Remanufacturing Inventory System with Demand-Dependent Returns: Optimality Analysis and Approximations

Chenxi Sun* Zhijie Tao† Xuefeng Gao ‡ Sean Zhou §

November 29, 2024

Abstract

We study an infinite-horizon periodic-review remanufacturing inventory system with random demand and product returns. The quantity of returned products each period depends on the historical demands following a distributed lag model. A firm operating the system remanufactures product returns into a serviceable product to fulfill customer demand. When needed, the serviceable product can also be manufactured/ordered. Manufacturing and remanufacturing have different lead times. The firm decides manufacturing quantity each period in order to minimize the expected long-run average cost of inventory holding, demand backlogging, and manufacturing. We first establish the existence of stationary optimal policy under the long-run average cost criterion using the vanishing discount factor approach together with a coupling argument. Via state space reduction, we further prove that the optimal policy is a so-called forecast-adjusted base-stock (FABS) policy when the maximum return lag is shorter than the manufacturing lead time. When the maximum return lag is longer than the manufacturing lead time, the optimal policy becomes state-dependent base-stock type. Due to the challenge of the computation of optimal solution and its implementation of the latter case, we adopt the FABS policy in general, provide an exact evaluation procedure via stationary analysis, and develop easily computable approximations for the optimal base-stock level. In a numerical study, we show that the FABS policy and the approximations are very effective; moreover, the FABS policy performs much better than a simple base-stock policy that does not incorporate forecast of product return, which demonstrate the value of return forecast.

Keywords: Remanufacturing, demand-dependent product return, optimal policy, approximations, long-run average cost

1 Introduction

Remanufacturing industry is an integral part of circular economy and is growing fast attributed to recent technical advances like additive manufacturing, data analytics, and internet of things as well

*CUHK Business School, The Chinese University of Hong Kong. E-mail: chenxi.sun@link.cuhk.edu.hk

†College of Business, Shanghai University of Finance and Economics. E-mail: tao.zhijie@mail.shufe.edu.cn

‡Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong. E-mail: xfgao@se.cuhk.edu.hk

§CUHK Business School, The Chinese University of Hong Kong. E-mail: zhoux@baf.cuhk.edu.hk

as increasing public awareness of sustainable development. Via a series of processes and new technologies, remanufacuring restores used or broken components/products returned by customers to like-new condition. The key industry sectors that practice remanufacturing include aerospace, automotive, heavy-duty and off-road equipment, and electronic and electrical equipment. In Europe, it is estimated that remanufacturing generates €30 billion annual turnover and employs roughly 190,000 people Parker et al. (2015). As noted in a recent New York Times article Rosen (2020), remanufacturing is different from refurbishing, a refurbished engine might be equivalent to one in excellent working condition but has already been in service for 30,000 miles whereas a remanufactured engine is equivalent to one that has not been in service. Customers would return their products because either the product is broken/replaced by a new one or the product is leased and it has to be returned at the end of the leasing period. By remanufacturing returned products, a company can save cost and energy of manufacturing from raw materials, generate additional revenue, and establish "environmental-friendly" images to the public.

One operational challenge of processing product returns faced by many remanufacturing programs is that the quantity and timing of returns are unknown and they are closely related to the historical demand of the product and how the product is used. A case studied in Bayiz and Tang (2004) consider a company selling thermoluminescent badges and find it useful to describe the quantity of returned badges as a function of previous demands. Clottey et al. (2012) report that product returns can be accurately estimated using certain statistical method based on historical demand information. The uncertainty associated with the quantity and the timing of product returns and their dependency on the historical demand cause great complexities in managing a remanufacturing inventory system. For example, a manufacturer of retreaded tyres, whose return depending on its sales (tyres can be retreaded up to 10 times). Ink/Toner cartridge manufacturers sell/remanufacture/resell their products, and the returns depend on the sales (toner cartridge can be remanufactured up to 4 times).

