Nonparametric Data-Driven Algorithms for Multiproduct Inventory Systems with Censored Demand

Cong Shi, Weidong Chen
Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, {shicong, aschenwd}@umich.edu
Izak Duenyas
Technology and Operations, Ross School of Business, University of Michigan, Ann Arbor, MI 48109, duenyas@umich.edu

We propose a nonparametric data-driven algorithm called DDM for the management of stochastic periodic-review multi-product inventory systems with a warehouse-capacity constraint. The demand distribution is not known a priori and the firm only has access to past sales data (often referred to as censored demand data). We measure performance of DDM through regret, the difference between the total expected cost of DDM and that of an oracle with access to the true demand distribution acting optimally. We characterize the rate of convergence guarantee of DDM. More specifically, we show that the average expected T-period cost incurred under DDM converges to the optimal cost at the rate of $O(1/\sqrt{T})$. Our asymptotic analysis significantly generalizes approaches used in Huh and Rusmevichientong (2009) for the uncapacitated single-product inventory systems. We also discuss several extensions and conduct numerical experiments to demonstrate the effectiveness of our proposed algorithm.

Key words: inventory, multi-product, censored demand, nonparametric algorithms, asymptotic analysis
Received July 2014; revisions received February 2015, June 2015, September 2015; accepted December 2015.

1. Introduction

The study of stochastic multi-product inventory systems dates back to Veinott (1965). Most, if not all, of the papers on stochastic multi-product inventory systems assume that the stochastic future demand is given by a specific exogeneous random variable, and the inventory decisions are made with full knowledge of the future demand distribution. However, in practice, the demand distribution is usually not known a priori. Even with past demand data (often censored) collected, the selection of the most appropriate distribution and its parameters remains difficult (see Huh and Rusmevichientong (2009), Huh et al. (2011), Besbes and Muharremoglu (2013) for more discussions on censored demand in inventory systems).

Model overview and research issue. In our periodic-review multi-product lost-sales inventory system over a finite horizon of T periods, the demands across periods $t = 1, \dots T$ are (i.i.d.) random

vectors $\mathbf{D}_t$ (with each component representing a different product), respectively. There is a joint warehouse-capacity constraint M imposed on the total number of products that can be held in inventory. The firm has no access to the true underlying demand distribution a priori, and can only observe sales data (i.e., censored demand) over time. We develop a nonparametric data-driven adaptive inventory control policy $\pi = (\mathbf{y}_t \mid t \geq 1)$ where the decision $\mathbf{y}_t$ represents the order-up-to level in period t. We measure performance of our proposed policy π through regret denoted by $\mathcal{R}_T \triangleq \mathcal{C}(\pi) - \mathcal{C}(\pi^*)$, where $\mathcal{C}(\pi)$ is the total expected cost of π and $\mathcal{C}(\pi^*)$ is the total expected cost of a *clairvoyant* optimal policy π^* with access to the true underlying demand distribution a priori. The research question is to devise an effective nonparametric data-driven policy π that drives the average regret $\mathcal{R}_T/T$ to zero with a fast convergence rate.

Main results and contributions. We propose a nonparametric data-driven algorithm called DDM for stochastic multi-product inventory systems with a warehouse-capacity constraint. We characterize the rate of convergence guarantee of DDM. More specifically, we show that the average regret $\mathcal{R}_T$ converges to zero at the rate of $O(1/\sqrt{T})$. Our algorithm DDM is a stochastic gradient descent type of algorithm, similar in spirit to Burnetas and Smith (2000), Kunnumkal and Topaloglu (2008) and Huh and Rusmevichientong (2009). The work closest to ours is Huh and Rusmevichientong (2009) who studied an uncapacitated inventory system with a single product. The novelty of our work lies in both algorithmic design and performance analysis of DDM. First, unlike the uncapacitated single-product case, the gradient estimator in DDM could be sometimes indeterminable in the presence of a warehouse-capacity constraint on multiple products. Second, the projection step in DDM has to factor in both positive inventory carry-over of all products and the warehouse-capacity constraint. To maintain feasibility of the solution in each step, we solve two additional optimization problems. The optimization problems can be efficiently solved by greedy algorithms, but the solution structure makes the asymptotic performance analysis invariably harder than that in the uncapacitated single-product case (where no optimization procedures are needed). The key technical challenge in our analysis is to derive an upper bound of the distance between the target order-up-to level and the actual implemented order-up-to level (due to the warehouse-capacity constraint and positive inventory carry-over from previous periods). Note that the upper bound on this distance function is almost immediate in the uncapacitated single-product case while the development of an upper bound is significantly more complex in our multi-product setting. Third, we relate the inventory process to a $GI/G/1$ queue. We then develop a stochastic dominance argument and invoke a classical result on the expected busy period in $GI/G/1$ queue due to Loulou (1978).

We compare the computational performance of DDM with several existing parametric and nonparametric approaches in the literature. Our results show that DDM outperforms these benchmark

algorithms in terms of both consistency and convergence rate. We also consider two interesting extensions, one with a more general warehouse-capacity constraint where different products may have different dimension or sizes, and the other one with discrete demand and order quantities.

Relevant literature. Our work is relevant to the following research streams.

Multi-product stochastic inventory systems. There is a large body of literature devoted to various classes of such problems. In this paper, we focus our attention on the classical stochastic multi-product inventory systems under a warehouse-capacity constraint, first studied by Veinott (1965). He provided conditions that ensure that the base-stock ordering policy is optimal in a periodic-review inventory system with a finite horizon. Subsequently, Ignall and Veinott (1969) showed that in the stationary demand case, a myopic ordering policy is optimal under certain mild conditions. Beyer et al. (2001, 2002) established the optimality of myopic policies in backlogged systems with separable costs by appealing to the sufficient condition provided by Ignall and Veinott (1969), which was further extended by Choi et al. (2005) under a relaxed demand assumption. Our work focuses on a nonparametric variant in which the demand distribution is not known a priori.

Nonparametric inventory systems. Burnetas and Smith (2000) developed a gradient descent type algorithm for ordering and pricing when inventory is perishable; they showed that the average profit converges to the optimal but did not establish the rate of convergence. Huh and Rusmevichientong (2009) proposed gradient descent based algorithms for lost-sales systems with censored demand. Subsequently, Huh et al. (2009) proposed algorithms for finding the optimal base-stock policy in lost-sales inventory systems with positive lead time. Huh et al. (2011) applied the concept of Kaplan-Meier estimator to devise another data-driven algorithm for censored demand. Other nonparametric approaches in the inventory literature include sample average approximation (SAA) (e.g., Kleywegt et al. (2002), Levi et al. (2007, 2015)) which uses the empirical distribution formed by *uncensored* samples drawn from the true distribution. Concave adaptive value estimation (e.g., Godfrey and Powell (2001), Powell et al. (2004)) successively approximates the objective cost function with a sequence of piecewise linear functions. The bootstrap method (e.g., Bookbinder and Lordahl (1989)) estimates the newsvendor quantile of the demand distribution. The infinitesimal perturbation approach (IPA) is a sampling-based stochastic gradient estimation technique that has been used to solve stochastic supply chain models (see, e.g., Glasserman (1991)). Maglaras and Eren (2015) employed maximum entropy distributions to solve a stochastic capacity control problem. For parametric approaches, such as Bayesian learning (see, e.g., Lariviere and Porteus (1999), Chen and Plambeck (2008)) or operational statistics (see, e.g., Liyanage and Shanthikumar (2005), Chu et al. (2008)) in stochastic inventory systems, we refer readers to Huh and Rusmevichientong (2009) for an excellent discussion of the key differences between nonparametric and parametric

approaches. Our paper contributes to the literature by studying multi-product inventory systems under a warehouse-capacity constraint, which is significantly more complex to analyze.

Online convex optimization. The aim of online convex optimization is to minimize the cumulative loss function defined over a convex compact set with online learning process since the optimizer does not know the (convex) objective function a priori (see Hazan (2015), Shalev-Shwartz (2012) for an overview). Zinkevich (2003) has shown that the average T-period cost using a gradient descent based algorithm converges to the optimal cost at the rate of $O(1/\sqrt{T})$. This result was further extended by Flaxman et al. (2005) in a bandit setting. Under additional technical assumptions, a modified algorithm by Hazan et al. (2006) achieves a faster convergence rate $O(\log T/T)$. Our problem differs from the conventional online convex optimization problems in that the target levels (or the iterates) may not be achieved due to policy-dependent dynamic inventory constraints.

Stochastic approximation. The proposed gradient descent type of algorithm also resembles the ones used in the Stochastic Approximation (SA) literature (see Nemirovski et al. (2009) and references therein), which should be carefully contrasted with ours. First, SA algorithms aim to solve a single-stage stochastic optimization problem by making successive experiments while the cost of experiments is ignored. On the other hand, our algorithm aims to minimize the cumulative loss suffered along the learning progress for a multi-stage closed-loop stochastic optimization problem. Putting into context, SA focuses on measuring the terminal regret $\mathbb{E}[\Pi(\mathbf{y}_T) - \Pi(\mathbf{y}^*)]$, whereas our algorithm focuses on measuring the cumulative loss over time $\mathbb{E}\left[\sum_{t=1}^{T}(\Pi(\mathbf{y}_t) - \Pi(\mathbf{y}^*))\right]$. Second, in the analysis of robust SA algorithms with general convex costs, the step size is chosen to be $O(1/\sqrt{t})$ to obtain a convergence rate of $O(1/\sqrt{t})$ in the terminal regret criterion by appropriately *averaging* the iterate solutions. The standard robust SA approaches cannot be adapted to our setting where the iterates cannot move "freely" due to policy-driven dynamic inventory constraints.

General notation. For any real vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, $\mathbf{y} \geq \mathbf{x}$ means component-wise greater or equal to; $\mathbf{x}^+ = (\max\{x^i, 0\})_{i=1}^n$; $|\mathbf{x}| = (|x^i|)_{i=1}^n$; the *join* operator $\mathbf{x} \vee \mathbf{y} = (\max\{x^i, y^i\})_{i=1}^n$; the *meet* operator $\mathbf{x} \wedge \mathbf{y} = (\min\{x^i, y^i\})_{i=1}^n$; for any integers t and s with $t \leq s$, $\mathbf{x}_{[t,s]} = \sum_{j=t}^{s} \mathbf{x}_j$ and $\mathbf{x}_{[t,s)} = \sum_{j=t}^{s-1} \mathbf{x}_j$; $||\cdot||$ or $||\cdot||_2$ means 2-norm; $||\cdot||_1$ means 1-norm. The notation $\triangleq$ means "is defined as".

2. Multi-Product Stochastic Inventory Systems

We consider a stochastic T-period n-product inventory system under a warehouse-capacity constraint M (e.g., Ignall and Veinott (1969), Beyer et al. (2001)). The firm has no knowledge of the true underlying demand distribution a priori, but can observe past sales data (i.e., censored demand data), and make adaptive inventory decisions based on the available information.

Random demand and regularity assumptions. For each period $t = 1, \ldots, T$ and each product $i = 1, \ldots n$, we denote the demand of product i in period t by a random variable D_t^i. For notational convenience, we use $\mathbf{D}_t = (D_t^1, \ldots, D_t^n)$ to denote the random demand vector in period t, and $\mathbf{d}_t = (d_t^1, \ldots, d_t^n)$ to denote their realizations.

ASSUMPTION 1. *We make the following assumptions and regularity conditions on demand.*

(a) For each product i, D_t^i is i.i.d. across time period t.

(b) For each product i and for each period t, D_t^i is independent (but not necessarily identically distributed) of D_s^j for all $j \neq i$ and $s = 1, \ldots, T$.

(c) For each product i and for each period t, D_t^i is a continuous random variable defined on a finite support $[0, M]$, whose CDF $F_{D^i}(\cdot)$ is differentiable and density $F'_{D^i}(x) > 0$ for all $x \in [0, M]$.

(d) For each product i and for each period t, $\mathbb{E}[D_t^i] \geq l$ for some real number $l > 0$.

Assumptions 1(a) and 1(b) assume some form of stationarity of demand, which is predominant in the nonparametric learning literature (see, e.g., Levi et al. (2007), Huh et al. (2009, 2011), Huh and Rusmevichientong (2009), Besbes and Muharremoglu (2013)). Assumption 1(c) ensures the per-period cost function defined in (3) is differentiable, finite-valued and strictly (jointly) convex, which guarantees a unique minimizer. Assumption 1(d) rules out degenerate demands.

System dynamics and objectives. Let $\mathbf{f}_t$ denote the information collected up to the beginning of period t, which includes all the realized demands and past decisions. A feasible *closed-loop* policy π is a sequence of functions $\mathbf{y}_t = \pi_t(\mathbf{x}_t, \mathbf{f}_t)$, $t = 1, \ldots, T$, mapping beginning inventory $\mathbf{x}_t$ and $\mathbf{f}_t$ (state) into ending inventory $\mathbf{y}_t$ (decision) while satisfying $\mathbf{y}_t \geq \mathbf{x}_t$ and the warehouse-capacity constraint (see Bertsekas (2000) for discussions on closed-loop optimization problems). Note that when the demand distribution is known a prior, it suffices to consider policies of the form $\mathbf{y}_t = \pi_t(\mathbf{x}_t)$, due to the assumed across-time independence of demands (see Bertsekas and Shreve (2007)).

Given a feasible policy π, we describe the sequence of events below. (Note that $\mathbf{x}_t^\pi$, $\mathbf{y}_t^\pi$ and $\mathbf{q}_t^\pi$'s are functions of π; for ease of presentation, we make their dependence on π implicit.)

(a) At the beginning of period t, the firm observes the starting inventory $\mathbf{x}_t = (x_t^1, \ldots, x_t^n)$.

(b) The firm decides to order $\mathbf{q}_t = (q_t^1, \ldots, q_t^n) \geq 0$, and the ending inventory $\mathbf{y}_t = \mathbf{x}_t + \mathbf{q}_t$, where $\mathbf{y}_t = (y_t^1, \ldots, y_t^n)$. We assume instantaneous replenishment. The total inventory level is restricted by a warehouse-capacity constraint (see Ignall and Veinott (1969)), i.e.,

$$\mathbf{y}_t \in \Gamma \triangleq \left\{ \mathbf{y}_t \in \mathbb{R}_+^n : \sum_{i=1}^n y_t^i \leq M \right\}. \tag{1}$$

(c) The demand $\mathbf{D}_t$ is realized, denoted by $\mathbf{d}_t$, which is satisfied to the maximum extent using on-hand inventory. Unsatisfied demand units are *lost*, and the firm only observes the *sales quantity* (or censored demand), i.e., $\min(d_t^i, y_t^i)$ for each product i in period t. The state transition can be written as $\mathbf{x}_{t+1} = (\mathbf{x}_t + \mathbf{q}_t - \mathbf{d}_t)^+ = (\mathbf{y}_t - \mathbf{d}_t)^+$.

(d) The production, overage and underage costs at the end of period t is then $\mathbf{c}\cdot\mathbf{q}_t + \mathbf{h}\cdot(\mathbf{y}_t - \mathbf{d}_t)^+ + \mathbf{p}\cdot(\mathbf{d}_t - \mathbf{y}_t)^+$, where $\mathbf{c} = (c^1, \ldots, c^n)$, $\mathbf{h} = (h^1, \ldots, h^n)$ and $\mathbf{p} = (p^1, \ldots, p^n)$ are the per-unit purchasing, holding and lost-sales penalty cost vectors, respectively. We note that the cost minimization model with lost-sales assumes that $\mathbf{p} \geq \mathbf{c}$ (see Zipkin (2000)) since the firm loses revenue and goodwill from the sale and the revenue has to be greater than the production cost. (Our approach also works for time-invariant random purchasing cost vector.)

