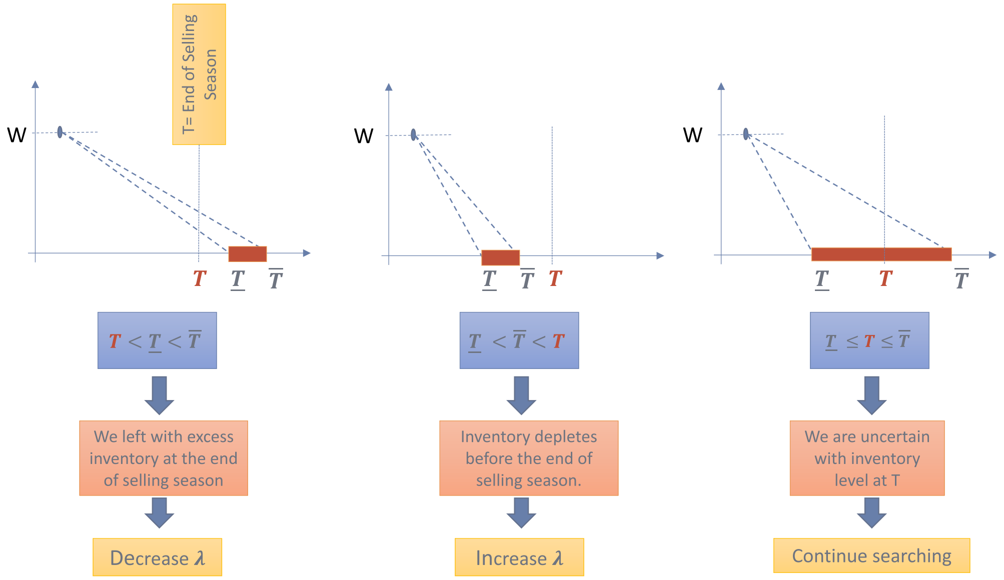
$$T = \frac{W}{\sum_i \mathbb{E}[\min(y_i^*(\lambda), D_{i,t})]}, \tag{4}$$

where the right-hand side of (4) corresponds to the number of periods required to deplete the total warehouse inventory. We refer the right-hand side of (4) as "time to hit zero" and denote it by T_0. Of course, in the learning algorithm, this is just a point estimate based on observed sales in each store i and is subject to estimation error because of a lack of information regarding the full demand distribution. Therefore, one needs to obtain an interval estimate for T_0 to capture the lack of knowledge on the demand distribution. Delegating the details to the analysis in Section 4.2, suppose that $[\underline{T}, \overline{T}]$ corresponds to the interval estimate for time to hit zero, which means that we have $T_0 \in [\underline{T}, \overline{T}]$ with high probability. Comparing this interval estimate to the remaining number of periods in planning horizon T, we can identify three cases:

- If $\underline{T} > T$, then it is likely that we cannot consume the warehouse inventory by the end of the planning horizon if we follow the current inventory ordering policy. So, in this case, we are motivated to decrease the Lagrange dual variable λ, which, in turn, induces us to order more inventory now rather than later.
- If $\overline{T} < T$, then it is likely that the central warehouse inventory will hit zero before the planning horizon ends. So, in this case, we are motivated to increase the Lagrange dual variable λ, which, in turn, induces ordering less inventory.
- If $\underline{T} \le T \le \overline{T}$, then we cannot clearly identify whether we can deplete the warehouse inventory by the end of the planning horizon. So, in order to improve the estimation accuracy, we keep the same value for λ and continue to collect more data (i.e., demand observation) from each store by following the current inventory-up-to levels.

We pictorially represent each case in three panels of Figure 1, respectively. Following the aforementioned procedure, once we update the estimate for the dual variable in the outer loop (indexed by $\tau = 0, 1, 2, \ldots$), we enter into the inner loop and iterate over T_τ periods in which we observe the realized sales and adjust order-up-to levels $y_i^*(\lambda)$ for each store. The outer and inner loops are then repeated as shown in Figure 2. To summarize, by increasing the duration of the search periods between two consecutive outer loops in a controlled

Figure 1. (Color online) Three Possible Cases Used for Updating λ in the Outer Loop


T= End of Selling Season
W
T $\underline{T}$ $\overline{T}$
$T < \underline{T} < \overline{T}$
We left with excess inventory at the end of selling season
Decrease λ
W
$\underline{T}$ $\overline{T}$ T
$\underline{T} < \overline{T} < T$
Inventory depletes before the end of selling season.
Increase λ
W
$\underline{T}$ T $\overline{T}$
$\underline{T} \leq T \leq \overline{T}$
We are uncertain with inventory level at T
Continue searching

fashion, we are able to develop the DBS algorithm and bound its regret.

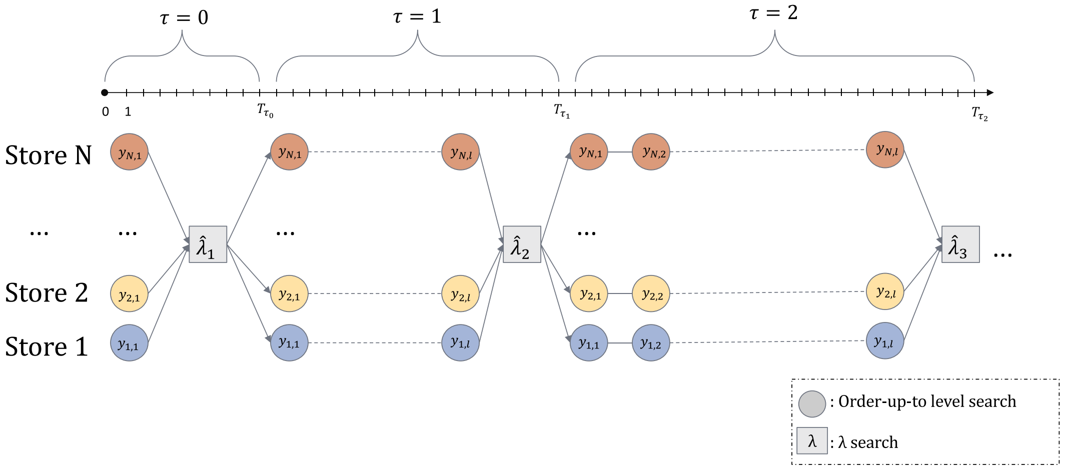
After this brief overview, we are now ready to explain the details of the DBS algorithm. In what follows, first, we present the steps conducted in the inner- and outer-loop stages and then characterize the regret bound in Theorem 1.

4.1.2. The Inner Loop of Estimating $y_i^*(\lambda)$. The idea of the inner loop is from Chen et al. (2020b), who study joint pricing and inventory control of a single product with censored demand. As we mention earlier, $y_i^*(\lambda)$ is given by minimizing the convex function $C_i(y_i;\lambda)$; then, we can use certain data-driven binary search to find the zero point of $C_i'(y_i;\lambda)$. More specifically, we have

$$C_i'(y_i;\lambda) = (h_i + b_i - c_i(\lambda))F^i(y_i) - (b_i - c_i(\lambda)),$$

and if we have multiple samples of $(h_i + b_i - c_i(\lambda))\mathbf{1}(D_{i,t} \leq y_i) - (b_i - c_i(\lambda))$, which are obviously observable regardless of demand censoring, we can have an unbiased estimation of $C_i'(y_i;\lambda)$ by taking the average of these samples. In particular, let l denote the index of an inner loop, let n_l be the current number of samples of

Figure 2. (Color online) Double Binary Search Algorithm



Notes. Arrows represent information flow. The search durations increase with time. The calculation for the order-up-to levels is done in a decentralized manner, whereas the update for the λ is conducted based on the whole data.

$(h_i + b_i - c_i(\lambda))\mathbf{1}(D_{i,t} \le y_{i,l}) - (b_i - c_i(\lambda))$ by applying the same $y_{i,l}$, and let $\hat{g}_l$ denote the average of these samples (i.e., the estimated $C_i'(y_{i,l};\lambda)$). By concentration inequality, with high probability, we have $C_i'(y_{i,l};\lambda) \in [\underline{g}_l, \overline{g}_l]$ with high probability, where $\underline{g}_l = \hat{g}_l - \mathcal{C}_1/\sqrt{n_l}$ and $\overline{g}_l = \hat{g}_l + \mathcal{C}_1/\sqrt{n_l}$ for some parameter $\mathcal{C}_1$ to be specified later. As a result, if $\underline{g}_l > 0$ ($\overline{g}_l < 0$), we have $y_{i,l}$ too small (large) and we go to the next iteration in binary search by setting $y_{i,l}$ as the new lower bound (upper bound) of the search range. Otherwise, we keep applying $y_{i,l}$. This process is visualized in Figure 3, and we refer to Algorithm 1 for the details of the inner loop subroutine.

