MICHIGAN ROSS

Working Paper

Certainty Equivalent Planning for Multi-Product Batch Differentiation: Analysis and Bounds

Hyun-Soo Ahn
Stephen M. Ross School of Business
University of Michigan

Stefanus Jasin
Stephen M. Ross School of Business
University of Michigan

Philip Kaminsky
Department of Industrial Engineering and Operations Research (IEOR)
University of California, Berkeley

Yang Wang
Department of Industrial Engineering and Operations Research (IEOR)
University of California, Berkeley

Ross School of Business Working Paper Series
Working Paper No. 1296
December 2015

Certainty Equivalent Planning for Multi-Product Batch Differentiation: Analysis and Bounds

Hyun-Soo Ahn · Stefanus Jasin
hsahn, sjasin@umich.edu
Stephen M. Ross School of Business, University of Michigan, Ann Arbor, MI 48109

Philip Kaminsky · Yang Wang
kaminsky, yangwang0803@berkeley.edu
Department of IEOR, University of California at Berkeley, Berkeley, CA 94720

We consider a multi-period planning problem faced by a firm that must coordinate the production and allocations of batches to end products for multiple markets. Motivated by a problem faced by a biopharmaceutical firm, we model this as a discrete-time inventory planning problem where in each period the firm must decide how many batches to produce and how to differentiate batches to meet demands for different end products. This is a challenging problem to solve optimally, so we derive a theoretical bound on the performance of a *Certainty Equivalent* (CE) control for this model, in which all random variables are replaced by their expected values and the corresponding deterministic optimization problem is solved. This is a variant of an approach that is widely used in practice. We show that while a CE control can perform very poorly in certain instances, a simple re-optimization of the CE control in each period can substantially improve both the theoretical and computational performance of the heuristic, and we bound the performance of this re-optimization. To address the limitations of CE control and provide guidance for heuristic design, we also derive performance bounds for two additional heuristic controls— (1) *Re-optimized Stochastic Programming* (RSP), which utilizes full demand distribution but limits the adaptive nature of decision dynamics, and (2) *Multi-Point Approximation* (MPA), which uses limited demand information to model uncertainty but fully capture the adaptive nature of decision dynamics. We show that although RSP in general outperforms the re-optimized CE control, the improvement is limited. On the other hand, with a carefully chosen demand approximation in each period, MPA can significantly outperform RSP. This suggests that, in our setting, explicitly capturing decision dynamics adds more value than simply capturing full demand information.

Key words: production planning; batch allocation; discrete time models; certainty equivalent control, re-optimization

1. Introduction

As product lines expands to target smaller, more segmented markets, the need to effectively determine stocking levels for individual end-products is becoming increasingly important. In this paper, we consider the production and inventory (SKUs) planning problem faced by a firm that must differentiate batches of a single intermediate product into market-specific end products to meet demand over a finite horizon. Since demand for a particular end product is small relative to intermediate batch size and batches must be entirely differentiated at one time, the determination of how many batches to differentiate, and how many of each end product to make, is critical for ensuring cost-effective operation. This model is motivated by an inventory planning problem in a leading "orphan drug" biopharmaceutical firm with which we have worked. This firm manufactures and distributes a biopharmaceutical treatment for a rare genetic disorder that leads to severe, potentially life threatening, symptoms in people who have the disorder. This is a so-called "orphan drug" – there is a relatively small set of patients who can benefit from the drug, and thus limited incentive for firms to invest in drug development. In many countries (including the United States), the government gives certain tax and patent incentives to encourage firms to develop orphan drugs. Due to the nature of the manufacturing process, biopharmaceutical are typically manufactured in large batches (this firm calls them "bulks"). In the case of this particular product, bulks can be stored in this intermediate form for up to a year. In order to distribute the product to the market, however, it must be "filled and labeled" for each end market—the intermediate form of the product must be packaged into different-sized vials, properly labeled, and packed into cartons for each country. Due to the way the product was approved in each market and because the contract manufacturing firms that complete this process do not have sufficient storage or tracking capabilities in place, an entire "bulk" must be filled and labeled at once. (In fact, we recommended that the firm relax this practice, and we were told that the firm would not seek suppliers with additional capabilities, as well as the necessary re-licensing, for this particular product, but that the firm plans to pursue this route for future products.) The firm sells this product in 19 different international markets, each of which has different labeling requirements; most importantly, once a vial is labeled for a particular market, it cannot be relabeled for another market. The nature of demand is such that no market consumes a “bulk’s worth” of products in any period, and some markets consume considerably less. At each decision epoch, the firm must decide how many bulks to order and how many vials to fill and label for each market.

Related literature. Our problem is essentially a centralized batch ordering and differentiation problem faced by a firm that must coordinate production of batches and allocations of batches to end products for different markets. Alternatively, it can be viewed as a multi-retailer system where, in each period, total orders from all retailers must be a multiple of an exogenously determined batch size. Various models related to batch production/ordering have been studied. Veinott (1965) studies the problem of batch ordering for a single retailer. For the backorder case, he shows that a (R, nQ) policy, in which the inventory level is raised to at least R by ordering the smallest multiple of Q whenever it falls below R, is optimal for both the finite and the infinite horizon problem. Axsäter (1993, 1995, 1998, 2000), Forsberg (1997), and Cachon (2001) study the batch ordering problem for two-echelon distribution systems with one warehouse and N retailers. The papers by Axsäter assume that both retailers and warehouse use a continuous review (Q, R) policy. Axsäter (1993) assumes N identical retailers and proposes both exact and approximate methods to evaluate and optimize the performance of the system. Axsäter (1998) extends these results to 2 non-identical retailers, Forsberg (1997) extends these results to N non-identical retailers with Poisson demand, and Axsäter (1995, 2000) considers compound Poisson demand. In all of these continuous time models, there is no allocation issue—all demands are filled using a first-come-first serve approach. Cachon (2001) relaxes the Poisson demand assumption and models the periodic review version of the system, adopting a (R, nQ) policy with random allocation at the warehouse. He characterizes the optimal reorder point at each retailer given a reorder point at the warehouse, and then searches for the optimal warehouse reorder point. Chen (2000) considers the batch production problem in a multi-echelon serial system (N stages) under periodic review and finds that a modified version of Veinott’s (1965) (R, nQ) policy is optimal, and Chao (2009) extends this model to allow fixed replenishment intervals (e.g., stage 1 can order every day, stage 2 can order every week, etc). The authors show that the system achieves the minimum expected average cost when the ordering times for all of the stages are synchronized. In all of this work, however, the batch ordering restriction is imposed on each individual retailer or stage. In contrast, in our problem, the batch ordering restriction is placed on the total orders from all retailers. The manufacturer utilizes this information to decide how many batches to produce.

Our problem is also related to the discrete time multi-retailer inventory model under limited resources, in which the allocation issue is explicitly studied: Given a scarce resource such as production capacity shared by multiple retailers, each of whom sells unique products, the decision maker must decide how to allocate production capacity in each period. DeCroix and Arreola-Risa (1998) characterize the optimal policy for homogeneous products and develop heuristics for the non-homogeneous case. Shaoxiang (2004) extends these results to two non-homogeneous products, and Janakiraman et al. (2009) further extends these results to more than two products and develop an asymptotically optimal policy. In all this work, there is a single capacity constraint on the resource; in contrast, the constraint in our setting also comes from the nature of batch ordering.

Our contribution. In contrast to much of the existing related literature, which primarily studies the structure of the optimal ordering/allocation policy, our primary objective in this paper is to explore the performance in our setting of a simple yet commonly used heuristic control, and analyze approaches for improving its performance. (For a special case, we also derive the structure of optimal allocation policy.) First, we provide a theoretical performance bound for *Certainty Equivalent* (CE) control relative to the optimal policy in our setting. In a CE approach, all random variables (i.e., random demands) are replaced by their expected values, and the resulting deterministic optimization is solved to determine the operating policy. In other words, the original stochastic dynamic problem is transformed into a deterministic optimization problem. Due to the challenges of estimating demand distributions in practice, as well as the challenges of solving the original stochastic dynamic problem, CE control and its variant have became popular approaches for solving industrial scale inventory problem (Treharne and Sox, 2002; Calmon, 2015). Despite its prevalence, however, we are not aware of a rigorous analysis of the theoretical performance of CE control in the inventory literature (see Sections 3 and 4 for more discussions). There is, of course, a deep literature focusing on deterministic inventory models, but this line of works tends to focus on solution approaches for these deterministic models and, for the most part, the quality of a deterministic model as an approximation for the related stochastic model is typically not rigorously addressed. In this paper, we show that, for our model, although CE control can perform very poorly if the planning horizon becomes long, periodic re-optimization of CE control improves this performance by dampening the impact of planning horizon on total costs. If, however, the size of demand variation is also relatively large, then re-optimization only has limited benefit and we need to apply more sophisticated heuristics. These results shed light on the appropriateness of the CE control approach in practice.

To address the limitation of the CE control, we additionally analyze two improvements to CE control: (1) *Re-optimized Stochastic Programming* (RSP) and (2) *Multi-Point Approximation* (MPA)—a full Dynamic Programming (DP) approach but with limited demand distribution information. Note that, for computational simplicity, in the CE approach we deliberately deemphasize two key elements of the original stochastic problem: *demand variability* (because we ignore the demand distribution) and *decision dynamics* (because we ignore the fact that future decisions should be contingent on current decisions, demand realizations, and system dynamics, and could potentially be captured in an adaptive way). Both RSP and MPA are intentionally chosen to highlight the potential improvement due to exploiting these elements. Unlike CE control, which only uses expected demand information, RSP uses complete demand distribution for calculating batch order and product allocation. However, it only partially captures decision dynamics via frequent re-optimizations. MPA, on the other hand, directly models full decision dynamics but only partially captures demand variability using a multi-point approximation instead of the complete demand distribution. The crucial and practically relevant question is this: When designing a heuristic control for an inventory problem, is it more important to capture demand variability or decision dynamics? We show that, for our model, although RSP improves on the performance of re-optimized CE control, the magnitude of this improvement is limited—RSP also performs poorly when the size of demand variation becomes large. Indeed, we show that the solution of RSP is identical to the solution of re-optimized CE control in some cases. This suggests that, in our setting, the benefit of including more granular demand information is already captured (at least, partially) by frequent re-optimizations. In contrast to RSP, with a carefully chosen demand approximation in each period, MPA exhibits stronger theoretical performance bound, even with only limited demand information. This highlights the importance in our model of explicitly modeling decision dynamics in order to get the most benefit—simply incorporating more granular demand information is not sufficient. Because solving large scale stochastic inventory problems to optimality is typically intractable in practice, designing computationally efficient heuristics with analytic performance bounds becomes important. In addition to providing guidance for constructing effective solutions for the specific model we are considering, the approach we take in this paper, comparing the two algorithms—RSP (capturing the fidelity of demand distribution but simplifying the decision dynamics) and MPA (capturing the decision dynamics but simplifying the demand distribution)—can be applied to other inventory problems to better understand the factors that lead to algorithm performance, which can ultimately used to improve algorithm design.

Organization of the paper. In Section 2, we formulate our model; in Sections 3 and 4, we introduce CE and analyze its performance; in Section 5, we discuss RSP and MPA; in Section 6, we present results of our computational experiments; in Section 7, we briefly discuss the performance of CE under a slightly modified modeling assumption; and finally, in Section 8, we conclude the paper.

2. Model

We discuss a discrete time model where a firm (a centralized decision maker) must satisfy demands in multiple markets (which we call retailers) through joint ordering and allocation decisions. In each period, the sum of allocated units across all end-product markets must equal the number of batches used in that period. Specifically, we consider a model with T discrete periods and m retailers, where the time periods are indexed by $t \in \{1, \dots, T\}$ and the retailers are indexed by $i \in \{1, \dots, m\}$. Demands across different periods are assumed to be independent and stationary. We assume that lead time is zero and unsatisfied demands are backordered. (See Remark 4 in Section 4.3 for a discussion of nonstationary demand and Section 7 for an extension to the case of lost sales. In general, our basic solution approach is straightforward to extend to deterministic lead times.) The following notations are used throughout the paper:

$$\begin{aligned}
D_{t,i} &= \text{demand faced by retailer } i \text{ in period } t \\
F_i(\cdot) &= \text{cumulative demand distribution for retailer } i \\
\mu_i &= \text{expected demand in a period faced by retailer } i \\
\Delta_{t,i} &= D_{t,i} - \mu_i \\
h_i &= \text{per unit holding cost for retailer } i \\
p_i &= \text{per unit penalty cost for retailer } i \\
B &= \text{batch size (i.e., the number of units in a batch)} \\
I^{\pi}_{t,i} &= \text{starting inventory for retailer } i \text{ at the beginning of period } t \text{ under policy } \pi \\
I_{1,i} &= \text{starting inventory for retailer } i \text{ at the beginning of period 1} \\
N^{\pi}_{t,i} &= \text{number of new units allocated to retailer } i \text{ in period } t \text{ under policy } \pi \\
Z^{\pi}_{t} &= \text{number of batches ordered in period } t \text{ under policy } \pi \\
C^{\pi} &= \text{total costs under policy } \pi
\end{aligned}$$

Note that $\Delta_{t,i} = D_{t,i} - \mu_i$ is the difference between the actual and the expected demand faced by retailer i in period t . Also, since total allocated units across all retailers must equal total units contained in the new batches, we must have: $\sum_{i=1}^{m} N^{\pi}_{t,i} = Z^{\pi}_{t} B$. For analytical tractability, although we require $Z^{\pi}_{t,i}$ to be a non-negative integer for all t and i, we allow $N^{\pi}_{t,i}$ to be a non-negative real number. Under the backorder assumption, the starting inventory level at retailer i at the beginning period $t+1$ under policy π is given by:

$$I^{\pi}_{t+1,i} = I^{\pi}_{t,i} + N^{\pi}_{t,i} - D_{t,i} = I_{1,i} + \sum_{s=1}^{t} N^{\pi}_{s,i} - \sum_{s=1}^{t} D_{s,i}. \tag{1}$$

2.1. The Stochastic Planning Problem

At the beginning of period t, upon observing the available inventories at all retailers, the firm first decides how many new batches to order. After ordering new batches, it must decide how many units of end product to allocate to each retailer (i.e., how many units to label and package for each market). Demands are then realized and inventories are consumed. Remaining units are held in inventory until the next period and unsatisfied demand is backordered.

Let Π denote the set of non-anticipating policies, i.e., the set of policies that determine how many new batches to order and how many units to allocate to each retailer in period t using only the accumulated information up to the beginning of period t. Let C^* denote the expected total costs under an optimal policy $\pi^* \in \Pi$. We can write C^* as follows:

$$C^* := \inf_{\pi \in \Pi} \sum_{t=1}^{T} \mathbf{E}\left[c\, Z_t^\pi\, B + \sum_{i=1}^{m} p_i (D_{t,i} - N_{t,i}^\pi - I_{t,i}^\pi)^+ + \sum_{i=1}^{m} h_i (N_{t,i}^\pi + I_{t,i}^\pi - D_{t,i})^+ \right] \tag{2}$$

$$\text{s.t.} \sum_{i=1}^{m} N_{t,i}^\pi = Z_t^\pi\, B,\ Z_t^\pi \in \mathbf{Z}^+,\ N_{t,i}^\pi \in \mathbf{R}^+ \qquad \forall t,\, i \tag{3}$$

$$I_{t+1,i}^\pi = I_{t,i}^\pi + N_{t,i}^\pi - D_{t,i} \qquad \forall t,\, i \tag{4}$$

$$I_{1,i}^\pi = I_{1,i} \qquad \forall i \tag{5}$$

where all the constraints must be satisfied almost surely (or with probability one). Let $I_t = (I_{t,i})$, $N_t = (N_{t,i})$, and $D_t = (D_{t,i})$ denote the vector of starting inventory levels, allocated units, and realized demands in period t, respectively. We can write the optimal control problem (2) using Bellman's equation as follows:

$$C_t(I_t) = \min_{Z_t, N_t \in \Omega(Z_t)} \{ cZ_tB + G(I_t + N_t) + \mathbf{E}\left[C_{t+1}(I_t + N_t - D_t)\right] \} \quad \text{for } t = 1, 2, ..., T, \text{ and}$$

$$C_{T+1}(I_{T+1}) = 0, \tag{6}$$

where $\Omega(Z) = \{N : N_i \in \mathbf{R}^+, \sum_{i=1}^m N_i = ZB\}$ and $G(y) = \mathbf{E}\left[\sum_i^m h_i(y - D)^+ + \sum_{i=1}^m p_i(D - y)^+\right]$.

In general, the joint ordering and allocation problem formulated in (6) is difficult to solve and the optimal policy is challenging to characterize. As might be expected given the batch production requirements, the optimal expected cost $C_t(I_t)$ is not convex in starting inventory levels, so a simple base-stock style policy is not likely to be optimal for this problem. It is also not difficult to find examples where the optimal decision as a function of inventory levels changes depending on the period for a given planning horizon, or depending on the planning horizon length for a given period. To calculate the optimal policy, it is therefore necessary to explicitly solve the entire dynamic programming (6), which is generally intractable due to the problem size. As an illustration, if demand is discrete and integral, full dynamic programming requires an exponential amount of space $O\left((I_{max} - I_{min})^m\right)$ to store the state information where I_{max} (I_{min}) is the maximum (minimum) possible inventory level. In a addition, since the problem not convex, to ensure global optimality, there are also an exponential number of decisions $O((Z_{max}B)^m)$ that need to be explored.

3. Certainty Equivalent Planning

In this section, we analyze the performance of the simple non-adaptive heuristic control we introduced above—*Certainty Equivalence* (CE)—in which all random demand variables are replaced by deterministic numbers and the resulting deterministic optimization problem is solved (Treharne and Sox, 2002). Although not always known by that name (e.g., it is sometimes called *Model Predictive Control*, see Ciocan and Farias, 2012; and indeed, it is sometimes naively employed by managers without any name at all), CE control is popular in practice because it addresses two complicating problems that arise when solving the original problem: (1) demand estimation is often challenging (for instance, the firm that motivates this project uses a one-point estimate of demand instead of the estimate of complete demand distribution, an approach which in our experience is not uncommon); (2) the optimal control problem, even if the distribution can be estimated, is difficult to solve (for a typical industrial-scale problem, even solving a deterministic version of the problem is already quite challenging). Naturally, these concerns have motivated many practitioners to use a heuristic control approach that can be implemented with little detailed demand information.