In this paper, we study a firm that collects returned products (also called cores) and remanufactures them into a serviceable product for meeting random customer demand over an infinite planning horizon. The quantity of returned product each period is random and dependent on the historical demand through the distributed lag model (Bayiz and Tang, 2004; Shang et al., 2020; Toktay et al., 2000). When there are no enough cores, the serviceable product can also be manufactured from raw materials (or ordered from external suppliers). For example, if the components/parts cannibalized from cores are insufficient, it is quite common in practice for a manufacturer, e.g., Caterpillar, to use new components/parts in remanufactured products[1]. Each period the firm receives and remanufactures all the cores while deciding manufacturing quantity if needed. Unused inventory at the end of each period incurs holding cost while excess demand is backlogged and incurs shortage cost. The firm aims to minimize the expected long-run average

[1] https://www.caterpillar.com/en/company/sustainability/remanufacturing.html

cost of the system.

The main contributions of the paper are summarized below.

First, we establish the existence of stationary optimal policy that minimizes the expected long-run average cost of the system using the vanishing discount factor approach together with a coupling argument. Different from the prior studies, the inventory dynamics and state transition are more complex in our system due to the consideration of correlated product demand and return. A direct formulation and analysis of the dynamic program will show that the optimal policy depends on a state variable that includes all the previous demands during the maximum lag of return (the maximum number of periods the product spends with customers before being returned), which largely hinders the implementation of the policy. Via a state transformation, we are able to reduce the dimensionality of the state space and provide a sharper characterization of the optimal manufacturing policy: When the maximum lag of return is shorter than the manufacturing lead time, the optimal policy is forecast-adjusted base-stock (FABS)–if the inventory position after remanufacturing the returned products and adjusted by return forecast at the beginning of each period is lower than a *constant* base-stock level, manufacture up to the base-stock level; otherwise, manufacture nothing. When the maximum return lag is longer than the manufacturing lead time, the optimal policy becomes state-dependent base-stock type, depending on the demands realized in the previous periods (from the most recent period backward covering the disparity between the maximum return lag and the manufacturing lead time).

Second, as the optimal policy may still be state-dependent in general, which is difficult to compute and implement, to simplify the control of a general remanufacturing system, we focus on the FABS policy and develop simple approximations for the optimal base-stock level. To this end, we conduct stationary analysis of the long-run average cost under a FABS policy, which involves the analysis of the overshooting process as the inventory position may exceed the base-stock level due to product returns, and employ heave-traffic analysis to obtain closed-form approximations of the base-stock level for the FABS policy. Numerically we show that the FABS policy performs close to the optimal one when the maximum return lag is longer than the manufacturing lead time. Moreover, the approximation of the base-stock level is effective, i.e., the resulting cost is close to the one under the optimal FABS policy. We further numerically demonstrate the superior performance of the FABS policy by comparing it with a simple base stock policy that does not incorporate forecast of product return.

1.1 Literature Review

This paper contributes to the literature of managing inventory systems with product returns, which dates back to Simpson (1978) who studies the optimal policy for a periodic-review inventory model with a single type of return. Several papers have extended Simpson (1978) in various ways. Decroix (2006) considers remanufacturing in a serial inventory system. Zhou et al. (2011) characterize optimal remanufacturing/manufacturing policies for systems with multiple types of returns. Tao et al. (2012) extend the model in Zhou et al. (2011) with random remanufacturing yield. Gong and Chao (2013) and Gong and Liu (2024) consider capacity constraints. When manufacturing and remanufacturing have different lead times, Xin and Yang (2022) show that a constant-order two-threshold policy is asymptotically optimal as the manufacturing lead time grows large. These papers focus on characterizing the optimal or asymptotically optimal policies under a major assumption that product returns and demands across different periods are independent.