Algorithm 1 (Inner Loop of Binary Search for Order-up-to Levels)

```
Input: dual parameter $\lambda$, search duration $\nu$, parameter $\mathcal{C}_1$, current time period $t$, store $i$
Output: $y_{i,l}, \hat{d}_{i,l} = d_l/n_l$ for $l$ with the largest $n_l$
function SearchOrderUpTo($i, t, \lambda, \nu, \mathcal{C}_1$):
    Set $L_l = \underline{D}$, $R_l = \overline{D}$, $y_{i,l} = \overline{D}/2$, $l = 0$, $g_l = 0$, $d_l = 0$, $n_l = 0$
    while $\nu \ge \sum_l n_l$ and sufficient capacity at the central warehouse do
        Observe $z_t = \min(D_{i,t}, y_{i,l})$, where $t$ is the index of the current time period
        $d_l \leftarrow d_l + z_t$,   $n_l \leftarrow n_l + 1$,   $g_l \leftarrow g_l - (h_i + b_i - c_i(\lambda))\,\mathbf{1}(D_{i,t} \le y_{i,l}) + (b_i - c_i(\lambda))$
        Set $\hat{g}_l = \frac{g_l}{n_l}$ and $[\underline{g}_l, \overline{g}_l] = \hat{g}_l \pm \frac{\mathcal{C}_1}{\sqrt{n_l}}$
        if $l < \lceil \log_2(\nu\overline{D}) \rceil$ and $\underline{g}_l > 0$ then
            $L_{l+1} \leftarrow y_{i,l}$, $R_{l+1} \leftarrow R_l$, $y_{i,l+1} \leftarrow \frac{L_{l+1}+R_{l+1}}{2}$, $n_{l+1} \leftarrow 0$
            $l \leftarrow l + 1$
        else if $l < \lceil \log_2(\nu\overline{D}) \rceil$ and $\overline{g}_l < 0$ then
            $L_{l+1} \leftarrow L_l$, $R_{l+1} \leftarrow y_{i,l}$, $y_{i,l+1} \leftarrow \frac{L_{l+1}+R_{l+1}}{2}$, $n_{l+1} \leftarrow 0$
            $l \leftarrow l + 1$
        end
        $t \leftarrow t + 1$
    end
end function.
```

4.1.3. The Outer Loop of Estimating λ^*.

Our main algorithm DBS consists of outer loops of finding optimal Lagrangian dual variable λ^*. Let τ denote the index of the outer loops. In each loop τ with dual variable $\hat{\lambda}_\tau$, which is an estimation of λ^*, we apply the subroutine Algorithm 1 to obtain approximated $y_i^*(\hat{\lambda}_\tau)$ for all $i \in [N]$. The idea of estimating λ^*, as discussed earlier, is based on the fact that $\tilde{V}^\lambda$ is a concave function of λ with a gradient equal to

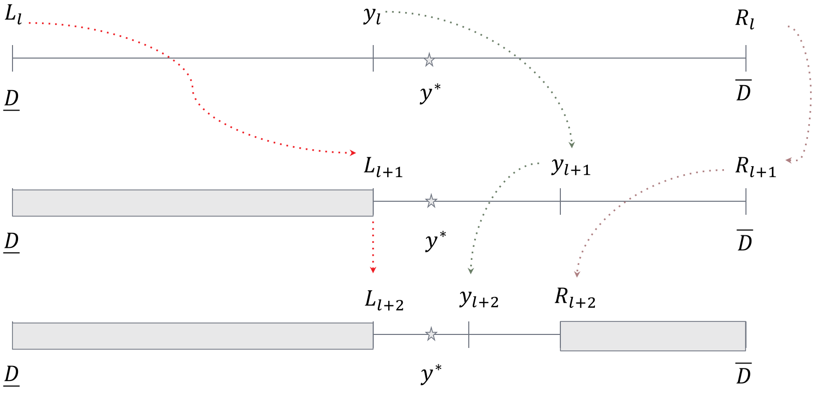
$$\sum_i \mathbb{E}[\min(y_i(\lambda^*), D_{i,t})] - \frac{W}{T},$$

according to lemma 10 of Miao et al. (2022) and Assumption 1. The method for estimating λ^* is still a binary search algorithm albeit different from the one in Algorithm 1. Recall that in (4), we define time to hit zero (denoted by T_0) by setting the gradient of $\tilde{V}^\lambda$ to zero and solving for T. Therefore, searching for λ that satisfies $T = T_0$ is equivalent to finding the zero of the (approximated) gradient of $\tilde{V}^\lambda$. More specifically, let $\hat{y}_{i;\tau}, \hat{d}_{i;\tau}$ be the output of subroutine Algorithm 1. Because $\hat{d}_{i;\tau}$ is the sample average of

$$\min(\hat{y}_{i;\tau}, D_{i,t}),$$

for all time periods t of applying base-stock level $\hat{y}_{i;\tau}$ in loop τ, and $\hat{y}_{i;\tau}$ approximates $y_i^*(\hat{\lambda}_\tau)$, we define $v_\tau = \sum_i \hat{d}_{i;\tau} - W/T$, which is an approximated gradient of $\tilde{V}^{\hat{\lambda}_\tau}$, and define the confidence interval around v_τ as $[\underline{v}_\tau, \overline{v}_\tau] = v_l \pm \mathcal{C}_2 N/\sqrt{\nu_\tau}$ for some parameter $\mathcal{C}_2$ such that the true gradient of $\tilde{V}^{\hat{\lambda}_\tau}$ is in $[\underline{v}_\tau, \overline{v}_\tau]$ with high probability. By concavity of $\tilde{V}^\lambda$, $\underline{v}_\tau \ge 0$ ($\overline{v}_\tau \le 0$) implies $\lambda^* \ge \hat{\lambda}_\tau$ ($\lambda^* \le \hat{\lambda}_\tau$), so we can cut the search range of λ^* in half. Otherwise, it means that λ^* is within a close range (with a similar length as $[\underline{v}_\tau, \overline{v}_\tau]$) of $\hat{\lambda}_\tau$, and we can immediately narrow our search range accordingly. The length of each outer loop τ is given by $\mathcal{C}_0\beta^\tau$, where $\mathcal{C}_0$ and β are

Figure 3. (Color online) Three Consecutive Steps of the Inner Loop



Note. After sampling enough data, the confidence interval for the derivative excludes zero, and the algorithm proceeds to the next step by halving the search range.

parameters to be determined later. We refer to Figure 4 for a graphic representation of the outer loop.

Note that the value of these learning parameters has to be determined in a nontrivial way, and we cannot directly use the same binary search as Algorithm 1 in order to achieve the optimal learning rate of λ^* (see Online Lemma EC.2). The detailed algorithm DBS is given in Algorithm 2.

Algorithm 2 (DBS)

```
Input: All model parameters for the OWMS problem; parameters $\mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3 > 0, \beta > 1$;
Initialize. $\tau = 0$; $t = 0$; $L_\tau = 0$; $U_\tau = \min_i(b_i - c_i)$; $\hat{\lambda}_\tau = 0$;
while $T$ periods have not been reached and warehouse did not run out of stock do
    Set $\nu_\tau = \mathcal{C}_0 \beta^\tau$;
    $\hat{y}_{i;\tau}, \hat{d}_{i;\tau} \leftarrow$ SearchOrderUpTo($i, t, \hat{\lambda}_\tau, \nu_\tau, \mathcal{C}_1$) for all $i \in [N]$; define $v_\tau = \sum_i \hat{d}_{i;\tau} - W/T$;
    Construct confidence interval $[\underline{v}_\tau, \overline{v}_\tau] = v_\tau \pm \mathcal{C}_2 N/\sqrt{\nu_\tau}$;
    if $\underline{v}_\tau \geq 0$ then
        Update $L_{\tau+1} \leftarrow \hat{\lambda}_\tau$, $U_{\tau+1} \leftarrow U_\tau$, $\hat{\lambda}_{\tau+1} \leftarrow (L_{\tau+1} + U_{\tau+1})/2$;
    end
    else if $\overline{v}_\tau \leq 0$ then
        Update $L_{\tau+1} \leftarrow L_\tau$, $U_{\tau+1} \leftarrow \hat{\lambda}_\tau$, $\hat{\lambda}_{\tau+1} \leftarrow (L_{\tau+1} + U_{\tau+1})/2$;
    end
    else
        Update $L_{\tau+1} \leftarrow \mathrm{Proj}_{[0, \min_i(b_i - c_i)]}(\hat{\lambda}_\tau - \mathcal{C}_3/\sqrt{\nu_\tau})$, $U_{\tau+1} \leftarrow \mathrm{Proj}_{[0, \min_i(b_i - c_i)]}(\hat{\lambda}_\tau + \mathcal{C}_3/\sqrt{\nu_\tau})$, $\hat{\lambda}_{\tau+1} \leftarrow (L_{\tau+1} + U_{\tau+1})/2$;
    end
    $\tau \leftarrow \tau + 1$;
    $t \leftarrow t + \nu_\tau$;
end.
```

The performance of our algorithm DBS is presented in the following theorem with its proof in the next section.