Given the fact that a CE-like approach is widely used in practice, an interesting set of questions arises. How much is actually lost if a heuristic control derived from a deterministic model such as CE is applied in a stochastic setting? (Common sense suggests that a deterministic model can be a poor approximation of a stochastic system.) Is there any setting in which a deterministic model is a good approximation of a stochastic system? If so, in what sense? (In a context beyond inventory problems, it is known that CE control can be optimal for some stochastic problems. One famous example is the so-called *Linear Quadratic Gaussian Control* problem (see Stengel, 1994). For most other known applications, CE control is typically suboptimal.)

Our results in this section show that CE control performs reasonably well when the size of demand variation is relatively small compared to its mean (i.e., there is a small coefficient of variation) and the planning horizon is short. However, as the problem size increases, the performance of CE control deteriorates at the rate of $T^{3/2}$ as the planning horizon gets longer. In this section, we characterize the performance of this deterministic heuristic.

To evaluate the performance of CE control, we define C^D as follows:

$$C^D \; := \min_{z,n} \sum_{t=1}^{T} \left[c\, z_t\, B + \sum_{i=1}^{m} p_i\, (\mu_i - x_{t,i} - n_{t,i})^+ + \sum_{i=1}^{m} h_i (x_{t,i} + n_{t,i} - \mu_i)^+ \right] \tag{7}$$

$$\text{s.t.} \; \sum_{i=1}^{m} n_{t,i} = z_t\, B, \; z_t \in \mathbf{Z}^+, \; n_{t,i} \in \mathbf{R}^+ \qquad \forall t,\, i \tag{8}$$

$$x_{t+1,i} = x_{t,i} + n_{t,i} - \mu_i \qquad \forall t,\, i \tag{9}$$

$$x_{1,i} = I_{1,i} \qquad \forall i \tag{10}$$

Note that (7) can be written as a mixed integer linear program (MILP). We first explore the relationship between C^D and C^*. In much of the CE literature (e.g., Jasin and Kumar, 2012; Ciocan and Farias, 2012), the optimal value of the deterministic problem serves as either a lower or upper bound for the optimal value of the original stochastic control problem. This allows the optimal value of the deterministic problem to be used as as a proxy for performance analysis of any heuristic strategy. Unfortunately, this is not the case here due to the integrality of z_t, i.e., $C^D \nleq C^*$. (A standard argument for proving either $C^D \leq C^*$ or $C^D \geq C^*$, depending on the context, is to apply Jensen's inequality and replace all random variables with their expected values. Since we still require the number of batches to be an integer, this strategy does not work. It is possible to further relax this assumption and allows z_t to be a real number, so we immediately get $C^D \leq C^*$. However, the resulting lower bound is too loose to be useful for performance benchmarking.)

To describe CE control, we let $z = z^D$ and $n = n^D$ denote an optimal solution of (7).

Certainty Equivalence - CE

1. At the beginning of period 1, solve C^D
2. For $t = 1, 2, ..., T$, do:
   - At the beginning of period t, order exactly z^D new batches
   - After the new batches arrive, allocate exactly $n_{t,i}^D$ units to retailer i

As we define it, CE control precludes any consideration of the starting inventory, or backorder levels, prior to making ordering and allocation decisions in each period because ordering and allocation decisions are directly dictated by z^D and n^D, regardless of the actual demand realizations. We define $\mathbf{E}[C^{CE}]$ to be the expected total costs associated with implementing CE control. The following results provide a bound for the regret introduced by implementing CE control:

THEOREM 1. *Let* $\sigma = \max_i \mathbf{E}[(D_{1,i} - \mu_i)^2]^{1/2}$. *Then,*

$$\mathbf{E}[C^{CE}] - C^* \leq 2\sigma(T+1)^{3/2} \left[\sum_{i=1}^{m} (p_i + h_i) \right].$$

Two comments are in order. First, due to the non-differentiability of $(\cdot)^+ = \max(\cdot, 0)$ in (7), the optimal solution z^D and n^D may *not* be unique. However, the bound in Theorem 1 holds regardless of the choice of optimal solution. Second, the performance of CE is proportional to the demand variability as measured by σ. If $\sigma = 0$, then CE control is optimal regardless of the planning horizon, T. When $\sigma > 0$, however, the bound in Theorem 1 depends not only on σ but also on $T^{3/2}$. The fact that we have $T^{3/2}$ in the bound, which is larger than T, should not be surprising (see Remark 1 for additional discussions). Moreover, this scaling factor is not an artifact of the proof. Our simulation results in Section 6 show that the relative regret of CE control quickly becomes worse as T increases. Indeed, this is the reason why CE control can perform very poorly for multi-period inventory problems even when demand variation is relatively small. This is in contrast to the performance of CE-type heuristics in other application areas such as revenue management and dynamic pricing, where the regret scales with $\sqrt{T}$ instead of $T^{3/2}$ (Gallego and van Ryzin, 1994; Jasin, 2014; Jasin and Kumar, 2013). The following result is a corollary of Theorem 1.

Corollary 1. *Suppose that demands are Poisson with $\mu_i = \mu^*$ for all i. Then, there exists a constant $M > 0$ independent of T and μ^* such that,*

$$\frac{\mathbf{E}[C^{CE}] - C^*}{C^*} \;\le\; M \left(\frac{T}{\mu^*} \right)^{1/2}.$$

The bound in Corollary 1 is proportional to $\sqrt{T/\mu^*}$. Thus, even for the case of Poisson demand, where the coefficient of variation goes to zero as the mean goes to infinity, T must be small relative to μ^* for CE control to be reasonably effective. While this may not be an issue for instances with very large μ^*, this result shows that the applicability of CE control is rather limited. To re-emphasize, in the context of our inventory problem, *CE control may perform poorly, even for instances with a small coefficient of variation, unless the planning horizon is also short.* This paints a rather bleak picture of the usefulness of deterministic approximation for multi-period stochastic inventory problems. The main culprit here is the manner in which randomness accumulates over time due to the per-period holding and penalty costs, which scale polynomially with T. This gives rise to an important question: Is it possible to construct an alternative heuristic control that retains the simplicity of CE control and yet is more effective than CE control, at least in the setting with a small coefficient of variation and typical industrial planning horizon (which is about 2 to 5 years, i.e., $T = 24$ to 60 if one period equals one month)? It turns out that simple re-optimization of CE control at the beginning of every period significantly reduces the dependency of relative regret on T. This makes CE control more practically appealing and also more amenable to problems with a longer planning horizon. We introduce and analyze this approach in the next section.

Remark 1 (On the Factor $T^{3/2}$ in Theorem 1). Suppose that $\mu_i = \mu^*$ for all i. Consider the most naive policy that does not order any batch until the end of the planning season. Aside from incurring ordering costs of order $T\mu^*$, this policy also incurs total average penalty costs of order

$T^2\mu^*$. Thus, CE shows an improvement in comparison to the most naive policy by reducing the dependency of regret on the length of planning season from T^2 to $T^{3/2}$.

4. Improving Certainty Equivalent Planning

We next discuss several modifications to the basic (static) CE control that can improve its performance. For analytical tractability, we focus our attention only on the class of policies that uses z^D as the ordering policy, but optimizes the allocation policy. (One might expect additional improvement if the ordering policy is also further optimized—this makes the analysis extremely challenging, but we computationally test this approach in Section 6.) For a given ordering policy, except for some special cases, the optimal allocation policy is difficult to determine when we have multiple retailers. In the case of homogeneous retailers with identical cost structures and i.i.d demands, however, we are able to completely characterize optimal allocation policy. We then prove that, as long as the magnitude of demand variation is not large compared to its mean, simple re-optimizations of CE control suffices to guarantee a significant improvement over the static CE control. This result gives credence to the practice of re-optimization that is often employed in industry. Moreover, our analysis of re-optimization also suggests a natural inventory-balancing policy that can be implemented in real-time. We discuss this at the end of this section.

Recall that z^D is found at the beginning of the horizon by solving for C^D. Let $\tilde{\Pi}$ denote the set of non-anticipating policies that use z^D as the ordering policy. Also, let J^π denote the total costs under policy $\pi \in \tilde{\Pi}$ and J^* denote the total costs under the optimal allocation policy for a given ordering policy z^D. We can write:

$$J^* := \inf_{\pi \in \tilde{\Pi}} \sum_{t=1}^{T} \mathbf{E}\left[c\, z_t^D\, B + \sum_{i=1}^{m} p_i(D_{t,i} - N_{t,i}^\pi - I_{t,i}^\pi)^+ + \sum_{i=1}^{m} h_i(N_{t,i}^\pi + I_{t,i}^\pi - D_{t,i})^+\right] \tag{11}$$

$$\text{s.t.} \sum_{i=1}^{m} N_{t,i}^\pi = z_t^D\, B,\ N_{t,i}^\pi \in \mathbf{R}^+ \qquad \forall t,\, i \tag{12}$$

$$I_{t+1,i}^\pi = I_{t,i}^\pi + N_{t,i}^\pi - D_{t,i} \qquad \forall t,\, i \tag{13}$$

$$I_{t,i}^\pi = I_{1,i} \qquad \forall i \tag{14}$$

Similar to C^* in Section 2.2, we can write J^* using Bellman's equation as follows:

$$J_t(I_t) = \min_{N_t \in \Omega(z_t^D)} \left\{ c\, z_t^D B + G(I_t + N_t) + \mathbf{E}\left[J_{t+1}(I_t + N_t - D_t) \right] \right\} \text{ for } t = 1, \ldots, T, \text{ and} \tag{15}$$

where $J_{T+1}(I_{T+1}) = 0$ $\Omega(\cdot)$ is as defined in Section 2.2.

4.1. Optimal Allocation Policy

In general, the optimal allocation policy that achieves J^* is challenging to find; doing so requires solving a full dynamic programming backward recursion, which is computationally intractable if either T, m, or the support for demand distribution is large. However, if all retailers are homogeneous with identical cost structure (i.e., $h_i = h^*$ and $p_i = p^*$ for all i) and i.i.d. demands (i.e., $D_{t,i} \sim D$ for all t and i), a simple and easy-to-implement allocation policy is optimal:

THEOREM 2. *Suppose that $I_1 = 0$, $p_i = p^*$ and $h_i = h^*$ for all i, and demand at all retailers in all periods is i.i.d. Then, the optimal allocation policy can be obtained as follows:*

(1) At the beginning of period t, sort all retailers from the smallest inventory level to the largest, $I_{t,1} \leq I_{t,2} \leq \ldots \leq I_{t,m}$. Let $\theta_t = \max\{k \in \mathbf{Z}^+ \mid \sum_{i=1}^{k} i(I_{t,i+1} - I_{t,i}) \leq z_t^D B\}$.

(2) Raise the inventory of retailers with $i \leq \theta_t$ to the same level, i.e.,

$$N_{t,i} + I_{t,i} = \frac{1}{\theta_t} \left(\sum_{i=1}^{\theta_t} I_{t,i} + z_t^D B \right) \qquad \forall i \leq \theta_t.$$

(3) Allocate nothing for all retailers with $i > \theta_t$ by setting $N_{t,i} = 0$, $\forall\, i > \theta_t$.

In other words, when all retailers are homogeneous, the optimal allocation policy is to balance the inventory levels in as many retailers as possible by allocating the new units starting with the lowest inventory retailers. If the retailers are not homogeneous (either in demand distribution or cost parameters), this allocation policy is no longer optimal. However, all is not lost. In Section 4.3, we will show that an inventory balancing policy similar to the one described in Theorem 2 is near-optimal in the non-homogeneous setting.

4.2. Re-optimized Certainty Equivalence

Motivated by our discusssions in Section 4.1, we now consider a simple heuristic control based on re-optimizing the deterministic counterpart of J^*. Re-optimizations have been shown to significantly improve the performance of CE-type heuristics in many application areas (Jasin, 2014; Jasin and Kumar, 2012; Reiman and Wang, 2008; and Ciocan and Farias, 2012). In the context of assemble-to-order system, Plambeck and Ward (2006) and Dogru et al. (2010) propose adaptive controls that utilize some forms of re-optimization. However, their results do not carry over to our setting for at least two reasons: (1) In the assembly-to-order setting considered in these papers, the firm first observes demand before making a decision while in our setting, the firm first makes a decision before observing demand; (2) in this assemble-to-order setting, the firm can make continuous adjustments, while in our setting, the firm is limited to making adjustments at the beginning of each period. In a standard inventory control setting, Secomandi (2008) analyzes the impact of re-optimization on performance. He shows that re-optimization does not always improve the original solution and provides sufficient conditions for re-optimization to guarantee a better result; however, he does not provide a theoretical performance bound on his approach. The lack of existing results in the literature is quite surprising given the practicality and prevalence of re-optimization-based heuristics in industry. In fact, most companies with which we have interacted employ a form of rolling horizon approach that periodically re-optimizes their planning models. Our results in this subsection contribute to the literature by characterizing the benefit of re-optimizations on model performance.

Define $J_t^D(I_t)$ as follows:

$$J_t^D(I_t) := \min_n \sum_{s=t}^{T} \left[c\, z_s^D\, B + \sum_{i=1}^{m} p_i\, (\mu_i - x_{s,i} - n_{s,i})^+ + \sum_{i=1}^{m} h_i (x_{s,i} + n_{s,i} - \mu_i)^+ \right] \tag{16}$$

$$\text{s.t.} \sum_{i=1}^{m} n_{s,i} = z_s^D\, B,\ n_{s,i} \in \mathbf{R}^+ \qquad \forall s,\, i \tag{17}$$

$$x_{s+1,i} = x_{s,i} + n_{s,i} - \mu_i \qquad \forall s,\, i \tag{18}$$

$$x_{t,i} = I_{t,i} \qquad \forall i \tag{19}$$

Let $n^{*t} = (n^{*t}_{s,i})_{s\geq t,\, i\geq 1}$ denote an optimal solution of $J^D_t(I_t)$. Note that n^{*t} is a function of I_t. However, for notational brevity, we will suppress its dependency on I_t. The complete description of the re-optimized CE (RCE) control is given below.

Re-optimized Certainty Equivalent - RCE

1. At the beginning of period 1, solve C^D
2. For $t = 1, 2, ..., T$, do:
   - At the beginning of period t, order exactly z^D new batches
   - Solve $J^D_t(I_t)$ and allocate exactly $n^{*t}_{t,i}$ units to retailer i
   - Update $I_{t+1} = I_t + n^{*t}_t - D_t$

In contrast to static CE control, which is implemented independent of demand realizations (as the policy allocates inventory according to n^D throughout the planning horizon), RCE incorporates realized demands and updated inventory/backorder level by re-optimizing the deterministic allocation problem at the beginning of every period. We now examine whether re-optimizations of CE control is sufficient to significantly improve the performance of static CE control.

The impact of re-optimization for a problem with general holding and penalty costs is difficult to analyze, primarily due to the non-differentiability of the function $(\cdot)^+ = \max(\cdot, 0)$. Moreover, the optimal solution of $J^D_t(I_t)$ may not be unique. This makes the task of analyzing the evolution of the re-optimized solution analytically intractable (see also Remark 2). However, we show that it is possible to theoretically characterize the benefit of re-optimization under a particular sequence of optimal solutions $n^{*1}_1, n^{*2}_2, ..., n^{*T}_T$ when either these solutions satisfy a certain condition (see Theorem 3) or all retailers are homogeneous with identical cost structure and i.i.d demands (see Theorem 4). In Section 4.3, we will argue that RCE can in fact be interpreted as a form of inventory-balancing policy. This observation is useful and can be used to motivate the development of optimal inventory-balancing policies in other inventory control problems. Let $D_{1:T}$ denote the vector of all realized demands in T periods. Define the hindsight total costs $J^H(D_{1:T})$ as follows:

$$J^H(D_{1:T}) \;:= \min_n \sum_{t=1}^{T} \left[c\, z^D_t\, B + \sum_{i=1}^{m} p_i\,(D_{t,i} - x_{t,i} - n_{t,i})^+ + \sum_{i=1}^{m} h_i(x_{t,i} + n_{t,i} - D_{t,i})^+ \right] \qquad (20)$$

$$\text{s.t.} \sum_{i=1}^{m} n_{t,i} = z_t^D B, \; n_{t,i} \in \mathbf{R}^+ \qquad \forall t, \, i \tag{21}$$

$$x_{t+1,i} = x_{t,i} + n_{t,i} - D_{t,i} \qquad \forall t, \, i \tag{22}$$

$$x_{1,i} = I_{1,i} \qquad \forall i \tag{23}$$

$J^H(D_{1:T})$ (or simply J^H) is thus the total costs if the firm has perfect knowledge of all future demands. Since we obviously cannot do better than the perfect hindsight policy, we immediately have $\mathbf{E}[J^H(D_{1:T})] \leq J^*$. The result below gives us a sense of the level of improvement that may result from periodic re-optimizations.