Assuming the salvage value of any left-over product at the end of planning horizon equals its production cost, the total expected cost incurred by π can be written as

$$\begin{aligned} \mathcal{C}(\pi) &= \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{c}\cdot(\mathbf{y}_t - \mathbf{x}_t) + \mathbf{h}\cdot(\mathbf{y}_t - \mathbf{D}_t)^+ + \mathbf{p}\cdot(\mathbf{D}_t - \mathbf{y}_t)^+\right] - \mathbb{E}[\mathbf{c}\cdot\mathbf{x}_{T+1}], \qquad (2)\\ &= -\mathbf{c}\cdot\mathbf{x}_1 + \sum_{t=1}^{T} \mathbb{E}\left[\mathbf{c}\cdot\mathbf{y}_t + (\mathbf{h}-\mathbf{c})\cdot(\mathbf{y}_t - \mathbf{D}_t)^+ + \mathbf{p}\cdot(\mathbf{D}_t - \mathbf{y}_t)^+\right], \end{aligned}$$

where the second equality follows from $\mathbf{x}_{t+1} = (\mathbf{y}_t - \mathbf{d}_t)^+$ and some simple algebra. If the underlying distribution $\mathbf{D}_t$ is given a priori, the stochastic inventory control problem specified above can be formulated using dynamic programming (see Beyer et al. (2001)) with state variables $\mathbf{x}_t$, control variables $\mathbf{y}_t$ (with $\mathbf{x}_t \leq \mathbf{y}_t \in \Gamma$), random disturbances $\mathbf{D}_t$, and state transition $\mathbf{x}_{t+1} = (\mathbf{y}_t - \mathbf{d}_t)^+$. It turns out that this problem is in fact "myopically" solvable, which is discussed next.

Clairvoyant optimal policy. We first characterize the clairvoyant optimal policy where the distribution of $\mathbf{D}_t$ is known a priori. We define $\Pi(\cdot)$ to be the per-period expected cost function,

$$\Pi(\mathbf{a}) = \Pi_t(\mathbf{a}) \triangleq \mathbb{E}\left[\mathbf{c}\cdot\mathbf{a} + (\mathbf{h}-\mathbf{c})\cdot(\mathbf{a} - \mathbf{D}_t)^+ + \mathbf{p}\cdot(\mathbf{D}_t - \mathbf{a})^+\right]. \qquad (3)$$

Let $\mathbf{y}^*$ be a unique critical (deterministic) vector defined by

$$\mathbf{y}^* \triangleq \arg\min_{\mathbf{a}\in\Gamma:\mathbf{a}\geq\mathbf{0}} \Pi(\mathbf{a}). \qquad (4)$$

THEOREM 1. *Under Assumption 1, when the demand distribution is known a priori, ordering up to* $\mathbf{y}^*$ *defined in* (4) *in each period is optimal, with expected per-period cost* $\Pi(\mathbf{y}^*)$.

Although not central to our main focus, the proof of Theorem 1 is quite involved, which relies on verifying a sufficient condition called *substitute property* provided by Ignall and Veinott (1969)

to establish the optimality of myopic policy. We relegate the full detailed proof to the Electronic Companion.

3. Nonparametric Data-Driven Inventory Control Policies

When the firm has no knowledge of the true underlying distribution of $\mathbf{D}_t$ a priori, we aim to find a provably good adaptive data-driven inventory control policy that makes the total expected system costs close to the optimal strategy. The proposed data-driven algorithm DDM maintains a vector triplet of sequences $(\mathbf{z}_t, \hat{\mathbf{y}}_t, \mathbf{y}_t)_{t\geq 0}$. The first sequence $(\mathbf{z}_t)_{t\geq 0}$ represents the *constraint-free target* inventory levels where the warehouse storage constraint is waived. The second sequence $(\hat{\mathbf{y}}_t)_{t\geq 0}$ represents the *target* inventory levels when the warehouse storage constraint is taken into account. However, the target inventory levels $(\hat{\mathbf{y}}_t)_{t\geq 0}$ may not be always feasible due to warehouse capacity constraint and positive inventory carry-over. Thus, we use the third sequence $(\mathbf{y}_t)_{t\geq 0}$ to represent the *actual implemented* inventory levels after ordering.

We first present a compact description of our data-driven multi-product algorithm (DDM).

Data-Driven Multi-product Algorithm (DDM).

Step 0. (Initialization.) Set the initial inventory levels $\mathbf{y}_0 = \hat{\mathbf{y}}_0 = \mathbf{z}_0$ to be any values within Γ and then set the initial values $t = 0$, $\tau_0 = 0$ and $k = 0$.

For each period $t = 0, \ldots, T-1$, repeat the following steps:

Step 1. (Setting the constraint-free and constrained target inventory levels.)

Case 1: If $\mathbf{y}_t \geq \hat{\mathbf{y}}_t$ (i.e., $y_t^i \geq \hat{y}_t^i$ for all $i = 1, \ldots, n$), the algorithm updates the constraint-free target inventory levels $\mathbf{z}_{t+1}$ by

$$\mathbf{z}_{t+1} = \hat{\mathbf{y}}_t - \eta_t \mathbf{G}_t(\hat{\mathbf{y}}_t), \text{ where } \eta_t = \left(\frac{\gamma M}{\sqrt{n} \cdot \max_i\{p^i - c^i, h^i\}} \right) \frac{1}{\sqrt{t}} \text{ for some } \gamma > 0 \tag{5}$$

for each product $i = 1, \ldots, n$, and the i^{th} component of $\mathbf{G}_t$ is defined as

$$G_t^i(\hat{\mathbf{y}}_t) = \begin{cases} h^i, & \text{if } \hat{y}_t^i > d_t^i, \\ -(p^i - c^i), & \text{if } \hat{y}_t^i \leq d_t^i. \end{cases} \tag{6}$$

Note that $\gamma = 1$ for achieving the tightest theoretical bound.

Then the algorithm sets the constrained target inventory levels $\hat{\mathbf{y}}_{t+1}$ by solving

$$\hat{\mathbf{y}}_{t+1} = \arg\min_{\mathbf{w}\in\Gamma} ||\mathbf{w} - \mathbf{z}_{t+1}||_2. \tag{7}$$

Record the break point $\tau_k := t$ and increase the value k by 1.

Case 2: Else if $\mathbf{y}_t \not\geq \hat{\mathbf{y}}_t$ (i.e., there exists an i such that $y_t^i < \hat{y}_t^i$), the algorithm keeps both the constraint-free and constrained target inventory levels unchanged, i.e., $\mathbf{z}_{t+1} = \mathbf{z}_t$ and $\hat{\mathbf{y}}_{t+1} = \hat{\mathbf{y}}_t$.

Step 2. (Solving for the actual implemented target inventory levels.)

Define the set J and its complement as

$$J \triangleq \left\{i : x_{t+1}^i > \hat{y}_{t+1}^i\right\}, \qquad \bar{J} \triangleq \left\{i : x_{t+1}^i \leq \hat{y}_{t+1}^i\right\}. \tag{8}$$

For each product $i \in J$, we set the actual implemented levels

$$y_{t+1}^i = x_{t+1}^i, \quad \text{if } x_{t+1}^i > \hat{y}_{t+1}^i. \tag{9}$$

If $\bar{J} \neq \emptyset$, then we set the actual implemented levels $\mathbf{y}_{t+1}$ by solving

$$\min \sum_{i \in \bar{J}} (\hat{y}_{t+1}^i - y_{t+1}^i)^2 \quad \text{s.t.} \quad \sum_{i \in \bar{J}} y_{t+1}^i \leq M - \sum_{j \in J} x_{t+1}^j, \quad y_{t+1}^i \geq x_{t+1}^i, \, \forall \, i \in \bar{J}. \tag{10}$$

This concludes the description of the algorithm.

3.1. Algorithm Overview of DDM and Properties

Step 1: (Stochastic Gradient Descent). Let $\mathcal{T} = \{\tau_0, \tau_1, \ldots, \tau_m\}$ with $\tau_m \leq T$, which is the set of break points of DDM. In each period $\tau_k + 1$ $(k = 1, \ldots, m)$, we update the constraint-free target levels $\mathbf{z}_{t+1}$ by a stochastic gradient descent step. Conceptually, we update the minimizer along the negative direction of the true gradient of $\Pi(\cdot)$. However, since the true cost function $\Pi(\cdot)$ is not available to us (without knowing the underlying demand distribution), we can only rely on the observed sales data $\mathbf{d}_t$ to provide us an estimator of the true gradient of $\Pi(\hat{\mathbf{y}}_t)$ at the points $\hat{\mathbf{y}}_t$. The estimator $G_t^i(\hat{\mathbf{y}}_t)$ defined in (5) can be computed using the sales (censored demand) data observed by the firm in period $t \in \mathcal{T}$. When $t \in \mathcal{T}$, we have $y_t^i \geq \hat{y}_t^i$ for all $i = 1, \ldots, n$. Hence, the event $\{\hat{y}_t^i \leq d_t^i\}$ is equivalent to the case where the ending inventory in period t is at most $y_t^i - \hat{y}_t^i$, which is an observable event; the event $\{\hat{y}_t^i > d_t^i\}$ is equivalent to the case where the ending inventory in period t is strictly greater than $y_t^i - \hat{y}_t^i$, which is also observable. In this case, $\mathbf{G}_t$ defined in (6) is an unbiased estimator of the true gradient $\nabla\Pi(\hat{\mathbf{y}}_t)$ at $\hat{\mathbf{y}}_t$, i.e., $\mathbb{E}[\mathbf{G}_t(\hat{\mathbf{y}}_t)] = \nabla\Pi(\hat{\mathbf{y}}_t)$, where the expectation is taken over the demand in period t. On the other hand, when $t \notin \mathcal{T}$, $G_t^i(\hat{\mathbf{y}}_t)$ may be *indeterminable* because the actual implemented inventory levels could fall below the target order-up-to levels. To be more specific, when $y_t^i < \hat{y}_t^i$ and $y_t^i \leq d_t$, the firm only observes the stockout

but not the lost-sales quantity. Therefore, the firm cannot distinguish between $y_t^i \le d_t < \hat{y}_t^i$ and $y_t^i < \hat{y}_t^i \le d_t$, and hence cannot determine the value of $G_t^i(\hat{\mathbf{y}}_t)$. In periods when $t \notin \mathcal{T}$, we keep the target order-up-to levels unchanged.

We then carry out a greedy projection of the constraint-free target inventory levels $\mathbf{z}_{t+1}$ onto the warehouse storage constraint set Γ via (7), more specifically,

$$\min \sum_{i=1}^{n} (\hat{y}_{t+1}^i - z_{t+1}^i)^2 \quad \text{s.t.} \quad \sum_{i=1}^{n} \hat{y}_{t+1}^i \le M, \quad \hat{y}_{t+1}^i \ge 0, \, \forall \, i. \tag{11}$$

We also make two simple observations that will be useful in Section 4. (a) A simple observation leads to the lower and upper bounds of z_{t+1}^i for each product i, i.e., $\hat{y}_t^i - \eta_t h^i \le z_{t+1}^i \le \hat{y}_t^i + \eta_t (p^i - c^i)$. In fact, z_{t+1}^i has to hit one of the two boundaries. (b) Another important observation is that when the product i in the first step updates its constraint-free target level z_{t+1}^i through a positive direction, i.e., $z_{t+1}^i = \hat{y}_t^i + \eta_t (p^i - c^i) \ge \hat{y}_t^i \ge 0$, we must have $\hat{y}_{t+1}^i \le z_{t+1}^i$. To see this, suppose otherwise $\hat{y}_{t+1}^i > z_{t+1}^i$, we can decrease $\hat{y}_{t+1}^i$ to z_{t+1}^i, thereby strictly improving the objective value of (11) while maintaining feasibility. On the other hand, when the product i in the first step updates its constraint-free target level z_{t+1}^i through a negative direction, we have $z_{t+1}^i = \hat{y}_t^i - \eta_t h^i \le \hat{y}_t^i$, Thus, this leads to the following property that will be useful in the performance analysis,

$$\hat{y}_{t+1}^i \le \hat{y}_t^i + \eta_t (p^i - c^i), \quad \forall \, i = 1, \ldots, n. \tag{12}$$

Step 2: (Maintaining Feasibility). The target inventory levels $\hat{\mathbf{y}}_{t+1}$ derived in the second step may not be achievable or implementable, due to the physical inventory carry-over and the warehouse capacity constraint. We then need to carry out an additional optimization procedure as follows. This step tries to order as many products as possible to reach the target level, and it is easy to solve quantitatively but hard to analyze. First we divide all the products into two groups, namely, the set J and its complement as defined in (8). We then have the following two cases.

Case 1. For each product $i \in J$, i.e., the beginning inventory level of product i is already greater than its target level. It is natural to not order any more product i and hence we follow (9).

Case 2. Now we focus on the set $\bar{J} \ne \emptyset$. Since the remaining inventory space now becomes $M - \sum_{j \in J} x_{t+1}^j$, we solve the optimization problem (10) to determine the actual implemented levels $\mathbf{y}_{t+1}$. Note that the optimization problem is well-defined since

$$M - \sum_{j \in J} x_{t+1}^j = M - \sum_{j \in J} (y_t^j - d_t^j)^+ \ge M - \sum_{j \in J} y_t^j \ge 0,$$

where the inequality follows from the fact that the algorithm keeps $\mathbf{y}_t \in \Gamma$.

The optimization (10) attempts to raise our inventory level as close as possible to the target inventory level $\hat{y}^i_{t+1}$ for each product $i \in \bar{J}$; however, it is possible that some of the products in $\bar{J}$ cannot hit the target level due to inventory constraints. Since we minimize the 2-norm type of objective function, it can be readily verified that the optimization (10) makes the *shortfalls* defined as $\hat{y}^i_{t+1} - y^i_{t+1}$ as even as possible across the products in the set $\bar{J}$.

Note that if the optimal objective value of (10) is equal to 0, then the algorithm goes to Case 1 in the next period and updates the target inventory levels. Otherwise it goes to Case 2 and maintains the target inventory levels; while maintaining these target levels, the inventory levels within J are decreasing and more inventory space is freed over time, and the shortfalls will decrease to zero.

4. Performance Analysis of the Nonparametric Data-Driven Algorithm

The regret of our data-driven algorithm, denoted by $\mathcal{R}_T$, is defined as the difference between the optimal clairvoyant cost (given the demand distribution a priori) and the cost incurred by our data-driven algorithm (which learns the demand distribution over time). That is, for any $T \geq 1$,

$$\mathcal{R}_T \triangleq \mathbb{E}\left[\sum_{t=1}^{T} \Pi(\mathbf{y}_t)\right] - \sum_{t=1}^{T} \Pi(\mathbf{y}^*),$$

where $\mathbf{y}_t$ are the actual implemented order-up-to levels of our nonparametric (closed-loop) algorithm DDM, and $\mathbf{y}^*$ is the clairvoyant optimal solution in (4).

Theorem 2 below states the main result in this paper.