Theorem 1. *Let π^{DBS} be the policy applied by the DBS algorithm with*

$$\mathcal{C}_0 \geq \max(4/\beta^2, 2\lceil \log_2(T\overline{D}) \rceil)$$
$$\mathcal{C}_1 \geq \sqrt{3/2}(h_i + b_i - c_i)\sqrt{\log(T)}$$
$$\mathcal{C}_2 \geq \left(\overline{D} + \frac{\mathcal{C}_1}{h_i \underline{f}}\right)\sqrt{2\log(NT)\lceil \log_2(T\overline{D}) \rceil}$$
$$\mathcal{C}_3 \geq \max\left(\sqrt{C_0}\min_i(b_i - c_i), 2\mathcal{C}_2 / \min_i \frac{h_i^2}{(h_i + b_i - c_i)^3 \overline{f}}\right)$$
$$\beta \in (1, 4].$$

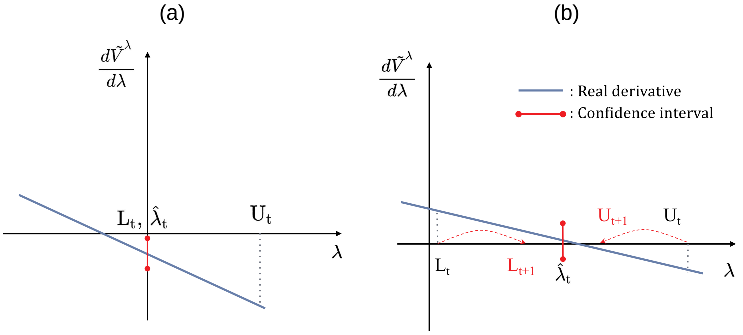
Under Assumptions 1 and 2, the regret of π^{DBS} satisfies that

$$\mathcal{R}(\pi^{\text{DBS}}) \leq \tilde{O}(N\sqrt{T}).$$

From Theorem 1, the performance of our algorithms with demand censoring is $\tilde{O}(N\sqrt{T})$. Therefore, the first question is whether this is the best rate we can achieve. First of all, it is unknown whether $\tilde{O}(N\sqrt{T})$ is optimal because we could not derive a matching lower bound of $\Omega(N\sqrt{T})$. The only regret lower bound we currently know is $\Omega(N \cdot \log(T))$, which is based on the result in Besbes and Muharremoglu (2013) given that there is more than enough inventory in the warehouse (a special case of our problem). Our conjecture is that $\Omega(N \cdot \log(T))$ is the optimal lower bound for our problem. The reason is that, in Chao et al. (2021), the authors prove that, when demand is known, policy readjustments of LaBS (the heuristic we use given learned demand distribution) can lead to a regret of $O(N \cdot \log(T))$. However, how to incorporate policy readjustments into our learning algorithm is challenging, and we, thus, leave this problem as a future research opportunity.

Another point to note is that the result in Theorem 1 is based on Assumption 2. Thus, we wonder what happens when this assumption is relaxed. Without loss of generality, let us assume $b_1 - c_1 \leq b_2 - c_2 \leq \cdots \leq b_N - c_N$ and denote $b'_1 < b'_2 < \cdots < b'_N$ the N' distinct values of

Figure 4. (Color online) Two Distinct Principles of the DBS Algorithm



Notes. (a) The algorithm starts with $\hat{\lambda} = 0$ to increase the effectiveness in the case of abundant inventory. If the value of λ^* is zero, the algorithm can easily commit to zero; otherwise, it can continue to search on the rest of the domain. (b) Different from the classical stochastic binary search, the DBS algorithm shrinks the search space even though the confidence interval does not have enough evidence for halving. (a) $\lambda^* = 0$, and the confidence interval already excluded zero. (b) The confidence interval cannot indicate a sign.

$b_i - c_i$. Suppose λ^* is in between some b'_k and b'_{k+1} (instead of between 0 and b'_1 as suggested in Assumption 2); we have the regret of our algorithm DBS equal to $\mathcal{R}(\pi^{\mathrm{DBS}}) \le \tilde{O}(N\sqrt{T} + N^3)$ with $\beta \in (1,2]$ and other parameters the same as the original (see Online Section EC.3 for details). To achieve this, we have to make the following changes. First, the definition of $[\underline{\nu}_\tau, \overline{\nu}_\tau]$ becomes $[\underline{\nu}_\tau, \overline{\nu}_\tau] = \nu_l \pm \mathcal{C}_2 N_\tau/\sqrt{\nu_\tau}$, where $N_\tau := |\{i \in [N] : c_i + \hat{\lambda}_\tau < b_i\}|$. Second, in line 2 in DBS, when we update $L_{\tau+1}$ and $U_{\tau+1}$, the deviation from $\hat{\lambda}_\tau$ is $\mathcal{C}_3 N_\tau/\sqrt{\nu_\tau}$ instead of $\mathcal{C}_3/\sqrt{\nu_\tau}$. Third, for the inner loop search of i with $c_i + \hat{\lambda}_\tau \ge b_i$, we just let $y_{i,t} = 0$ during the whole inner loop as the ordering cost is already more than the lost-sale cost. A crucial step in proving Theorem 1 is Online Lemma EC.2, which proves the convergence of $\hat{\lambda}_\tau$ to λ^*. It is important to show that this convergence is independent of N, and this is precisely the place where we use Assumption 2. In Online Section EC.3, we prove a slightly modified version of Online Lemma EC.2 in the relaxed setting.

4.2. Main Steps of the Proof of Theorem 1

By our definition of regret, we can write

$$\mathcal{R}(\pi^{\mathrm{DBS}}) \le V^{DBS} - V^{DBS}_{unlim} + V^{DBS}_{unlim} - \tilde{V}^{\lambda^*}, \tag{5}$$

where V^{DBS}_{unlim} is the cost of applying Algorithm 2 in a hypothetical system with the same cost structure as the original OWMS problem except without inventory constraint in the central warehouse. Therefore, our regret is bounded by $V^{DBS} - V^{DBS}_{unlim}$ and $V^{DBS}_{unlim} - \tilde{V}^{\lambda^*}$, respectively.

4.2.1. Bound of $V^{DBS} - V^{DBS}_{unlim}$.

Let us first consider the first term. Note that the discrepancy between $V^{DBS} - V^{DBS}_{unlim}$ comes only from the possible lost sales because of stockout in the central warehouse. As a result,

$$\begin{aligned} V^{DBS} - V^{DBS}_{unlim} &\le \bar{b}\mathbb{E}\left[\left(\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t} - I_{i,t})^+ - W\right)^+\right] \\ &\le \bar{b}\mathbb{E}\left[\left|\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t} - I_{i,t})^+ \right.\right. \\ &\qquad \left.\left. - \sum_{i=1}^{N} T \cdot \mathbb{E}[\min\{y_i^*(\lambda^*), D_{i,1}\}]\right|\right], \end{aligned} \tag{6}$$

where the second inequality is from lemma 10 in Miao et al. (2022) because $\sum_{i=1}^{N} T \cdot \mathbb{E}[\min\{y_i^*(\lambda^*), D_{i,1}\}] = \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbb{E}[(y_i^*(\lambda^*) - I_{i,t})^+] - \sum_{i=1}^{N} \mathbb{E}[(y_i^*(\lambda^*) - D_{i,T})^+] \le W$, $\bar{b} = \max_i b_i$, and $y_{i,t}$ denotes the actual base-stock level from algorithm DBS. Denote $Z_{i,t} = (y_{i,t} - I_{i,t})^+$ as the ordering in time t at store i and $\bar{\tau}$ the index of the last outer loop. By our algorithm design of exponentially increasing cycles τ, we have

$$\begin{aligned} &\mathbb{E}\left[\left|\sum_{i=1}^{N}\sum_{t=1}^{T}(y_{i,t} - I_{i,t})^+ - \sum_{i=1}^{N} T \cdot \mathbb{E}[\min\{y_i^*(\lambda^*), D_{i,1}\}]\right|\right] \\ &= \mathbb{E}\left[\left|\sum_{i=1}^{N}\sum_{\tau=0}^{\bar{\tau}}\sum_{t\in\mathcal{T}_\tau} Z_{i,t} - \sum_{i=1}^{N}\sum_{\tau=0}^{\bar{\tau}} \nu_\tau \cdot \mathbb{E}[\min\{y_i^*(\lambda^*), D_{i,1}\}]\right|\right] \\ &\le \sum_{i=1}^{N}\sum_{\tau=0}^{\bar{\tau}} \mathbb{E}\left[\left|\sum_{t\in\mathcal{T}_\tau} Z_{i,t} - \nu_\tau \cdot \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}]\right|\right] \\ &\quad + \sum_{i=1}^{N}\sum_{\tau=0}^{\bar{\tau}} \mathbb{E}[|\nu_\tau \cdot \mathbb{E}[\min\{y_i^*(\lambda^*), D_{i,1}\}] \\ &\qquad - \nu_\tau \cdot \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}]|] \\ &\le \sum_{\tau=0}^{\bar{\tau}}\left(K_1 N \log(NT)\sqrt{\mathcal{C}_0}\beta^{\tau/2} + K_2 N\right) \\ &\quad + \sum_{i=1}^{N}\sum_{\tau=0}^{\bar{\tau}} \mathbb{E}[|\nu_\tau \cdot \mathbb{E}[\min\{y_i^*(\lambda^*), D_{i,1}\}] \\ &\qquad - \nu_\tau \cdot \mathbb{E}[\min\{y_i^*(\hat{\lambda}_\tau), D_{i,1}\}]|] \\ &\le \sum_{\tau=0}^{\bar{\tau}}\left(K_1 N \log(NT)\sqrt{\mathcal{C}_0}\beta^{\tau/2} + K_2 N + N\mathcal{C}_0\mathcal{C}_3\beta^{\tau/2+3/2}\right) \\ &\le K_1 \log(NT) N\sqrt{T} + K_2 N\sqrt{T} + \mathcal{C}_0\mathcal{C}_3\beta^{3/2} N\sqrt{T}, \end{aligned}$$

where $\mathcal{T}_\tau$ is the set of time periods in cycle τ so that $\nu_\tau = |\mathcal{T}_\tau|$ and $\bar{\tau} \le \lceil \log_\beta(T) \rceil$ is the index of the last cycle. The second inequality is from Online Lemma EC.1(iv). The constants K_1 and K_2 are given in Online Lemma EC.1(iv). The third inequality is from Online Lemma EC.2. Combining this with (6), we have that

$$\begin{aligned} V^{DBS} - V^{DBS}_{unlim} &\le K_1 \bar{b} \log(NT) N\sqrt{T} + K_2 \bar{b} N\sqrt{T} \\ &\quad + \mathcal{C}_0\mathcal{C}_3\bar{b}\beta^{3/2} N\sqrt{T} + N\overline{D}. \end{aligned} \tag{7}$$