THEOREM 3. *Let* $\sigma = \max_i \mathbf{E}[(D_{1,i} - \mu_i)^2]^{1/2}$. *Suppose that* $I_1 = 0$ *and there exists an optimal solution* n^{*1} *and a constant* $\varphi > 0$ *such that* $n_{t,i}^{*1} \geq \varphi$ *and* $\left| \frac{1}{t} \sum_{s=1}^{t} n_{s,i}^{*1} - \mu_i \right| \geq \varphi$ *for all* t *and* i. *Let* J^{CE} *and* J^{RCE} *be the expected total costs under the CE and RCE controls, respectively, and define* $\mathcal{A} = \{\sum_{i=1}^{m} |\Delta_{t,i}| < \varphi/2, \; \forall t\}$. *Then,*

$$\mathbf{E}[(J^{CE} - J^H)\, \mathbf{1}\{\mathcal{A}\}] \;\leq\; 2\sigma(T+1)^{3/2} \left[\sum_{i=1}^{m} (p_i + h_i) \right].$$

Moreover, there exists a sequence of optimal solutions $n_2^{*2}, ..., n_T^{*T}$ *such that*

$$\mathbf{E}[(J^{RCE} - J^H)\, \mathbf{1}\{\mathcal{A}\}] \;\leq\; 2\sigma T \left[\sum_{i=1}^{m} (p_i + h_i) \right].$$

Note that, under CE control, we simply apply allocation policy n_t^D during period t as in Theorem 1. Thus, the fact that the bound for CE control is of order $T^{3/2}\sigma$ is not surprising. In contrast, the bound for RCE is only of order $T\sigma$, which means that re-optimizations improve the performance guarantee of CE control by reducing the effect that planning horizon has on regret from $T^{3/2}$ to T, at least in the set $\mathcal{A}$ where total demand variation during each period is relatively small compared to the number of allocated units. If σ is small compared to φ, then $\mathcal{A}$ happens with high probability. In such a case, we can properly say that periodic re-optimizations improve the performance of static CE control with a high probability. The conditions $n_{t,i}^{*1} \geq \varphi$ and $\left| \frac{1}{t} \sum_{s=1}^{t} n_{s,i}^{*1} - \mu_i \right| \geq \varphi$ in Theorem 3 simply mean that in a deterministic world, we always allocate a positive number of products to each retailer at every period, and that the starting inventory level at each retailer at the beginning of period $t > 1$ (i.e., $\sum_{s=1}^{t-1} n_{s,i}^{*1} - (t-1)\mu_i$) is always either strictly positive or strictly negative. These conditions are not as strong as they appear—they can be easily satisfied especially when the batch size is sufficiently large. Finally, note that the probability of event $\mathcal{A}$ is a function of T. Without further assumptions on the cost structure and demand distribution, it is not immediately clear from Theorem 3 alone how long the planning horizon can be before the benefit of re-optimizations start to diminish. Per our discussions in Section 3, T must be much smaller than μ^* for static CE control to perform sufficiently well. The next result shows that RCE clearly outperforms CE for a wide range of T values.

THEOREM 4. *Suppose that the following conditions hold: $I_1 = 0$, $p_i = p^*$ and $h_i = h^*$ for all i, demands are i.i.d with mean μ^* and standard deviation σ^*, and $z_t^D > 0$ for all t. Define $\hat{\mathcal{A}} := \{\sum_{i=1}^{m} |\Delta_{t,i}| \le B/(2m)\ \forall t\}$. There exists a sequence of optimal solutions $n_1^{*1}, n_2^{*2}, ..., n_T^{*T}$ such that*

$$\mathbf{E}[(J^{CE} - J^H)\mathbf{1}\{\hat{\mathcal{A}}\}] \le 2m(p^* + h^*)(T+1)^{3/2}\sigma^* \quad \textit{and}$$
$$\mathbf{E}[(J^{RCE} - J^H)\mathbf{1}\{\hat{\mathcal{A}}\}] \le 2m(p^* + h^*)T\sigma^*.$$

Moreover, if demands are Poisson, $B > \mu^*$, and $T = o(e^{\mu^*/(192m^4)})$, there exists a constant $M > 0$ independent of T and μ^* such that, for all large μ^*,

$$\frac{\mathbf{E}[J^{CE}] - J^*}{J^*} \le \frac{M\sqrt{T}}{\sqrt{\mu^*} + \sqrt{T}} \quad \textit{and} \quad \frac{\mathbf{E}[J^{RCE}] - J^*}{J^*} \le \frac{M}{\sqrt{\mu^*} + \sqrt{T}}.$$

The setting in Theorem 4 is *not* a special case of the setting in Theorem 3. (In Theorem 4, we do not require $|\frac{1}{t}\sum_{s=1}^{t} n_{s,i}^{*1} - \mu_i| \ge \varphi$ for some $\varphi > 0$ for all t and i; in fact, it is possible that $|\frac{1}{t}\sum_{s=1}^{t} n_{s,i}^{*1} - \mu_i| = 0$ for all t and i. Mathematically, we do not need this condition because we assume that all retailers are homogeneous and demands are i.i.d.) Thus, the result of Theorem 4 cannot be seen as a corollary of Theorem 3. The bound for CE control in Theorem 4 is similar to the bound for CE control in Corollary 1. Although the bound holds for $T = o(e^{\mu^*/(192m^4)})$ (i.e., T can be very large), T must be much smaller than μ^* to guarantee the effectiveness of CE. In contrast, the bound for RCE is almost independent of T—as long as $T = o(e^{\mu^*/(192m^4)})$, which can be much larger than μ^*, the relative regret of RCE decreases to 0 at a rate that is (roughly speaking) proportional to $1/\sqrt{\mu^*}$ as $\mu^* \to \infty$. (Unlike with CE control, in the case of RCE, the additional $\sqrt{T}$ in the bound also helps speed up the convergence. However, since a typical planning horizon extends about 2 to 5 years, if one period is one month (i.e., $T = 24$ to 60) and μ^* is, at least, on the order of hundreds or thousands, the greatest impact on performance comes from $\sqrt{\mu^*}$ instead of $\sqrt{T}$.) Practically, this means that re-optimizations not only yield a stronger performance guarantee, but also allow a much longer planning horizon. If T is larger than $o(e^{\mu^*/(192m^4)})$, then it becomes necessary to also re-optimize the ordering decision z_t in addition to the allocation decisions $\{n_{t,i}\}$ to maintain a good performance. This alternative re-optimization policy, which essentially re-optimizes the whole integer program instead of a linear program, can also be used to address the case where demand variation is relatively large compared to its mean (in contrast to Poisson demand in Theorem 4). We computationally test the performance of this approach in Section 6.

Remark 2 (On the Nonuniqueness of the Optimal Solution). Although we only prove the results in Theorems 3 and 4 for a particular choice of optimal solution, we conjecture that the non-uniqueness of the optimal solution is not detrimental to the performance of this approach (as in the context of revenue management; see Jasin and Kumar, 2013, for results). Indeed, if we use a differentiable convex cost functions for holding and penality costs that incur in each period instead of the linear holding and penalty cost functions, the resulting deterministic problem is differentiable and its optimal solution is unique. In such a setting, it can be shown that the bounds in Theorems 3 and 4 still hold. This suggests that the bounds in Theorems 3 and 4 are not simply an artifact of a particular choice of optimal solution.

Remark 3 (On the Use of Poisson Demand in Theorem 4). Instead of using Poisson demand in Theorem 4, it is also possible to use Normal demand to better highlight the impact of B, σ, and μ on performance. Suppose that demands are all Normal with mean μ^* and standard deviation σ^*. If $\sigma^* = o(B)$ and $T = o(e^{B^2/(8m^4(\sigma^*)^2)})$, it can be shown using arguments similar to those used in the proof of Theorem 4 that there exists a constant $M > 0$ independent of T, μ^*, σ^*, and B such that $\frac{\mathbf{E}[J^{ROPA} - J^*]}{J^*} \leq \frac{M\sigma^*}{\mu^* + \sigma^*\sqrt{T}}$. Thus, it is not necessary that $B > \mu^*$ as long as σ^* is small compared to B.

4.3. Inventory Balancing

There is a relationship between the optimal allocation policy derived in Theorem 2 and the proposed solution used in Theorems 3 and 4. The proof of Theorem 3 proceeds by constructing an optimal solution n^{*t}. To be precise, using dual arguments (i.e., the sufficiency of Karush-Kuhn-Tucker (KKT) conditions for optimality in linear programs), we show that if we use

$$n^{*s}_{s,i} \;=\; n^{*1}_{s,i} + \Delta_{s-1,i} - \frac{1}{m}\sum_{j=1}^{m} \Delta_{s-1,j}$$

for all $s \leq t-1$, then $n^{*t}_{t,i} = n^{*1}_{t,i} + \Delta_{t-1,i} - \frac{1}{m}\sum_{j=1}^{m} \Delta_{t-1,j}$ and $n^{*t}_{s,i} = n^{*1}_{s,i}$ for all $s > t$ is optimal for $J^D_t(I_t)$ on $\mathcal{A}$. Similarly, the proof of Theorem 4 proceeds by constructing an optimal solution n^{*t}. However, instead of using duality arguments, we use convexity arguments. (We cannot use the same duality arguments employed in Theorem 3 because the conditions required for Theorem 4 may not hold.) We show that if we use

$$n^{*s}_{s,i} \;=\; \frac{z^D_s B}{m} + \Delta_{s-1,i} - \frac{1}{m}\sum_{j=1}^{m} \Delta_{s-1,j}$$

for $s \leq t-1$, then $n^{*t}_{t,i} = \frac{z^D_t B}{m} + \Delta_{t-1,i} - \frac{1}{m}\sum_{j=1}^{m} \Delta_{t-1,j}$ and $n^{*t}_{s,i} = \frac{z^D_s B}{m}$ for all $s > t$ is optimal for $J^D_t(I_t)$ on $\hat{\mathcal{A}}$. (If B is large and $\sigma = o(B)$, it can be shown that $\hat{\mathcal{A}}$ happens with a high probability.) Thus, in both Theorems 3 and 4, the proposed optimal solution is of the form $n^{*t}_{t,i} = n^{*1}_{t,i} + \Delta_{t-1,i} - \frac{1}{m}\sum_{j=1}^{m} \Delta_{t-1,j}$. It is not difficult to check that this solution corresponds to a particular inventory-balancing policy. Let $I^D_t := \sum_{s=1}^{t-1} n^{*1}_s - (t-1)\mu$ denote the inventory level at the beginning of period t under the deterministic system with $I_1 = 0$. Also, let $\hat{I}^D_t := I^D_t + n^{*1}_t$. ($\hat{I}^D_t$ can be interpreted as the after-allocation target inventory level in period t.) Under RCE, we can write $I_{t,i} = \sum_{s=1}^{t-1} n^{*s}_{s,i} - \sum_{s=1}^{t-1} D_{s,i} = I^D_{t,i} - \frac{1}{m}\sum_{j=1}^{m}\sum_{s=1}^{t-2} \Delta_{s,j} - \Delta_{t-1,i}$. So, allocating $n^{*t}_{t,i}$ units to retailer i at period t immediately brings the inventory level to $I_{t,i} + n^{*t}_{t,i} = I^D_{t,i} + n^{*1}_{t,i} - \frac{1}{m}\sum_{j=1}^{m}\sum_{s=1}^{t-1} \Delta_{s,j} =$

$\hat{I}_{t,i}^{D} - \frac{1}{m}\sum_{j=1}^{m}\sum_{s=1}^{t-1}\Delta_{s,j}$. This means that our proposed solution balances the inventory level at all retailers by the same offset relative to the deterministic target level. In the case of Theorem 2, since all retailers are homogeneous, there are uniform target levels, so this is equivalent to bringing the inventory level at each retailer to the same value.

Remark 4 (On The Case of Nonstationary Demand). The arguments in the proof of Theorem 4 can be generalized to a setting where demands across different periods are independent but nonstationary. Let $\mu_{i,t} := \mathbf{E}[D_{t,i}]$. If $\min_{t,i}\left\{\frac{z_t^D B}{m} + \mu_{t,i} - \frac{1}{m}\sum_{j=1}^{m}\mu_{t,j}\right\} := \frac{\varphi}{m} > 0$, we can use:

$$n_{t,i}^{*t} = \frac{z_t^D B}{m} + \left(\mu_{t,i} - \frac{1}{m}\sum_{j=1}^{m}\mu_{t,j}\right) + \left(\Delta_{t-1,i} - \frac{1}{m}\sum_{j=1}^{m}\Delta_{t-1,j}\right)$$

and $n_{s,i}^{*t} = \frac{z_t^D B}{m} + \mu_{s,i} - \frac{1}{m}\sum_{j=1}^{m}\mu_{s,j}$ for all i and $s > t$ as an optimal solution of $J_t^D(I_t)$ on $\bar{\mathcal{A}} := \{\sum_{i=1}^{m}|\Delta_{t,i}| \le \varphi/(2m)\ \forall t\}$. As long as φ is at least of the same scale as μ (i.e., the variation in $\mu_{t,i}$ for each i is not too large), it can be shown that $\bar{\mathcal{A}}$ happens with a high probability. Hence, the results of Theorem 4 still hold.

5. Beyond Certainty Equivalence

Erring on the side of simplicity, CE control deliberately ignores two key elements of the original stochastic problem: the magnitude of *demand variation* (as captured in σ) and the *decision dynamics* (adaptive decisions contingent on realized demand, as captured in T). The combined impact of these two elements shows up in the bound—the regret of CE control scales linearly with σ and polynomially with T. In Section 4, we proved that simple periodic re-optimizations improve the performance of CE control by reducing the dependency of its regret on T from polynomial to linear. This suggests several follow-up questions: Is it possible to further reduce the dependency of regret on either σ or T from linear to sublinear? Is it more important to incorporate more demand information or decision dynamics? In this section, we analyze two improvements on CE control: (1) *Re-optimized Stochastic Programming* (RSP) and (2) *Multi-Point Approximation* (MPA). RSP uses knowledge of the full demand distribution but only partially deals with decision dynamics via re-optimizations; in contrast, MPA directly models full decision dynamics (i.e., it solves a complete dynamic programming), but only partially captures demand variation by generalizing the one-point approximation used in static CE control and RCE to a multi-point approximation of demand. We show that although RSP improves the performance of RCE, in general, its regret still scales linearly with both σ and T. This result has an important implication: to significantly reduce the dependency of regret on σ and T from linear to sublinear in our setting, it appears to be necessary to explicitly model decision dynamics.

5.1. Re-optimized Stochastic Programming

We proceed in two stages as follows: In stage 1, we solve C^D to calculate the number of new batches to order at the beginning of each period; in stage 2, instead of re-optimizing $J_t^D(I_t)$ as in the case of RCE, we re-optimize $J_t^S(I_t)$ defined below:

$$J_t^S(I_t) := \min_n \sum_{\xi=t}^{T} \left\{ cz_t^D B + \mathbf{E}\left[\sum_{i=1}^{m} p_i \left(\sum_{s=t}^{\xi} D_{s,i} - \sum_{s=t}^{\xi} n_{s,i} - I_{t,i} \right)^+ + \sum_{i=1}^{m} h_i \left(I_{t,i} + \sum_{s=t}^{\xi} n_{s,i} - \sum_{s=t}^{\xi} D_{s,i} \right)^+ \right] \right\} \tag{24}$$

$$\text{s.t.} \sum_{i=1}^{m} n_{s,i} = z_s^D B,\ n_{s,i} \in \mathbf{R}^+ \qquad \forall s,\, i \tag{25}$$

Observe that $J_t^S(\cdot)$ is a stochastic program. If demand is continuous and $F_i(\cdot)$ is differentiable and strictly positive on $(0,\infty]$ for all i, it is not difficult to show that the objective function in (24) is twice differentiable and *strongly convex* on $(0,\infty]^m$. Thus, an interior optimal solution of $J_t^S(I_t)$ is also a unique optimal solution of $J_t^S(I_t)$. Let $n^{St} = (n_{s,i}^{St})_{s\geq t,\, i\geq 1}$ denote the optimal solution of $J_t^S(I_t)$. (We suppress the notational dependency of n^{St} on I_t.) The complete description of RSP is given below:

Re-optimized Stochastic Programming - RSP

---

1. At the beginning of period 1, solve C^D
2. For $t = 1, 2, ..., T$, do:
   - At the beginning of period t, order exactly z_t^D new batches
   - Solve $J_t^S(I_t)$ and allocate exactly $n_{t,i}^{St}$ units to retailer i
   - Update $I_{t+1} = I_t + n_t^{St} - D_t$

Although our numerical results in Section 6 show that RSP consistently performs better than RCE (by up to 10%), it is not easy to analytically characterize this improvement. Interestingly, it is possible to show that the computed allocation under RSP is sometimes the same as the computed allocation under RCE (see Lemma 1 below). This suggests that the benefit of including full demand distribution is already captured (at least, partially) by simple re-optimizations of CE control.

LEMMA 1. *Suppose that $I_1 = 0$, $p_i = p^*$ and $h_i = h^*$ for all i, demands are i.i.d and their common cdf is differentiable and strictly positive on $(0,\infty]$, and $z_t^D > 0$ for all t. Let $\hat{\mathcal{A}} := \{\sum_{i=1}^m |\Delta_{t,i}| \leq B/(2m)\ \forall t\}$. Then, on $\hat{\mathcal{A}}$, the optimal allocation under RSP at period t is given by:*

$$n_{t,i}^{St} \ = \ \frac{z_t^D B}{m} + \Delta_{t-1,i} - \frac{1}{m}\sum_{j=1}^{m} \Delta_{t-1,j}.$$

Recall from Section 4.3 that $n_{t,i}^{St} = \frac{z_t^D B}{m} + \Delta_{t-1,i} - \frac{1}{m}\sum_{j=1}^m \Delta_{t-1,j}$ is the constructed optimal allocation under RCE for period t. Thus, despite the fact that RSP uses the full demand distribution, the regret associated with RSP in general is still $O(\sigma T)$ (because the bounds in Theorem 4 also hold for RSP). We conclude that, at best, RSP only provides limited improvement over RCE.

5.2. Multi-Point Approximation

In the previous subsection, we saw that incorporating full demand distribution alone is not sufficient to improve significantly on RCE. We now consider the impact of explicitly modeling decision dynamics. The essence of MPA is the use of multi-point demand approximations, to capture some demand variation, along with full dynamic programming to fully exploit decision dynamics. There are potentially many ways of doing this; here, we will only discuss one such approach. For each i, let S_i denote the support of $D_{t,i}$ and $\hat{D}_{t,i}$ denote the approximation of $D_{t,i}$. We consider an approximation of the following form: There exists a partition $\{\Omega_{t,i}^k\}$ (i.e., $\cup_k\ \Omega_{t,i}^k = S_i$) and a mapping $m_{t,i} : \{\Omega_{t,i}^k\} \to \{v_{t,i}^k\}$ such that $D_{t,i} \in \Omega_{t,i}^k$ is approximated (or represented) by $\hat{D}_{t,i} = v_{t,i}^k$. Note that, by construction, $P(\hat{D}_{t,i} = v_{t,i}^k) = P(D_{t,i} \in \Omega_{t,i}^k)$. For example, if $v_{t,i}^k = \mu_i = \mathbf{E}[D_{t,i}]$ for all k, t, and i, then we have the CE approximation.