THEOREM 2. *Under Assumption 1, the average regret $\mathcal{R}_T/T$ of our data-driven algorithm DDM approaches 0 at the rate of $1/\sqrt{T}$. That is, there exists some constant K, such that for any $T \geq 1$,*

$$\frac{1}{T}\mathcal{R}_T \triangleq \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} \Pi(\mathbf{y}_t)\right] - \Pi(\mathbf{y}^*) \leq \frac{K}{\sqrt{T}},$$

where $\mathbf{y}_t$ are actual implemented order-up-to levels of our nonparametric (closed-loop) algorithm DDM, and $\mathbf{y}^*$ is the clairvoyant optimal solution in (4).

It is known that in the general convex case (without assuming smoothness and strong convexity), this rate of $O(1/\sqrt{T})$ is unimprovable (see, e.g., Theorem 3.2. of Hazan (2015)). Our key contribution here is to establish this best possible rate even with inventory and capacity constraints (i.e., the iterates cannot move "freely" due to policy-driven dynamic inventory constraints).

Then the proof of Theorem 2 is the direct consequence of the following two key lemmas.

LEMMA 1. *For any $T \geq 1$, there exists a constant $K_1 \in \mathbb{R}$ such that*

$$\Delta_1(T) = \mathbb{E}\left[\sum_{t=1}^{T} \Pi(\hat{\mathbf{y}}_t) - \sum_{t=1}^{T} \Pi(\mathbf{y}^*)\right] \leq K_1\sqrt{T},$$

where $\hat{\mathbf{y}}_t$ are target order-up-to levels of DDM, and $\mathbf{y}^*$ is the clairvoyant optimal solution in (4).

LEMMA 2. *For any $T \geq 1$, there exists some constant $K_2 \in \mathbb{R}$ such that*

$$\Delta_2(T) = \mathbb{E}\left[\sum_{t=1}^{T} \Pi(\mathbf{y}_t) - \sum_{t=1}^{T} \Pi(\hat{\mathbf{y}}_t)\right] \leq K_2\sqrt{T},$$

where $\mathbf{y}_t$ and $\hat{\mathbf{y}}_t$ are actual implemented and target order-up-to levels of DDM, respectively.

4.1. Bound on Δ_1 - Online Convex Optimization (Proof of Lemma 1)

The proof of Lemma 1 builds upon the ideas and techniques used in online convex optimization (see, e.g., Zinkevich (2003) and Flaxman et al. (2005)). It is shown the cost function $\Pi(\cdot)$ is jointly convex, and $G(\cdot)$ is an unbiased estimator of the true expected gradient of $\Pi(\cdot)$ under censored demand within the set of breakpoints. In addition, this gradient estimator is bounded, i.e., $||G(\cdot)||_2^2 \leq n(\max_i\{p^i - c^i, h^i\})^2$. We relegate the proof of Lemma 1 to the Electronic Companion.

4.2. Bound on Δ_2 - Stochastic Dominance and a GI/G/1 Queue (Proof of Lemma 2)

The main focus of this paper is to establish the result in Lemma 2. First we derive a bound of the gap between the cost functions associated with the actual implemented level $\mathbf{y_t}$ and the desired target level $\hat{\mathbf{y}}_t$, using the distance function $|\mathbf{y}_t - \hat{\mathbf{y}}_t|$.

LEMMA 3. *The difference in cost functions*

$$\mathbb{E}[\Pi(\mathbf{y_t}) - \Pi(\hat{\mathbf{y}}_t)] \leq \mathbb{E}\left[(\mathbf{h} \vee (\mathbf{p} - \mathbf{c})) \cdot |\mathbf{y}_t - \hat{\mathbf{y}}_t|\right].$$

Given Lemma 3, we need to develop an upper bound on the distance function $|\mathbf{y}_t - \hat{\mathbf{y}}_t|$, which is the crux of our performance analysis. Lemmas 4 and 5 below play a major role in the development of such an upper bound. Their proof strategy relies heavily on the construction of DDM and also the structural properties of optimization problems (10) and (11), which is quite involved.

Lemma 4 below provides an upper bound on the distance function for products in the set J in which the beginning inventory level already exceeds the target order-up-to level.

LEMMA 4. *In each period $t+1$, we bound the distance function for all $i \in J \triangleq \{i : x^i_{t+1} > \hat{y}^i_{t+1}\}$.*

$$\sum_{i\in J} |y^i_{t+1} - \hat{y}^i_{t+1}| \le \sum_{i\in J} |y^i_t - \hat{y}^i_t| + \eta_t \left(\sum_{i\in J} h^i + \sum_{j\in \bar{J}} (p^j - c^j) \right) - \sum_{i\in J} d^i_t.$$

Lemma 5 below provides an upper bound on the distance function for products in the complement set $\bar{J}$ in which the beginning inventory level is below the target order-up-to level. If this is the case for all products, i.e., all products belong to $\bar{J}$, then the target levels can always be achieved. If not, we solve (10) to re-distribute our target levels such that the difference between the target level and the actual implemented level is as even as possible across different products.

LEMMA 5. *In each period $t+1$, we bound the distance function for all $i \in \bar{J} \triangleq \{i : x^i_{t+1} \le \hat{y}^i_{t+1}\}$ as follows. If $J = \emptyset$, we have $\sum_{i\in\bar{J}} |\hat{y}^i_{t+1} - y^i_{t+1}| = 0$. Otherwise, if $J \neq \emptyset$, we have*

$$\sum_{i\in \bar{J}} |\hat{y}^i_{t+1} - y^i_{t+1}| \le \sum_{i\in \bar{J}} |\hat{y}^i_t - y^i_t| + \eta_t \sum_{i\in\bar{J}} (p^i - c^i) - \sum_{j\in J} d^j_t.$$

With the upper bounds on the distance function in two mutually exclusive sets J and $\bar{J}$ obtained from Lemmas 4 and 5, we provide an overarching upper bound in Lemma 6.

LEMMA 6. *In each period $t+1$, we bound the sum of distance functions as follows.*

$$\sum_{i=1}^n |y^i_{t+1} - \hat{y}^i_{t+1}| \le \left(\sum_{i=1}^n |y^i_t - \hat{y}^i_t| + \eta_t \left(\sum_{i=1}^n \left(h^i + 2\left(p^i - c^i\right)\right)\right) - \min_{j=1,\dots,n} d^j_t \right)^+.$$

Next, we wish to find a stochastic process that can be used to bound the sum of distance functions. It is now convenient to introduce the notion of *stochastic order* and *convex order* (see Shaked and Shanthikumar (2007)). Consider two random variables X and Y. X is said to be stochastically smaller than Y (denoted by $X \le_{st} Y$) if $\mathbb{P}(X > x) \le \mathbb{P}(Y > x), \forall x \in \mathbb{R}$. Also, X is said to be smaller than Y in the convex order (denoted as $X \le_{cx} Y$) if $\mathbb{E}[\phi(X)] \le \mathbb{E}[\phi(Y)]$ for all convex functions $\phi : \mathbb{R} \to \mathbb{R}$, provided the expectations exist. Note that convex order is weaker, i.e., $X \le_{st} Y \Rightarrow X \le_{cx} Y$.

Next, corresponding to the sum of distance functions, we consider a stochastic process $(Z_t \mid t \ge 0)$

$$Z_{t+1} = \left[Z_t + \frac{S_t}{\sqrt{t}} - \tilde{D}_t \right]^+, \qquad Z_0 = 0,$$

where $S_t \triangleq \sum_{i=1}^n \left(h^i + 2(p^i - c^i)\right)$, and $\tilde{D}_t$ is a random variable satisfying $\tilde{D}_t \le_{st} D^j_t$, $\forall j$.

LEMMA 7. *The total expected distance function*

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n}|y_t^i-\hat{y}_t^i|\right]\le\mathbb{E}\left[\sum_{t=1}^{T}Z_t\right],$$

where Z_{t+1}^i is a stochastic process defined above.

We observe that the stochastic process Z_t is very similar to a $GI/G/1$ queue, except that the service time is scaled by $1/\sqrt{t}$ in each period t. Now consider a $GI/G/1$ queue $(W_n \mid n\ge 0)$ defined by the following Lindley's equation: $W_0=0$, and

$$W_{t+1}=[W_t+S_t-\tilde{D}_t]^+, \tag{13}$$

where the sequences S_t and $\tilde{D}_t$ consist of independent and identically distributed random variables. Let $\tau_0=0$, $\tau_1=\inf\{t\ge 1: W_t=0\}$ and for $k\ge 1$, $\tau_{k+1}=\inf\{t>\tau_k: W_t=0\}$. Let $B_k=\tau_k-\tau_{k-1}$. The random variable W_t is the waiting time of the t^{th} customer in the $GI/G/1$ queue, where the inter-arrival time between the t^{th} and $t+1^{th}$ customers is distributed as $\tilde{D}_t$, and the service time is distributed as S_t. Then, B_k is the length of the k^{th} busy period. Let $\rho=\mathbb{E}[S_1]/\mathbb{E}[\tilde{D}_1]$ represent the system utilization. It is well-known that in a $GI/G/1$ queue, if $\rho\le 1$, then the queue is stable and the random variable B_k is independent and identically distributed. Note that this stability condition $\rho\le 1$ can always be satisfied by appropriately scaling the units of cost parameters.

We invoke the following result from Loulou (1978) to bound $\mathbb{E}[B]$, the expected busy period of a $GI/G/1$ queue with inter-arrival distribution D_n and service distribution S_n.

THEOREM 3 **(Loulou 1978)**. *Let $X_n=S_n-D_n$, and $\alpha=-\mathbb{E}[X_1]$. Let σ^2 be the variance of X_1. If $\mathbb{E}[X_1^3]=\beta<\infty$, and $\rho<1$,*

$$\mathbb{E}[B]\le\frac{\sigma}{\alpha}\exp\left(\frac{6\beta}{\sigma^3}+\frac{\alpha}{\sigma}\right).$$

We can now obtain an upper bound on our expected busy period $\mathbb{E}[B]$ for the stochastic process W_t defined in (13), by setting $X_1=\sum_{i=1}^n(h^i+2(p^i-c^i))-\tilde{D}_1$ (whose expectation is negative since $\rho\le 1$).

With the explicit form of $\mathbb{E}[B]$, Lemma 8 gives the upper bound of our distance function below. The idea is to connect the upper-bounding stochastic process (which evolves as a $GI/G/1$ queue) with the expected busy period of this queue (where there exists an explicit upper bound that does not depend on the time horizon T).

LEMMA 8. *The total expected distance function*

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n}|y_t^i-\hat{y}_t^i|\right]\le 2\mathbb{E}[B]S\sqrt{T},$$

where $\mathbb{E}[B] \leq \frac{\sigma}{\alpha}\exp\left(\frac{6\beta}{\sigma^3}+\frac{\alpha}{\sigma}\right)$, *and* $S = \sum_{i=1}^{n}\left(h^i + 2\left(p^i - c^i\right)\right)$.

The proof of Lemma 2 then follows from Lemma 3 and Lemma 8, with their proofs provided in the Electronic Companion.

5. Extensions

Improving the convergence rate. If we change Assumption 1(c) slightly to enforce a uniform lower bound $\delta > 0$ on the density of demand, i.e., $F'_{D^i}(x) \geq \delta > 0$ for all $x \in [0, M]$ and all $i = 1, \ldots, n$, and also change the step size $\eta_t = O(1/t)$ in the algorithm, one can readily show that the cost function is δ-strongly convex, and the rate of convergence of DDM can be improved to $O(\log T/T)$.

Different product dimensions or sizes. Our basic model (defined in Section 2) assumes that all products have exactly the same dimension or sizes. However, in general, different products may have different dimension or sizes. Let $v^1, v^2, \ldots, v^n$ denote the sizes of the different products, and

$$\mathbf{y}_t \in \Gamma \triangleq \left\{ \mathbf{y}_t \in \mathbb{R}^n_+ : \sum_{i=1}^{n} v^i y_t^i \leq M \right\}, \tag{14}$$

By a simple cost tranformation, we show in the Electronic Companion that our algorithm DDM (now defined in terms of transformed variables) and its performance analysis remain the same.

Discrete demand and order quantities. In practice, the demand and ordering quantities are often integers. We provide a modified algorithm (denoted by DDM-Discrete) in the Electronic Companion to handle such discrete cases, which achieves the same convergence rate $O(1/\sqrt{T})$ with the aid of lost-sales indicators (i.e., the firm knows whether lost-sales has occurred in each period).

6. Numerical Experiments

We compare the performance of DDM with several existing parametric and nonparametric approaches in the literature (briefly described below). Our results show that DDM outperforms these benchmark algorithms in terms of both consistency and convergence rate. We relegate the detailed experimental setup and numerical results (figures) to the Electronic Companion.

1. **Algorithm a1 (Known Distribution): Clairvoyant Optimal Policy.**
2. **Algorithm a2 (Uncensored): Uncensored SAA.** This is a sample average approximation (SAA) algorithm with uncensored demand (a hypothetical situation). The target inventory level is the quantile of the empirical demand distribution using uncensored demand data.
3. **Algorithm b1 (Parametric): MLE Censored.** Assuming the correct parametric form has been pre-specified, this parametric policy uses censored demand data to construct maximum likelihood estimators (MLE) for the parameters in the demand distribution.

4. **Algorithm c1 (Nonparametric): Burnetas-Smith (B-S) Policy.** The B-S policy is a nonparametric policy which was developed by Burnetas and Smith (2000).
5. **Algorithm c2 (Nonparametric): CAVE Policy.** The CAVE policy, developed by Godfrey and Powell (2001), is a nonparametric approach by approximating the underlying objective function using a series of piecewise linear functions.

Comparison with parametric MLE algorithms. The numerical results are presented in Figure EC.1. Our results indicate that DDM performs very well, and is consistent (i.e., it converges to the optimal solution). In contrast, MLE Censored is significantly slower than DDM, and also suffers from inconsistency, i.e., it often fails to converge to the optimal solution. This is due to a spiral-down effect. More specifically, if the initial inventory level is lower than the optimal target level, the censored demand is likely to give an even lower estimate in the next period (because the firm cannot observe the lost-sales quantity). Then the target inventory level will be set lower and lower, resulting in divergent cost. The consistency of MLE Censored hinges on the (almost) perfect initial estimation of target levels, which is often impractical. In fact, in three of the examples in Figure EC.1, the MLE Censored algorithm did not converge; in the only one where it did converge, we actually picked starting target levels close enough to the optimal levels so that it would converge, which of course would not be possible in practice.

Comparison with nonparametric algorithms. The numerical results are presented in Figure EC.2. Our results show that DDM consistently outperforms both the B-S policy and the CAVE policy. We also find out that the B-S policy has an extremely slow convergence rate while the CAVE policy is much faster but still slower than DDM. Figure EC.2 also displays the performance of the Uncensored SAA policy (assuming the uncensored demand information). It is interesting to note that the DDM policy performs very close to the Uncensored SAA policy in all of our examples.

Extreme cases with uneven lost-sales penalty costs. DDM performs consistently very well for extreme cases with some pathological parameters (see Figure EC.3).

Supplemental Material

An electronic companion to this paper is available at http://or.journal.informs.org/.

Acknowledgments

The authors are grateful to the area editor Professor Chung-Piaw Teo, the anonymous associate editor, and the three anonymous referees for their detailed comments and suggestions, which have helped to significantly improve both the content and the exposition of this paper. We thank Huseyin Topaloglu for sharing the CAVE implementation. This research is partially supported by NSF grants CMMI-1362619 and CMMI-1451078.