4.2.2. Bound of $V^{DBS}_{unlim} - \tilde{V}^{\lambda^*}$.

For the part $V^{DBS}_{unlim} - \tilde{V}^{\lambda^*}$, we first note that,

$$\begin{aligned} V^{DBS}_{unlim} &= \sum_{i=1}^{N}\sum_{\tau=0}^{\bar{\tau}}\sum_{t\in\mathcal{T}_\tau} \mathbb{E}[C_i(\max(y_{i,t}, I_{i,t}); \hat{\lambda}_\tau)] \\ &\le \sum_{i=1}^{N}\sum_{\tau=0}^{\bar{\tau}}\left(\sum_{t\in\mathcal{T}_\tau} C_i(y_{i,t}; \hat{\lambda}_\tau) + \bar{b}\overline{D}\lceil \log_2(\nu_\tau \overline{D}) \rceil\right) + wW \\ &\quad - \sum_{i=1}^{N}\sum_{\tau=0}^{\bar{\tau}}\sum_{t\in\mathcal{T}_\tau} \mathbb{E}[\hat{\lambda}_\tau (y_{i,t} - I_{i,t})^+], \end{aligned}$$

where the inequality is because there are at most $\lceil \log_2(\nu_\tau \overline{D}) \rceil$ updates of $y_{i,t}$ for $t \in \mathcal{T}_\tau$ and the assumption that $D_{i,t} \geq \underline{D} > 0$. As a result, we have

$$
\begin{aligned}
&V^{DBS}_{unlim} - \tilde{V}^{\lambda^*}\\
&\leq \sum_{i=1}^{N}\sum_{\tau=0}^{\overline{\tau}}\sum_{t\in\mathcal{T}_\tau}(C_i(y_{i,t};\hat{\lambda}_\tau) - C_i(y_i^*(\hat{\lambda}_\tau);\hat{\lambda}_\tau))\\
&\quad + \sum_{i=1}^{N}\sum_{\tau=0}^{\overline{\tau}}\nu_\tau(C_i(y_i^*(\hat{\lambda}_\tau);\hat{\lambda}_\tau) - C_i(y_i^*(\lambda^*);\lambda^*))\\
&\quad + \sum_{i=1}^{N}\sum_{\tau=0}^{\overline{\tau}}\Bigg(\nu_\tau\lambda^*\mathbb{E}[\min\{y_i^*(\lambda^*),D_{i,t}\}]\\
&\quad - \sum_{t\in\mathcal{T}_\tau}\mathbb{E}[\hat{\lambda}_\tau(y_{i,t}-I_{i,t})^+]\Bigg) + \overline{b}\overline{D}\lceil\log_2(T\overline{D})\rceil N. \qquad (8)
\end{aligned}
$$

Therefore, we need to bound the three summations, respectively.

First, we have, with probability at least $1 - O(T^{-1})$,

$$
\begin{aligned}
&\sum_{i=1}^{N}\sum_{\tau=0}^{\overline{\tau}}\sum_{t\in\mathcal{T}_\tau}(C_i(y_{i,t};\hat{\lambda}_\tau) - C_i(y_i^*(\hat{\lambda}_\tau);\hat{\lambda}_\tau))\\
&\leq \sum_{\tau=0}^{\overline{\tau}}N\log_2(\nu_\tau\overline{D})(\mathcal{C}_1^2\underline{f}^{-1}h_i\log(\nu_\tau) + (h_i+b_i-c_i) + \overline{b}\overline{D})\\
&\leq \log_2(T\overline{D})(\mathcal{C}_1^2\underline{f}^{-1}h_i\log(T) + (h_i+b_i-c_i) + \overline{b}\overline{D})N, \qquad (9)
\end{aligned}
$$

according to Online Lemma EC.1.(i) by applying a union bound over all τ.

Second, with probability at least $1 - O(T^{-1})$,

$$
\begin{aligned}
&\sum_{i=1}^{N}\sum_{\tau=0}^{\overline{\tau}}\nu_\tau(C_i(y_i^*(\hat{\lambda}_\tau);\hat{\lambda}_\tau) - C_i(y_i^*(\lambda^*);\lambda^*))\\
&\leq \sum_{i=1}^{N}\sum_{\tau=0}^{\overline{\tau}}\nu_\tau L\mathbb{E}[|\hat{\lambda}_\tau - \lambda^*|]\\
&\leq \sum_{i=1}^{N}\sum_{\tau=0}^{\overline{\tau}}\mathcal{C}_0\mathcal{C}_3L\beta^{\tau/2+3/2}\\
&\leq \mathcal{C}_0\mathcal{C}_3LN\sqrt{T}, \qquad (10)
\end{aligned}
$$

where the first inequality is from Lipschitz continuity of $C_i(y_i^*(\lambda);\lambda)$ with respect to λ, and the second inequality is from Online Lemma EC.2.

Third, with probability at least $1 - O(T^{-1})$,

$$
\begin{aligned}
&\sum_{i=1}^{N}\sum_{\tau=0}^{\overline{\tau}}\Bigg(\nu_\tau\lambda^*\mathbb{E}[\min\{y_i^*(\lambda^*),D_{i,t}\}] - \sum_{t\in\mathcal{T}_\tau}\mathbb{E}[\hat{\lambda}_\tau(y_{i,t}-I_{i,t})^+]\Bigg)\\
&= \sum_{i=1}^{N}\sum_{\tau=0}^{\overline{\tau}}(\nu_\tau\lambda^*\mathbb{E}[\min\{y_i^*(\lambda^*),D_{i,t}\}]\\
&\qquad - \nu_\tau\mathbb{E}[\hat{\lambda}_\tau\min\{y_i^*(\hat{\lambda}_\tau),D_{i,t}\}])\\
&\quad + \sum_{i=1}^{N}\sum_{\tau=0}^{\overline{\tau}}\Bigg(\nu_\tau\mathbb{E}[\hat{\lambda}_\tau\min\{y_i^*(\hat{\lambda}_\tau),D_{i,t}\}]\\
&\qquad - \sum_{t\in\mathcal{T}_\tau}\mathbb{E}[\hat{\lambda}_\tau(y_{i,t}-I_{i,t})^+]\Bigg)\\
&\leq \sum_{i=1}^{N}\sum_{\tau=0}^{\overline{\tau}}\nu_\tau L\mathbb{E}[|\hat{\lambda}_\tau - \lambda^*|]\\
&\quad + \sum_{i=1}^{N}\sum_{\tau=0}^{\overline{\tau}}\mathbb{E}\left[\hat{\lambda}_\tau\left(\nu_\tau\min\{y_i^*(\hat{\lambda}_\tau),D_{i,t}\} - \sum_{t\in\mathcal{T}_\tau}Z_{i,t}\right)\right]\\
&\leq \sum_{i=1}^{N}\sum_{\tau=0}^{\overline{\tau}-1}\mathcal{C}_0\mathcal{C}_3L\beta^{\tau/2+3/2} + \sum_{\tau=0}^{\overline{\tau}}NK_1\log(NT)\sqrt{\nu_\tau} + NK_2\\
&\leq \mathcal{C}_0\mathcal{C}_3LN\sqrt{T} + \mathcal{C}_0NK_1\log(NT)\sqrt{T} + NK_2, \qquad (11)
\end{aligned}
$$

where the first inequality is from Lipschitz continuity, the second inequality is from Online Lemmas EC.2 and EC.1(iv) with union bound over τ, and constants K_1 and K_2 are given in Online Lemma EC.1(iv). Combining (8)–(11), we have

$$
\begin{aligned}
&V^{DBS}_{unlim} - \tilde{V}^{\lambda^*}\\
&\leq \log_2(T\overline{D})(\mathcal{C}_1^2\underline{f}^{-1}h_i\log(T) + (h_i+b_i-c_i) + \overline{b}\overline{D})N\\
&\quad + \overline{b}\overline{D}\lceil\log_2(T\overline{D})\rceil N + 2\mathcal{C}_0\mathcal{C}_3LN\sqrt{T} + \mathcal{C}_0NK_1\log(NT)\sqrt{T}\\
&\quad + NK_2. \qquad (12)
\end{aligned}
$$

In the end, this theorem is proved by combining (5), (6), and (12).

5. Extensions of the Proposed Algorithm

Before we numerically evaluate the performance of the DBS algorithm based on both synthetic and real data sets, we discuss two modifications. Recall that we developed the DBS algorithm under the assumption that the demand information is censored, and underlying demand follows a continuous and bounded distribution. In this section, we discuss how the DBS would be modified if we have uncensored demand, discrete demand distribution, and unbounded demand distribution.