Now, consider the following optimization problem:

$$\hat{C}^* := \inf_{\pi\in\Pi} \sum_{t=1}^{T} \mathbf{E}\left[c\, Z_t^\pi\, B + \sum_{i=1}^{m} p_i\ (\hat{D}_{t,i} - N_{t,i}^\pi - \hat{I}_{t,i}^\pi)^+ + \sum_{i=1}^{m} h_i\ (\hat{I}_{t,i}^\pi + N_{t,i}^\pi - \hat{D}_{t,i})^+\right] \tag{26}$$

$$\text{s.t. } \sum_{i=1}^{m} N_{t,i}^\pi = Z_t^\pi\, B,\ Z_t^\pi \in \mathbf{Z}^+,\ N_{t,i}^\pi \in \mathbf{R}^+ \qquad \forall t,\, i \tag{27}$$

$$I_{t+1,i}^\pi = I_{t,i}^\pi + N_{t,i}^\pi - \hat{D}_{t,i} \qquad \forall t,\, i \tag{28}$$

$$I_{1,i}^\pi = I_{1,i} \qquad \forall i \tag{29}$$

where the expectation is taken with respect to the induced probability distribution for $\{\hat{D}_{t,i}\}$. Let $\hat{\pi}^* = (\hat{\pi}_1^*, \cdots, \hat{\pi}_T^*)$ denote the optimal policy of $\hat{C}^*$. Since $\hat{\pi}^*$ is a policy defined in a "virtual" world where demands are realized according to $\hat{D}$ instead of D, it is not immediately clear how to translate $\hat{\pi}^*$ into a policy $\pi^R = (\pi_1^R, \cdots, \pi_T^R)$ to be implemented in the "real" world where demands are realized according to D. (The superscript "R" stands for "real".) Here, we will focus on the following policy translation scheme:

$$\pi_1^R = \hat{\pi}_1^* \quad \text{and} \quad \pi_t^R(D_1, ..., D_{t-1}) = \hat{\pi}_t^*(I_t^{\hat{\pi}^*})\ \ \forall\, t > 1$$

where $I_{t+1,i}^{\hat{\pi}^*} = I_{t,i}^{\hat{\pi}^*} + N_{t,i}^{\hat{\pi}^*} - \hat{D}_{t,i}$ (i.e., $I_t^{\hat{\pi}^*}$ is the virtual inventory level at the beginning of period t under policy $\hat{\pi}^*$ and demand realizations $\hat{D}_1, \ldots, \hat{D}_{t-1}$). Under policy π, at the beginning of period t, we first calculate the virtual inventory level at each retailer; next, we order exactly $Z_t^{\hat{\pi}^*}$ new batches and allocate exactly $N_{t,i}^{\hat{\pi}^*}$ units to retailer i. So, we respond as if demands are generated according to $\hat{D}$ instead of D. We state our result below.

Theorem 5. *Let $\theta_{t,i} := \mathbf{E}[(D_{t,i} - \hat{D}_{t,i})^2]^{1/2}$. Then,*

$$\mathbf{E}\left[C^{\pi^R}\right] - C^* \ \le\ 2\sum_{i=1}^{m}(p_i + h_i)\left[\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\theta_{s,i}^2\right)^{1/2}\right].$$

Theorem 5 is the generalization of Theorem 1. (If $\hat{D}_{t,i} = \mu_i$, then $\theta_{t,i} \le \sigma$. So, we completely recover the bound in Theorem 1.) It highlights the value of information in a multi-period inventory control problem; in particular, it shows that it is most beneficial to use a more refined demand approximation during earlier instead of later periods. To illustrate this, suppose that $D_{t,i}$ is uniformly distributed on $[L_i, U_i]$. If we use a $(T-t+1)^{1/2+\alpha}-$demand approximation for $D_{t,i}$ for some $\alpha > 0$ (i.e., by using $(T-t+1)^{1/2+\alpha}$ points in $[L_i, U_i]$), then $\mathbf{E}[C^{\pi^R}] - C^* = O(\sigma T^{1-\alpha})$. Note that, as α becomes large, the regret decreases to 0.

Remark 5 (Computational Complexity of MPA). Despite the promising bound in Theorem 5, MPA solves a full dynamic programming problem. Thus, it is computationally much more expensive than either RCE or RSP. One potential way to mitigate this computational burden is to use a form of *rollout* algorithm with *limited lookahead* (Bertsekas, 2013; Goodson et al., 2015). The analysis of a rollout algorithm for an undiscounted finite-horizon stochastic inventory problem is an open research problem. As it is possibly a very challenging task, we leave this for future research pursuit—our purpose in this paper is simply to highlight the potential benefit of including more decision dynamics in designing a heuristic control.

6. Computational Experiments

We computationally test the performance of CE, RCE, and RSP. In addition, we also consider jointly re-optimizing both the ordering and allocation decisions instead of the allocation decision alone under RCE and RSP; this essentially amounts to re-optimizing the entire integer program. We call the resulting heuristics RCE-IP and RSP-IP, respectively. We use an industrial-size example of 8 non-homogeneous retailers with $\mu = [5000, 3000, 2000, 1000, 500, 300, 200, 100]$, $h = [2, 2, 2, 2, 3, 3, 4, 4]$, $p = [8, 5, 9, 6, 10, 5, 11, 5]$, and $B = 10,000$. We run four different experiments: In the first experiment demands are Poisson and, in the last three experiments, demands are Normal with standard deviations equal to 5%, 15%, and 25% of their mean, respectively. The first two experiments represent the case of "small" demand variation and the last two experiments represent the case of "large" demand variation. For each of four experiments, we run 40 Monte Carlo simulations and average the results. All experiments are run using MATLAB R2010b with Intel Core i7-5820K CPU. We report the percentage regret for each heuristic control relative to the hindsight policy (see (20)) in Tables 1 and 2.

Table 1 Percentage regret

	Poisson			5% Normal		
T	CE	RCE	RCE-IP	CE	RCE	RCE-IP
1	4.81	4.81	4.81	10.01	10.01	10.01
2	4.97	4.10	3.66	10.61	8.83	7.78
3	5.22	3.56	3.56	10.43	7.22	7.22
4	6.47	3.99	3.99	12.40	7.69	7.69
5	7.72	4.33	4.33	14.90	8.28	8.28
6	8.00	4.18	4.18	16.60	8.53	8.23
7	8.05	4.18	3.83	18.25	9.69	7.74
8	8.50	4.26	3.95	18.31	8.86	7.55
9	9.08	4.33	4.04	19.47	9.09	7.87
10	9.88	4.47	4.20	21.12	9.22	8.11
11	10.06	4.30	4.06	22.40	9.78	8.26
12	9.76	4.19	3.92	21.84	9.38	7.83
13	10.12	4.26	4.00	22.44	9.23	7.93
14	10.58	4.33	4.08	23.43	9.29	8.06
15	11.20	4.39	4.16	24.77	9.46	8.16
16	11.43	4.31	4.10	27.32	11.13	8.14
17	11.23	4.20	3.99	25.88	9.78	7.83
18	11.64	4.26	4.06	26.40	9.63	7.83
19	12.15	4.34	4.14	27.05	9.66	7.92
20	12.66	4.34	4.15	28.00	9.86	7.93

Table 2 Percentage regret

	15% Normal					25% Normal				
T	CE	RCE	RCE-IP	RSP	RSP-IP	CE	RCE	RCE-IP	RSP	RSP-IP
1	27.23	27.23	27.23	23.96	23.96	35.12	35.12	35.12	32.15	32.15
2	34.70	30.53	26.52	22.39	20.09	49.00	45.80	41.37	31.79	31.51
3	31.08	23.81	23.75	13.46	13.96	47.71	39.55	39.14	25.14	26.52
4	33.32	23.07	23.15	13.21	13.76	53.49	40.81	40.05	26.19	27.33
5	37.79	23.52	23.38	14.56	14.72	60.98	41.84	40.13	28.74	27.64
6	45.25	26.33	23.14	18.42	14.96	71.97	47.49	40.60	37.10	28.61
7	53.39	33.25	24.18	23.77	15.11	80.96	53.94	39.88	42.51	27.42
8	52.34	30.42	23.41	20.67	14.19	80.90	51.62	39.33	40.08	27.41
9	52.09	29.07	23.61	19.20	14.17	85.03	52.07	39.34	40.73	26.43
10	53.95	29.13	23.78	19.40	14.23	91.18	54.47	39.03	43.96	25.92
11	59.27	30.53	23.72	21.96	14.53	102.42	60.68	39.58	51.54	26.32
12	61.18	33.58	23.27	23.27	14.22	102.28	60.38	39.28	48.37	25.96
13	60.39	31.43	23.00	20.99	13.93	106.09	61.21	39.99	48.95	26.50
14	61.45	30.50	22.85	20.22	13.61	111.77	63.16	40.53	51.30	26.72
15	64.49	30.97	23.29	20.78	13.92	119.83	67.23	40.96	56.46	27.00
16	68.99	32.54	23.08	23.07	13.86	128.52	71.93	40.78	62.85	27.31
17	70.27	33.95	23.13	23.56	13.84	127.47	71.31	41.35	60.30	27.48
18	69.76	32.48	22.73	22.03	13.46	129.98	71.73	40.79	60.56	26.91
19	71.71	32.27	22.95	21.84	13.29	135.03	72.31	40.50	61.85	26.68
20	74.69	32.99	23.03	23.07	13.50	142.75	75.57	40.84	66.20	26.68

Table 3 Solution time

	15% Normal				
	CE	RCE	RCE-IP	RSP	RSP-IP
Average solving time (sec)	0.0151	0.2113	2.7021	1.0937	3.3325

As suggested by Theorems 1 and 4, the regret associated with CE control gets worse rapidly as the length of planning horizon increases. The performance of RCE appears to be quite stable for the case of Poisson demand and 5% Normal. As the size of demand variation becomes large (e.g., 15% and 25% Normal), the benefit of re-optimizations slowly diminishes; this is also as expected. (Per Remark 3, as σ^* increases, $T = o(e^{B^2/(8m^4(\sigma^*)^2)})$ becomes smaller. So, it becomes necessary to also re-optimize the ordering decision.) In all four cases, RCE-IP further reduces the regret of RCE and also helps stabilize the performance, which highlights the benefit of jointly re-optimizing the ordering and allocation decision. Clearly, RSP-IP significantly reduces the regret of RCE-IP. Despite this, as discussed in Section 5.1, its relative regret is still of the same order as the size of demand variation. (Note that, although RSP improves the performance of RCE, it does not always perform better than RCE-IP.)

In Table 3, we report the solution time for each heuristic control for one of our experimental cases: Normal demand with standard deviation equals to 15% of the mean. The recorded time is the average solution time for one 20-period simulation. Although CE control requires solving an integer program, the average solution time is fast, even for industrial-scale problems. The other four approaches lead to a significant increase in solution time, although resulting times are by no means unreasonable for practical purposes. All of the algorithms listed above, however, are significantly more computationally efficient than solving the original problem to optimality. In fact, even solving a problem that is 100 times smaller directly using the original dynamic programming takes several hours. This highlights the relative solution speed of our heuristic algorithms.

7. The Case of Lost Sales

We now briefly consider the case of lost sales. We employ the same optimal formulation as in (2), with two exceptions: the constant p_i will now be interpreted as the lost sales penalty for retailer

i and the inventory level evolves according to the formula $I^{\pi}_{t+1,i} = (I^{\pi}_{t,i} + N^{\pi}_{t,i} - D_{t,i})^+$ instead of $I^{\pi}_{t+1,i} = I^{\pi}_{t,i} + N^{\pi}_{t,i} - D_{t,i}$. Alternatively, we can also write $I^{\pi}_{t+1,i} = \left[\max_{1\le\xi\le t}\sum_{s=\xi}^{t}(N^{\pi}_{s,i} - D_{s,i})\right]^+$. Define $\tilde{C}^D$ as follows:

$$\tilde{C}^D \;:= \min_{z,n}\; \sum_{t=1}^{T}\left[c\, z_t\, B + \sum_{i=1}^{m} p_i\,(\mu_i - x_{t,i} - n_{t,i})^+ + \sum_{i=1}^{m} h_i(x_{t,i} + n_{t,i} - \mu_i)^+\right] \tag{30}$$

$$\text{s.t.} \; \sum_{i=1}^{m} n_{t,i} = z_t\, B,\; z_t \in \mathbf{Z}^+,\; n_{t,i} \in \mathbf{R}^+ \qquad \forall t,\, i \tag{31}$$

$$x_{t+1,i} = (x_{t,i} + n_{t,i} - \mu_i)^+ \qquad \forall t,\, i \tag{32}$$

$$x_{1,i} = I_{1,i} \qquad \forall i \tag{33}$$

As with C^D, $\tilde{C}^D$ can be formulated as a mixed integer linear program (MILP). Let $z = \tilde{z}^D$ and $n = \tilde{n}^D$ denote an optimal solution of (30). Under CE control, we order exactly $\tilde{z}^D_t$ new batches at the beginning of period t and allocate exactly $\tilde{n}^D_{t,i}$ units to retailer i during period t. Let $\tilde{C}^*$ and $\mathbf{E}[\tilde{C}^{CE}]$ denote the expected total costs under the optimal policy and CE control, respectively.

Theorem 6. *Let* $\sigma = \max_i \mathbf{E}[(D_{1,i} - \mu_i)^2]^{1/2}$. *Then,*

$$\mathbf{E}[\tilde{C}^{CE}] - \tilde{C}^* \;\le\; 2\sigma \sum_{i=1}^{m}\left[(T+1)^{1/2}p_i + \frac{2(T+1)^{3/2}}{3}h_i\right].$$

In contrast to the bound in Theorem 1 where both h_i and p_i are multiplied by $T^{3/2}$, in Theorem 6, the term h_i is still multiplied by $T^{3/2}$ but the term p_i is only multiplied by $T^{1/2}$. This suggests that CE control should perform better in a lost sales system than in a backorder system, especially when p_i is much larger than h_i.

8. Closing Remarks

In practice, firms often solve planning problems by replacing random variables representing future demand with deterministic demand estimates and firms often use a rolling horizon approach to implement these solutions. In this paper, we considered an inventory planning problem from the biopharmaceutical industry involving batch production and allocation, and analyzed a variety of heuristic controls that solve the Certainty Equivalent version of this planning problem, both one time and in a rolling horizon setting. We characterized the performance of these heuristic controls, finding that the performance of this deterministic approximation decreases with coefficient of variation and horizon length, but that implementation of a rolling horizon re-optimization approach can significantly increase performance. We also explored heuristic controls that either use more demand information or more decision dynamics. As expected, these heuristic controls perform better. However, in our setting, we found that using additional demand information has limited value while using more decision dynamics potentially leads to greater improvements, at the expense of increased computational time. A natural way to mitigate this computational burden is to implement a rollout algorithm with limited lookahead (Bertsekas, 2013; Goodson et al., 2015). We are not aware of results in the literature characterizing the theoretical performance of this type of rollout algorithm for undiscounted inventory problems, and we intend to explore this challenging problem in the future.

References

Axsäter, S. 1993. Exact and approximate evaluation of batch-ordering policies for two-level inventory systems. Operations Research, 41(4): 777-785.

Axsäter, S. 1995. Approximate evaluation of batch-ordering policies for a one-warehouse, N non-identical retailer system under compound poisson demand. Naval Research Logistics, 42(5): 807-819.

Axsäter, S. 1998. Evaluation of installation stock based (R,Q)-Policies for two-level inventory systems with Poisson demand. Operations Research, 46(3-supplement-3): 135-S145.

Axsäter, S. 2000. Exact analysis of continuous review (R,Q) policies in two-echelon inventory systems with compound Poisson demand. Operations Research, 48(5): 686-696.

Bertsekas, Dimitri P. 2013. Rollout algorithms for discrete optimization: a survey. Handbook of Combinatorial Optimization. Springer New York, 2989-3013.

Cachon, G. P. 2001. Exact evaluation of batch-ordering inventory policies in two-echelon supply chains with periodic review. Operations Research, 49(1): 79-98.

Calmon, Andre du Pin. 2015. Reverse Logistics for Consumer Electronics: Forecasting Failures, Managing Inventory, and Matching Warranties. MIT Doctoral Dissertation.

Chao, X., Zhou, S.X. 2009. Optimal policy for a multiechelon inventory system with batch ordering and fixed replenishment intervals. Operations Research, 57(2): 377-390.

Chen, F. 2000. Optimal policies for multi-echelon inventory problems with batch ordering. Operations Research, 48(3): 376-389.

Ciocan, D.F, Vivek F. Farias. 2012. Model Predictive Control for Dynamic Resource Allocation. Mathematics of Operations Research, 37(3): 501-525.

DeCroix, G. A., Arreola-Risa, A. 1998. Optimal production and inventory policy for multiple products under resource constraints. Management Science, 44(7): 950-961.

Dogru, M. K., Reiman, M. I., Wang, Q. 2010. A stochastic programming based inventory policy for assemble-to-order systems with application to the W model. Operations Research, 58(4): 849-864.

Forsberg, R. 1997. Exact evaluation of (R, Q)-policies for two-level inventory systems with Poisson demand. European Journal of Operational Research, 96(1): 130 - 138.

Gallego, G., G. van Ryzin. 1997. A Multiproduct Dynamic Pricing Problem and Its Applications to Network Yield Management. Operations Research, 45(1): 24-41.

Goodson, Justin C., Barrett W. Thomas, Jeffrey W. Ohlmann. 2015. A Rollout Algorithm Framework for Heuristic Solutions to Finite-Horizon Stochastic Dynamic Programs. Working paper. Accessed via http://www.slu.edu/ goodson/papers/GoodsonRolloutFramework.pdf

Janakiraman, G., Nagarajan, M., Veeraraghavan, S. 2009. Simple policies for capacitated multi-product inventory systems. submitted to Management Science.

Jasin, S., Kumar, S. 2012. A Re-solving Heuristic with Bounded Revenue Loss for Network Revenue Management with Customer Choice. Mathematics of Operations Research, 37(2): 313-345.

Jasin, S., Kumar, S. 2013. Analysis of deterministic LP-based booking limit and bid price controls for revenue management. Operations Research, 61(6): 1312-1320.

Jasin, S. 2014. Reoptimization and self-adjusting price control for network revenue management. Operations Research, 62(5): 1168-1178.

Plambeck, E. L., Ward, A. R. 2006. Optimal control of a high-volume assemble-to-order system. Mathematics of Operations Research, 31(3): 453-477.

Reiman, M. I., Q. Wang. 2008. An Asymptotically Optimal Policy for a Quantity-Based Network Revenue Management Problem. Mathematics of Operations Research, 33(2): 257-282.