References

Bertsekas, D. P. 2000. *Dynamic Programming and Optimal Control*. 2nd ed. Athena Scientific.

Bertsekas, D. P., S. E. Shreve. 2007. *Stochastic Optimal Control: The Discrete-Time Case*. Athena Scientific.

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

Beyer, D., S. P. Sethi, R. Sridhar. 2001. Stochastic multi-product inventory models with limited storage. *Journal of Optimization Theory and Applications* **111** 553–588.

Beyer, D., S. P. Sethi, R. Sridhar. 2002. Average-cost optimality of a base-stock policy for a multi-product inventory model with limited storage. G. Zaccour, ed., *Decision & Control in Management Science*, *Advances in Computational Management Science*, vol. 4. Springer US, 241–260.

Bookbinder, J. H., A. E. Lordahl. 1989. Estimation of inventory re-order levels using the bootstrap statistical procedure. *IIE Transactions* **21**(4) 302–312.

Boyd, S., L. Vandenberghe. 2004. *Convex Optimization*. Cambridge University Press, New York, NY, USA.

Burnetas, A. N., C. E. Smith. 2000. Adaptive ordering and pricing for perishable products. *Operations Research* **48**(3) 436–443.

Chen, L., E. L. Plambeck. 2008. Dynamic inventory management with learning about the demand distribution and substitution probability. *Manufacturing & Service Operations Management* **10**(2) 236–256.

Choi, J., J. J. Cao, H. E. Romeijn, J. Geunes, S. X. Bai. 2005. A stochastic multi-item inventory model with unequal replenishment intervals and limited warehouse capacity. *IIE Transactions* **37**(12) 1129–1141.

Chu, L. Y., J. G. Shanthikumar, Z.-J. M. Shen. 2008. Solving operational statistics via a bayesian analysis. *Operations Research Letters* **36**(1) 110 – 116.

Flaxman, A. D., A. T. Kalai, H. B. McMahan. 2005. Online convex optimization in the bandit setting: Gradient descent without a gradient. *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*. SODA '05, 385–394.

Glasserman, P. 1991. *Gradient Estimation Via Perturbation Analysis*. Kluwer international series in engineering and computer science: Discrete event dynamic systems, Springer.

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

Hazan, E. 2015. Introduction to online convex optimization. Book Draft. Computer Science, Princeton University. Available at http://ocobook.cs.princeton.edu/OCObook.pdf.

Hazan, E., A. Kalai, S. Kale, A. Agarwal. 2006. Logarithmic regret algorithms for online convex optimization. *In 19th COLT*. 499–513.

Huh, W. H., P. Rusmevichientong. 2009. A non-parametric asymptotic analysis of inventory planning with censored demand. *Mathematics of Operations Research* **34**(1) 103–123.

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

Huh, W. T., G. Janakiraman, J. A. Muckstadt, P. Rusmevichientong. 2009. An adaptive algorithm for finding the optimal base-stock policy in lost sales inventory systems with censored demand. *Mathematics of Operations Research* **34**(2) 397–416.

Ignall, E., A. F. Veinott. 1969. Optimality of myopic inventory policies for several substitute products. *Management Science* **15**(5) 284–304.

Kleywegt, A. J., A. Shapiro, T. Homem-de Mello. 2002. The sample average approximation method for stochastic discrete optimization. *SIAM J. on Optimization* **12**(2) 479–502.

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

Lariviere, M. A., E. L. Porteus. 1999. Stalking information: Bayesian inventory management with unobserved lost sales. *Management Science* **45**(3) 346–363.

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

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

Liyanage, L. H., J. G. Shanthikumar. 2005. A practical inventory control policy using operational statistics. *Operations Research Letters* **33**(4) 341 – 348.

Loulou, R. 1978. An explicit upper bound for the mean busy period in a GI/G/1 queue. *Journal of Applied Probability* **15**(2) 452–455.

Maglaras, C., S. Eren. 2015. A maximum entropy joint demand estimation and capacity control policy. *Production and Operations Management* **24**(3) 438–450.

Nemirovski, A., A. Juditsky, G. Lan, A. Shapiro. 2009. Robust stochastic approximation approach to stochastic programming. *SIAM J. on Optimization* **19**(4).

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

Shaked, M., J. G. Shanthikumar. 2007. *Stochastic Orders*. Springer Series in Statistics, Physica-Verlag.

Shalev-Shwartz, S. 2012. Online learning and online convex optimization. *Found. Trends Mach. Learn.* **4**(2) 107–194.

Veinott, Jr., A. F. 1965. Optimal policy for a multi-product, dynamic, nonstationary inventory problem. *Management Science* **12**(3) pp. 206–222.

Zinkevich, M. 2003. Online convex programming and generalized infinitesimal gradient ascent. Tom Fawcett, Nina Mishra, eds., *Proceedings of the 20th International Conference on Machine Learning (ICML)*. AAAI Press, Cambridge, MA, USA, 928–936.

Zipkin, P. 2000. *Foundations of Inventory Management*. McGraw-Hill, New York.

Brief Bio:

Cong Shi is an assistant professor in the Department of Industrial and Operations Engineering at the University of Michigan. His research lies in stochastic optimization and online learning theory with applications to inventory and supply chain management, and revenue management. He won first prize in the 2009 George Nicholson Student Paper Competition.

Weidong Chen is a Ph.D. candidate in the Department of Industrial and Operations Engineering at the University of Michigan. His research lies in stochastic optimization and online learning theory with applications to inventory and supply chain management, and revenue management.

Izak Duenyas is the Donald C. Cook Professor of Business Administration and Professor of Technology and Operations at the Ross School of Business. His recent research interests are in pricing and revenue management, sourcing and procurement, and production, inventory and capacity control.

This page is intentionally blank. Proper e-companion title page, with INFORMS branding and exact metadata of the main paper, will be produced by the INFORMS office when the issue is being assembled.

Electronic Companion to "Nonparametric Data-Driven Algorithms for Multi-Product Inventory Systems with Censored Demand"

Cong Shi*, Weidong Chen*, Izak Duenyas†

* Industrial and Operations Engineering, University of Michigan, {shicong, aschenwd}@umich.edu

† Technology and Operations, Ross School of Business, University of Michigan, duenyas@umich.edu

EC.1. Clairvoyant Optimal Policy – Proof of Theorem 1

Based on (3), we define a *myopic* feasible (closed-loop) policy $\bar{\pi}$ as a sequence of functions $\bar{\mathbf{y}}_t = \bar{\pi}_t(\mathbf{x}_t)$, $t = 1, \ldots, T$, mapping beginning inventory (state) $\mathbf{x}_t$ into ending inventory (decision) $\bar{\mathbf{y}}_t$, which also "myopically" minimizes per-period cost $\Pi_t(\cdot)$ with beginning inventory $\mathbf{x}_t$, i.e.,

$$\bar{\mathbf{y}}_t(\mathbf{x}_t) \triangleq \arg \min_{\mathbf{a} \in \Gamma : \mathbf{a} \geq \mathbf{x}_t} \Pi_t(\mathbf{a}). \tag{EC.1}$$

The above feasible policy $\bar{\pi}$ is myopic, because it only optimizes per-period cost in each period (the immediate reward). This is in contrast with standard dynamic programming or approximate dynamic programming approaches. To ease the presentation of establishing optimality of $\bar{\pi}$, following Ignall and Veinott (1969), we keep $\mathbf{x}_t$, $\bar{\mathbf{y}}_t$, Π_t time-generic, i.e.,

$$\bar{\mathbf{y}}(\mathbf{x}) \triangleq \arg \min_{\mathbf{a} \in \Gamma : \mathbf{a} \geq \mathbf{x}} \Pi(\mathbf{a}). \tag{EC.2}$$

It is important to see that $\bar{\mathbf{y}}(\mathbf{x})$ is the unique minimizer of (EC.2), due to Assumption 1 ensuring strict (joint) convexity of $\Pi(\mathbf{y})$ over the feasible region, and the fact that the constraint set is affine (see Boyd and Vandenberghe (2004)).

LEMMA EC.1. *The optimization problem defined in* (EC.2) *has a unique minimizer* $\bar{\mathbf{y}}(\mathbf{x})$.

Proof of Lemma EC.1. Due to Assumption 1, the cost function $\Pi(\cdot)$ is differentiable and finite-valued. The derivatives inside expectation are bounded, and also the expectation is a multiple integration over finite ranges. Hence this guarantees the validity of interchange between differentiation and expectation.

Next we argue that $\Pi(\cdot)$ are strictly (jointly) convex over the feasible region. For all i and j,

$$\frac{\partial^2 \Pi(\mathbf{a})}{\partial (a^i)^2} = (h^i + p^i - c^i) F'_{D^i}(a^i) > 0; \quad \frac{\partial^2 \Pi(\mathbf{a})}{\partial a^i \partial a^j} = 0,$$

where Assumption 1(c) ensures $F'_{D^i}(a^i) > 0$ for all $a^i \in [0, M]$. Hence, the Hessian matrix is positive definte (with all strictly positive eigenvalues) over the entire feasible region, ensuring Π to be strictly (jointly) convex.

Now consider the optimization problem (with a given starting inventory $\mathbf{x}$) defined in (EC.2). Since $\Pi(\mathbf{y})$ is strictly (jointly) convex and the constraint set is affine, $\bar{\mathbf{y}}(\mathbf{x})$ is the unique minimizer. (See Boyd and Vandenberghe (2004) for discussions of unique minimizer in convex optimization problems and also Example 5.4.). **Q.E.D.**

Next we shall show that the myopic policy $\bar{\pi}$ defined above is optimal. Ignall and Veinott (1969) provided a sufficient condition called *substitute property* (together with two mild regularity assumptions) under which the myopic policy is optimal.

Definition EC.1 (Substitute property). For any inventory levels $\mathbf{x}, \tilde{\mathbf{x}} \in \Gamma$,

$$\text{if } \mathbf{x} \geq \tilde{\mathbf{x}}, \text{ then } \bar{\mathbf{y}}(\mathbf{x}) - \mathbf{x} \leq \bar{\mathbf{y}}(\tilde{\mathbf{x}}) - \tilde{\mathbf{x}}.$$

Definition EC.2 (Regularity conditions in Ignall and Veinott (1969)). The two regularity conditions in Ignall and Veinott (1969) are: (a) $\mathbf{x} \leq \mathbf{x}' \leq \bar{\mathbf{y}}(\mathbf{x})$ implies $\bar{\mathbf{y}}(\mathbf{x}) = \bar{\mathbf{y}}(\mathbf{x}')$ for $\mathbf{x}, \mathbf{x}' \in \Gamma$; (b) The state transition permits either pure, partial, or no backlogging (lost-sales).

The regularity condition (a) is satisfied by $\bar{\mathbf{y}}(\mathbf{x})$ being the unique minimizer of (EC.2) by Lemma EC.1, and the regularity condition (b) is immediate since we consider a standard lost-sales model.

We can now proceed to establish the optimality of myopic policies for the multi-product lost-sales system by showing that the sufficient condition (substitute property) given above holds for our system.

Proposition EC.1. *Under Assumption 1, when the demand distribution is known a priori, the myopic ordering policy defined in* (EC.1) *is optimal for the multi-product lost-sales inventory systems.*

To prove Proposition EC.1, we need to derive several important properties of the myopic policy. Now consider the two possible starting inventory levels $\mathbf{x}$ and $\tilde{\mathbf{x}}$, with $\mathbf{x} \geq \tilde{\mathbf{x}}$. For notational (superscript) convenience, we use θ instead of $\mathbf{y}^*$ to be the global minimizer of $\Pi(\cdot)$ over Γ. Recall that $\theta = \mathbf{y}^* \triangleq \arg\min_{\mathbf{a} \in \mathbf{\Gamma}} \Pi(\mathbf{a})$, and also the myopic order-to-up level $\bar{\mathbf{y}}(\mathbf{x}) \triangleq \arg\min_{\mathbf{a} \in \mathbf{\Gamma}: \mathbf{a} \geq \mathbf{x}} \Pi(\mathbf{a})$. For simplicity, we define the boundary of our warehouse storage constraint,

$$\partial\Gamma \triangleq \left\{ \mathbf{y} \in \mathbb{R}^n_+ : \sum_{i=1}^n y^i = M \right\}.$$

Note that $\mathbf{y} \in \partial\Gamma$ means that the total order-up-to levels have reached the total storage limit M. If $\mathbf{y} \notin \partial\Gamma$, then the warehouse storage constraint is not tight.

Now denote the j^{th} partial derivative of $\Pi(\cdot)$ by $\Pi'_j(\cdot)$. We then develop some useful properties of the myopic order-up-to levels $\bar{\mathbf{y}}(\cdot)$.

LEMMA EC.2. *Let $\mathbf{x} \in \Gamma$ and θ be the global minimizer of $\Pi(\cdot)$ over Γ,*

(i) $x^j \geq \theta^j \Rightarrow \bar{y}^j(\mathbf{x}) = x^j$.

(ii) $x^j \leq \theta^j \Rightarrow \bar{y}^j(\mathbf{x}) \leq \theta^j$.

Proof of Lemma EC.2. The proof is straightforward. Statement (i) holds because $x^j \geq \theta^j$ (the starting inventory is higher than the global minimizer) for product j, it is sub-optimal to order any more product j. Statement (ii) holds because if $x^j \leq \theta^j$ (the starting inventory is lower than the global minimizer), it is sub-optimal to raise the inventory above the global minimizer. **Q.E.D.**

LEMMA EC.3. *Let $\mathbf{x} \in \Gamma$ and θ be the global minimizer of $\Pi(\cdot)$ over Γ,*

(i) $\theta \in \partial\Gamma \Rightarrow \bar{\mathbf{y}}(\mathbf{x}) \in \partial\Gamma$;

(ii) $\bar{\mathbf{y}}(\mathbf{x}) \notin \partial\Gamma, x^j \leq \theta^j \Rightarrow \bar{y}^j(\mathbf{x}) = \theta^j$.

In Lemma EC.3, statement (i) states that if the global minimizer occupies the entire storage space, then the myopic order-up-to levels will also occupy the entire storage space. This is because our myopic policy will always order as much as possible to approach the global minimizer. Statement (ii) states that if the total myopic order-up-to level has not reached the storage limit M, then if $x^j \leq \theta^j$, the myopic policy will raise inventory level for product j to the global minimizer θ^j.

Proof of Lemma EC.3. We prove (i) by contradiction. Suppose that $\theta \in \partial\Gamma$ and $\bar{\mathbf{y}}(\mathbf{x}) \notin \partial\Gamma$, then

$$\sum_{i=1}^{n} \theta^i = M \qquad \text{and} \qquad \sum_{i=1}^{n} \bar{y}^i(\mathbf{x}) < M.$$

It is obvious that there exists at least one j such that $\bar{y}^j(\mathbf{x}) < \theta^j$. Since θ minimizes $\Pi(\cdot)$ over Γ, it is clear that θ either reaches the global minimizer of $\Pi(\cdot)$ over the entire real line $\mathbb{R}$ or is smaller than it due to the storage constraint, so the derivative $\Pi'_j(\theta) \leq 0$. Therefore, since $\Pi(\cdot)$ is strictly convex,

$$\Pi'_j(\bar{\mathbf{y}}(\mathbf{x})) < \Pi'_j(\theta) \leq 0.$$

On the other hand, since $\bar{\mathbf{y}}(\mathbf{x}) \notin \partial\Gamma$ and $\bar{\mathbf{y}}(\mathbf{x})$ is a minimizer of $\Pi(\cdot)$ over set $\{\mathbf{y} \mid \mathbf{y} \geq \mathbf{x}, \mathbf{y} \in \Gamma\}$, it is clear that $\bar{\mathbf{y}}(\mathbf{x})$ either reaches θ or is greater than it because of the initial on-hand inventory, so $\Pi'_j(\bar{\mathbf{y}}(\mathbf{x})) \geq 0$, which results in a contradiction, thereby proving (i).