The work on inventory systems with stochastic and correlated demands and returns is rather limited due to its complexity. Motivated by the supply chain of Kodak's single-use camera, Toktay et al. (2000) use a closed queueing network model to investigate inventory policies that minimize the procurement, inventory holding, and lost sales costs of a remanufacturing system. Kiesmüller and Van Der Laan (2001) adopt a base-stock policy for a remanufacturing inventory system with product returns that depend on the historical demand. They show that the system behaves quite differently from the system that the returns and demand are independent and neglecting such dependency may result in poor performance. Tao and Zhou (2014) introduce a balancing policy for a finite-horizon remanufacturing system that allows return to depend on the past demands and show that the expected cost under this balancing policy is at most twice of the optimal one. Miao (2023) and Bu et al. (2023) both consider return processes that are correlated with past sales. Miao (2023) studies a finite-horizon system with zero lead time, introducing a heuristic whose cost is shown to be at most four times of the optimal cost. Bu et al. (2023) study a lost-sales inventory system using two policies: a base-stock policy and a myopic policy. They demonstrate that both policies are asymptotically optimal as the unit penalty cost approaches infinity. Our work differs from these studies in that we characterize the optimal policy for the remanufacturing inventory system under the long-run average cost criterion and derive simple approximations for the optimal base-stock level of the FABS policy.

Our paper proves the existence of stationary optimal long-run average policy for a remanufacturing inventory system with correlated demands and returns. There is a stream of literature discussing the conditions for verifying the long-run average optimality of Markov decision processes (MDP), see, e.g., Schal (1993), Huh et al. (2011), Feinberg et al. (2012) and Feinberg and Lewis (2018). Feinberg and Lewis (2018) provide sufficient conditions to prove that the optimal policy for the discounted infinite-horizon problem converges to the optimal policy for the long-run average problem via the vanishing discount factor approach. Our proof validates the conditions in Feinberg and Lewis (2018) in our system, which is nontrivial given the high-dimensional state space in our problem with correlated returns over time. More recently, Avci et al. (2020) consider an inventory system with Markov-modulated demand by assuming current demand state is only partially observable and show the existence of stationary optimal policy under the long-run average cost. Similarly, Bai et al. (2023) analyze a partially observable lost-sales inventory system where the inventory level is only observed when it reaches zero, and they establish the long-run average optimality. Different from the papers above, we consider remanufacturing of product returns and so the inventory dynamics are more complex.

Our paper employs some results in heavy-traffic analysis to develop approximations for the optimal base-stock level of the FABS policy, so it is broadly related to the prior studies on heavy traffic approximation for production-inventory systems. For example, Toktay and Wein (2001) consider a production-inventory system where the demand process is governed by a Martingale Model of Forecast Evolution (MMFE) and they apply heavy traffic approximations to analyze the forecast-corrected base-stock policy. Plambeck and Ward (2006) study the optimal control of a high-volume assemble-to-order system and prove the asymptotic optimality of their proposed policy in the heavy-traffic regime. Ata and Barjesteh (2023) provide an approximate analysis of dynamic pricing, outsourcing, and scheduling policies for a multiclass make-to-stock manufacturing system in heavy traffic. Gao and Huang (2024) study a model slightly more general than that in Ata and Barjesteh (2023) and prove the asymptotic optimality of their proposed policy in the heavy-traffic regime.

The rest of the paper is organized as follows. In Section 2, we introduce the model. In Section 3, we conduct a state transformation that helps reduce the dimensionality of the state space. In Section 4, we characterize the optimal policy. In Section 5, we analyze the FABS policy and derive approximations for optimal base stock levels. In Section 6, we conduct a numerical study to demonstrate the effectiveness of the FABS policy and the approximations. Section 7 concludes. All the proofs are relegated to the Appendix. Throughout the paper, $y^+ = \max\{y, 0\}$, $y^- = \max\{-y, 0\}$, $y \wedge y' = \min\{y, y'\}$ and $y \vee y' = \max\{y, y'\}$ for any y, $y' \in \mathcal{R}$.