Shaoxiang, C. 2004. The optimality of hedging point policies for stochastic two-product flexible manufacturing systems. Operations Research, 52(2): 312-322.

Secomandi, N. 2008. An Analysis of the Control-Algorithm Re-solving Issue in Inventory and Revenue Management. Manufacturing & Service Operations Management 10(3): 468-483

Stengel, Robert F. 1994. Optimal Control and Estimation. Dover Publications. Reissue edition.

Treharne, James T., Charles R. Sox. 2012. Adaptive inventory control for nonstationary demand and partial information. Management Science, 48(5): 607-624.

Veinott, A. F. 1965. The Optimal inventory policy for batch ordering. Operations Research, 13(3): 424-432.

APPENDIX

Proof of Theorem 1. Define $W^{\pi}_{t+1,i} = W^{\pi}_{t,i} + N^{\pi}_{t,i} - \mu_i$, where $W^{\pi}_1 = I^{\pi}_1 = I_1$. Observe that we can write: $W^{\pi}_{t,i} = I_{1,i} + \sum_{s=1}^{t-1} N^{\pi}_{s,i} - \sum_{s=1}^{t-1} \mu_i$ and $I^{\pi}_{t,i} = I_{1,i} + \sum_{s=1}^{t-1} N^{\pi}_{s,i} - \sum_{s=1}^{t-1} D_{s,i}$. So, $I^{\pi}_{t,i} = W^{\pi}_{t,i} - \sum_{s=1}^{t-1} \Delta_{s,i}$, where $\Delta_{s,i} = D_{s,i} - \mu_i$. We now proceed to prove Theorem 1 in three steps.

Step 1

We first compute an upper bound for $C^D - C^*$. We claim that

$$C^* \geq C^D - \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i + h_i)\left(\sum_{s=1}^{t}\Delta_{s,i}\right)^+\right].$$

This is not difficult to show. For any policy $\pi \in \Pi$, we can bound:

$$\begin{aligned}
&\sum_{t=1}^{T}\left[c\,Z^{\pi}_t\,B + \sum_{i=1}^{m} p_i(D_{t,i} - N^{\pi}_{t,i} - I^{\pi}_{t,i})^+ + \sum_{i=1}^{m} h_i(N^{\pi}_{t,i} + I^{\pi}_{t,i} - D_{t,i})^+\right] \\
&\quad = \sum_{t=1}^{T}\left[c\,Z^{\pi}_t\,B + \sum_{i=1}^{m}(p_i + h_i)(I^{\pi}_{t,i} + N^{\pi}_{t,i} - D_{t,i})^+ - \sum_{i} p_i(I^{\pi}_{t,i} + N^{\pi}_{t,i} - D_{t,i})\right] \\
&\quad = \sum_{t=1}^{T}\left[c\,Z^{\pi}_t\,B + \sum_{i=1}^{m}(p_i + h_i)(I^{\pi}_{t+1,i})^+ - \sum_{i=1}^{m} p_i I^{\pi}_{t+1,i}\right]
\end{aligned}$$

$$
\begin{aligned}
&\geq \sum_{t=1}^{T}\left[c\,Z_t^{\pi}\,B+\sum_{i=1}^{m}(p_i+h_i)(W_{t+1,i}^{\pi})^+-\sum_{i=1}^{m}p_iW_{t+1,i}^{\pi}\right]\\
&\quad -\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\left(\sum_{s=1}^{t}\Delta_{s,i}\right)^+ +\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{s=1}^{t}p_i\Delta_{s,i}\\
&\geq C^D-\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\left(\sum_{s=1}^{t}\Delta_{s,i}\right)^+ +\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{s=1}^{t}p_i\Delta_{s,i},
\end{aligned}
$$

where the first inequality holds because the identity $I_{t,i}^{\pi}=W_{t,i}^{\pi}-\sum_{s=1}^{t-1}\Delta_{s,i}$ implies $(I_{t,i}^{\pi})^+\geq(W_{t,i}^{\pi})^+-(\sum_{s=1}^{t-1}\Delta_{s,i})^+$ and the second inequality follows by definition of C^D. Taking expectation on both sides and minimizing the sum in the left side of the inequality over $\pi\in\Pi$ yields the result.

Step 2

We now compute an upper bound for $\mathbf{E}[C^{CE}-C^D]$. We claim that

$$
\mathbf{E}[C^{CE}]-C^D\leq\mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\left(-\sum_{s=1}^{t}\Delta_{s,i}\right)^+\right].
$$

This can be shown using similar arguments as in Step 1. Let $I_{t+1,i}=I_{t,i}+n_{t,i}^D-D_{t,i}$ and $x_{t+1,i}=x_{t,i}+n_{t,i}^D-\mu_i$ (with $x_1=I_1$). Since $I_{t,i}=x_{t,i}-\sum_{s=1}^{t-1}\Delta_{s,i}$, we can bound:

$$
\begin{aligned}
\mathbf{E}[C^{CE}]&=\sum_{t=1}^{T}\mathbf{E}\left[c\,z_t^D\,B+\sum_{i=1}^{m}h_i(I_{t+1,i})^++\sum_{i}p_i(-I_{t+1,i})^+\right]\\
&=\sum_{t=1}^{T}\mathbf{E}\left[c\,z_t^D\,B+\sum_{i=1}^{m}(p_i+h_i)(I_{t+1,i})^+-\sum_{i}^{m}p_iI_{t+1,i}\right]\\
&\leq\sum_{t=1}^{T}\mathbf{E}\left[c\,z_t^D\,B+\sum_{i=1}^{m}(p_i+h_i)(x_{t+1,i})^+-\sum_{i}^{m}p_ix_{t+1,i}\right]\\
&\qquad+\mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\left(-\sum_{s=1}^{t}\Delta_{s,i}\right)^+\right]+\mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{s=1}^{t}p_i\Delta_{s,i}\right]\\
&=C^D+\mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\left(-\sum_{s=1}^{t}\Delta_{s,i}\right)^+\right].
\end{aligned}
$$

The inequality follows because $I_{t,i}=x_{t,i}-\sum_{s=1}^{t-1}\Delta_{s,i}$ implies $(I_{t,i})^+\leq(x_{t,i})^++(-\sum_{s=1}^{t-1}\Delta_{s,i})^+$.

Step 3

Putting the bounds from Steps 1 and 2 together, we conclude that

$$
\mathbf{E}[C^{CE}]-C^*=\mathbf{E}[C^{CE}]-C^D+C^D-C^*
$$

$$
\begin{aligned}
&\le \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\left(-\sum_{s=1}^{t}\Delta_{s,i}\right)^{+}\right] + \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\left(\sum_{s=1}^{t}\Delta_{s,i}\right)^{+}\right] \\
&\le 2\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\mathbf{E}\left[\left(\sum_{s=1}^{t}\Delta_{s,i}\right)^{2}\right]^{1/2} \\
&\le 2\sigma\,(T+1)^{3/2}\left[\sum_{i=1}^{m}(p_i+h_i)\right].
\end{aligned}
$$

This completes the proof of Theorem 1. ■

Proof of Corollary 1. Since each fulfilled demand incurs at least an ordering cost and each unfulfilled demand incurs at least a penalty cost, we can roughly bound $C^* \ge \mathbf{E}[\sum_{t=1}^{T}\sum_{i=1}^{m}\min\{c,p_i\} D_{t,i}] = T\sum_{i=1}^{m}\min\{c,p_i\}\mu_i$. Putting this together with the bound in Theorem 1 and the fact that $\sigma = \sqrt{\mu^*}$ (because demand is Poisson) completes the proof. ■

Proof of Theorem 2. The proof proceeds in two steps. The first step shows the general structure of the allocation policy: raise the inventory for a subset of retailers to the same level and allocate nothing to the rest of them. The second step shows the retailers we raise to the same level are the retailers with lowest inventory.

Step 1

Define $Y_{t,i} = I_{t,i} + N_{t,i}$ and $Y_t = (Y_{t,i})$. The optimal allocation in period t given I_t and z_t^D can be characterized by the first-order condition of (15): there exists λ, μ_i satisfies

$$\nabla_{Y_{t,i}} G(Y_{t,i}) + \nabla_{Y_{t,i}} \mathbf{E}_D\{J_{t+1}(Y_t - D_t)\} + \lambda - \mu_i = 0 \ \forall i \tag{34}$$

$$(Y_{t,i} - I_{t,i})\mu_i = 0 \ \forall i \tag{35}$$

$$\sum_{i=1}^{m}(Y_{t,i} - I_{t,i}) = z_t^D B \tag{36}$$

$$\mu_i \ge 0 \ \forall i \tag{37}$$

Notice the objective function in (15) is convex, and in $\Omega(Z)$ equality constraints are affine functions and inequality constraints are convex as well. Therefore the first-order conditions are sufficient for optimality.

Since $J_t(I_t)$ is always feasible, there must be a solution that satisfies (34) - (37), say $(\bar{Y}_{t,1}, \bar{Y}_{t,2}, \ldots, \bar{Y}_{t,m})$. Given $(\bar{Y}_{t,1}, \bar{Y}_{t,2}, \ldots, \bar{Y}_{t,m})$, retailers can be divided into two subsets: one set of retailers with $\mu_i > 0$ and one set of retailers with $\mu_i = 0$. Let $\mathcal{A} = \{i \,|\, \mu_i = 0\}$ and $k = |\mathcal{A}|$.

Case 1: For all retailers with $\mu_i > 0$

By (35) we have $Y_{t,i} = I_{t,i}$. Allocate nothing for those retailers.

Case 2: For all retailers with $\mu_i = 0$

(34) - (37) reduce to:

$$\nabla_{Y_{t,i}} g(Y_{t,i}) + \nabla_{Y_{t,i}} \mathbf{E}_D\{J_{t+1}(Y_t - D_t)\} + \lambda = 0 \;\; \forall i \in \mathcal{A} \tag{38}$$

$$\sum_{i \in \mathcal{A}} (Y_{t,i} - I_{t,i}) = z_t^D B \tag{39}$$

where $g(y) = \mathbf{E}_D\left[(h_i(y-D)^+ + p_i(y-D)^-)\right]$ is the single retailer inventory cost function. Suppose the solution for $\forall i \in \mathcal{A}$ is $(\bar{Y}_{t,1}, \bar{Y}_{t,2}, \ldots, \bar{Y}_{t,k})$ which satisfies (38) and (39). Notice that $\mathbf{E}_D\{J_{t+1}(Y_t - D_t)\}$ is symmetric in $Y_{t,i}$ and D_t, and $g(\cdot)$ is the same for all retailers.

To see why, consider a simple example with discrete demand and $T = 2, k = 2$, let P_s be the probability of demand scenario s and write (38)-(39) in extensive form.

$$\nabla_{Y_{1,1}} g(Y_{1,1}) + \sum_s P_s(c + \nabla_{Y_{1,1}} g(Y_{1,1} - d_{1,1}^s + N_{2,1}^s)) = -\lambda \tag{40}$$

$$\nabla_{Y_{1,2}} g(Y_{1,2}) + \sum_s P_s(c + \nabla_{Y_{1,2}} g(Y_{1,2} - d_{1,2}^s + N_{2,2}^s)) = -\lambda \tag{41}$$

$$Y_{1,1} + Y_{1,2} = z_1^D B + I_{1,1} + I_{1,2} \tag{42}$$

$$N_{2,1}^s + N_{2,2}^s = z_2^D B \;\; \forall s \tag{43}$$

where $d_{t,i}^s$ is the demand and $N_{t,i}^s$ is the allocated units for retailer i in period t under scenario s. Since demand D is i.i.d. for all retailers, demand scenarios must be symmetric. If $(\bar{Y}_{1,1}, \bar{Y}_{1,2})$ satisfies (40)-(43), by exchanging $N_{2,1}^s$ and $N_{2,2}^s$, $(\bar{Y}_{1,2}, \bar{Y}_{1,1})$ also satisfies (40)-(43).

More generally, if there is a vector $(\bar{Y}_{t,1}, \bar{Y}_{t,2}, \ldots, \bar{Y}_{t,k})$ that satisfies (38)-(39), then any permutation of $(\bar{Y}_{t,1}, \bar{Y}_{t,2}, \ldots, \bar{Y}_{t,k})$ also satisfies (38)-(39), which is optimal. By the convexity of $J_t(I_t)$, the solution $(\frac{1}{k}\sum_{i=1}^{k} \bar{Y}_{t,i}, \frac{1}{k}\sum_{i=1}^{k} \bar{Y}_{t,i}, \ldots, \frac{1}{k}\sum_{i=1}^{k} \bar{Y}_{t,i})$, which is the convex combination of the permutations of $(\bar{Y}_{t,1}, \bar{Y}_{t,2}, \ldots, \bar{Y}_{t,k})$, must be optimal as well, which means raising inventory of all retailers with $\mu_i = 0$ to the same level is optimal. By (39), this level is

$$N_{t,i} + I_{t,i} = \frac{1}{k}\left(\sum_{i \in \mathcal{A}} I_{t,i} + z_t^D B\right) \quad \forall i \in \mathcal{A}$$

Step 2

Next, we'll show that the set of retailers in $\mathcal{A}$ is indeed the θ retailers with lowest starting inventory level, i.e. $\mathcal{A} = \{i \,|\, i \leq \theta\}$ and $k = \theta$. From now on, we will drop the subscript t if it is not ambiguous. First, we define a total cost function $\Gamma(\cdot)$ for a single retailer to be

$$\Gamma(Y_i) = c(Y_i - I_i) + g(Y_i) + \mathbf{E}_D\{\Gamma(Y_i - D_i)\}$$

Notice $\Gamma(\cdot)$ is convex. Next assume retailers have pre-allocation inventory level $\{I_1, I_2, \ldots, I_m\}$ and remember these levels are sorted from smallest to largest. Denotes the order-up-to level in Theorem 2 to be Y^*

$$Y^* = \frac{1}{\theta}\left(\sum_{i=1}^{\theta} I_i + zB\right) \quad \forall i \leq \theta$$

and let the cost of the policy that raise $\{I_1, I_2, \ldots, I_\theta\}$ to the same level Y^* to be P^*. Now consider two cases:

Case 1: Raise $\theta' \neq \theta$ retailers with the lowest starting inventory to the same level.

If $\theta' > \theta$ then (36) is infeasible. Now, without loss of generality, assume $\theta - 1$ retailers with inventory $\{I_1, I_2, \ldots, I_{\theta-1}\}$ are raised to the same level $\bar{Y}$. Let the cost of this policy to be P'.

$$\begin{aligned} P' &= (\theta - 1)\Gamma(\bar{Y}) + \Gamma(I_\theta) + \sum_{i>\theta} \Gamma(I_i) \\ &\geq \theta\,\Gamma(Y^*) + \sum_{i>\theta} \Gamma(I_i) \\ &= P^* \end{aligned}$$

where the inequality holds because of the convexity of $\Gamma(\cdot)$. So P^* dominates P'.

Case 2: Raise inventory for θ arbitrarily selected retailers to the same level,

Without loss of generality, suppose we skip retailer j and raise the retailers with inventory

$\{I_1, I_2, \ldots, I_{j-1}, I_{j+1}, \ldots, I_\theta, I_{\theta+1}\}$ to the same level $\bar{Y}$. It is easy to check we have $\bar{Y} \geq I_{\theta+1} \geq Y^* \geq I_j$. Let the cost of this policy to be $P^{'}$, then by convexity $\Gamma(\cdot)$ we have

$$\begin{aligned} P^{'} &= \theta\Gamma(\bar{Y}) + \Gamma(I_j) + \sum_{i>\theta+1} \Gamma(I_i) \\ &\geq \theta\Gamma(Y^*) + \Gamma(I_{\theta+1}) + \sum_{i>\theta+1} \Gamma(I_i) \\ &= P^* \end{aligned}$$

which again P^* dominates $P^{'}$.

Thus, raising the θ retailers with the lowest inventory level dominates any other possible set of retailers, so we have $\mathcal{A} = \{i \,|\, i \leq \theta\}$. Since there is no other allocation outperforms the one in Theorem 3, the expected total cost under the allocation policy in Theorem 2 must be J^*. ■.

Proof of Theorem 3. We proceed in several steps.