5.1. Algorithm Under Uncensored Demand Observations

Even though real life is characterized by censored demand cases, we briefly describe the case of uncensored demand. Recall that the main challenge under the censored demand case is the inability to estimate both order-up-to levels and the dual variable in an unbiased fashion. To address this challenge, we developed the DBS method in which we actively search for the optimal dual variable in the outer loop and order-up-to level in the inner loop. However, in the case of uncensored demand, we can eliminate both inner and outer loops and simultaneously generate unbiased estimates for the dual variable and order-up-to levels from demand observations. Delegating the details to the online appendix, we can achieve this by constructing an empirical estimation $\hat{F}^i_t(\cdot)$ of the demand distribution $F^i(\cdot)$. More specifically, we have

$$\hat{F}^i_t(d) := \frac{\sum_{s=1}^{t-1} \mathbf{1}(D_{i,s} \le d)}{t-1}.$$

According to the Dvoretzky–Kiefer–Wolfowitz inequality (see Massart 1990), as t grows, $\hat{F}^i_t(\cdot)$ becomes more and more accurate (see Lemma 1).

Lemma 1 (Dvoretzky–Kiefer–Wolfowitz Inequality). *For any $\epsilon \ge 0$ and any period $t \in [T]$,*

$$P\left(\sup_{d \in [\underline{D}, \overline{D}]} |\hat{F}^i_{t+1}(d) - F^i(d)| \ge \epsilon \right) \le 2\exp(-2t\epsilon^2). \quad (13)$$

Once we have this empirical demand distribution, without having to search for T_τ periods, we can estimate $y^*_i(\lambda)$ directly for any λ by solving the following:

$$\hat{y}_{i,t}(\lambda) := \inf\{y : \hat{F}^i_t(y) \ge \kappa_i(\lambda)\}. \quad (14)$$

Similarly, we can estimate the dual variable $\hat{\lambda}_t$ by solving the following "empirical" version of the complementary slackness:

$$\lambda \left(\sum_i \hat{\mathbb{E}}[\min(\hat{y}_{i,t}(\lambda), \hat{D}_{i,t})] - \frac{W}{T} \right) = 0, \quad (15)$$

where $\hat{D}_{i,t}$ is the demand with empirical distribution and $\hat{\mathbb{E}}[\cdot]$ represents the expectation according to the empirical distribution. Because the left-hand side of (15) is a left-continuous stepwise function, we let

$$\hat{\lambda}_t := \sup\left\{ \lambda \ge 0 : \sum_i \hat{\mathbb{E}}[\min(\hat{y}_{i,t}(\lambda), \hat{D}_{i,t})] - \frac{W}{T} \ge 0 \right\}. \quad (16)$$

The modified algorithm, which we refer to as empirical distribution algorithm for full demand (EDAF), then simply in each period updates the empirical distribution and generates estimates for order-up-to levels and the dual variable by directly solving (14) and (16), respectively. Because of space limitations, we present the detailed algorithm in Online Section EC.4 and state its performance herein (with its proof in Online Section EC.4 as well). Moreover, we conduct extra numerical experiments of EDAF in Online Section EC.7.2.

Proposition 1. *Let π^{EDAF} be the policy constructed by the EDAF algorithm. Under Assumptions 1 and 2, the regret of π^{EDAF} satisfies that*

$$\mathcal{R}(\pi^{\text{EDAF}}) \le \tilde{O}(N\sqrt{T}).$$

5.2. Algorithm with Discrete Demand Distribution

Recall that, in Assumption 1, we consider that the underlying demand is characterized by a continuous distribution. In this section, we have the following modified version of Assumption 1.

Assumption 3. *Demand $D_{i,t} \in J := \{\underline{D}, \underline{D}+1, \ldots, \overline{D}\}$, where $0 < \underline{D} < \overline{D} < \infty$ are two integers. Its density function satisfies $0 < \underline{f} \le m^i(d) \le \bar{f} < \infty$ for all $i \in [N]$ and $d \in J$, where $m^i(\cdot)$ is the probability mass function.*

Various issues are different under the discrete demand assumption. First, under any dual variable λ, the cost function $C_i(y;\lambda)$ is a piecewise linear and convex function with J defined as the "turning points" (i.e., the value of y that is the intersection of two pieces of linear function). This structure is shown in Figure 5. It is highly likely that the minimizer of $C_i(y;\lambda)$ is not unique in this case. In particular, we have $y^*_i(\lambda)$ defined as

$$\min\{y : (h_i + b_i - c_i(\lambda))F^i(y) > b_i - c_i(\lambda)\},$$

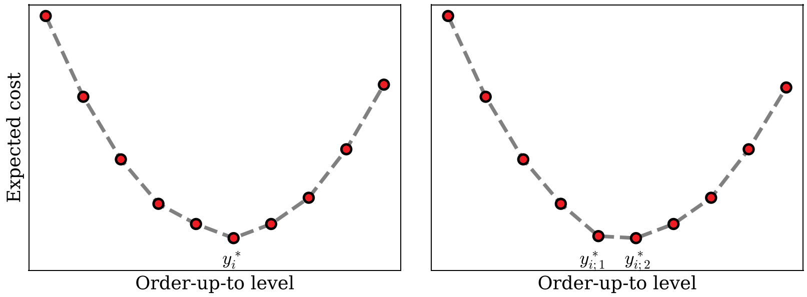
if there is no y satisfying $(h_i + b_i - c_i(\lambda))F^i(y) = b_i - c_i(\lambda)$; otherwise, $y^*_i(\lambda)$ is somewhere in between $y^*_{i;1}(\lambda) = \min\{y : (h_i + b_i - c_i(\lambda))F^i(y) = b_i - c_i(\lambda)\}$ and $y^*_{i;2}(\lambda) = \min\{y : (h_i + b_i - c_i(\lambda))F^i(y) > b_i - c_i(\lambda)\}$ so that the complementary slackness condition is satisfied. Second, it is easy to see that V^λ is also piecewise linear and concave in λ.

Our original algorithm DBS and its analysis cannot be easily extended to the case of discrete demand because of these peculiarities. In technical analysis, there are some significant differences. First, the DBS seeks to locate the unique zero of the gradient of $\tilde{V}\lambda$ in order to discover the best λ^* for the case of continuous demand, for which the gradient is defined as

$$\sum_i \mathbb{E}[\min(y^*_i(\lambda), D_{i,t})] - \frac{W}{T},$$

for unique $y^*_i(\lambda)$. However, the reason we can achieve the optimal convergence rate of $\hat{\lambda}_\tau$ to λ^* is that V^λ (for continuous demand) is continuously differentiable and strongly concave, but neither of these properties applies to discrete demand. Second, for a fixed λ, $y^*_i(\lambda)$ is the unique minimizer of $C_i(y;\lambda)$, which is strongly convex for continuous demand. Consequently, the convergence

Figure 5. (Color online) Discrete Cost Function



Notes. On the left, the unique minimizer of the expected cost function can easily be found. On the right, it is difficult to detect the minimizer.

of our inner loop search relies on these properties, which do not apply to discrete demand either.

As a result, in order to obtain a similar $\tilde{O}(N\sqrt{T})$ regret upper bound, we must devise a new algorithm that employs significantly different approaches and exploits the discrete structure. More specifically, we apply an initial sampling that allows us to simulate all policies and eliminate suboptimal policies accurately. On the warehouse constraint, we perform a binary search over the set of all optimal order-up-to levels to find the one that satisfies complementary slackness. Based on these principles, we develop a modified algorithm, referred to as the discrete demand (DD) algorithm. Because the modified DD algorithm is rather complex, we defer its technical analysis to Online Section EC.5 and numerical experiments to Online Section EC.7.4.

5.3. Algorithm Under Unbounded Demand

In Assumption 1, we assume that the demand is bounded in $[\underline{D}, \overline{D}]$ almost surely. In this section, we explain how the algorithm's initial parameters should be adjusted and the effect of these changes on the regret when the demand is unbounded, that is, $\overline{D} = +\infty$ and $\underline{D} = 0$.

There are two main purposes for these assumptions. First, $\underline{D} > 0$ is to guarantee that the inventory goes down to zero in a finite number of time periods when there is no replenishment. Therefore, even if the support of demand distribution includes zero, as long as the mean demand is nonzero, that is, $\mathbb{E}[D_{i,t}] > 0$ (which is true in most applications) and $D_{i,t}$ is sub-Gaussian, we can still show that any finite inventory is consumed within $O(\log(NT))$ time periods with probability at least $1 - O(N^{-1}T^{-1})$. As a result, relaxing $\underline{D} > 0$ only has an impact on our final regret in the order of a logarithmic factor of T. Second, $\overline{D} < \infty$ is to ensure that the optimal order-up-to level $y_i^*(\lambda^*)$ is finite. When the demand is unbounded, we can still ensure $y_i^*(\lambda^*) \le \overline{y}$ for some constant $\overline{y}$ under various conditions (which are discussed later).

To summarize, after relaxing the boundedness assumption, Assumption 1 can be relaxed as follows.