2 The Model

Consider a remanufacturer or a manufacturer's remanufacturing division (referred to as the firm hereafter) who receives product returns and remanufactures them to meet customer demand (e.g., orders from service requirement of certain equipment/product) over a planing horizon with infinite number of periods, indexed by $n = 1, 2 \ldots$. In each period, some customers who have bought/used the product earlier may return their products to the firm, either due to the end of use or end of the life cycle. The returned products are also referred to as the *cores*. The firm remanufactures the cores and puts them into inventory for serving future customer demand. If the firm anticipates the amount of remanufactured product is not enough, additional units can be manufactured (e.g., new parts, new cells for batteries) or ordered (e.g., from a core broker). We also refer to the remanufactured/manufactured product as the *serviceable product*. Demand for the serviceable product is assumed to be i.i.d. across different periods. Product returns are random in each period and depend on the previous demands (e.g., Goh and Varaprasad 1986, Kelle and Silver 1987, Toktay et al. 2000, Kiesmüller and Van Der Laan 2001, Clottey et al. 2012, Tao et al. 2012). Unused serviceable inventory in one period is carried over to the next while unsatisfied demand is backlogged. Inventory holding and demand backlogging incur costs. The firm aims to minimize the expected long-run average cost of the system.

The sequence of events within each period is as follows: First, at the beginning of period n, the firm receives the serviceable product manufactured in the previous periods. Next, the firm receives random returns R_n and remanufactures all of them into serviceable products, which is a common practice when dealing with high-value products or components. Without loss of generality, we assume the remanufacturing lead time, denoted as L_r, is zero. If not, we can redefine the return process as $R'_n = R_{n-L_r}$ for $L_r > 0$. Following this, the firm decides the quantity Q_n to manufacture from raw materials with unit cost k, subject to a positive manufacturing lead time, denoted as L. This reflects the practical scenario where manufacturing new products often takes longer than remanufacturing, especially when new parts must be sourced from external suppliers. Finally, random customer demand for the product D_n is realized and filled by the serviceable product inventory to the largest extent possible. The mean and variance of demand are $\mathbb{E}[D_n] = \lambda$ and $Var[D_n] = \sigma^2$, respectively, for $n = 1, 2, ...$ If the end-of-period inventory level I_n is greater than zero, the firm incurs a holding cost h per unit; otherwise the unmet demand is fully backlogged and the firm incurs a unit backlogging cost b.

For many remanufacturing programs, the product returns depend closely on the previous demands of the product. For example, the cores like the engines of Caterpillar or John Deere are those purchased/leased previously by customers. Hence, we assume R_n is correlated with $D_t, t = n-1, \ldots, n-\tau$:

$$R_n = p_1 D_{n-1} + p_2 D_{n-2} + \ldots + p_\tau D_{n-\tau} + \epsilon_n, \tag{1}$$

where we call τ the *maximum return lag* and $p_t, t = 1, \ldots, \tau$, represents the proportion of demand that will return t periods later. Here, p_t is usually called the *reaction coefficient*, ϵ_n is the noise term and we assume that $\{\epsilon_n\}$ is a sequence of i.i.d. random variables with mean zero and variance κ^2, which are independent of D_n for any n. We assume $\sum_{t=1}^{\tau} p_t < 1$, as the average return should not exceed the original demands and there is often attrition due to quality issues. The model (1) is also known as the *distributed lag model*, and there are various methods to estimate the value of $p_t, t = 1, \ldots, \tau$ through historical demand and return data, see, for example, Toktay et al. (2000) and Clottey et al. (2012). Such a return model is among the most popular ones in the literature on forecasting of product returns (Shang et al., 2020).

Let IP_n denote the serviceable product inventory position (= on hand inventory + pipeline inventory − demand backlog) after remanufacturing the cores received R_n but before manufacturing new product at period n. Then

$$IP_{n+1} = IP_n + Q_n - D_n + R_{n+1}. \tag{2}$$

Let I_n be the inventory level at the end of period n. Given a manufacturing lead time of L, we have,

$$I_{n+L} = IP_n + Q_n - D[n, n+L] + R[n+1, n+L], \tag{3}$$

where $D[n, n+L] := D_n + D_{n+1} + \cdots + D_{n+L}$ and $R[n+1, n+L] := R_{n+1} + R_{n+2} + \cdots + R_{n+L}$ represent the demand and product returns during the lead time, respectively. Let $\hat{\boldsymbol{D}}_n := (D_n, D_{n-1}, ..., D_{n-\tau+2})$, representing the previous $\tau - 1$ periods of demand including that in period n, where D_{n-i} is the demand at period $n-i$, for $i = 0, 1, ..., \tau - 2$. Define $g(y)$ as the total holding and backlogging costs when the inventory level at the end of a period is y, such that $g(y) = hy^+ + by^-$. For $n \geq L+1$, the system cost in period n consists of inventory holding and backlogging costs incurred in period n, and the manufacturing cost incurred in period $n-L$, i.e., $kQ_{n-L} + g(I_n)$. The firm aims to minimize the long-run expected average system cost.