Step 1

Define n^H as follows: $n^H_{t,i} = n^{*1}_{t,i} + \Delta_{t,i} - \frac{1}{m}\sum_{j=1}^m \Delta_{t,j}$. We claim that if $D_{1:T} \in \mathcal{A}$, then n^H is an optimal allocation for J^H. To see this, first note that, using $x_{t+1,i} = I_{1,i} + \sum_{s=1}^t n_{s,i} - t\mu_i$ (because $x_{s+1,i} = x_{s,i} + n_{s,i} - \mu_i$), optimization $J_1^D(I_1)$ can be written as:

$$J_1^D(I_1) = \min_n \sum_{t=1}^T \left[c\, z_t^D B + \sum_{i=1}^m p_i\, y_{t,i} + \sum_{i=1}^m h_i \theta_{t,i} \right] \tag{44}$$

$$\text{s.t. } y_{t,i} \geq t\mu_i - \sum_{s=1}^t n_{s,i} - I_{1,i} \qquad \forall t,\, i \tag{45}$$

$$y_{t,i} \geq 0 \qquad \forall t,\, i \tag{46}$$

$$\theta_{t,i} \geq I_{1,i} + \sum_{s=1}^t n_{s,i} - t\mu_i \qquad \forall t,\, i \tag{47}$$

$$\theta_{t,i} \geq 0 \qquad \forall t,\, i \tag{48}$$

$$\sum_{i=1}^m n_{t,i} = z_t^D B \qquad \forall t \tag{49}$$

$$n_{t,i} \geq 0 \qquad \forall t,\, i \tag{50}$$

By Karush-Kuhn-Tucker (KKT) conditions, there exists dual variables $\lambda^1_{t,i}$, $\lambda^2_{t,i}$, $\lambda^3_{t,i}$, $\lambda^4_{t,i}$, ξ_t, and $\Omega_{t,i}$ corresponding to constraints (45)-(50) such that

$$p_i = \lambda^1_{t,i} + \lambda^2_{t,i} \qquad \forall t,\, i$$

$$
\begin{aligned}
h_i &= \lambda^3_{t,i} + \lambda^4_{t,i} \qquad \forall t,\, i \\
0 &= -\sum_{s=t}^{T} \lambda^1_{t,i} + \sum_{s=t}^{T} \lambda^3_{t,i} + \xi_t - \Omega_{t,i} \qquad \forall t,\, i \\
0 &= \lambda^1_{t,i} \left[y_{t,i} - t\mu_i + \sum_{s=1}^{t} n_{s,i} + I_{1,i} \right] \qquad \forall t,\, i \\
0 &= \lambda^2_{t,i}\, y_{t,i} \qquad \forall t,\, i \\
0 &= \lambda^3_{t,i} \left[\theta_{t,i} - I_{1,i} - \sum_{s=1}^{t} n_{s,i} + t\mu_i \right] \qquad \forall t,\, i \\
0 &= \lambda^4_{t,i}\, \theta_{t,i} \qquad \forall t,\, i \\
0 &= \Omega_{t,i}\, n_{t,i} \qquad \forall t,\, i \\
&\lambda^1_{t,i} \geq 0,\ \lambda^2_{t,i} \geq 0,\ \lambda^3_{t,i} \geq 0,\ \lambda^4_{t,i} \geq 0 \qquad \forall t,\, i.
\end{aligned}
$$

Since we assume that $I_1 = 0$ and $\sum_{s=1}^{t} n^{*1}_{s,i} - t\mu_i$ is either strictly positive or strictly negative for all t and i, we immediately have $\Omega_{t,i} = 0$ for all t and i. Now, to show that n^H is optimal for the hindsight problem on $\mathcal{A}$, it is sufficient that we show: (1) $n^H_{t,i} \geq 0$ for all t and i, (2) $\sum_{i=1}^{m} n^H_{t,i} = z^D_t B$ for all t, and (3) $I_{1,i} + \sum_{s=1}^{t} n^H_{s,i} - \sum_{s=1}^{t} D_{t,i}$ has the same sign (i.e., strictly positive or strictly negative) as $I_{1,i} + \sum_{s=1}^{t} n^{*1}_{s,i} - t\mu_i$ for all t and i. (If these conditions are satisfied, then we can use $\lambda^1_{t,i}$, $\lambda^2_{t,i}$, $\lambda^3_{t,i}$, $\lambda^4_{t,i}$, ξ_t, and $\Omega_{t,i}$ from $J^D_1(I_1)$ as dual variables for the hindsight problem. Since KKT conditions are both necessary and sufficient for optimality in linear program, we can then conclude that n^H is optimal.) But, conditions (1)-(3) immediately follow from the definition of n^H and $\mathcal{A}$. This completes the proof.

Step 2

We will now prove that $\mathbf{E}[(J^{CE} - J^H)\, \mathbf{1}\{\mathcal{A}\}] \;\leq\; 2\sigma(T+1)^{3/2} \left[\sum_{i=1}^{m}(p_i + h_i)\right]$. This is not difficult to show. Using n^H as the optimal solution for the hindsight problem, we can write: $I^H_{t,i} = \sum_{s=1}^{t} n^{*1}_{s,i} + \sum_{s=1}^{t} \Delta_{s,i} - \frac{1}{m} \sum_{s=1}^{t} \sum_{j=1}^{m} \Delta_{s,j} - \sum_{s=1}^{t} D_{s,i}$. Moreover, we also have: $I^{CE}_t = \sum_{s=1}^{t} n^{*1}_s - \sum_{s=1}^{t} D_s$. So, on $\mathcal{A}$, we can bound:

$$
J^{CE} - J^H = \sum_{t=1}^{T} \sum_{i=1}^{m} p_i \left[\left(\sum_{s=1}^{t} D_{s,i} - \sum_{s=1}^{t} n^{*1}_{s,i} \right)^{+} - \left(\sum_{s=1}^{t} D_{s,i} - \sum_{s=1}^{t} n^{*1}_{s,i} - \sum_{s=1}^{t} \Delta_{s,i} + \frac{1}{m} \sum_{s=1}^{t} \sum_{j=1}^{m} \Delta_{s,j} \right)^{+} \right]
$$

$$+\sum_{t=1}^{T}\sum_{i=1}^{m} h_i \left[\left(\sum_{s=1}^{t} n_{s,i}^{*1} - \sum_{s=1}^{t} D_{s,i} \right)^{+} - \left(\sum_{s=1}^{t} n_{s,i}^{*1} + \sum_{s=1}^{t} \Delta_{s,i} - \frac{1}{m}\sum_{s=1}^{t}\sum_{j=1}^{m} \Delta_{s,j} - \sum_{s=1}^{t} D_{s,i} \right)^{+} \right]$$

$$\leq \sum_{t=1}^{T}\sum_{i=1}^{m} (p_i + h_i) \left[|\Delta_{t,i}| + \frac{1}{m}\sum_{s=1}^{t}\sum_{j=1}^{m} |\Delta_{s,j}| \right].$$

The result immediately follows because $\mathbf{E}[|\Delta_{t,i}|\,\mathbf{1}\{\mathcal{A}\}] \leq \mathbf{E}[|\Delta_{t,i}|] \leq \sigma$ for all t and i.

Step 3

We now argue that $\mathbf{E}[(J^{RCE} - J^{H})\,\mathbf{1}\{\mathcal{A}\}] \leq 2\sigma T\,[\sum_{i=1}^{m}(p_i + h_i)]$. This requires that we first study the evolution of the re-optimized solution. However, since the solution of $J_t^D(I_t)$ may not be unique, we will only prove the result under a particular sequence of optimal solution $\{n_t^{*t}\}$.

Define n^{*t} for $t > 1$ as follows: $n_{t,i}^{*t} = n_{t,i}^{*1} + \Delta_{t-1,i} - \frac{1}{m}\sum_{j=1}^{m} \Delta_{t-1,j}$ and $n_{s,i}^{*t} = n_{s,i}^{*1}$ for $s > t$. Suppose that $D_{1:T} \in \mathcal{A}$. The following can be shown: If we use $n_s = n_s^{*s}$ for all $s \leq t-1$, then (1) the starting inventory level for retailer i at the beginning of period t is given by $I_{t,i} = \sum_{s=1}^{t-1} n_{s,i}^{*1} - (t-1)\mu_i - \Delta_{t-1,i} - \frac{1}{m}\sum_{s=1}^{t-2}\sum_{j=1}^{m} \Delta_{s,j}$ and (2) n^{*t} is an optimal solution for $J_t^D(I_t)$. These can be proved by induction. We start with $t = 2$. (The case $t = 1$ is trivially true.) Note that we can write: $I_{2,i} = I_{1,i} + n_1^{*1} - D_{1,i} = n_{1,i}^{*1} - \mu_i - \Delta_{1,i}$. At the beginning of period 2, we have to solve the following linear program:

$$\begin{aligned}
J_2^D(I_2) \;&= \min_{n} \; \sum_{t=2}^{T} \left[c\, z_t^D\, B + \sum_{i=1}^{m} p_i\, y_{t,i} + \sum_{i=1}^{m} h_i \theta_{t,i} \right] \\
\text{s.t.} \;\; & y_{t,i} \geq (t-1)\mu_i - \sum_{s=2}^{t} n_{s,i} - I_{2,i} \qquad \forall t \geq 2,\, i \\
& y_{t,i} \geq 0 \qquad \forall t \geq 2,\, i \\
& \theta_{t,i} \geq I_{2,i} + \sum_{s=2}^{t} n_{s,i} - (t-1)\mu_i \qquad \forall t \geq 2,\, i \\
& \theta_{t,i} \geq 0 \qquad \forall t \geq 2,\, i \\
& \sum_{i=1}^{m} n_{t,i} = z_t^D B \qquad \forall t \geq 2 \\
& n_{t,i} \geq 0 \qquad \forall t \geq 2,\, i
\end{aligned}$$

It is not difficult to show using similar dual arguments as in Step 1 that n^{*2} is an optimal solution for $J_2^D(I_2)$. In particular, on $\mathcal{A}$, all the three conditions in Step 1 still hold: $n_{t,i}^{*2} \geq 0$ for all $t \geq 2$ and i, $\sum_{i=1}^{m} n_{t,i}^{*2} = z_t^D B$ for all $t \geq 2$, and $\sum_{s=2}^{t} n_{s,i}^{*2} - (t-1)\mu_i$ has the same sign (i.e., strictly positive or strictly negative) as $\sum_{s=2}^{t} n_{s,i}^{*1} - (t-1)\mu_i$ for all $t \geq 2$ and i. This allows us to use the same dual variables that correspond to the constraints in $J_2^D(I_1 + n_1^{*1} - \mu)$ for $J_2^D(I_2)$; therefore, by the sufficiency of KKT conditions, we conclude that n^{*2} is optimal for $J_2^D(I_2)$. This is our base case. Now, suppose that (1) and (2) hold for all $s \leq t-1$. We want to show that they still hold for $s = t$. By induction hypothesis, we can write: $I_{t,i} = I_{t-1,i} + n_{t-1,i}^{*\,t-1} - D_{t-1,i} = \left(\sum_{s=1}^{t-2} n_{s,i}^{*1} - (t-2)\mu_i - \Delta_{t-2,i} - \frac{1}{m}\sum_{s=1}^{t-3}\sum_{j=1}^{m} \Delta_{s,j}\right) + \left(n_{t-1,i}^{*1} + \Delta_{t-2,i} - \frac{1}{m}\sum_{j=1}^{m} \Delta_{t-2,j}\right) - D_{t-1,i} = \sum_{s=1}^{t-1} n_{s,i}^{*1} - (t-1)\mu_i - \Delta_{t-1,i} - \frac{1}{m}\sum_{s=1}^{t-2}\sum_{j=1}^{m} \Delta_{s,j}$. So, (1) holds. At the beginning of period t, we have to solve the following linear program:

$$
\begin{aligned}
J_t^D(I_t) \;&= \min_n \sum_{s=t}^{T} \left[c\, z_s^D\, B + \sum_{i=1}^{m} p_i\, y_{s,i} + \sum_{i=1}^{m} h_i \theta_{s,i} \right] \\
\text{s.t.}\;\; & y_{s,i} \geq (t-s+1)\mu_i - \sum_{r=s}^{t} n_{r,i} - I_{s,i} \qquad \forall s \geq t,\, i \\
& y_{s,i} \geq 0 \qquad \forall s \geq t,\, i \\
& \theta_{s,i} \geq I_{s,i} + \sum_{r=s}^{t} n_{r,i} - (t-s+1)\mu_i \qquad \forall s \geq t,\, i \\
& \theta_{s,i} \geq 0 \qquad \forall s \geq t,\, i \\
& \sum_{i=1}^{m} n_{s,i} = z_s^D B \qquad \forall s \geq t \\
& n_{s,i} \geq 0 \qquad \forall s \geq t,\, i
\end{aligned}
$$

By similar arguments as before, it is not difficult to check that, on $\mathcal{A}$, we have: $n_{s,i}^{*t} \geq 0$ for all $s \geq t$ and i, $\sum_{i=1}^{m} n_{s,i}^{*t} = z_s^D B$ for all $s \geq t$, and $\sum_{r=s}^{t} n_{r,i}^{*t} - (t-s+1)\mu_i$ has the same sign (i.e., strictly positive or strictly negative) as $\sum_{r=s}^{t} n_{r,i}^{*1} - (t-s+1)\mu_i$ for all $s \geq t$ and i. This allows us to use the same dual variables that correspond to the constraints in $J_t^D(I_1 + \sum_{s=1}^{t-1} n_s^{*1} - (t-1)\mu)$ for $J_t^D(I_t)$; hence, by the sufficiency of KKT conditions, we conclude that n^{*t} is optimal for $J_t^D(I_t)$. This completes the induction.

We now make two important observations: under RCE (i.e., using $n_s = n_s^{*s}$ for $s \leq t$) we have $I_{t+1,i}^{RCE} = \sum_{s=1}^{t} n_{s,i}^{*1} - t\mu_i - \Delta_{t,i} - \frac{1}{m}\sum_{s=1}^{t-1}\sum_{j=1}^{m} \Delta_{s,j}$. In contrast, under the perfect hindsight policy in Step 1, it is not difficult to check that $I^H_{t+1,i} = \sum_{s=1}^t n^{*1}_{s,i} - t\mu_i - \frac{1}{m}\sum_{s=1}^t \sum_{j=1}^m \Delta_{s,j}$. So, $I^{RCE}_{t+1,i} = I^H_{t+1,i} - \Delta_{t,i} + \frac{1}{m}\sum_{j=1}^m \Delta_{t,j}$. This implies:

$$J^{RCE} - J^H \;\le\; \sum_{t=1}^T \sum_{i=1}^m (p_i + h_i)\left[|\Delta_{t,i}| + \frac{1}{m}\sum_{j=1}^m |\Delta_{t,j}|\right].$$

As in Step 2, the result follows because $\mathbf{E}[|\Delta_{t,i}|\,\mathbf{1}\{\mathcal{A}\}] \le \mathbf{E}[|\Delta_{t,i}|] \le \sigma$ for all t and i. ■

Proof of Theorem 4. We proceed in several steps.

Step 1

Similar to Theorem 3, we first argue that $\mathbf{E}[(J^{CE} - J^H)\,\mathbf{1}\{\hat{\mathcal{A}}\}] \le 2m(p^* + h^*)(T+1)^{3/2}\sigma^*$ and $\mathbf{E}[(J^{RCE} - J^H)\,\mathbf{1}\{\hat{\mathcal{A}}\}] \;\le\; 2m(p^* + h^*)T\sigma^*$. Define n^H as follows:

$$n^H_{t,i} \;=\; \frac{z^D_t B}{m} + D_{t,i} - \frac{1}{m}\sum_{j=1}^m D_{t,j} \;=\; \frac{z^D_t B}{m} + \Delta_{t,i} - \frac{1}{m}\sum_{j=1}^m \Delta_{t,j}.$$

We claim that if $D_{1:T} \in \hat{\mathcal{A}}$, then n^H is an optimal allocation for J^H. To see this, simply note that

$$\begin{aligned}
&\sum_{t=1}^T \left[c\, z^D_t\, B + \sum_{i=1}^m p^*\,(D_{t,i} - x_{t,i} - n_{t,i})^+ + \sum_{i=1}^m h^*(x_{t,i} + n_{t,i} - D_{t,i})^+\right]\\
&\quad= \sum_{t=1}^T \left[c\, z^D_t\, B + \sum_{i=1}^m p^* \left(\sum_{s=1}^t D_{s,i} - \sum_{s=1}^t n_{s,i}\right)^+ + \sum_{i=1}^m h^* \left(\sum_{s=1}^t n_{s,i} - \sum_{s=1}^t D_{s,i}\right)^+\right]\\
&\quad\ge \sum_{t=1}^T \left[c\, z^D_t\, B + \sum_{i=1}^m p^* \left(\frac{1}{m}\sum_{s=1}^t\sum_{j=1}^m D_{s,j} - \frac{1}{m}\sum_{s=1}^t\sum_{j=1}^m n_{s,j}\right)^+ + \sum_{i=1}^m h^* \left(\frac{1}{m}\sum_{s=1}^t\sum_{j=1}^m n_{s,j} - \frac{1}{m}\sum_{s=1}^t\sum_{j=1}^m D_{s,j}\right)^+\right]\\
&\quad= \sum_{t=1}^T \left[c\, z^D_t\, B + \sum_{i=1}^m p^* \left(\frac{1}{m}\sum_{s=1}^t\sum_{j=1}^m D_{s,j} - \frac{1}{m}\sum_{s=1}^t z^D_s B\right)^+ + \sum_{i=1}^m h^* \left(\frac{1}{m}\sum_{s=1}^t z^D_s B - \frac{1}{m}\sum_{s=1}^t\sum_{j=1}^m D_{s,j}\right)^+\right],
\end{aligned}$$

where the first equality follows by the definition of x_t, the first inequality follows by the convexity of $(\cdot)^+$, and the last inequality follows by the definition of n_t and z^D_t. Since $I_{t+1,i} = \sum_{s=1}^t n_{s,i} - \sum_{s=1}^t D_{s,i}$, it is not difficult to check that the above lower bound is achieved by setting $n_t = n^H_t$. Moreover, on $\hat{\mathcal{A}}$, we have $n^H_t > 0$ for all t. So, n^H is an optimal feasible solution for J^H.

We will now prove that $\mathbf{E}[(J^{CE} - J^H)\,\mathbf{1}\{\hat{\mathcal{A}}\}] \;\le\; 2m(p^* + h^*)(T+1)^{3/2}\sigma^*$. This is not difficult to show. Note that, by similar arguments as above, it can be shown that $n^D_{t,i} = z^D_t B/m$ for all i is an optimal solution of C^D. Using $I^{CE}_t = \sum_{s=1}^t n^D_s - \sum_{s=1}^t D_s$, we can write:

$$J^{CE} = \sum_{t=1}^{T} \left[c\, z_t^D B + \sum_{i=1}^{m} p^* \left(D_{t,i} - \frac{1}{m} \sum_{s=1}^{t} z_s^D B \right)^+ + \sum_{i=1}^{m} h^* \left(\frac{1}{m} \sum_{s=1}^{t} z_s^D B - D_{t,i} \right)^+ \right].$$

So, on $\hat{\mathcal{A}}$, we can bound:

$$\begin{aligned} J^{CE} - J^H &= \sum_{t=1}^{T} \sum_{i=1}^{m} p^* \left[\left(D_{t,i} - \frac{1}{m} \sum_{s=1}^{t} z_s^D B \right)^+ - \left(\frac{1}{m} \sum_{s=1}^{t} \sum_{j=1}^{m} D_{s,j} - \frac{1}{m} \sum_{s=1}^{t} z_s^D B \right)^+ \right] \\ &\quad + \sum_{t=1}^{T} \sum_{i=1}^{m} h^* \left[\left(\frac{1}{m} \sum_{s=1}^{t} z_s^D B - D_{t,i} \right)^+ - \left(\frac{1}{m} \sum_{s=1}^{t} z_s^D B - \frac{1}{m} \sum_{s=1}^{t} \sum_{j=1}^{m} D_{s,j} \right)^+ \right] \\ &\le \sum_{t=1}^{T} \sum_{i=1}^{m} (p^* + h^*) \left[|\Delta_{t,i}| + \frac{1}{m} \sum_{s=1}^{t} \sum_{j=1}^{m} |\Delta_{s,j}| \right]. \end{aligned}$$

The result follows because $\mathbf{E}[|\Delta_{t,i}|\, \mathbf{1}\{\hat{\mathcal{A}}\}] \le \mathbf{E}[|\Delta_{t,i}|] \le \sigma^*$ for all t and i.