To prove (ii), we observe from the contraposition of (i), i.e., $\bar{\mathbf{y}}(\mathbf{x}) \notin \partial\Gamma \Rightarrow \theta \notin \partial\Gamma$. Then for any product j, $\bar{y}^j(\mathbf{x})$ is not restricted by the storage constraint, and thus if $\theta^j \geq x^j$, then θ^j can always be reached, implying that $\bar{y}^j(\mathbf{x}) = \theta^j$. This completes the proof. **Q.E.D.**

LEMMA EC.4. $\bar{y}^j(\mathbf{x}) > x^j \Rightarrow \Pi'_j(\bar{\mathbf{y}}(\mathbf{x})) = \min_i \Pi'_i(\bar{\mathbf{y}}(\mathbf{x}))$

Lemma EC.4 states that if a product is ordered, then the marginal cost of any additional ordering must be equal across the products. Intuitively, if the marginal cost of ordering this product is higher than others, we can always reduce the quantity of this product and order more of the other products. The rigorous proof is as follows.

Proof of Lemma EC.4. We prove this result by contradiction. Suppose that there exists an i, $1 \le i \le n$, such that $\Pi'_i(\bar{\mathbf{y}}(\mathbf{x})) < \Pi'_j(\bar{\mathbf{y}}(\mathbf{x}))$. Then, for a sufficiently small $\epsilon > 0$, $(\bar{y}^1(\mathbf{x}), ...\bar{y}^j(\mathbf{x}) - \epsilon, ..., \bar{y}^i(\mathbf{x}) + \epsilon, ..., \bar{y}^n(\mathbf{x})) \in \Gamma$, and we have

$$\begin{aligned} &\Pi(\bar{\mathbf{y}}(\mathbf{x})) - \Pi(\bar{y}^1(\mathbf{x}), ...\bar{y}^j(\mathbf{x}) - \epsilon, ..., \bar{y}^i(\mathbf{x}) + \epsilon, ..., \bar{y}^n(\mathbf{x})) \\ =\ &\epsilon(\Pi'_j(\bar{\mathbf{y}}(\mathbf{x})) - \Pi'_i(\bar{\mathbf{y}}(\mathbf{x}))) + o(\epsilon^2) > 0, \end{aligned}$$

which contradicts to the fact that $\bar{\mathbf{y}}(\mathbf{x})$ minimizes $\Pi(\cdot)$ over set $\{\mathbf{y} \mid \mathbf{y} \ge \mathbf{x}, \mathbf{y} \in \Gamma\}$. **Q.E.D.**

Now, we are ready to prove Proposition EC.1.

Proof of Proposition EC.1. To establish the optimality of myopic policies for the multi-product lost-sales system, it suffices to verify that the substitute property (EC.1) holds, i.e., for any inventory levels $\mathbf{x}, \tilde{\mathbf{x}} \in \Gamma$, if $\mathbf{x} \ge \tilde{\mathbf{x}}$, then $\bar{\mathbf{y}}(\mathbf{x}) - \mathbf{x} \le \bar{\mathbf{y}}(\tilde{\mathbf{x}}) - \tilde{\mathbf{x}}$.

We know that the myopic order-up-to levels $\bar{y}^j(\mathbf{x}) \ge x^j$ for any product j if $\mathbf{x} \in \Gamma$. Similarly, $\bar{y}^j(\tilde{\mathbf{x}}) \ge \tilde{x}^j$ for any product j if $\tilde{\mathbf{x}} \in \Gamma$. Now if $\bar{y}^j(\mathbf{x}) = x^j$, then we have

$$0 = \bar{y}^j(\mathbf{x}) - x^j \le \bar{y}^j(\tilde{\mathbf{x}}) - \tilde{x}^j.$$

Thus, it suffices to prove that $\bar{y}^j(\mathbf{x}) \le \bar{y}^j(\tilde{\mathbf{x}})$, whenever $\bar{y}^j(\mathbf{x}) > x^j$. We have to consider three cases as follows.

Case (a). First, if both $\bar{\mathbf{y}}(\mathbf{x}) \notin \partial\Gamma$ and $\bar{y}^j(\tilde{\mathbf{x}}) \notin \partial\Gamma$, then it follows from Lemma EC.2 and Lemma EC.3 that

$$\bar{y}^j(\mathbf{x}) = \max\{\theta^j, x^j\}, \qquad \bar{y}^j(\tilde{\mathbf{x}}) = \max\{\theta^j, \tilde{x}^j\}, \qquad \forall\, j.$$

Then $\bar{y}^j(\mathbf{x}) = \bar{y}^j(\tilde{\mathbf{x}})$ and the result follows immediately.

Case (b). Second, if $\bar{\mathbf{y}}(\mathbf{x}) \in \partial\Gamma$ but $\bar{y}^j(\tilde{\mathbf{x}}) \notin \partial\Gamma$, then by Lemma EC.2 (ii) and Lemma EC.3 (ii), we have $\bar{y}^j(\mathbf{x}) \le \theta^j = \bar{y}^j(\tilde{\mathbf{x}})$, and the result also follows immediately. It is impossible for the case where $\bar{\mathbf{y}}(\mathbf{x}) \notin \partial\Gamma$ and $\bar{y}^j(\tilde{\mathbf{x}}) \in \partial\Gamma$ to happen. To see this, if such case exists, then we can always find some j such that for $x^j > \tilde{x}^j$, $\bar{y}^j(\tilde{\mathbf{x}}) > \bar{y}^j(\mathbf{x})$. However, by Lemma EC.2 (ii) and Lemma EC.3 (ii), we know that $\bar{y}^j(\mathbf{x}) \le \theta^j = \bar{y}^j(\tilde{\mathbf{x}})$, which results in a contradiction.

Case (c). Third, we need to analyze the remaining case where $\bar{\mathbf{y}}(\mathbf{x}) \in \partial\Gamma$ and $\bar{y}^j(\tilde{\mathbf{x}}) \in \partial\Gamma$, i.e.,

$$\sum_{j=1}^{n} \bar{y}^j(\mathbf{x}) = \sum_{j=1}^{n} \bar{y}^j(\tilde{\mathbf{x}}) = M. \tag{EC.3}$$

We partition all the products into three sets as follows,

$$I^a = \{k : \bar{y}^k(\mathbf{x}) > x^k\},\ I^b = \{k : \bar{y}^k(\mathbf{x}) = x^k \cap \Pi'_k(\bar{y}^k(\mathbf{x}) \le 0\},\ I^c = \{k : \Pi'_k(\bar{y}^k(\mathbf{x}) > 0\}.$$

Note that these three sets are disjoint and the union of them is exhaustive.

Now we focus on the set I^c first and let $j \in I^c$. Then we have $\bar{y}^j(\mathbf{x}) \ge \max\{\tilde{x}^j, \theta^j\}$. By Lemma EC.2, it is clear that $\bar{y}^j(\tilde{\mathbf{x}}) \le \max\{\tilde{x}^j, \theta^j\}$. Hence, $\bar{y}^j(\tilde{\mathbf{x}}) - \bar{y}^j(\mathbf{x}) \le 0$ for all $j \in I^c$. Together with (EC.3), we know that

$$\sum_{j \in I^a \cup I^b}^{n} \left(\bar{y}^j(\tilde{\mathbf{x}}) - \bar{y}^j(\mathbf{x})\right) \ge 0.$$

If $\bar{y}^m(\tilde{\mathbf{x}}) = \bar{y}^m(\mathbf{x})$ for all $m \in I^a \cup I^b$, then the result follows immediately. Now consider the case where there exists a product $m \in I^a \cup I^b$ such that $\bar{y}^m(\tilde{\mathbf{x}}) > \bar{y}^m(\mathbf{x})$. This implies that $\bar{y}^m(\tilde{\mathbf{x}}) > \bar{y}^m(\mathbf{x}) \ge x^m \ge \tilde{x}^m \ge 0$. By Lemma EC.4, we have $\Pi'_m(\bar{\mathbf{y}}(\tilde{\mathbf{x}})) = \min_i \Pi'_i(\bar{\mathbf{y}}(\tilde{\mathbf{x}}))$. Moreover, due to the strict convexity of $\Pi(\cdot)$, then we have

$$\min_i \Pi'_i(\bar{\mathbf{y}}(\tilde{\mathbf{x}})) = \Pi'_m(\bar{\mathbf{y}}(\tilde{\mathbf{x}})) > \Pi'_m(\bar{\mathbf{y}}(\mathbf{x})). \tag{EC.4}$$

To complete the proof, it suffices to show that for any product $j \in I^a$, $\bar{y}^j(\tilde{\mathbf{x}}) \ge \bar{y}^j(\mathbf{x})$. Now suppose there exists a product $n \in I^a$ such that $\bar{y}^n(\tilde{\mathbf{x}}) < \bar{y}^n(\mathbf{x})$. It is clear that $\bar{y}^n(\mathbf{x}) > \bar{y}^n(\tilde{\mathbf{x}}) \ge 0$. By Lemma EC.4, we have $\Pi'_n(\bar{\mathbf{y}}(\mathbf{x})) = \min_i \Pi'_i(\bar{\mathbf{y}}(\mathbf{x}))$. Moreover, due to the strict convexity of $\Pi(\cdot)$, then we have

$$\min_i \Pi'_i(\bar{\mathbf{y}}(\mathbf{x})) = \Pi'_n(\bar{\mathbf{y}}(\mathbf{x})) > \Pi'_n(\bar{\mathbf{y}}(\tilde{\mathbf{x}})). \tag{EC.5}$$

Note that (EC.4) implies that $\Pi'_n(\bar{\mathbf{y}}(\tilde{\mathbf{x}})) > \Pi'_m(\bar{\mathbf{y}}(\mathbf{x}))$ but (EC.5) implies that $\Pi'_m(\bar{\mathbf{y}}(\mathbf{x})) > \Pi'_n(\bar{\mathbf{y}}(\tilde{\mathbf{x}}))$, which results in a contradiction. This completes the proof. **Q.E.D.**

Equipped with Proposition EC.1, we are ready to prove Theorem 1.

Proof of Theorem 1. Proposition EC.1 fully characterizes the structural properties of optimal policies as follows. Let $\mathbf{y}^*$ be a unique critical (deterministic) vector defined by in (4). Then a clairvoyant optimal policy π^* is characterized as follows:

(a) If the beginning inventory level of product i is above its individual base-stock level (i.e., the i^{th} component of $\mathbf{y}^*$), then this product is not ordered in the period.

(b) If this product i is ordered in the period, the ending inventory level (after ordering) does not exceed its individual base-stock level (i.e., the i^{th} component of $\mathbf{y}^*$).

(c) If there is enough storage space to bring all products (whose inventory levels are below their individual base-stock levels) up to their base-stock levels, then such an order is optimal. Otherwise, the ending inventory levels takes up all the available storage space.

Thus, the stationary multi-period inventory problem is analytically equivalent to the single-period problem, and ordering up to $\mathbf{y}^*$ in each period is also optimal for this problem. Clearly, once we start below $\mathbf{y}^*$, and order up to $\mathbf{y}^*$, we remain at or below $\mathbf{y}^*$ thereafter; in such a case, the expected cost incurred in each period is $\Pi(\mathbf{y}^*)$. **Q.E.D.**

EC.2. Proof of Lemma 1 - Online Convex Optimization

Proof of Lemma 1. Due to convexity of the cost function $\Pi(\mathbf{y})$, we have

$$\mathbb{E}\left[\Pi(\hat{\mathbf{y}}_t) - \Pi(\mathbf{y}^*)\right] \leq \mathbb{E}\left[\nabla\Pi(\hat{\mathbf{y}}_t)(\hat{\mathbf{y}}_t - \mathbf{y}^*)\right]. \tag{EC.6}$$

Note that the subgradient $\nabla\Pi(\hat{\mathbf{y}}_t)$ defines the supporting hyperplane of Π at the point $\hat{\mathbf{y}}_t$.

For any period $t \in \mathcal{T}$, i.e., in the set of break points, we can obtain the upper bound of the second moment difference between our target inventory level and the optimal target inventory level.

$$\begin{aligned}
\mathbb{E}||\hat{\mathbf{y}}_{t+1} - \mathbf{y}^*||^2 &\leq \mathbb{E}||\mathbf{z}_{t+1} - \mathbf{y}^*||^2 \\
&= \mathbb{E}||\hat{\mathbf{y}}_t - \eta_t G_t(\hat{\mathbf{y}}_t) - \mathbf{y}^*||^2 \\
&= \mathbb{E}||\hat{\mathbf{y}}_t - \mathbf{y}^*||^2 + \eta_t^2\mathbb{E}||G_t(\hat{\mathbf{y}}_t)||^2 - 2\eta_t\mathbb{E}[G_{t_n}(\hat{\mathbf{y}}_t)(\hat{\mathbf{y}}_t - \mathbf{y}^*)],
\end{aligned} \tag{EC.7}$$

where the first inequality follows the optimization (7) and the Pythagorean Theorem since

$$||\mathbf{z}_{t+1} - \mathbf{y}^*||^2 = ||\hat{\mathbf{y}}_{t+1} - \mathbf{y}^*||^2 + ||\mathbf{z}_{t+1} - \hat{\mathbf{y}}_{t+1}||^2$$

by property of the 2-norm projection; the first equality follows from the definition of $\mathbf{z}_{t+1}$; the second equality follows from a simple binomial expansion.

We can also re-write $\mathbb{E}[G_t(\hat{\mathbf{y}}_t)(\hat{\mathbf{y}}_t - \mathbf{y}^*)]$ by taking conditional expectation on the value of $\hat{\mathbf{y}}_t$,

$$\begin{aligned}
\mathbb{E}\left[G_t(\hat{\mathbf{y}}_t)(\hat{\mathbf{y}}_t - \mathbf{y}^*)\right] &= \mathbb{E}\left[\mathbb{E}\left[G_t(\hat{\mathbf{y}}_t)(\hat{\mathbf{y}}_t - \mathbf{y}^*)|\hat{\mathbf{y}}_t\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[G_t(\hat{\mathbf{y}}_t)|\hat{\mathbf{y}}_t\right](\hat{\mathbf{y}}_t - \mathbf{y}^*)\right] \\
&= \mathbb{E}\left[\nabla\Pi(\hat{\mathbf{y}}_t)(\hat{\mathbf{y}}_t - \mathbf{y}^*)\right],
\end{aligned} \tag{EC.8}$$

where the first equality holds because $\mathbf{y}^*$ does not relate with $\hat{\mathbf{y}}_t$; the last equality follows from the fact that G_t is an unbiased estimator of the true gradient $\nabla\Pi$.