Assumption 1′. *Demand $D_{i,t} \in [0, \infty)$ almost surely, and it is sub-Gaussian with parameter σ (i.e., $\mathbb{E}[\exp(sD_{i,t})] \le \exp(\sigma^2 s^2/2)$ for all $s \in \mathbb{R}$). Its distribution $F^i : [0, \infty) \to [0, 1]$ is continuous, and it has a probability density function f^i with $0 < \underline{f} \le f_i(d) \le \overline{f} < \infty$ for all $i \in [N]$ and $d \in [0, \overline{y}]$, where $\overline{y}$ is any constant satisfying $\max_{i\in[N]} y_i^*(\lambda^*) \le \overline{y}$. The parameters $\overline{y}, \underline{f}, \overline{f}$ are known.*

Now, in what follows, we discuss two special cases under which we can establish a finite bound $\overline{y}$ on the inventory order-up-to level, which is known to the retailer.

Example 1. By definition,

$$y_i^*(\lambda^*) = (F^i)^{-1}\left(\frac{b_i - c_i - \lambda^*}{h_i + b_i - c_i - \lambda^*}\right)$$

$$\le \max_{i\in[N]} (F^i)^{-1}\left(\frac{b_i - c_i}{h_i + b_i - c_i}\right) < \infty$$

because $h_i > 0$ for all $i \in [N]$; thus, $\overline{y}$ is any upper bound of $\max_{i\in[N]}(F^i)^{-1}((b_i - c_i)/(h_i + b_i - c_i))$. In many applications, although we do not know the specific distribution of F^i, it is not impossible to establish a finite upper bound on $(F^i)^{-1}((b_i - c_i)/(h_i + b_i - c_i))$. For instance, consider that F^i belongs to some parametric distribution family (e.g., (truncated) normal distribution). Then, one can establish a finite upper bound on $(F^i)^{-1}((b_i - c_i)/(h_i + b_i - c_i))$ if the parameters are known to be bounded in a certain compact set based on historical data and insights.

Example 2. Recall that we need to satisfy complementary slackness, which requires that

$$\sum_{i\in[N]} \mathbb{E}[\min\{y_i^*(\lambda^*), D_{i,1}\}] \le W/T.$$

According to Assumption 1′, it is not difficult to show that $\mathbb{E}[\min\{y_i^*(\lambda^*), D_{i,1}\}] \ge \underline{f}(y_i^*(\lambda^*))^2/2$. Therefore, another choice of $\overline{y}$ is $\overline{y} = \sqrt{2W/(T\underline{f})}$. Note that, in many real practices, W scales at most linearly with T, and thus, $\overline{y}$ can be considered as a constant.

With this modified assumption, the algorithm and its analysis remain the same except we essentially replace every $\overline{D}$ with $\overline{y}$. We give the performance of the algorithm with these modifications in Corollary 1 and the proof of Corollary 1 in Online Section EC.6.

Corollary 1. *Let π^{DBS} be the policy applied by the DBS algorithm with an ordering upper bound $\overline{y}$ and*

$$\mathcal{C}_0 \ge \max(4/\beta^2, 2\lceil\log_2(T\overline{y})\rceil)$$

$$\mathcal{C}_1 \ge \sqrt{3/2}(h_i + b_i - c_i)\sqrt{\log(T)}$$

$$\mathcal{C}_2 \ge \left(\overline{y} + \frac{\mathcal{C}_1}{h_i\underline{f}}\right)\sqrt{2\log(NT)\lceil\log_2(T\overline{y})\rceil}$$

$$\mathcal{C}_3 \ge \max\left(\sqrt{\mathcal{C}_0}\min_i(b_i - c_i), 2\mathcal{C}_2/\min_i\frac{h_i^2}{(h_i + b_i - c_i)^3\overline{f}}\right)$$

$$\beta \in (1, 4].$$

Under Assumptions 1′ and 2, the regret of π^{DBS} satisfies that

$$\mathcal{R}(\pi^{\mathrm{DBS}}) \le \tilde{O}(N\sqrt{T}).$$

6. Numerical Study

In this section, we present the performance of the DBS algorithm in two groups of experiments. In Experiment 1, using synthetic data, we perform a sensitivity analysis with respect to different model parameters. In Experiment 2, we make use of a unique data set collected from our partner fast-fashion company and test the performance of the DBS algorithm on six sets of products grouped according to their price and sales volume. In both Experiments 1 and 2, the algorithms run in 100 different random seeds, and their average values and standard errors are reported.

As for the benchmark algorithms, we use the explore-then-commit (Exp) algorithm, which has two phases. In the exploration phase, the Exp algorithm raises the inventory levels to $\overline{D}$ in the stores to get uncensored demand samples. In the exploitation phase, it applies the best policy calculated from these unbiased data. As shown in Online Section EC.7.1, there is a trade-off between the exploration and exploitation phases. Whereas the Exp algorithm learns the true demand distribution more and more accurately as the length of the exploration phase increases, its performance degrades because it does not spend enough time to exploit the policy tuned for true demand parameters. Therefore, we pick three different exploration durations (i.e., $T^{1/2}$, $T^{2/3}$, and $T^{3/4}$) that yield the best compromise between demand accuracy and performance. For a better comparison of the performances in different scenarios, we define the relative regret for a policy π as

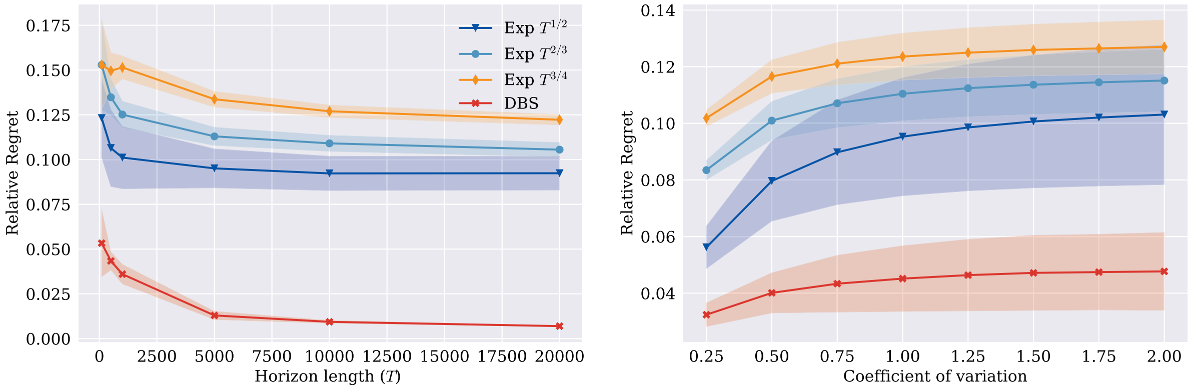
$$\mathcal{RR}(\pi) = \frac{V^{\pi} - \tilde{V}^{\lambda^*}}{\tilde{V}^{\lambda^*}}.$$

6.1. Experiment 1: Synthetic Data

In this experiment, we test the empirical performance of the DBS algorithm and perform its sensitivity analysis with respect to various input parameters, such as the length of the horizon, coefficient of variation of demand distribution, central warehouse capacity, cost parameters (holding and lost-sales cost), and number of stores. For the base instance, we assume that the system has two stores with the following cost parameters as $h_i = 6$, $b_i = 60$, and $c_i = 0.5$. For the baseline demand, we use synthetic demand data generated from a truncated normal distribution with domain $[0, 175]$, mean $\mu = 50$, and standard deviation $\sigma = 50$. We fix horizon length to be $T = 1{,}000$ and the central warehouse capacity to be $\frac{T \cdot \mu \cdot N}{2}$. We then change parameters one at a time for sensitivity analysis, and the results are summarized as follows.

6.1.1. Impact of Length of Horizon *T*. Our first experiment is to evaluate the impact of the length of the horizon T on the relative regret $\mathcal{RR}(\pi)$. The result is summarized in the left panel of Figure 6, which shows

Figure 6. (Color online) Change in Relative Regret with Respect to Horizon Length (Left) and Coefficient of Variation (right)



Note. The shades of the curves represent standard errors.

clearly that the longer the horizon, the lower the relative regret for all algorithms. It is also evident from Figure 6 that the DBS algorithm performs significantly better than the Exp algorithms. The gap between the DBS and Exp algorithms is driven by two factors. First, throughout the planning horizon, the DBS algorithm adjusts its estimation of the underlying demand distribution and applies the best ordering policy given the estimated demand parameters, whereas the Exp algorithm only explores the demand for a fixed number of periods and does not change its estimation during the exploitation phase. Second, the DBS algorithm considers the impact of central warehouse capacity on its ordering decisions by estimating Lagrangian dual variables, whereas the Exp algorithms order myopically without considering warehouse capacity, which can be particularly suboptimal when the warehouse inventory level is not very high.

6.1.2. Impact of Coefficient of Demand Variation. To investigate how demand variation affects learning, we alter the standard deviation of the demand distribution, keeping its mean constant. As seen in the right panel of Figure 6, the relative regret grows with the coefficient of variation for all algorithms. This is intuitive because learning from demand distributions with high volatility is more difficult. We can also see that the DBS algorithm's performance is significantly more robust than the performance of Exp algorithms as the DBS algorithm utilizes the data more efficiently by continuously learning the underlying demand and adjusting the inventory ordering policies accordingly.