We now argue that $\mathbf{E}[(J^{RCE} - J^H)\, \mathbf{1}\{\hat{\mathcal{A}}\}] \le 2m(p^* + h^*)T\sigma^*$. Define n^{*t} as follows: $n^{*1}_{t,i} = \frac{z_t^D B}{m}$ for all t, $n^{*t}_{t,i} = \frac{z_t^D B}{m} + \Delta_{t-1,i} - \frac{1}{m} \sum_{j=1}^{m} \Delta_{t-1,j}$, and $n^{*t}_{s,i} = \frac{z_s^D B}{m}$ for all $s > t$. Suppose that $D_{1:T} \in \hat{\mathcal{A}}$. The following can be shown: If we use $n_s = n_s^{*s}$ or all $s \le t-1$, the inventory level at the beginning of period t is given by

$$I_{t,i} = \frac{1}{m} \sum_{s=1}^{t-1} z_s^D B - (t-1)\mu^* - \Delta_{t-1,i} - \frac{1}{m} \sum_{s=1}^{t-2} \sum_{j=1}^{m} \Delta_{s,j}$$

and n^{*t} is an optimal solution of $J_t^D(I_t)$. These can be proved by induction. We start with $t=2$. (The case $t=1$ is trivial.) At the beginning of period 2, we have $I_{2,i} = I_{1,i} + n_1 - D_{1,i} = \frac{z_1^D}{m} - \mu^* - \Delta_{1,i}$. To show that n^{*2} is optimal for $J_2^D(I_2)$, note that, by convexity of $(\cdot)^+$, we can bound:

$$\begin{aligned} &\sum_{t=2}^{T} \left[c\, z_t^D B + \sum_{i=1}^{m} p^* (\mu^* - I_{t,i} - n_{t,i})^+ + \sum_{i=1}^{m} h^* (I_{t,i} + n_{t,i} - \mu^*)^+ \right] \\ &\quad = \sum_{t=2}^{T} \left[c\, z_t^D B + \sum_{i=1}^{m} p^* \left((t-1)\mu^* - \sum_{s=2}^{t} n_{s,i} - I_{2,i} \right)^+ + \sum_{i=1}^{m} h^* \left(I_{2,i} + \sum_{s=2}^{t} n_{s,i} - (t-1)\mu^* \right)^+ \right] \\ &\quad \ge \sum_{t=2}^{T} \left[c\, z_t^D B + \sum_{i=1}^{m} p^* \left((t-1)\mu^* - \frac{1}{m} \sum_{s=2}^{t} \sum_{j=1}^{m} n_{s,j} - \frac{1}{m} \sum_{j=1}^{m} I_{2,j} \right)^+ \right. \\ &\qquad \left. + \sum_{i=1}^{m} h^* \left(\frac{1}{m} \sum_{j=1}^{m} I_{2,j} + \frac{1}{m} \sum_{s=2}^{t} \sum_{j=1}^{m} n_{s,j} - (t-1)\mu^* \right)^+ \right] \end{aligned}$$

$$= \sum_{t=2}^{T} \left[c\, z_t^D B + \sum_{i=1}^{m} p^* \left((t-1)\mu^* - \frac{1}{m}\sum_{s=2}^{t} z_s^D B - \frac{1}{m}\sum_{j=1}^{m} I_{2,j} \right)^+ + \sum_{i=1}^{m} h^* \left(\frac{1}{m}\sum_{j=1}^{m} I_{2,j} + \frac{1}{m}\sum_{s=2}^{t} z_s^D B - (t-1)\mu^* \right)^+ \right].$$

Recursively solving $(t-1)\mu^* - \sum_{s=2}^{t} n_{s,i} - I_{2,i} = (t-1)\mu^* - \frac{1}{m}\sum_{s=2}^{t} z_s^D B - \frac{1}{m}\sum_{j=1}^{m} I_{2,j}$ yields $n_{s,i} = n_{s,i}^{*2}$ for all $s \geq 2$ and i. Since n^{*2} exactly achieves the lower bound, it must be optimal. (Since $z_t^D \geq 1$ for all t, $n^{*2} > 0$ on $\hat{\mathcal{A}}$; so, it is a feasible optimal solution.) This is our base case. Now, suppose that the conditions hold for all $s \leq t-1$. We want to show that they also hold for period t. By induction hypothesis, $I_{t,i} = I_{t-1,i} + n_{t-1,i}^{*\,t-1} - D_{t-1,i} = \left(\frac{1}{m}\sum_{s=1}^{t-2} z_s^D B - (t-2)\mu^* - \Delta_{t-2,i} - \frac{1}{m}\sum_{s=1}^{t-3}\sum_{j=1}^{m} \Delta_{s,j} \right) + \left(\frac{z_{t-1}^D B}{m} + \Delta_{t-2,i} - \frac{1}{m}\sum_{j=1}^{m} \Delta_{t-2,j} \right) - D_{t-1,i} = \frac{1}{m}\sum_{s=1}^{t-1} z_s^D B - (t-1)\mu^* - \Delta_{t-1,i} - \frac{1}{m}\sum_{s=1}^{t-2}\sum_{j=1}^{m} \Delta_{s,j}$. Using similar convexity arguments as above, it is not difficult to check that n^{*t} is optimal for $J_t^D(I_t)$. This completes the induction.

Putting our results together, we have $I_{t+1,i}^{RCE} = \frac{1}{m}\sum_{s=1}^{t} z_s^D B - t\mu^* - \Delta_{t,i} - \frac{1}{m}\sum_{s=1}^{t-1}\sum_{j=1}^{m} \Delta_{s,j}$ and $I_{t+1,i}^{H} = \frac{1}{m}\sum_{s=1}^{t} z_s^D B - t\mu^* - \frac{1}{m}\sum_{s=1}^{t}\sum_{j=1}^{m} \Delta_{s,j}$. So, $I_{t+1,i}^{RCE} = I_{t+1,i}^{H} - \Delta_{t,i} + \frac{1}{m}\sum_{j=1}^{m} \Delta_{t,j}$. This implies: $J^{RCE} - J^H \leq \sum_{t=1}^{T}\sum_{i=1}^{m} (p^* + h^*) \left[|\Delta_{t,i}| + \frac{1}{m}\sum_{j=1}^{m} |\Delta_{t,j}| \right]$ on $\hat{\mathcal{A}}$. Taking expectation yields $\mathbf{E}[(J^{RCE} - J^H)\,\mathbf{1}\{\hat{\mathcal{A}}\}] \leq 2m(p^* + h^*)T\sigma^*$.

Step 2

We claim that $P(\hat{\mathcal{A}}) \geq 1 - 2mTe^{-B^2/(16m^4\mu^*)}$ for all sufficiently large B^2/μ^*. First, note that:

$$P(\hat{\mathcal{A}}) = \prod_{t=1}^{T} P\left(2m\sum_{i=1}^{m} |\Delta_{t,i}| < B \right) \geq \prod_{t=1}^{T} \left[1 - \sum_{i=1}^{m} P\left(|\Delta_{t,i}| \geq \frac{B}{2m^2} \right) \right].$$

By an exponential tail bound for Poisson random variable, it can be shown that $P(|\Delta_{t,i}| \geq B/2m^2) \leq 2e^{-B^2/(16m^4\mu^*)}$ (see below). Thus, by Bernoulli's inequality, we can write $P(\hat{\mathcal{A}}) \geq (1 - 2m\, e^{-B^2/(16m^4\mu^*)})^T \geq 1 - 2mTe^{-B^2/(16m^4\mu^*)}$ for all sufficiently large B^2/μ^*.

Proof of an exponential tail bound for Poisson random variable. By Markov's inequality, $P(|\Delta_{t,i}| > B/2m^2) \leq \mathbf{E}[e^{r|\Delta_{t,i}|}]/e^{rB/2m^2}$ for all $r > 0$. By moment generating function of Poisson distribution, as long as $r < 1$ is sufficiently small, $\mathbf{E}[e^{r|\Delta_{t,i}|}] \leq \mathbf{E}[e^{r\Delta_{t,i}}] + \mathbf{E}[e^{-r\Delta_{t,i}}] = e^{\mu^*(e^r - 1 - r)} +$

$e^{\mu^*(e^{-r}-1+r)} \leq 2e^{\mu^* r^2}$. (The last inequality holds because $e^r - 1 - r \leq r^2$ for all small r.) This implies $P(|\Delta_{t,i}| > B/2m^2) \leq 2e^{\mu^*(r^2-(B/2\mu^* m^2)r)}$ for all sufficiently small $r > 0$. Minimizing the bound over $r > 0$, yields $r = B/(4\mu^* m^2)$. Since $B/(4\mu^* m^2)$ is small for all large μ^*, we can use the above bound and get $P(|\Delta_{t,i}| > B/2m^2) \leq 2e^{-B^2/(16m^4\mu^*)}$.

Step 3

We now put the results of Steps 1 and 2 together. From Step 1,

$$\begin{aligned} J^* &\geq \mathbf{E}[J^H] \\ &\geq \sum_{t=1}^{T} \mathbf{E}\left[c\, z_t^D\, B + \sum_{i=1}^{m} p^* \left(\frac{1}{m}\sum_{s=1}^{t}\sum_{j=1}^{m} D_{s,j} - \frac{1}{m}\sum_{s=1}^{t} z_s^D B \right)^+ + \sum_{i=1}^{m} h^* \left(\frac{1}{m}\sum_{s=1}^{t} z_s^D B - \frac{1}{m}\sum_{s=1}^{t}\sum_{j=1}^{m} D_{s,j} \right)^+ \right]. \end{aligned}$$

Since $\sum_{t=1}^{T} z_t^D B \geq TB \geq T\mu^*$ (because we assume that $z_t^D > 0$) and, for each t, either we have $\sum_{s=1}^{t} z_s^D B \geq tm\mu^*$ or $\sum_{s=1}^{t} z_s^D B < tm\mu^*$, we can further bound

$$J^* \geq cT\mu^* + \frac{\min\{p^*, h^*\}}{m} \sum_{t=1}^{T} \min\left\{ \mathbf{E}\left[\left(\sum_{s=1}^{t}\sum_{j=1}^{m} D_{s,j} - tm\mu^* \right)^+ \right], \mathbf{E}\left[\left(tm\mu^* - \sum_{s=1}^{t}\sum_{j=1}^{m} D_{s,j} \right)^+ \right] \right\}.$$

The expectations inside the $\min\{\cdot\}$ operator are of order $\sqrt{tm\mu^*}$. This means that there exists a constant $M' > 0$ independent of $T > 0$ such that, for all large μ^*, we have

$$J^* \;\geq\; cT\mu^* + M'\sqrt{\mu^*}\, T^{3/2}.$$

Note that, on $\hat{\mathcal{A}}^c$, we can loosely bound J^π with $\sum_{t=1}^{T} c z_t^D B + \sum_{t=1}^{T}\sum_{i=1}^{m} T(p^* + h^*)D_{t,i}$ (i.e., each unit of demand incur both holding and penalty cost T times). Since $\mathbf{E}[D_{t,i}\mathbf{1}\{\hat{\mathcal{A}}^c\}] \leq \mathbf{E}[D_{t,i}^2]^{1/2}$ $\mathbf{E}[\mathbf{1}\{\hat{\mathcal{A}}^c\}^2]^{1/2} \leq 2\mu^* P(\hat{\mathcal{A}}^c)^{1/2}$ (because $\mathbf{E}[D_{t,i}^2] = Var(D_{t,i}) + \mathbf{E}[D_{t,i}]^2 = \mu^* + (\mu^*)^2 \leq 4(\mu^*)^2$ for all large μ^*), using the exponential tail bound proved in Step 2, we can bound $\mathbf{E}[(J^\pi - J^*)\,\mathbf{1}\{\hat{\mathcal{A}}^c\}] \leq 4m^2\mu^*(p^* + h^*)T^3 e^{-B^2/(32m^4\mu^*)} \leq 4m^2\mu^*(p^* + h^*)T^3 \;\; e^{-\mu^*/(32m^4)} \leq 4m^2(p^* + h^*)$ (because $T^3 = o(e^{\mu^*/(64m^4)})$). Putting these together with the bounds in Step 1, we conclude that, for all large μ^*,

$$\frac{\mathbf{E}[J^{RCE}] - J^*}{J^*} \;\leq\; \frac{2m(p^* + h^*)T\sqrt{\mu^*} + 4m^2(p^* + h^*)}{cT\mu^* + M'\sqrt{\mu^*}\, T^{3/2}} \;\leq\; \frac{M}{\sqrt{\mu^*} + \sqrt{T}}$$

for some $M > 0$. This completes the proof. ■

Proof of Lemma 1. We prove by induction. Let $F_{t,s}(\cdot)$ denote the cdf of $\sum_{\xi=t}^{s} D_{\xi,i}$. Define:

$$G_t(n_t; I_t) \;:=\; \sum_{\xi=t}^{T}\sum_{i=1}^{m}\mathbf{E}\left[p^*\left(\sum_{s=t}^{\xi} D_{s,i} - \sum_{s=t}^{\xi} n_{s,i} - I_{t,i}\right)^+ + h^*\left(I_{t,i} + \sum_{s=t}^{\xi} n_{s,i} - \sum_{s=t}^{\xi} D_{s,i}\right)^+\right].$$

The following expression is useful for the proof:

$$\frac{\partial G_t}{\partial n_{s,i}} \;=\; \sum_{k=s}^{T}\left[(h^*+p^*)F_{t,k}\left(\sum_{\xi=t}^{k} n_{\xi,i} + I_{t,i}\right) - p^*\right] \quad \text{for all } s \geq t. \tag{51}$$

We now proceed in two steps.

Step 1

In this step, we show that the result is true for $t = 1$. (This is our base case.) Consider $J_1^S(0)$. By (51) and KKT conditions, there exists dual variables $v_{s,i}^1 \geq 0$ and w_s^1 corresponding to constraints $n_{s,i} \geq 0$ and $\sum_{i=1}^{m} n_{n_{s,i}} = z_s^D B$, respectively, such that, for all i, we have:

$$(h^*+p^*)F_{1,T}\left(\sum_{\xi=1}^{T} n_{\xi,i}\right) - p^* \;=\; v_{T,i}^1 + w_T^1 \tag{52}$$

$$(h^*+p^*)F_{1,T-1}\left(\sum_{\xi=1}^{T-1} n_{\xi,i}\right) - p^* \;=\; v_{T-1,i}^1 + w_{T-1}^1 - (v_{T,i}^1 + w_T^1) \tag{53}$$

$$\vdots$$

$$(h^*+p^*)F_{1,1}\left(n_{1,i}\right) - p^* \;=\; v_{1,i}^1 + w_1^1 - \sum_{s=2}^{T}(v_{s,i}^1 + w_s^1) \tag{54}$$

$$v_{s,i}^1 \cdot n_{s,i} \;=\; 0 \quad \forall\, s \tag{55}$$

Let $\theta_{T,i}^1 := v_{T,i}^1 + w_T^1$ and $\theta_{s,i}^1 := v_{s,i}^1 + w_s^1 - \sum_{\xi=s+1}^{T}(v_{\xi,i}^1 + w_\xi^1)$ for $s \leq T-1$. We claim that $n_{s,i} = z_s^D B/m$ is the unique optimal solution of $J_1^S(0)$. To prove this, note that, if we set $v_{s,i}^1 = 0$ for all s and i, the variable $\theta_{s,i}^1$ is independent of i. By abuse of notation, let $\theta_{s,i}^1 = \theta_s^1$ for all s and i. By (52), $\sum_{\xi=1}^{T} n_{\xi,i} = F_{1,T}^{-1}\left(\frac{p^*+\theta_T^1}{p^*+h^*}\right)$. Taking the sum over all i's gives $\sum_{\xi=1}^{T} z_\xi^D B = m\, F_{1,T}^{-1}\left(\frac{p^*+\theta_T^1}{p^*+h^*}\right)$. So,

$$\sum_{\xi=1}^{T} n_{\xi,i} = \sum_{\xi=1}^{T}\frac{z_\xi^D B}{m}.$$

Similarly, by (53) and (54), for $s \leq T-1$, we have:

$$\sum_{\xi=1}^{s} z_\xi^D B = m\, F_{1,s}^{-1}\left(\frac{p^*+\theta_s^1}{p^*+h^*}\right) \quad \text{and} \quad \sum_{\xi=1}^{s} n_{\xi,i} = \sum_{\xi=1}^{s}\frac{z_\xi^D B}{m}.$$

We conclude that

$$n_{s,i} = \frac{z_s^D B}{m} \quad \text{and} \quad \theta_s^1 = (p^* + h^*) F_{1,s} \left(\sum_{\xi=1}^{s} \frac{z_\xi^D B}{m} \right) - p^* \quad \text{for all } s,$$

from which the constants $w_1^1, w_2^1, ..., w_T^1$ can be calculated properly. (Since $p^* + \theta_s^1 > 0$, the term $F_{1,s}^{-1}\left(\frac{p^*+\theta_s^1}{p^*+h^*}\right)$ is well-defined.) We have just shown that there exists dual variables $v_{s,i}^1 \geq 0$ and w_s^1 that not only satisfy KKT conditions but also yield $n_{s,i} = z_s^D B/m$ for all s and i. Our result for $t=1$ follows by the sufficiency of KKT conditions for optimality in a strongly convex optimization.