Combining (EC.7) and (EC.8), it is clear that

$$\mathbb{E}[\nabla\Pi(\hat{\mathbf{y}}_t)(\hat{\mathbf{y}}_t-\mathbf{y}^*)] \le \frac{1}{2\eta_t}\left(\mathbb{E}||\hat{\mathbf{y}}_t-\mathbf{y}^*||^2-\mathbb{E}||\hat{\mathbf{y}}_{t+1}-\mathbf{y}^*||^2\right)+\frac{\eta_t}{2}\mathbb{E}||G_t(\hat{\mathbf{y}}_t)||^2. \tag{EC.9}$$

Without loss of generality, let $\mathcal{T}=\{\tau_0,\dots,\tau_k\}$ with $\tau_0=0$ and $\tau_k=T$. By the construction of DDM,

$$\mathbb{E}\left[\sum_{t=1}^{T}\Pi(\hat{\mathbf{y}}_t)-\sum_{t=1}^{T}\Pi(\mathbf{y}^*)\right]=\mathbb{E}\left[\sum_{s=0}^{k-1}\sum_{t=\tau_s+1}^{\tau_{s+1}}(\Pi(\hat{\mathbf{y}}_t)-\Pi(\mathbf{y}^*))\right]\le\frac{M}{l}\cdot\mathbb{E}\left[\sum_{s=1}^{k}(\Pi(\hat{\mathbf{y}}_{\tau_s})-\Pi(\mathbf{y}^*))\right],$$

where the inequality follows from the fact that the time between any two consecutive break points cannot exceed the time for a "fictitious" system with M inventory units for each product $i=1,\dots,n$ to become empty along every sample path. The expectation of the latter (which is independent of $\hat{\mathbf{y}}_t$) is upper bounded by M/l.

It then suffices to bound the term $\mathbb{E}\left[\sum_{s=1}^{k}(\Pi(\hat{\mathbf{y}}_{\tau_s})-\Pi(\mathbf{y}^*))\right]$. Now, by summing both sides of (EC.6) over periods τ_1 to τ_k,

$$\begin{aligned}
&\mathbb{E}\left[\sum_{s=1}^{k}(\Pi(\hat{\mathbf{y}}_{\tau_s})-\Pi(\mathbf{y}^*))\right]\le\sum_{s=1}^{k}\mathbb{E}\left[\nabla\Pi(\hat{\mathbf{y}}_{\tau_s})(\hat{\mathbf{y}}_{\tau_s}-\mathbf{y}^*)\right] \qquad \text{(EC.10)}\\
&\le\sum_{s=1}^{k}\left(\frac{1}{2\eta_{\tau_s}}\left(\mathbb{E}||\hat{\mathbf{y}}_{\tau_s}-\mathbf{y}^*||^2-\mathbb{E}||\hat{\mathbf{y}}_{\tau_s+1}-\mathbf{y}^*||^2\right)+\frac{\eta_{\tau_s}}{2}\mathbb{E}||G_{\tau_s}(\hat{\mathbf{y}}_{\tau_s})||^2\right)\\
&=\sum_{s=1}^{k}\left(\frac{1}{2\eta_{\tau_s}}\left(\mathbb{E}||\hat{\mathbf{y}}_{\tau_s}-\mathbf{y}^*||^2-\mathbb{E}||\hat{\mathbf{y}}_{\tau_{s+1}}-\mathbf{y}^*||^2\right)+\frac{\eta_{\tau_s}}{2}\mathbb{E}||G_{\tau_s}(\hat{\mathbf{y}}_{\tau_s})||^2\right)\\
&=\frac{1}{2\eta_{\tau_1}}\mathbb{E}||\hat{\mathbf{y}}_{\tau_1}-\mathbf{y}^*||^2-\frac{1}{2\eta_{\tau_k}}\mathbb{E}||\hat{\mathbf{y}}_{\tau_{k+1}}-\mathbf{y}^*||^2+\frac{1}{2}\sum_{s=2}^{k}\left(\frac{1}{\eta_{\tau_s}}-\frac{1}{\eta_{\tau_{s-1}}}\right)\mathbb{E}||\hat{\mathbf{y}}_{\tau_s}-\mathbf{y}^*||^2\\
&\quad+\sum_{s=1}^{k}\eta_{\tau_s}\frac{\mathbb{E}||G_{\tau_s}(\hat{\mathbf{y}}_{\tau_s})||^2}{2}\\
&\le 2M^2\left(\frac{1}{2\eta_{\tau_1}}+\frac{1}{2}\sum_{s=2}^{k}\left(\frac{1}{\eta_{\tau_s}}-\frac{1}{\eta_{\tau_{s-1}}}\right)\right)+\frac{n(\max_i\{p^i-c^i,h^i\})^2}{2}\sum_{s=1}^{k}\eta_{\tau_s}\\
&=\frac{M^2}{\eta_{\tau_k}}+\frac{n(\max_i\{p^i-c^i,h^i\})^2}{2}\sum_{s=1}^{k}\eta_{\tau_s},
\end{aligned}$$

where the first and second inequalities follows from (EC.6) and (EC.9), respectively; the first equality holds since $\hat{\mathbf{y}}_{\tau_s+1}=\hat{\mathbf{y}}_{\tau_{s+1}}$ by the construction of DDM; the last inequality follows from the fact that for any $\mathbf{x},\mathbf{y}\in\Gamma$,

$$||\mathbf{x}-\mathbf{y}||_2^2\le||\mathbf{x}||_2^2+||\mathbf{y}||_2^2\le||\mathbf{x}||_1^2+||\mathbf{y}||_1^2\le 2M^2.$$

Putting everything together, we have

$$\mathbb{E}\left[\sum_{t=1}^{T}\Pi(\hat{\mathbf{y}}_t)-\sum_{t=1}^{T}\Pi(\mathbf{y}^*)\right]\le\frac{M}{l}\left(\frac{M^2}{\eta_T}+\frac{n(\max_i\{p^i-c^i,h^i\})^2}{2}\sum_{s=1}^{k}\eta_{\tau_s}\right). \tag{EC.11}$$

Note that we have chosen our step size "optimally" as

$$\eta_t=\left(\frac{\gamma M}{\sqrt{n}\cdot\max_i\{p^i-c^i,h^i\}}\right)\frac{1}{\sqrt{t}}\text{ for some }\gamma>0,$$

so that

$$\sum_{s=1}^{k}\eta_{\tau_s}\le\sum_{t=1}^{T}\eta_t=\left(\frac{\gamma M}{\sqrt{n}\cdot\max_i\{p^i-c^i,h^i\}}\right)\sum_{t=1}^{T}\frac{1}{\sqrt{t}}\le\left(\frac{\gamma M}{\sqrt{n}\cdot\max_i\{p^i-c^i,h^i\}}\right)2\sqrt{T}. \tag{EC.12}$$

Plugging (EC.12) and η_T into (EC.11) yields the result with the constant term

$$K_1=(\gamma+\gamma^{-1})M^2l^{-1}\sqrt{n}\cdot\max_i\{p^i-c^i,h^i\}.$$

Note that putting $\gamma=1$ gives the tightest bound. This completes the proof. **Q.E.D.**

EC.3. Proof of Lemma 2 - Stochastic Dominance and a GI/G/1 Queue

Proof of Lemma 3. By the definition of the per-period cost function in (3), it follows that

$$\begin{aligned}&\mathbb{E}[\Pi(\mathbf{y}_t)-\Pi(\hat{\mathbf{y}}_t)]\le\mathbb{E}\left[\mathbf{c}\cdot(\mathbf{y}_t-\hat{\mathbf{y}}_t)\right]+\mathbb{E}\left[(\mathbf{h}-\mathbf{c})\cdot(\mathbf{y}_t-\hat{\mathbf{y}}_t)^+\right]+\mathbb{E}\left[\mathbf{p}\cdot(\hat{\mathbf{y}}_t-\mathbf{y}_t)^+\right]\\&=\mathbb{E}\left[\mathbf{h}\cdot(\mathbf{y}_t-\hat{\mathbf{y}}_t)^+\right]+\mathbb{E}\left[(\mathbf{p}-\mathbf{c})\cdot(\hat{\mathbf{y}}_t-\mathbf{y}_t)^+\right]\le\mathbb{E}\left[(\mathbf{h}\vee(\mathbf{p}-\mathbf{c})\cdot|\mathbf{y}_t-\hat{\mathbf{y}}_t|\right],\end{aligned}$$

where the last inequality follows from various operators defined at the end of Section 1. **Q.E.D.**

Proof of Lemma 4. **Case 1.** We first consider time period $t\in\mathcal{T}=\{\tau_0,\ldots,\tau_k\}$, which belongs to the set of break points in DDM. Due to the construction of DDM, we update the target levels at $t+1$ only if $t\in\mathcal{T}$. For each product $i\in J$, i.e., $x^i_{t+1}>\hat{y}^i_{t+1}$, we have $y^i_{t+1}=x^i_{t+1}>\hat{y}^i_{t+1}\ge 0$ by (9). This implies that $y^i_{t+1}>0$, and by the lost-sales system dynamics, we have

$$x^i_{t+1}=(y^i_t-d_t)^+=y^i_t-d_t>0. \tag{EC.13}$$

The next key step is to compare the target level $\hat{y}^i_{t+1}$ with the constraint-free target level z^i_{t+1}. First, notice that when $y^i_t-d^i_t>0$, the algorithm updates the constraint-free target level in a negative direction, i.e.,

$$z^i_{t+1}=\hat{y}^i_t-\eta_t h^i<\hat{y}^i_t. \tag{EC.14}$$

Second, by the important property (12) of our algorithm, we have

$$\hat{y}_{t+1}^j \le \hat{y}_t^j + \eta_t \left(p^j - c^j\right), \quad \forall\, j = 1, \ldots, n.$$

Thus, the maximum positive displacement of $\hat{\mathbf{y}}_{t+1}$ from $\hat{\mathbf{y}}_t$ (excluding the set J) is

$$\sum_{j\in\bar{J}} \left(\hat{y}_{t+1}^j - \hat{y}_t^j\right) \le \sum_{j\in\bar{J}} \eta_t \left(p^j - c^j\right). \tag{EC.15}$$

Now, to draw a relation between $\hat{y}_{t+1}^i$ and z_{t+1}^i, there are two cases.

Subcase 1a. In the first case where $\sum_{j=1}^n \hat{y}_{t+1}^j \ge \sum_{j=1}^n \hat{y}_t^j$, we must have

$$\sum_{i\in J} \left(z_{t+1}^i - \hat{y}_{t+1}^i\right) < \sum_{i\in J} \left(\hat{y}_t^i - \hat{y}_{t+1}^i\right) \le \sum_{j\in\bar{J}} \left(\hat{y}_{t+1}^j - \hat{y}_t^j\right) \le \sum_{j\in\bar{J}} \eta_t \left(p^j - c^j\right), \tag{EC.16}$$

where the first inequality follows from (EC.14); the second inequality follows from $\sum_{j=1}^n \hat{y}_{t+1}^j \ge \sum_{j=1}^n \hat{y}_t^j$; and the third inequality follows from (EC.15).

Subcase 1b. In the second case where $\sum_{j=1}^n \hat{y}_{t+1}^j < \sum_{j=1}^n \hat{y}_t^j \le M$, the warehouse storage constraint is not tight (i.e., the constraint-free target levels are in the interior of Γ), and by the optimization procedure (10), $z_{t+1}^j = \hat{y}_{t+1}^j$ for all $j = 1, \ldots, n$. Thus, we have

$$z_{t+1}^i - \hat{y}_{t+1}^i = \hat{y}_{t+1}^i - \hat{y}_{t+1}^i = 0, \quad i = 1, \ldots, n. \tag{EC.17}$$

Combining the above two cases and using the relations (EC.16) and (EC.17), we can then obtain an upper bound for our distance function as follows,

$$\begin{aligned}
\sum_{i\in J} |y_{t+1}^i - \hat{y}_{t+1}^i| &= \sum_{i\in J} \left(x_{t+1}^i - \hat{y}_{t+1}^i\right) = \sum_{i\in J} \left(y_t^i - d_t^i - \hat{y}_{t+1}^i\right) \\
&\le \sum_{i\in J} \left(y_t^i - d_t^i - z_{t+1}^i\right) + \sum_{j\in\bar{J}} \eta_t \left(p^j - c^j\right) \\
&= \sum_{i\in J} \left(y_t^i - \hat{y}_t^i\right) + \eta_t \left(\sum_{i\in J} h^i + \sum_{j\in\bar{J}} \left(p^j - c^j\right)\right) - \sum_{i\in J} d_t^i, \\
&\le \sum_{i\in J} \left|y_t^i - \hat{y}_t^i\right| + \eta_t \left(\sum_{i\in J} h^i + \sum_{j\in\bar{J}} \left(p^j - c^j\right)\right) - \sum_{i\in J} d_t^i,
\end{aligned}$$

where the first equality follows from the fact that $i \in J$ and the construction of our algorithm (9); the second equality is due to (EC.13); the first inequality follows from (EC.16) and (EC.17), and the third equality follows from (EC.14). Now we have completed the proof for Case 1.

Case 2. We then consider time period $t \notin \mathcal{T}$, which does not belong to the set of break points in DDM. According to the construction of DDM, the target order-up-to levels are kept unchanged,

i.e., $\hat{y}^i_{t+1} = \hat{y}^i_t$ for all $i = 1, \ldots, n$ and for all $t \notin \mathcal{T}$. we can similarly obtain an upper bound for our distance function as follows,

$$\begin{aligned}\sum_{i\in J} |y^i_{t+1} - \hat{y}^i_{t+1}| &= \sum_{i\in J} \left(x^i_{t+1} - \hat{y}^i_{t+1}\right) = \sum_{i\in J} \left(y^i_t - d^i_t - \hat{y}^i_{t+1}\right) \\ &= \sum_{i\in J} \left(y^i_t - d^i_t - \hat{y}^i_t\right) \le \sum_{i\in J} \left|y^i_t - \hat{y}^i_t\right| - \sum_{i\in J} d^i_t,\end{aligned}$$

where the first equality follows from the fact that $i \in J$ and the construction of our algorithm (9); the second equality is due to (EC.13); the third equality follows from $\hat{y}^i_{t+1} = \hat{y}^i_t$. Now we have completed the proof for Case 2. **Q.E.D.**

Proof of Lemma 5. **Case 1.** We first consider time period $t \in \mathcal{T} = \{\tau_0, \ldots, \tau_k\}$, which belongs to the set of break points in DDM. Due to the construction of DDM, we update the target levels at $t+1$ only if $t \in \mathcal{T}$. For each product $i \in \bar{J}$, i.e., $x^i_{t+1} \le \hat{y}^i_{t+1}$, recall that we need to solve the optimization problem (10) to determine our actual implemented levels $\mathbf{y}_{t+1}$. That is,

$$\min \sum_{i\in\bar{J}} (\hat{y}^i_{t+1} - y^i_{t+1})^2 \quad \text{s.t.} \quad \sum_{i\in\bar{J}} y^i_{t+1} \le M - \sum_{j\in J} x^j_{t+1}, \quad y^i_{t+1} \ge x^i_{t+1}, \, \forall\, i \in \bar{J}.$$

It is straightforward to see that $\hat{y}^i_{t+1} \ge y^i_{t+1}$ for each product $j \in \bar{J}$. To see this, suppose otherwise $\hat{y}^i_{t+1} < y^i_{t+1}$; we can always lower the value of y^i_{t+1} to $\hat{y}^i_{t+1}$ strictly improving the objective value while maintaining feasibility.

Now there are three sub-cases.