Formally, this problem can be formulated as a discrete-time Markov decision process (MDP). The state of the MDP is represented by $(IP_n, \hat{\boldsymbol{D}}_{n-1})$, where IP_n is the inventory position following the dynamics described in equation (2), and $\hat{\boldsymbol{D}}_{n-1}$ is a vector containing $\tau - 1$ realized demands before period n, i.e., $\hat{\boldsymbol{D}}_{n-1} = (D_{n-1}, ..., D_{n-\tau+1})$. The action in period n is the manufacturing quantity $Q_n \geq 0$. For the state dynamics, we note that

$$\begin{aligned}
&IP_{n+1} = IP_n + Q_n - D_n + R_{n+1} = IP_n + \sum_{i=1}^{\tau-1} p_{i+1} D_{n-i} - (1-p_1) D_n + Q_n + \epsilon_{n+1}, \\
&\hat{\boldsymbol{D}}_n = (D_n, D_{n-1}, ..., D_{n-\tau+2}),
\end{aligned}$$

where $(D_{n-1}, ..., D_{n-\tau+2})$ is determined by $\hat{\boldsymbol{D}}_{n-1}$. Hence, one can spell out the transition probability function $\mathcal{q}((IP_{n+1}, \hat{\boldsymbol{D}}_n)|(IP_n, \hat{\boldsymbol{D}}_{n-1}), Q_n)$ given the distribution of D_n and ϵ_{n+1} (if necessary). Moreover, the single-period cost function $c(IP_n, \hat{\boldsymbol{D}}_{n-1}, Q_n)$ is given by,

$$c(IP_n, \hat{\boldsymbol{D}}_{n-1}, Q_n) = kQ_n + \mathbb{E}\left[g(I_{n+L}) | (IP_n, \hat{\boldsymbol{D}}_{n-1}, Q_n) \right]. \tag{4}$$

To facilitate our analysis later, we make the following assumption which states that the one-period demand and the noise in the return model are bounded random variables (which can be either discrete or continuous). A distinctive feature of our system is the inclusion of exogenous historical demands in the state space. The states across different time are correlated, and part of the state is independent of the manufacturing/ordering policy. Hence, we introduce the boundedness assumption for technical purposes. Similar assumptions have been made in the related literature.

For example, Beyer and Sethi (1997) consider a finite collection of possible demand states under the Markovian demands. Similarly, Huh et al. (2011) explore a finite-state Markov-modulated environment.

Assumption 1. *There exist constants $\underline{D}, \bar{D}, \underline{\epsilon}$ and $\bar{\epsilon}$, such that $D_1 \in [\underline{D}, \bar{D}]$ and $\epsilon_1 \in [\underline{\epsilon}, \bar{\epsilon}]$.*

By Assumption 1, we have the state space $\mathcal{S} = \mathbb{R} \times [\underline{D}, \bar{D}]^{\tau-1}$. The objective is to minimize the long-run average expected costs per unit time. Specifically, for a given initial state $(x, \hat{\boldsymbol{d}})$ where $\hat{\boldsymbol{d}} = (d_0, d_{-1}, ..., d_{2-\tau})$ and a feasible (stationary) policy π, let

$$\begin{aligned}\Phi^\pi(x, \hat{\boldsymbol{d}}) &= \liminf_{N\to\infty} \frac{1}{N}\mathbb{E}^\pi\Big[\sum_{n=1}^{N} c(IP_n, \hat{\boldsymbol{D}}_{n-1}, Q_n)\Big| IP_1 = x, \hat{\boldsymbol{D}}_0 = \hat{\boldsymbol{d}}\Big] \\ &= \liminf_{N\to\infty} \frac{1}{N}\mathbb{E}^\pi\left[\sum_{n=1}^{N}\Big(kQ_n + hI_{n+L}^+ + bI_{n+L}^-\Big)\Big| IP_1 = x, \hat{\boldsymbol{D}}_0 = \hat{\boldsymbol{d}}\right].\end{aligned} \tag{5}$$