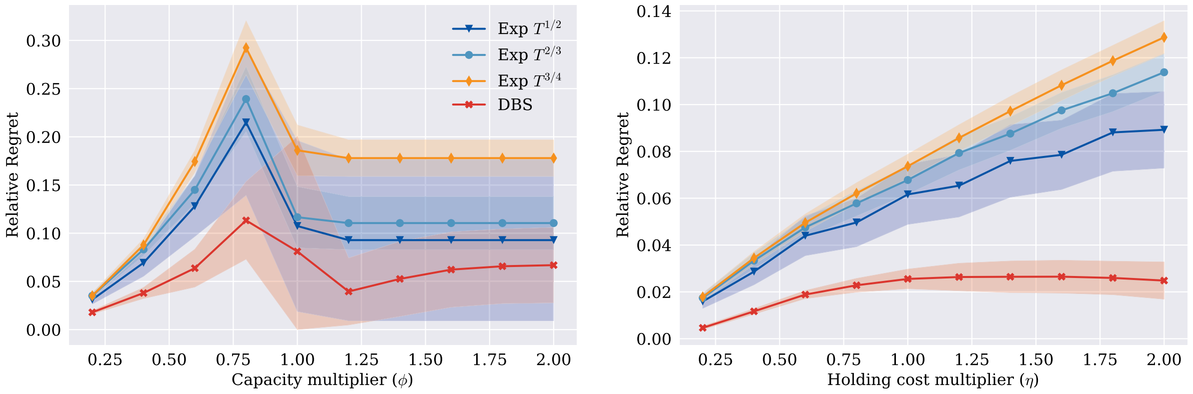
6.1.3. Impact of the Warehouse Capacity. In this experiment, we evaluate the performance of algorithms with respect to different warehouse capacities. In particular, we let the initial inventory be $\phi \cdot T \cdot \mu \cdot N$ and change the values of ϕ (referred to as capacity multiplier). The left panel in Figure 7 shows that relative regret grows first and subsequently declines with increasing capacity under all algorithms. This is due to a trade-off between the cost of holding and of lost sales. The system suffers more from excessive lost sales costs when capacity is insufficient. As capacity increases, the holding cost grows, but the cost of lost sales decreases. Because having too much capacity has no effect on policy, relative regrets of all algorithms remain constant once the capacity becomes sufficiently large (i.e., 1.4 times the baseline capacity).

6.1.4. Impact of Holding Cost h_i. As the holding cost increases, the learning process becomes more expensive because ordering more to receive uncensored demand observations hurts the retailer. As shown in the right panel of Figure 7, the regret of the Exp algorithms increases almost linearly in holding cost, whereas the regret of the DBS algorithm increases sublinearly with holding cost and remains almost constant (even slightly decreases) when h_i is large. This suggests that, compared with the Exp algorithms, the DBS is able to learn the underlying demand without having to increase the order quantity.

6.1.5. Impact of Lost-Sales Cost b_i. As can be seen in the left panel of Figure 8, as lost sales cost increases, the relative regrets of all algorithms decrease. The rationale behind this seemingly counterintuitive observation comes from the fact that the optimal order-up-to levels under the optimal policy increase in b_i, which, in turn, makes the exploration phase less costly for both the Exp and the DBS algorithms. As such, whereas the DBS still outperforms Exp, their differences become smaller with higher lost-sales costs.

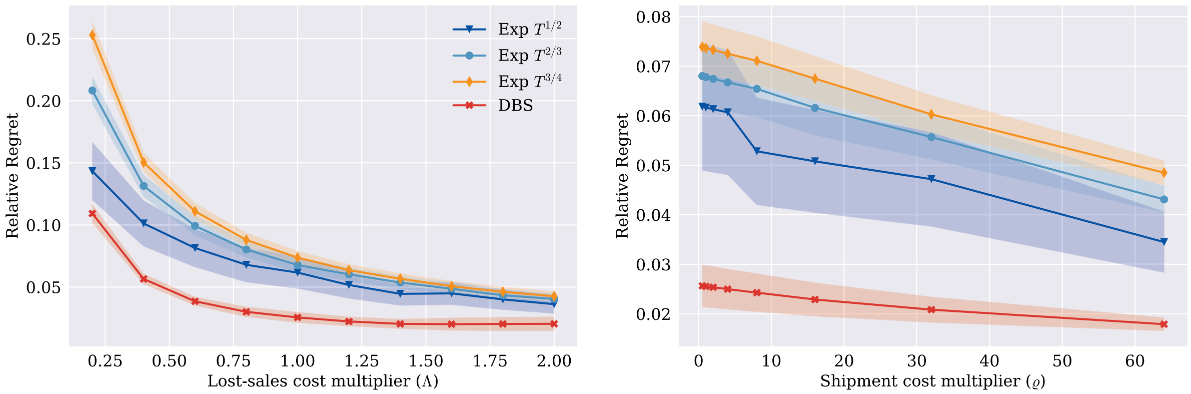
6.1.6. Impact of Shipment Cost c_i. Similar to the previous case, the increase in shipping cost leads to lower relative regret for both DBS and Exp as can be seen in the right panel of Figure 8. The rationale behind this is

Figure 7. (Color online) Change in Relative Regret with Respect to Capacity (Left) and Holding Cost (Right)



Note. The shades of the curves represent standard errors.

Figure 8. (Color online) Change in Relative Regret with Respect to Lost Sales Cost (Left) and Shipment Cost (Right)



Note. The shades of the curves represent standard errors.

because the increase of the benchmark objective $\tilde{V}^{\lambda^*}$ outgrows the increase of regret $\mathcal{R}(\pi)$. That said, our algorithm DBS again is significantly better than Exp algorithms.

6.1.7. Impact of the Number of Stores *N*. In this experiment, we explore the impact of the number of stores on the relative regrets of DBS and Exp algorithms. Table 2 shows that the relative regret is constant for both DBS and Exp as the number of stores increases. This result is consistent with our theoretical analysis that shows that cumulative regret grows linearly in N, which suggests that the relative regret is independent of N.

6.2. Experiment 2: Real Data

In this set of experiments, using the data set collected from our partner fast-fashion company, we test the performance of the DBS algorithm. The fast fashion retailer with which we partnered operates several stores in the European region. They have a large selection of fashion items for women, men, and children as well as a large selection of accessories. They, as do other fast-fashion firms, release new products on a regular basis, making inventory management particularly critical. The data this retailer has provided us includes detailed information on weekly sales and inventories at the SKU level in more than 300 stores between January 2016 and June 2019. The data also contains detailed information on each product, such as initial purchase cost, price, and inventory allocations between regional central warehouses and stores. Because of the extensive nature of the data set, in our numerical experiments, we focus only on the data for women's clothing between July 2018 and June 2019. To set up a computationally tractable instance, we select five stores and choose 125 products that are quite representative of the entire set of products sold in these stores. Please refer to Table 3 for the summary statistics of the data set.

Table 2. Average Relative Regret with Number of Stores

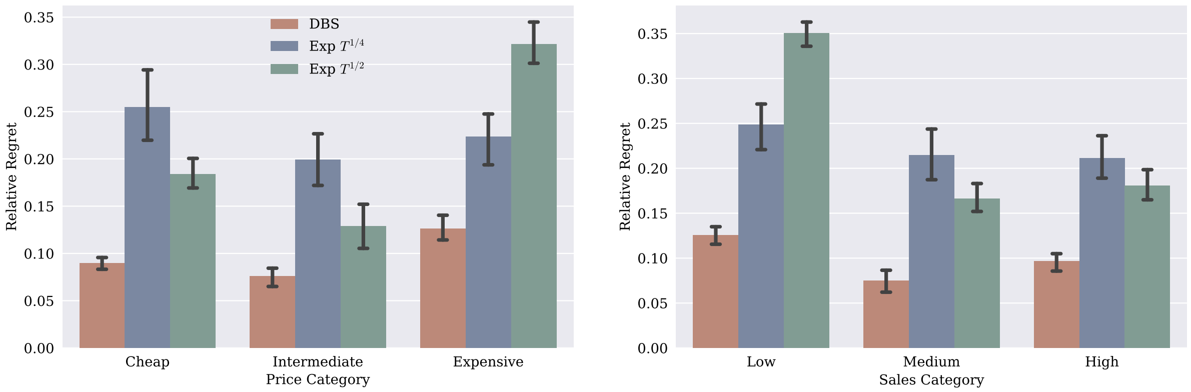
Number of stores	2	4	6	8	10
Exp $T^{1/2}$	0.060	0.060	0.060	0.060	0.060
Exp $T^{2/3}$	0.068	0.068	0.068	0.068	0.068
Exp $T^{3/4}$	0.074	0.074	0.074	0.074	0.074
DBS	0.028	0.025	0.024	0.024	0.024

We stratify these products into three groups with respect to their prices: low price if the price is between the minimum and 25th percentile, intermediate price if the price is between the 25th and 75th percentiles, and high price if the price is between the 75th percentile and maximum price. Similarly, we create three groups for the sales based on the same percentiles: low, medium, and high sales, respectively. We calculated the lost-sales cost based on the difference between price and purchase cost. Following inventory literature that considers the holding cost as the opportunity cost of capital tied in inventories, we estimated the holding cost by multiplying the cost of the product with the average cost of capital. Using both parametric and nonparametric methods, we estimated demand distribution parameters from sales data. We provide more details on the parameter estimation in Online Section EC.7.3. Finally, for each instance, we calibrated the model parameters as mentioned and calculated the relative regret performance of the DBS and two variants of Exp algorithms. The results of the experiment can be found in Figure 9.