Subcase 1a. The simplest case is when $J = \emptyset$, then (10) reduces to

$$\min \sum_{i=1}^{n} (\hat{y}^i_{t+1} - y^i_{t+1})^2 \quad \text{s.t.} \quad \sum_{i\in\bar{J}} y^i_{t+1} \le M, \quad y^i_{t+1} \ge x^i_{t+1}, \, \forall\, i \in \bar{J}.$$

Since $\hat{\mathbf{y}}_{t+1} \in \Gamma$, we have $y^i_{t+1} = \hat{y}^i_{t+1}$ for each product $i = 1, \ldots, n$, and thus the distance function is zero for each product $i = 1, \ldots, n$.

Subcase 1b. The second case is when upon solving $\mathbf{y}_{t+1}$, the warehouse storage constraint is not tight, i.e.,

$$\sum_{i\in\bar{J}} y^i_{t+1} < M - \sum_{j\in J} x^j_{t+1}. \tag{EC.18}$$

Then we claim that

$$\sum_{i\in\bar{J}} \hat{y}^i_{t+1} < M - \sum_{j\in J} x^j_{t+1}, \tag{EC.19}$$

We argue the claim by contradiction. Suppose otherwise that

$$\sum_{i\in\bar{J}} \hat{y}^i_{t+1} \geq M - \sum_{j\in J} x^j_{t+1} > \sum_{i\in\bar{J}} y^i_{t+1}.$$

Then there must exist a product k such that $y^k_{t+1} < \hat{y}^k_{t+1}$, you can always increase y^k_{t+1} by

$$\epsilon \triangleq M - \sum_{j\in J} x^j_{t+1} - \sum_{i\in\bar{J}} y^i_{t+1} > 0$$

to make the warehouse storage constraint tight, thereby strictly reducing the optimal objective value. This contradicts the optimality of $\mathbf{y}_{t+1}$ in (10).

Thus, by (EC.18) and (EC.19), we have $y^i_{t+1} = \hat{y}^i_{t+1}$ for each product $i \in \bar{J}$, and the distance function is zero for each product $i = 1, \ldots, n$.

Subcase 1c. The third case is much more involved. That is, upon solving $\mathbf{y}_{t+1}$, the warehouse storage constraint becomes tight, i.e.,

$$\sum_{i\in\bar{J}} y^i_{t+1} = M - \sum_{j\in J} x^j_{t+1},$$

and the set $J \neq \emptyset$. We can then rewrite the optimization problem (10) as follows,

$$\min \sum_{i\in\bar{J}} (\hat{y}^i_{t+1} - y^i_{t+1})^2 \quad \text{s.t.} \quad \sum_{i\in\bar{J}} \left(\hat{y}^i_{t+1} - y^i_{t+1}\right) = \sum_{i\in\bar{J}} \hat{y}^i_{t+1} - M + \sum_{j\in J} x^j_{t+1}, \quad y^i_{t+1} \geq x^i_{t+1}, \forall\, i \in \bar{J}.$$

We then bound the distance function as follows,

$$\begin{aligned}
\sum_{i\in\bar{J}} \left|\hat{y}^i_{t+1} - y^i_{t+1}\right| &= \sum_{i\in\bar{J}} \hat{y}^i_{t+1} - M + \sum_{j\in J} x^j_{t+1} = \sum_{i\in\bar{J}} \hat{y}^i_{t+1} - M + \sum_{j\in J} (y^j_t - d^j_t)^+ \\
&= \sum_{i\in\bar{J}} \hat{y}^i_{t+1} - M + \sum_{j\in J} (y^j_t - d^j_t) = \sum_{i\in\bar{J}} \hat{y}^i_{t+1} - \left(M - \sum_{j\in J} y^j_t\right) - \sum_{j\in J} d^j_t \\
&\leq \sum_{i\in\bar{J}} \hat{y}^i_{t+1} - \sum_{i\in\bar{J}} y^i_t - \sum_{j\in J} d^j_t = \sum_{i\in\bar{J}} \left(\hat{y}^i_{t+1} - y^i_t\right) - \sum_{j\in J} d^j_t \\
&\leq \sum_{i\in\bar{J}} \left(\hat{y}^i_t + \eta_t(p^i - c^i) - y^i_t\right) - \sum_{j\in J} d^j_t \\
&\leq \sum_{i\in\bar{J}} \left(\hat{y}^i_t - y^i_t\right) + \sum_{i\in\bar{J}} \eta_t(p^i - c^i) - \sum_{j\in J} d^j_t \\
&\leq \sum_{i\in\bar{J}} \left|\hat{y}^i_t - y^i_t\right| + \eta_t \sum_{i\in\bar{J}} (p^i - c^i) - \sum_{j\in J} d^j_t,
\end{aligned}$$

where the first equality is because the warehouse storage constraint becomes tight; the second equality is due to the system dynamics; the third equality is because $j \in J$ implies that $x^j_{t+1} > 0$,

and hence the plus sign can be removed; the first inequality is due to the fact that

$$\sum_{j\in\bar{J}} y_t^j + \sum_{j\in J} y_t^j = \sum_{j=1}^{n} y_t^j \le M;$$

and the second inequality follows from the important property (12) of our algorithm. Now we have completed the proof for Case 1.

Case 2. We then consider time period $t \notin \mathcal{T}$, which does not belong to the set of break points in DDM. Due to the construction of DDM, the target order-up-to levels are kept unchanged, i.e., $\hat{y}_{t+1}^i = \hat{y}_t^i$ for all $i = 1,\dots,n$ and for all $t \notin \mathcal{T}$. Also, if $t \notin \mathcal{T}$, then the optimization problem (10) has a nonzero objective value, which suggests that the warehouse storage constraint has to be tight. We then similarly bound the distance function as follows,

$$\begin{aligned}
\sum_{i\in\bar{J}} \left|\hat{y}_{t+1}^i - y_{t+1}^i\right| &= \sum_{i\in\bar{J}} \hat{y}_{t+1}^i - M + \sum_{j\in J} x_{t+1}^j = \sum_{i\in\bar{J}} \hat{y}_{t+1}^i - M + \sum_{j\in J} (y_t^j - d_t^j)^+ \\
&= \sum_{i\in\bar{J}} \hat{y}_{t+1}^i - M + \sum_{j\in J} (y_t^j - d_t^j) = \sum_{i\in\bar{J}} \hat{y}_{t+1}^i - \left(M - \sum_{j\in J} y_t^j\right) - \sum_{j\in J} d_t^j \\
&\le \sum_{i\in\bar{J}} \hat{y}_{t+1}^i - \sum_{i\in\bar{J}} y_t^i - \sum_{j\in J} d_t^j = \sum_{i\in\bar{J}} \left(\hat{y}_{t+1}^i - y_t^i\right) - \sum_{j\in J} d_t^j \\
&= \sum_{i\in\bar{J}} \left(\hat{y}_t^i - y_t^i\right) - \sum_{j\in J} d_t^j \le \sum_{i\in\bar{J}} \left|\hat{y}_t^i - y_t^i\right| - \sum_{j\in J} d_t^j,
\end{aligned}$$

where we used the same arguments as in Subcase 1c and also the fact that $\hat{y}_{t+1}^i = \hat{y}_t^i$ for all $i = 1,\dots,n$ if $t \notin \mathcal{T}$. Now we have completed the proof for Case 2. **Q.E.D.**

Proof of Lemma 6. By Lemma 4, we have

$$\sum_{i\in J} |y_{t+1}^i - \hat{y}_{t+1}^i| \le \left(\sum_{i\in J} \left|y_t^i - \hat{y}_t^i\right| + \eta_t \left(\sum_{i\in J} h^i + \sum_{j\in\bar{J}} \left(p^j - c^j\right)\right) - \sum_{i\in J} d_t^i\right)^+,$$

and by Lemma 5, we have

$$\sum_{i\in\bar{J}} |y_{t+1}^i - \hat{y}_{t+1}^i| \le \left(\sum_{i\in\bar{J}} \left|y_t^i - \hat{y}_t^i\right| + \eta_t \sum_{i\in\bar{J}} (p^i - c^i) - \min_{j=1,\dots,n} d_t^j\right)^+.$$

Combining the above two inequalities yields the result. **Q.E.D.**

Proof of Lemma 7. By Lemma 6, for each period $t+1$, the sum of distance functions

$$\sum_{i=1}^{n} |y_{t+1}^i - \hat{y}_{t+1}^i| \le \left(\sum_{i=1}^{n} \left|y_t^i - \hat{y}_t^i\right| + \eta_t \sum_{i=1}^{n} \left(h^i + 2\left(p^i - c^i\right)\right) - \min_{j=1,\dots,n} d_t^j\right)^+.$$

In addition, we know that $\sum_{i=1}^{n}|y_0^i-\hat{y}_0^i|=0$ (since the policy starts with zero inventory). Thus, by the definition of the stochastic process Z_{t+1}, it is clear that $\sum_{i=1}^{n}|y_{t+1}^i-\hat{y}_{t+1}^i|\leq_{st} Z_{t+1}$. This implies that $\sum_{i=1}^{n}|y_{t+1}^i-\hat{y}_{t+1}^i|\leq_{cx} Z_{t+1}$, and then the result follows immediately. **Q.E.D.**

Proof of Lemma 8. By Lemma 7, it suffices to show that $\mathbb{E}\left[\sum_{t=1}^{T}Z_t\right]\leq 2\mathbb{E}[B]S\sqrt{T}$. Recall that

$$Z_{t+1}=\left[Z_t+\frac{S_t}{\sqrt{t}}-\tilde{D}_t\right]^+,\qquad Z_0=0.$$

Let the random variable $l(t)$ denote the index k in which B_k contains t, and it is clear that

$$Z_t\leq\sum_{s=1}^{t}\frac{S_s}{\sqrt{s}}\mathbb{1}\left[s\in B_{l(t)}\right]\ \text{a.s.}$$

By summing Z_t over periods 1 to T and taking expectation, we have

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}Z_t\right]&\leq\mathbb{E}\left[\sum_{t=1}^{T}\sum_{s=1}^{t}\frac{S_s}{\sqrt{s}}\mathbb{1}\left[s\in B_{l(t)}\right]\right]\leq\mathbb{E}\left[\sum_{s=1}^{T}\frac{S_s}{\sqrt{s}}\sum_{t=1}^{T}\mathbb{1}\left[s\in B_{l(t)}\right]\right]\\&=\mathbb{E}\left[\sum_{s=1}^{T}\frac{S_s}{\sqrt{s}}B_{l(s)}\right]=\sum_{s=1}^{T}\frac{1}{\sqrt{s}}\mathbb{E}[B_1]S\leq 2\mathbb{E}[B]S\sqrt{T}.\end{aligned}$$

This completes the proof. **Q.E.D.**

Proof of Lemma 2. Combining Lemma 3 and Lemma 8, we have

$$\begin{aligned}\Delta_2(T)&=\mathbb{E}\left[\sum_{t=1}^{T}(\Pi(\mathbf{y}_t)-\Pi(\hat{\mathbf{y}}_t))\right]\leq\mathbb{E}\left[\sum_{t=1}^{T}(\mathbf{h}\vee(\mathbf{p}-\mathbf{c}))\cdot|\mathbf{y}_t-\hat{\mathbf{y}}_t|\right]\\&\leq\max_i\{p^i-c^i,h^i\}\mathbb{E}\left[\sum_{t=1}^{T}|\mathbf{y}_t-\hat{\mathbf{y}}_t|\right]=\max_i\{p^i-c^i,h^i\}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{n}|y_t^i-\hat{y}_t^i|\right]\\&\leq\max_i\{p^i-c^i,h^i\}\left(2\sqrt{T}\mathbb{E}[B]S\right)=\left(2\max_i\{p^i-c^i,h^i\}\mathbb{E}[B]S\right)\sqrt{T}.\end{aligned}$$

Recall that $\mathbb{E}[B]\leq\frac{\sigma}{\alpha}e^{\frac{6\beta}{\sigma^3}+\frac{\alpha}{\sigma}}$ and $S=\sum_{i=1}^{n}\left(h^i+2\left(p^i-c^i\right)\right)$. Setting the constant

$$K_2=2\max_i\{p^i-c^i,h^i\}\frac{\sigma}{\alpha}e^{\frac{6\beta}{\sigma^3}+\frac{\alpha}{\sigma}}\left\{\sum_{i=1}^{n}\left(h^i+2\left(p^i-c^i\right)\right)\right\}.$$

yields the result. This completes the proof. **Q.E.D.**

EC.4. Extensions

We present the detailed arguments to the following extensions of our model.

EC.4.1. Different Product Dimensions or Sizes

We prove that by a cost transformation, this general model reduces to the basic model and hence our results remain the same. We define new decision variables as follows,

$$\tilde{y}_t^i = v^i y_t^i, \qquad \tilde{x}_t^i = v^i x_t^i, \qquad \tilde{q}_t^i = v^i q_t^i,$$

for $i = 1, \ldots, n$. In addition, we appropriately scale the demand and cost parameters as follows,

$$\tilde{D}_t^i = v^i D_t^i, \qquad \tilde{c}^i = c^i / v^i, \qquad \tilde{h}^i = h^i / v^i, \qquad \tilde{p}^i = p^i / v^i$$

for $i = 1, \ldots, n$. With the above transformation, the cost of a feasible policy π under the new warehouse-capacity constraint (14) can be transformed as follows,

$$\begin{aligned} \mathcal{C}(\pi) &= \mathbb{E}\left[\sum_{t=1}^{T} \mathbf{c}\cdot(\mathbf{y}_t - \mathbf{x}_t) + \mathbf{h}\cdot(\mathbf{y}_t - \mathbf{D}_t)^+ + \mathbf{p}\cdot(\mathbf{D}_t - \mathbf{y}_t)^+\right] - \mathbb{E}[\mathbf{c}\cdot\mathbf{x}_{T+1}], \qquad \text{(EC.20)} \\ &= \mathbb{E}\left[\sum_{t=1}^{T} \tilde{\mathbf{c}}\cdot(\tilde{\mathbf{y}}_t - \tilde{\mathbf{x}}_t) + \tilde{\mathbf{h}}\cdot(\tilde{\mathbf{y}}_t - \tilde{\mathbf{D}}_t)^+ + \tilde{\mathbf{p}}\cdot(\tilde{\mathbf{D}}_t - \tilde{\mathbf{y}}_t)^+\right] - \mathbb{E}[\tilde{\mathbf{c}}\cdot\tilde{\mathbf{x}}_{T+1}]. \end{aligned}$$

Moreover, it is clear that the new constraint defined in (14) is equivalent to

$$\tilde{\mathbf{y}}_t \in \Gamma \triangleq \left\{\tilde{\mathbf{y}}_t \in \mathbb{R}_+^n : \sum_{i=1}^{n} \tilde{y}_t^i \le M\right\},$$

which has the same form as in the original constraint defined in (1). Hence, this more general model has been reduced to the basic model. Our data-driven algorithm (now defined in terms of transformed variables) and its performance analysis remain the same.

EC.4.2. Discrete Demand and Ordering Quantities

Data-Driven Multi-product Algorithm for Discrete Demands (DDM-Discrete).

Step 0. (Initialization.) Set the initial inventory levels $\mathbf{y}_0 = \hat{\mathbf{y}}_0 = \bar{\mathbf{y}}_0$ to be any non-negative integer values within Γ and then set the initial values $t = 0$, $\tau_0 = 0$ and $k = 0$.

For each period $t = 0, \ldots, T-1$, repeat the following steps:

Step 1. (Setting the constraint-free and constrained target inventory levels.)