Step 2

Now, suppose that the formula given in the lemma holds for all $t \leq t'$. We want to show that it also holds for $t = t'+1$. Note that, since the formula for n^{St} is exactly the constructed optimal solution in Theorem 4, by the same arguments as in the proof of Theorem 4, we have:

$$I_{t'+1,i} = \frac{1}{m} \sum_{s=1}^{t'} z_s^D B - t'\mu^* - \Delta_{t',i} - \frac{1}{m} \sum_{s=1}^{t'-1} \sum_{j=1}^{m} \Delta_{s,j}.$$

Consider $J_{t'+1}^S(I_{t'+1})$. By KKT conditions, there exists dual variables $v_{s,i}^{t'+1} \geq 0$ and $w_s^{t'+1}$ corresponding to constraints $n_{s,i} \geq 0$ and $\sum_{i=1}^m n_{n_{s,i}} = z_s^D B$, respectively, such that, for all i, we have:

$$(h^* + p^*) F_{t'+1,T} \left(I_{t'+1,i} + \sum_{\xi=t'+1}^{T} n_{\xi,i} \right) - p^* \;=\; v_{T,i}^{t'+1} + w_T^{t'+1} \tag{56}$$

$$(h^* + p^*) F_{t'+1,T-1} \left(I_{t'+1,i} + \sum_{\xi=t'+1}^{T-1} n_{\xi,i} \right) - p^* \;=\; v_{T-1,i}^{t'+1} + w_{T-1}^{t'+1} - (v_{T,i}^{t'+1} + w_T^{t'+1}) \tag{57}$$

$$\vdots$$

$$(h^* + p^*) F_{t'+1,t'+1} \left(I_{t'+1,i} + n_{t'+1,i} \right) - p^* \;=\; v_{t'+1,i}^{t'+1} + w_{t'+1}^{t'+1} - \sum_{s=t'+2}^{T} (v_{s,i}^{t'+1} + w_s^{t'+1}) \tag{58}$$

$$v_{s,i}^{t'+1} \cdot n_{s,i} \;=\; 0 \;\; \forall s \tag{59}$$

Arguing as in Step 1, setting $v_{s,i}^{t'+1} = 0$ for all $s \geq t'+1$ and i, yields:

$$I_{t'+1,i} + \sum_{\xi=t'+1}^{s} n_{\xi,i} \;=\; \sum_{j=1}^{m} \frac{I_{t'+1,j}}{m} + \sum_{\xi=t'+1}^{s} \frac{z_\xi^D B}{m} \quad \text{for all } s \geq t'+1 \text{ and } i.$$

Simple algebra gives:

$$n_{t'+1,i} = \frac{z_{t'+1}^D B}{m} - I_{t'+1,i} + \sum_{j=1}^{m} \frac{I_{t'+1,j}}{m} = \frac{z_{t'+1}^D B}{m} + \Delta_{t',i} - \frac{1}{m} \sum_{j=1}^{m} \Delta_{t',j} \quad \text{and}$$

$$n_{s,i} = \frac{z_s^D B}{m} \quad \text{for all } s > t' + 1.$$

This completes the induction. ■

Proof of Theorem 5. The proof is similar to the proof of Theorem 1. Define $W^\pi_{t+1,i} = W^\pi_{t,i} + N^\pi_{t,i} - \hat{D}_{t,i}$, where $W^\pi_1 = I^\pi_1 = I_1$. Observe that we can write: $W^\pi_{t,i} = I_{1,i} + \sum_{s=1}^{t-1} N^\pi_{s,i} - \sum_{s=1}^{t-1} \hat{D}_{s,i}$ and $I^\pi_{t,i} = I_{1,i} + \sum_{s=1}^{t-1} N^\pi_{s,i} - \sum_{s=1}^{t-1} D_{s,i}$. So, $I^\pi_{t,i} = W^\pi_{t,i} - \sum_{s=1}^{t-1} \tilde{\Delta}_{s,i}$, where $\tilde{\Delta}_{s,i} = D_{s,i} - \hat{D}_{s,i}$. We now proceed in three steps.

Step 1

We first compute an upper bound for $\hat{C}^* - C^*$. We claim that

$$C^* \geq \hat{C}^* - \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i + h_i)\left(\sum_{s=1}^{t}\tilde{\Delta}_{s,i}\right)^+\right].$$

This is not difficult to show. For any policy $\pi \in \Pi$, we can bound:

$$\begin{aligned}
&\sum_{t=1}^{T}\left[c\, Z^\pi_t\, B + \sum_{i=1}^{m} p_i (D_{t,i} - N^\pi_{t,i} - I^\pi_{t,i})^+ + \sum_{i=1}^{m} h_i (N^\pi_{t,i} + I^\pi_{t,i} - D_{t,i})^+\right] \\
&\quad = \sum_{t=1}^{T}\left[c\, Z^\pi_t\, B + \sum_{i=1}^{m}(p_i + h_i)(I^\pi_{t,i} + N^\pi_{t,i} - D_{t,i})^+ - \sum_{i}^{m} p_i (I^\pi_{t,i} + N^\pi_{t,i} - D_{t,i})\right] \\
&\quad = \sum_{t=1}^{T}\left[c\, Z^\pi_t\, B + \sum_{i=1}^{m}(p_i + h_i)(I^\pi_{t+1,i})^+ - \sum_{i=1}^{m} p_i I^\pi_{t+1,i}\right] \\
&\quad \geq \sum_{t=1}^{T}\left[c\, Z^\pi_t\, B + \sum_{i=1}^{m}(p_i + h_i)(W^\pi_{t+1,i})^+ - \sum_{i=1}^{m} p_i W^\pi_{t+1,i}\right] \\
&\qquad\quad - \sum_{t=1}^{T}\sum_{i=1}^{m}(p_i + h_i)\left(\sum_{s=1}^{t}\tilde{\Delta}_{s,i}\right)^+ + \sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{s=1}^{t} p_i \tilde{\Delta}_{s,i},
\end{aligned}$$

where the first inequality holds because the identity $I^\pi_{t,i} = W^\pi_{t,i} - \sum_{s=1}^{t-1}\Delta_{s,i}$ implies $(I^\pi_{t,i})^+ \geq (W^\pi_{t,i})^+ - (\sum_{s=1}^{t-1}\Delta_{s,i})^+$. Taking expectation on both sides, minimizing the sum in the right side of the inequality over $\pi \in \Pi$ followed by minimizing the sum in the left side of the inequality yields the result.

Step 2

We now compute an upper bound for $\mathbf{E}\left[C^{\pi^R} - \hat{C}^*\right]$. We claim that

$$\mathbf{E}\left[C^{\pi^R}\right] - \hat{C}^* \le \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\left(-\sum_{s=1}^{t}\tilde{\Delta}_{s,i}\right)^+\right].$$

This can be shown using similar arguments as in Step 1. Let $I_{t+1,i} = I_{t,i} + n_{t,i}^{\pi^R} - D_{t,i}$ and $x_{t+1,i} = x_{t,i} + n_{t,i}^{\pi^R} - \tilde{D}_{t,i}$, with $x_1 = I_1$. Since $I_{t,i} = x_{t,i} - \sum_{s=1}^{t-1}\tilde{\Delta}_{s,i}$, we can bound:

$$\begin{aligned}
\mathbf{E}\left[C^{\pi^R}\right] &= \sum_{t=1}^{T}\mathbf{E}\left[c\, z_t^{\pi^R} B + \sum_{i=1}^{m} h_i (I_{t+1,i})^+ + \sum_{i}^{m} p_i(-I_{t+1,i})^+\right] \\
&= \sum_{t=1}^{T}\mathbf{E}\left[c\, z_t^{\pi^R} B + \sum_{i=1}^{m}(p_i+h_i)(I_{t+1,i})^+ - \sum_{i}^{m} p_i I_{t+1,i}\right] \\
&\le \sum_{t=1}^{T}\mathbf{E}\left[c\, z_t^{\pi^R} B + \sum_{i=1}^{m}(p_i+h_i)(x_{t+1,i})^+ - \sum_{i}^{m} p_i x_{t+1,i}\right] \\
&\qquad + \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\left(-\sum_{s=1}^{t}\tilde{\Delta}_{s,i}\right)^+\right] + \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}\sum_{s=1}^{t} p_i\tilde{\Delta}_{s,i}\right] \\
&= \hat{C}^* + \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\left(-\sum_{s=1}^{t}\tilde{\Delta}_{s,i}\right)^+\right].
\end{aligned}$$

The inequality follows because $z^{\pi^R} = z^{\hat{\pi}^*}$ and $I_{t,i} = x_{t,i} - \sum_{s=1}^{t-1}\tilde{\Delta}_{s,i}$ implies $(I_{t,i})^+ \le (x_{t,i})^+ + (-\sum_{s=1}^{t-1}\tilde{\Delta}_{s,i})^+$.

Step 3

Putting the bounds from Steps 1 and 2 together, we conclude that

$$\begin{aligned}
\mathbf{E}\left[C^{\pi^R}\right] - C^* &= \mathbf{E}\left[C^{\pi^R}\right] - \hat{C}^* + \hat{C}^* - C^* \\
&\le \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\left(-\sum_{s=1}^{t}\tilde{\Delta}_{s,i}\right)^+\right] + \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\left(\sum_{s=1}^{t}\tilde{\Delta}_{s,i}\right)^+\right] \\
&\le 2\sum_{t=1}^{T}\sum_{i=1}^{m}(p_i+h_i)\mathbf{E}\left[\left(\sum_{s=1}^{t}\tilde{\Delta}_{s,i}\right)^2\right]^{1/2} \\
&\le 2\sum_{i=1}^{m}(p_i+h_i)\left[\sum_{t=1}^{T}\left(\sum_{s=1}^{t}\theta_{s,i}^2\right)^{1/2}\right].
\end{aligned}$$

This completes the proof of Theorem 5. ■

Proof of Theorem 6. Define $W_{t+1,i}^{\pi} = W_{t,i}^{\pi} + N_{t,i}^{\pi} - \mu_i$, where $W_1^{\pi} = I_1^{\pi} = I_1$. Also, define $e_{t+1,i} = (e_{t,i} + \Delta_{t,i})^+$ and $v_{t+1,i} = (v_{t,i} - \Delta_{t,i})^+$, where $e_1 = v_1 = 0$.

Step 1

We first show that $I^\pi_{t,i} \geq W^\pi_{t,i} - e_{t,i}$ for all t and i. This can be proved by induction. The inequality obviously holds for $t = 1$. Now, suppose that $I^\pi_{t,i} \geq W^\pi_{t,i} - e_{t,i}$ for some $t > 1$, we want to show that it also holds for $t+1$. But, $W^\pi_{t+1,i} = W^\pi_{t,i} + N^\pi_{t,i} - \mu_i \leq I^\pi_{t,i} + N^\pi_{t,i} - \mu_i + e_{t,i} = I^\pi_{t,i} + N^\pi_{t,i} - D_{t,i} + e_{t,i} + \Delta_{t,i} \leq (I^\pi_{t,i} + N^\pi_{t,i} - D_{t,i})^+ + (e_{t,i} + \Delta_{t,i})^+ = I^\pi_{t+1,i} + e_{t+1,i}$. This completes the induction.

We claim that

$$\tilde{C}^* \geq \tilde{C}^D - \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m} h_i\, e_{t+1,i} + \sum_{i=1}^{m} p_i\, e_{T+1,i}\right].$$

Let Π be the set of non-anticipating policies. For any policy $\pi \in \Pi$, we can bound:

$$\begin{aligned}
&\sum_{t=1}^{T}\left[c\, Z^\pi_t\, B + \sum_{i=1}^{m} p_i (D_{t,i} - N^\pi_{t,i} - I^\pi_{t,i})^+ + \sum_{i=1}^{m} h_i (N^\pi_{t,i} + I^\pi_{t,i} - D_{t,i})^+\right] \\
&\quad= \sum_{t=1}^{T}\left[c\, Z^\pi_t\, B + \sum_{i=1}^{m} (p_i + h_i)(I^\pi_{t,i} + N^\pi_{t,i} - D_{t,i})^+ - \sum_{i} p_i (I^\pi_{t,i} + N^\pi_{t,i} - D_{t,i})\right] \\
&\quad= \sum_{t=1}^{T}\left[c\, Z^\pi_t\, B + \sum_{i=1}^{m} (p_i + h_i) I^\pi_{t+1,i} - \sum_{i=1}^{m} p_i (I^\pi_{t,i} + N^\pi_{t,i} - D_{t,i})\right] \\
&\quad= \sum_{t=1}^{T} c\, Z^\pi_t\, B + \sum_{t=1}^{T}\sum_{i=1}^{m} h_i I^\pi_{t+1,i} + \sum_{i=1}^{m} p_i (I^\pi_{T+1,i} - I^\pi_{1,i}) - \sum_{t=1}^{T}\sum_{i=1}^{m} p_i (N^\pi_{t,i} - D_{t,i}) \\
&\quad\geq \sum_{t=1}^{T} c\, Z^\pi_t\, B + \sum_{t=1}^{T}\sum_{i=1}^{m} h_i W^\pi_{t+1,i} + \sum_{i=1}^{m} p_i (W^\pi_{T+1,i} - W^\pi_{1,i}) - \sum_{t=1}^{T}\sum_{i=1}^{m} p_i (N^\pi_{t,i} - D_{t,i}) \\
&\qquad - \sum_{t=1}^{T}\sum_{i=1}^{m} h_i\, e_{t+1,i} - \sum_{i=1}^{m} p_i\, e_{T+1,i} \\
&\quad= \sum_{t=1}^{T}\left[c\, Z^\pi_t\, B + \sum_{i=1}^{m} p_i (\mu_i - N^\pi_{t,i} - W^\pi_{t,i})^+ + \sum_{i=1}^{m} h_i (N^\pi_{t,i} + W^\pi_{t,i} - \mu_i)^+\right] \\
&\qquad - \sum_{t=1}^{T}\sum_{i=1}^{m} h_i\, e_{t+1,i} - \sum_{i=1}^{m} p_i\, e_{T+1,i} + \sum_{t=1}^{T}\sum_{i=1}^{m} p_i \Delta_{t,i} \\
&\quad\geq \tilde{C}^D - \sum_{t=1}^{T}\sum_{i=1}^{m} h_i\, e_{t+1,i} - \sum_{i=1}^{m} p_i\, e_{T+1,i} + \sum_{t=1}^{T}\sum_{i=1}^{m} p_i \Delta_{t,i},
\end{aligned}$$

where the last inequality follows by definition of $\tilde{C}^D$. Taking expectation on both sides and minimizing the sum in the left side of the inequality over $\pi \in \Pi$ yields the result.

Step 2

Let $I_{t+1,i} = (I_{t,i} + \tilde{n}^D_{t,i} - D_{t,i})^+$ and $x_{t+1,i} = (x_{t,i} + \tilde{n}^D_{t,i} - \mu_i)^+$ (with $x_1 = I_1$). Note that $I_{t,i} \le x_{t,i} + v_{t,i}$ for all t and i. This can be proved by induction. The inequality obviously holds at $t = 1$. Now, suppose that $I_{t,i} \le x_{t,i} + v_{t,i}$ for some $t > 1$. Then, $I_{t+1,i} = (I_{t,i} + \tilde{n}^D_{t,i} - D_{t,i})^+ \le (x_{t,i} + v_{t,i} + \tilde{n}^D_{t,i} - D_{t,i})^+ \le x_{t+1,i} + v_{t+1,i}$. This completes the inducion.

We now compute an upper bound for $\mathbf{E}[\tilde{C}^{CE} - \tilde{C}^D]$. We claim that

$$\mathbf{E}[\tilde{C}^{CE}] - \tilde{C}^D \le \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m} h_i v_{t+1,i} + \sum_{i=1}^{m} p_i v_{T+1,i}\right].$$

This can be shown using similar arguments as in Step 1. Observe that

$$\begin{aligned}
\mathbf{E}[\tilde{C}^{CE}] &= \sum_{t=1}^{T} \mathbf{E}\left[c\,\tilde{z}^D_t\,B + \sum_{i=1}^{m} h_i(I_{t,i} + \tilde{n}^D_{t,i} - D_{t,i})^+ + \sum_{i=1}^{m} p_i(D_{t,i} - I_{t,i} - \tilde{n}^D_{t,i})^+\right] \\
&= \sum_{t=1}^{T} \mathbf{E}\left[c\,\tilde{z}^D_t\,B + \sum_{i=1}^{m}(p_i + h_i)I_{t+1,i} - \sum_{i=1}^{m} p_i(I_{t,i} + \tilde{n}^D_{t,i} - D_{t,i})\right] \\
&= \mathbf{E}\left[\sum_{t=1}^{T} c\,\tilde{z}^D_t\,B + \sum_{t=1}^{T}\sum_{i=1}^{m} h_i I_{t+1,i} + \sum_{i=1}^{m} p_i(I_{T+1,i} - I_{1,i}) - \sum_{t=1}^{T}\sum_{i=1}^{m} p_i(\tilde{n}^D_{t,i} - D_{t,i})\right] \\
&\le \mathbf{E}\left[\sum_{t=1}^{T} c\,\tilde{z}^D_t\,B + \sum_{t=1}^{T}\sum_{i=1}^{m} h_i x_{t+1,i} + \sum_{i=1}^{m} p_i(x_{T+1,i} - x_{1,i}) - \sum_{t=1}^{T}\sum_{i=1}^{m} p_i(\tilde{n}^D_{t,i} - \mu_i)\right] \\
&\qquad + \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m} h_i v_{t+1,i} + \sum_{i=1}^{m} p_i v_{T+1,i}\right] \\
&= \sum_{t=1}^{T} \mathbf{E}\left[c\,\tilde{z}^D_t\,B + \sum_{i=1}^{m} h_i(x_{t,i} + \tilde{n}^D_{t,i} - \mu_i)^+ + \sum_{i}^{m} p_i(\mu_i - x_{t,i} - \tilde{n}^D_{t,i})^+\right] \\
&\qquad + \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m} h_i v_{t+1,i} + \sum_{i=1}^{m} p_i v_{T+1,i}\right] \\
&= \tilde{C}^D + \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m} h_i v_{t+1,i} + \sum_{i=1}^{m} p_i v_{T+1,i}\right].
\end{aligned}$$

Step 3

Putting the bounds from Steps 1 and 2 together, we have:

$$\begin{aligned}
\mathbf{E}[\tilde{C}^{CE}] - \tilde{C}^* &= \mathbf{E}[\tilde{C}^{CE}] - \tilde{C}^D + \tilde{C}^D - \tilde{C}^* \\
&\le \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m} h_i v_{t+1,i} + \sum_{i=1}^{m} p_i v_{T+1,i}\right] + \mathbf{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{m} h_i\, e_{t+1,i} + \sum_{i=1}^{m} p_i\, e_{T+1,i}\right].
\end{aligned}$$

Now, since $\{\Delta_{t,i}\}$ forms a Martingale with respect to the natural filtration, by Doob's Maximal inequality, $\mathbf{E}[e_{t,i}] \le 2\sigma\sqrt{t}$ and $\mathbf{E}[v_{t,i}] \le 2\sigma\sqrt{t}$. Applying integral comparison with $\int \sqrt{t}\,dt$ yields the result. This completes the proof of Theorem 6. ■