Table 3. Summary Statistics of the Data of 125 Products Used in Real Data Experiments

	Sales	Inventory	Price
Mean	18.1	86.8	83.3
Standard deviation	16.3	98.4	82.8
Minimum	0.0	0.0	5.6
25%	8	17	17.8
50%	13	62	36.4
75%	25	106	173.2
Maximum	113	550	247.7

Figure 9. (Color online) Real Data Experiment Results



Notes. On the left, products are grouped by their prices. On the right, products are grouped by their sales volume.

We make several observations. First, the duration of exploration in the Exp algorithm affects its performance. Quite notably, the optimal exploration duration depends on the product category, which implies that there is no fixed duration that works for all types of product and sales categories. Second, in all the experiments, the DBS performed better than the Exp algorithms. This implies that the DBS algorithm is able to adjust the learning period optimally based on the cost and demand characteristics. This suggests that the DBS algorithm yields more robust performance with respect to changes in business parameters. Finally, in both price and sales categories, the relative regrets of DBS for the middle categories are lower than those for the low and high categories.

7. Conclusion

Motivated by our collaboration with one of the largest fast-fashion retailers in Europe, this paper studies the OWMS problem under censored demand with a priori unknown demand distribution, which is the first among the literature to the best of our knowledge. In particular, the decision maker has to jointly make inventory allocation decisions from the central warehouse to different stores and learn the unknown demand faced by stores on the fly. The main challenge under censored demand is to estimate the demand distribution in an unbiased fashion. To tackle this challenge, we propose an algorithm named DBS, which adaptively learns the demand and decides inventory allocation policy by optimizing a Lagrange dual variable of the inventory constraint in the central warehouse. This algorithm DBS is proved to have regret $\tilde{O}(N\sqrt{T})$. We then discuss three extensions of the DBS algorithm to consider discrete, uncensored, and unbounded demand cases. We show that their regrets are also $\tilde{O}(N\sqrt{T})$. We performed simulations leveraging both synthetic data and real data from a fast-fashion firm to illustrate the effectiveness of our algorithm DBS. Our numerical experiments show that our algorithm performs better than the benchmark algorithm and is more robust to changes in business parameters.

Finally, we discuss some possible future work directions. First, in this paper, we assume i.i.d. demand over time for each store, whereas in reality, the demand distribution is usually changing over time. How to model this nonstationarity and design learning algorithms will be interesting future research. Second, in some cases, fixed replenishment costs are charged. Clark and Scarf (1962) show that, with fixed costs, characterizing the optimal policy is difficult (see Chao et al. 2021 for a recent study on a multiwarehouse multistore system with fixed costs). It would be interesting to develop a solution for the fixed cost setting with demand learning. Additionally, although the one-warehouse multistore system is a classic model in inventory management, demand learning in a general model for multiple warehouses would also be interesting. In that case, our benchmark would not be generalizable for multiwarehouses because of the curse of dimensionality. Hence, designing a new algorithm with provable performance is a nontrivial task, and we leave it as a future research direction.

References

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

Ban GY (2020) Confidence intervals for data-driven inventory policies with demand censoring. *Oper. Res.* 68(2):309–326.

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

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

Bu J, Simchi-Levi D, Wang L (2020) Offline pricing and demand learning with censored data. Preprint, submitted July 1, https://dx.doi.org/10.2139/ssrn.3619625.

Caro F, Gallien J (2010) Inventory management of a fast-fashion retail network. *Oper. Res.* 58(2):257–273.

Chao X, Jasin S, Miao S (2021) Adaptive algorithms for multiwarehouse multi-store inventory system with lost sales and fixed replenishment cost. Preprint, submitted July 19, https://dx.doi.org/10.2139/ssrn.3888794.

Chen B, Chao X, Ahn HS (2019) Coordinating pricing and inventory replenishment with nonparametric demand learning. *Oper. Res.* 67(4):1035–1052.

Chen B, Chao X, Shi C (2021) Nonparametric learning algorithms for joint pricing and inventory control with lost sales and censored demand. *Math. Oper. Res.* 46(2):726–756.

Chen B, Chao X, Wang Y (2020a) Data-based dynamic pricing and inventory control with censored demand and limited price changes. *Oper. Res.* 68(5):1445–1456.

Chen B, Wang Y, Zhou Y (2020b) Optimal policies for dynamic pricing and inventory control with nonparametric censored demands. Preprint, submitted December 17, https://dx.doi.org/10.2139/ssrn.3750413.

Chen B, Simchi-Levi D, Wang Y, Zhou Y (2020c) Dynamic pricing and inventory control with fixed ordering cost and incomplete demand information. Preprint, submitted June 24, https://dx.doi.org/10.2139/ssrn.3632475.

Chen L (2010) Bounds and heuristics for optimal Bayesian inventory control with unobserved lost sales. *Oper. Res.* 58(2):396–413.

Cheung WC, Simchi-Levi D (2019) Sampling-based approximation schemes for capacitated stochastic inventory control models. *Math. Oper. Res.* 44(2):668–692.

Clark AJ, Scarf H (1960) Optimal policies for a multi-echelon inventory problem. *Management Sci.* 6(4):475–490.

Clark AJ, Scarf H (1962) Approximate solutions to a simple multi-echelon inventory problem. *Studies in Applied Probability and Management Science*, 88–110.

Davenport T, Guszcza J, Smith T, Jackson BS (2019) 2021 inventory optimization survey report. Deloitte Consulting LLP. Accessed July 15, 2023, https://www2.deloitte.com/us/en/insights/topics/analytics/insight-driven-organization.html.

enVista (2021) 2021 inventory optimization survey report. National Retail Federation. Accessed July 15, 2023, https://nrf.com/research/member-submitted/2021-inventory-optimization-survey-report.

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

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

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

Jackson PL (1988) Stock allocation in a two-echelon distribution system or "what to do until your ship comes in." *Management Sci.* 34(7):880–895.

Jackson PL, Muckstadt JA, Li Y (2019) Multiperiod stock allocation via robust optimization. *Management Sci.* 65(2):794–818.

Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* 53(282):457–481.

Lei YM, Liu S, Jasin S, Vakhutinsky A (2020) On the joint inventory and pricing control for a one-warehouse multi-store problem with lost sales: Spiraling phenomena and a near-optimal heuristic. Preprint, submitted January, https://dx.doi.org/10.2139/ssrn.3688561.

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

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

Marklund J, Rosling K (2012) Lower bounds and heuristics for supply chain stock allocation. *Oper. Res.* 60(1):92–105.

Massart P (1990) The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *Ann. Probab.* 18(3):1269–1283.

McGavin EJ, Schwarz LB, Ward JE (1993) Two-interval inventory-allocation policies in a one-warehouse *n*-identical-retailer distribution system. *Management Sci.* 39(9):1092–1107.

McGavin EJ, Ward JE, Schwarz LB (1997) Balancing retailer inventories. *Oper. Res.* 45(6):820–830.

Miao S, Jasin S, Chao X (2022) Asymptotically optimal Lagrangian policies for multi-warehouse, multi-store systems with lost sales. *Oper. Res.* 70(1):141–159.

Nambiar M, Simchi-Levi D, Wang H (2020) Dynamic inventory allocation with demand learning for seasonal goods. *Production Oper. Management* 30(3):750–765.

Nicasio F (2021) Retail analytics: How to use data to win more sales and customers. Vend. Accessed July 15, 2023, https://www.vendhq.com/blog/how-retailers-can-use-data-to-boost-productivity-customer-service-sales.

Petro G (2021) Supply chain blues: Challenges, shortages and disruptions ahead. 4 minutes to a better retail industry. https://www.forbes.com/sites/gregpetro/2021/10/08/supply-chain-blues-challenges-shortages-and-disruptions-ahead——-5-minutes-to-a-better-retail-industry-read-on.

Porteus EL (2002) *Foundations of Stochastic Inventory Theory* (Stanford University Press, Stanford).

Qin H, Simchi-Levi D, Wang L (2019) Data-driven approximation schemes for joint pricing and inventory control models. Preprint, submitted March 25, https://dx.doi.org/10.2139/ssrn.3354358.

Scarf H (1959) Bayes solutions of the statistical inventory problem. *Ann. Math. Statist.* 30(2):490–508.

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

Yuan H, Luo Q, Shi C (2021) Marrying stochastic gradient descent with bandits: Learning algorithms for inventory systems with fixed costs. *Management Sci.* 67(10):6089–6115.

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

Zipkin PH (2000) *Foundations of Inventory Management* (McGraw-Hill, New York).

Recep Yusuf Bekci is an assistant professor in the Department of Management Sciences at the University of Waterloo. His research focuses on learning and optimization problems in revenue management and inventory control, and the development of algorithms for operational efficiency and strategic decision making.

Mehmet Gümüş is a professor of operations management and the academic director for the Masters of Management in Analytics Program at the Desautels Faculty of Management at McGill University. In his research, he explores the impact of customer behavior and information asymmetry on supply chain management, dynamic pricing, and risk management.

Sentao Miao is an assistant professor in the Leeds School of Business at the University of Colorado Boulder. His research interests are mainly in developing efficient and effective learning and optimization algorithms with various applications in operations management, such as dynamic pricing, assortment selection, and inventory control.