Case 1: If $\mathbf{y}_t \ge \hat{\mathbf{y}}_t$, the algorithm updates the constraint-free target inventory levels $\mathbf{z}_{t+1}$ by (5); however, for each product $i = 1, \ldots, n$, the i^{th} component of $\hat{\mathbf{G}}_t$ is defined as

$$\hat{G}_t^i(\hat{\mathbf{y}}_t) = \begin{cases} -p^i + c_t^i + (h^i + p^i - c^i)\cdot \mathbb{1}(d_t^i \le \hat{y}_t^i), & \text{if } \hat{y}_t^i = \lfloor \bar{y}_t^i \rfloor, \\ -p^i + c_t^i + (h^i + p^i - c^i)\cdot \mathbb{1}(d_t^i \le \hat{y}_t^i - 1), & \text{if } \hat{y}_t^i = \lceil \bar{y}_t^i \rceil, \end{cases} \qquad \text{(EC.21)}$$

which is the right derivative of Π at $\lfloor \bar{y}_t^i \rfloor$, i.e., the slope of Π at $\bar{y}_t^i$ for the piece-wise linear cost.

Then the algorithm obtains an intermediate (continuous) target level $\bar{\mathbf{y}}_{t+1}$ by solving $\bar{\mathbf{y}}_{t+1} = \arg\min_{\mathbf{w}\in\Gamma} ||\mathbf{w} - \mathbf{z}_{t+1}||_2$. Then set the constrained (discrete) target inventory levels $\hat{\mathbf{y}}_{t+1}$ by probabilistic rounding. That is, if $\bar{y}_{t+1}^i$ is already an integer, set $\hat{y}_{t+1}^i = \bar{y}_{t+1}^i$; otherwise, we flip a (biased) coin with probability $\hat{y}_{t+1}^i - \lfloor \bar{y}_{t+1}^i \rfloor$ of heads. Set $\hat{y}_{t+1}^i = \lceil \bar{y}_{t+1}^i \rceil$ if the outcome is head and set $\hat{y}_{t+1}^i = \lfloor \bar{y}_{t+1}^i \rfloor$ if the outcome is tail.

Record the break point $\tau_k := t+1$ and increase the value k by 1.

Case 2: Else if $\mathbf{y}_t \not\geq \hat{\mathbf{y}}_t$, the algorithm keeps both the constraint-free and constrained target inventory levels unchanged, i.e., $\mathbf{z}_{t+1} = \mathbf{z}_t$ and $\hat{\mathbf{y}}_{t+1} = \hat{\mathbf{y}}_t$.

Step 2. (Solving for the actual implemented target inventory levels.)

Define the set J and its complement as $J \triangleq \left\{i : x_{t+1}^i > \hat{y}_{t+1}^i\right\}$ and $\bar{J} \triangleq \left\{i : x_{t+1}^i \leq \hat{y}_{t+1}^i\right\}$. For each product $i \in J$, we set the actual implemented levels $y_{t+1}^i = x_{t+1}^i$ if $x_{t+1}^i > \hat{y}_{t+1}^i$.

If $\bar{J} \neq \emptyset$, then we set the actual implemented levels $\mathbf{y}_{t+1}$ by solving

$$\min \sum_{i\in\bar{J}} (\hat{y}_{t+1}^i - y_{t+1}^i)^2 \quad \text{s.t.} \quad \sum_{i\in\bar{J}} y_{t+1}^i \leq M - \sum_{j\in J} x_{t+1}^j,\ y_{t+1}^i \geq x_{t+1}^i,\ y_{t+1}^i \in \mathbb{Z}^+,\ \forall\, i \in \bar{J}. \tag{EC.22}$$

This concludes the description of the algorithm.

Note that the key differences between DDM-Discrete and DDM are in Step 1 – defining a modified gradient $\hat{\mathbf{G}}_t$, and probabilistic rounding. In order to establish our performance guarantee, we need a *lost-sales indicator* for whether lost-sales has occurred in each period t, thereby determining the value of $\mathbb{1}(\hat{y}_t^i \leq d_t^i)$. Without this indicator, it is not sufficient to obtain an unbiased estimator for the right derivative of our cost function $\Pi(\cdot)$ by observing the past sales quantities. The decision maker no longer has access to a local (stochastic) direction of cost improvement. For example, if the computed target inventory level is 15.5 for some product i and we round it down to 15, even an infinite number of sales observations would not allow the decision maker to obtain an estimate of the slope of $\Pi^i(\cdot)$ at 15.5. This is because if the demand turns out to be exactly 15, the unbiased gradient should be h^i since our target level 15.5 is higher than 15; however, if the demand turns out to be 16, then the unbiased gradient should be $-p^i + c^i$. In both cases, we observe zero inventory but cannot determine if the demand is strictly greater than 15 without a lost-sales indicator.

However, with access to this lost-sales indicator, we can construct such an estimator $\hat{\mathbf{G}}_t$ defined in (EC.21), which is unbiased when $t \in \mathcal{T}$. Then the proofs of bounds Δ_1 and Δ_2 are almost identical to the ones used in DDM as long as the warehouse-capacity M is also an integer. Hence we are able to extend our results to the discrete demand and inventory case as stated in the theorem below.

THEOREM EC.1. *Assume that the clairvoyant optimal solution in* (EC.2) *is unique in the discrete demand case. With access to the lost-sales indicator, the average regret* $\mathcal{R}_T/T$ *of our data-driven*

algorithm for discrete demand case (DDM-Discrete) approaches 0 *at the rate of* $1/\sqrt{T}$. *That is, there exists some constant* K, *such that for any* $T \geq 1$,

$$\frac{1}{T}\mathcal{R}_T \triangleq \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}\Pi(\mathbf{y}_t)\right] - \Pi(\mathbf{y}^*) \leq \frac{K}{\sqrt{T}}.$$

To the best of our knowledge, the availability of the lost-sales indicator has been assumed in all nonparametric studies that analyze newsvendor-type problems with an unknown discrete demand distribution (see, e.g., Huh and Rusmevichientong (2009) and Besbes and Muharremoglu (2013)). They showed that active exploration plays a much stronger role in the discrete case (compared to the continuous case). However, the need for active exploration disappears as soon as a lost-sales indicator (that records whether demand was censored or not) becomes available, in addition to the censored demand samples. The access to this indicator allows the decision maker to obtain a noisy signal about the potential need for an upward correction.

EC.5. Numerical Experiments

We present the setup of our numerical experiments and also the numerical results.

EC.5.1. Experimental Setup

For each experiment, we specify a (hindsight) demand distribution with cumulative distribution function $F(\cdot)$. The lost-sales penalty cost p^i for each product i is randomly drawn from the interval $[70, 90]$ and the purchasing cost c^i for each product i is randomly drawn from the interval $[55, 65]$. We then set the holding cost $h^i = 0.02c^i$ for each product i (see, e.g., Zipkin (2000)). To compare the cost under each algorithm, we evaluate each algorithm on $N = 200$ randomly generated problem instances. Each problem instance consists of independent demand samples and parameters over a time horizon of 500 periods, unless specified otherwise. For each algorithm π, we compute the average cost till period t, which is given by

$$\frac{1}{N}\sum_{j=1}^{N}\frac{1}{t}\sum_{s=1}^{t}\tilde{\Pi}_{j,s}(y_{j,s}^{\pi}),$$

where the one-period cost $\tilde{\Pi}_{j,s}(y)$ in period s of the problem instance j is given by

$$\tilde{\Pi}_{j,s}(y) = \sum_{i=1}^{n} c^i y_{j,s}^{i,\pi} + (h^i - c^i)(y_{j,s}^{i,\pi} - d_{j,s}^i)^+ + p^i(d_{j,s}^i - y_{j,s}^{i,\pi})^+,$$

where $d_{j,s}^i$ is the demand realization for product i in period s of the problem instance j, and $y_{j,s}^{i,\pi}$ is the corresponding order-up-to level computed by each algorithm π.

EC.5.2. Numerical Results

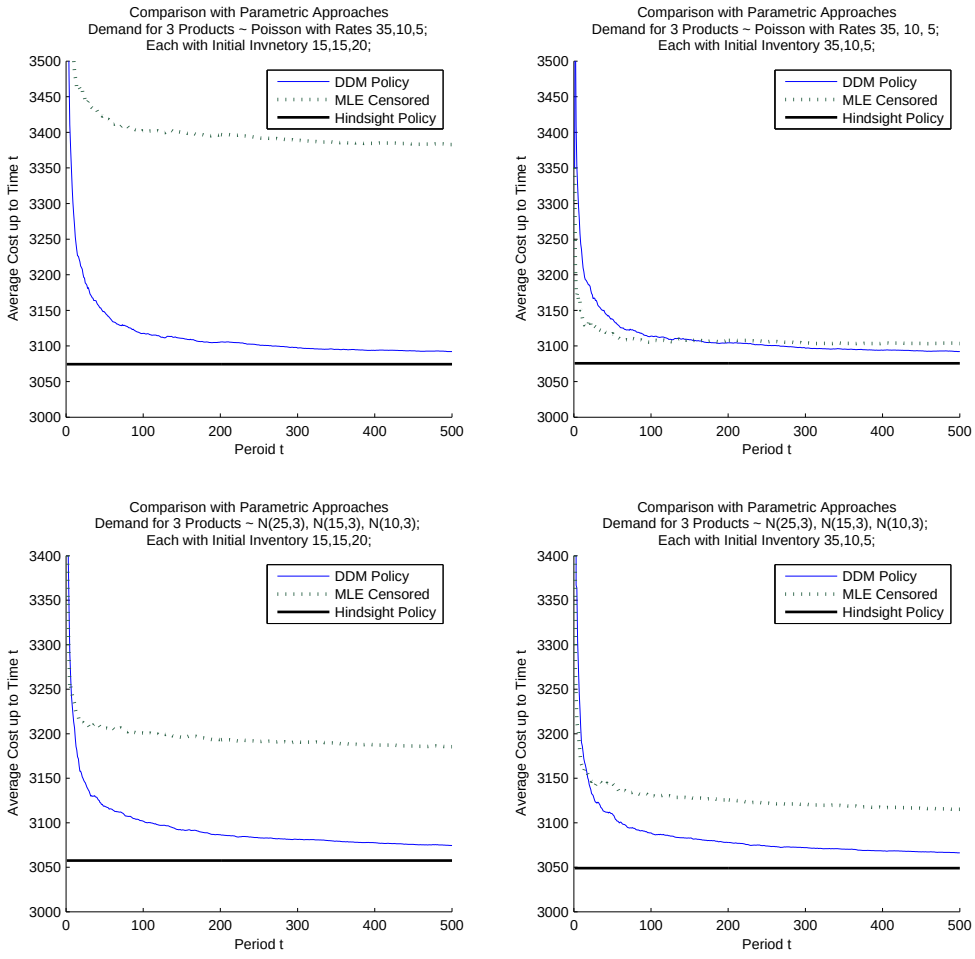

Figure EC.1 Comparison with parametric approaches.

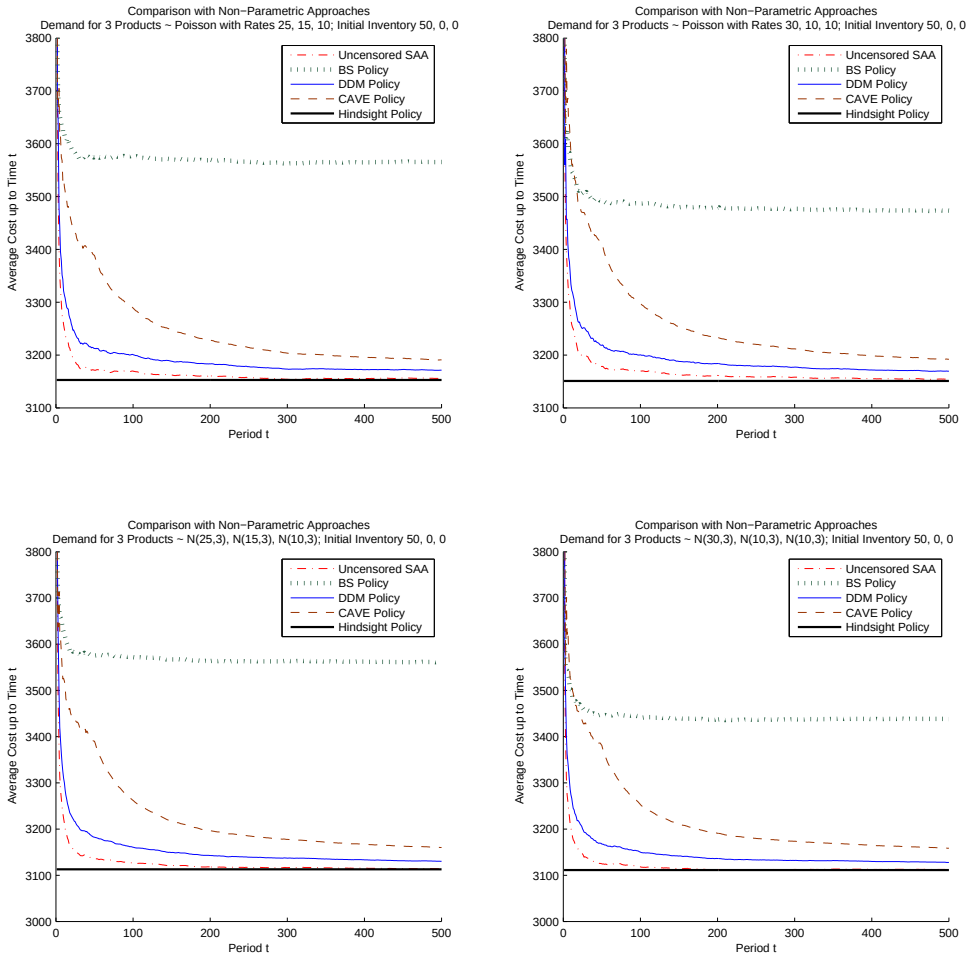

Figure EC.2 Comparison with nonparametric approaches.

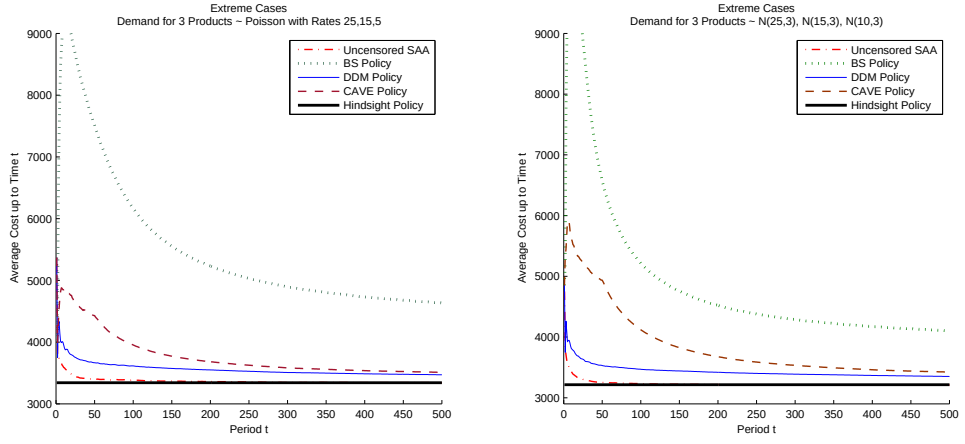

Figure EC.3 Extreme cases with uneven lost-sales penalty costs $\mathbf{p} = [400, 80, 80]$.