The firm aims to find a (stationary) policy π^* such that it is average-cost optimal, that is,

$$\Phi^{\pi^*}(x, \hat{\boldsymbol{d}}) = \inf_\pi \Phi^\pi(x, \hat{\boldsymbol{d}}), \quad \text{for any } (x, \hat{\boldsymbol{d}}). \tag{6}$$

The average-cost MDP (6) described above features an uncountably infinite state space, a non-compact action space, and an unbounded one-period cost function. A priori, the existence of a stationary optimal policy is not clear, though there are general sufficient conditions (which can be nontrivial to verify) for the existence under certain assumptions (see, e.g. Feinberg and Lewis (2018)). In addition, the dimension of the state space is τ, which can be large in practice. Hence, even we assume the existence of stationary optimal policies, directly computing such policies based on the Bellman equation is challenging numerically due to the curse of dimensionality. In what follows, we address these issues. In particular, we will show the existence of stationary optimal policies and characterize the optimal policy via a state transformation. In the next section, we first analyze the associated infinite-horizon discounted problem and present an important result on the state space dimensionality reduction.

3 State Space Dimensionality Reduction

In this section, we first examine the structure of the optimal policy for the finite-horizon problem, and we show that by incorporating past demand information, the state space $\mathcal{S}$ with dimension τ can be transformed into a reduced-size state space of dimension $\max\{\tau - L, 1\}$. We then demonstrate that this result extends to the infinite-horizon discounted problem. Such results, combined with the vanishing discount factor approach, provide insights about the structure of the optimal policy for the average-cost MDP we consider.

For a finite-horizon N, let $V_{n,\alpha}(x, \hat{\boldsymbol{d}})$ be the optimal discounted costs from period n to period N, with discount factor $\alpha \in [0, 1)$. Specifically, for $(x, \hat{\boldsymbol{d}}) \in \mathcal{S}$,

$$V_{n,\alpha}(x, \hat{\boldsymbol{d}}) = \inf_{\pi} \ \mathbb{E}^{\pi} \left[\sum_{t=n}^{N} \alpha^{t-1} \ c(IP_t, \hat{\boldsymbol{D}}_{t-1}, Q_t) \ | IP_n = x, \hat{\boldsymbol{D}}_{n-1} = \hat{\boldsymbol{d}} \right], \tag{7}$$

and we set $V_{N+1,\alpha}(x, \hat{\boldsymbol{d}}) = 0$. We also define the infinite-horizon optimal discounted cost $V_{\alpha}(x, \hat{\boldsymbol{d}})$ as

$$V_{\alpha}(x, \hat{\boldsymbol{d}}) = \inf_{\pi} \ \mathbb{E}^{\pi} \left[\sum_{n=1}^{\infty} \alpha^{n-1} c(IP_n, \hat{\boldsymbol{D}}_{n-1}, Q_n) \Big| IP_1 = x, \hat{\boldsymbol{D}}_0 = \hat{\boldsymbol{d}} \right]. \tag{8}$$

We first show that the single-period cost function can be expressed in an alternative form. Define

$$r(\hat{\boldsymbol{d}}) := \sum_{j=1}^{\tau-1} \beta_j d_{1-j} + \sum_{j=1}^{L} (\gamma_j + 1) \mathbb{E}[D_{t+L-j}] = \sum_{j=1}^{\tau-1} \beta_j d_{1-j} + \sum_{j=1}^{L} (\gamma_j + 1) \lambda, \tag{9}$$

where $\beta_j = \sum_{i=j+1}^{\tau \wedge (j+L)} p_i$ and $\gamma_j = \sum_{i=1}^{\tau \wedge j} p_i - 1$. From the definition of $c(x, \hat{\boldsymbol{d}}, q)$ in (4), equation (3) and the definition of R_n in (1), we can immediately obtain the following result.

Lemma 1. *The single-period cost function can be rewritten as:*

$$c(x, \hat{\boldsymbol{d}}, q) = kq + \ell(x + q + r(\hat{\boldsymbol{d}})), \tag{10}$$

where

$$\ell(z) = \mathbb{E} \left[g \left(z + \sum_{j=1}^{L} \gamma_j D_{t+L-j} - D_{t+L} + \sum_{j=1}^{L} \epsilon_{t+j} - \sum_{j=1}^{L} (\gamma_j + 1) \lambda \right) \right]. \tag{11}$$

This simplification reveals that the impact of past demand information $\hat{\boldsymbol{d}}$ on the single-period cost function is completely captured by $r(\hat{\boldsymbol{d}})$, which is linear in $\hat{\boldsymbol{d}}$. Consequently, the optimality equation for the finite-horizon problem (7) can be written as, for $n = 1, 2, ..., N$,

$$V_{n,\alpha}(x, \hat{\boldsymbol{d}}) = \min_{y \geq x} \left\{ k(y - x) + \ell(y + r(\hat{\boldsymbol{d}})) + \alpha \mathbb{E} \left[V_{n+1,\alpha}(y - D_n + R_{n+1}, \hat{\boldsymbol{D}}_n) | \hat{\boldsymbol{D}}_{n-1} = \hat{\boldsymbol{d}} \right] \right\}. \tag{12}$$

One can interpret $r(\cdot)$ in (9) in the following way. Let $r_{n,n+i}$ be the forecast of product returns in period $n + i$ given the available information about the demand history $\mathcal{F}_{n-1}$ at the beginning of period n, that is,

$$r_{n,n+i} = \mathbb{E}[R_{n+i} | \mathcal{F}_{n-1}].$$

Let d_n denote the actual demand realization in period n. From the return-demand relationship in (1), we can obtain

$$r_{n,n+i} = \begin{cases} p_1\mathbb{E}[D_{n+i-1}] + \ldots + p_i\mathbb{E}[D_n] + p_{i+1}d_{n-1} + \ldots + p_\tau d_{n-\tau+i}, & 1 \le i < \tau, \\ p_1\mathbb{E}[D_{n+i-1}] + \ldots + p_\tau\mathbb{E}[D_{n-\tau+i}], & i \ge \tau, \end{cases} \tag{13}$$

One can easily see that

$$r(\hat{\boldsymbol{D}}_{n-1}) = \sum_{i=1}^{L} r_{n,n+i}.$$

Hence, $r(\hat{\boldsymbol{D}}_{n-1})$ can be interpreted as a forecasting model to predict total returns during the manufacturing lead time. Intuitively, if anticipating some returns in the future periods, particularly during the manufacturing lead time, the firm would adjust its manufacturing quantity to avoid unnecessary costs related to inventory holding and backlogging.

Next, we transform the state of the MDPs (with state space $\mathcal{S}$) based on $r(\cdot)$ in (9) to incorporate the information used in forecasting returns. Define

$$z_n = IP_n + r(\hat{\boldsymbol{D}}_{n-1}), \tag{14}$$

which we refer to as *the adjusted inventory position*. The adjusted inventory position z_n consists of the inventory position IP_n and the predicted returns during the manufacturing lead time. We have the following result.

Proposition 2. *The state space $\mathcal{S}$ can be simplified as follows:*

(1) If $L \ge \tau - 1$, the new state in period n is $z_n \in \mathbb{R}$.

(2) If $L < \tau - 1$, the new state in period n is $(z_n, \boldsymbol{D}_{n-1})$, where $\boldsymbol{D}_{n-1} = (D_{n-1}, ..., D_{n-\tau+L+1})$. The new state space $\mathbb{X} = \mathbb{R} \times [\underline{D}, \bar{D}]^{\tau-L-1}$.

Having transformed the states, we can now establish the equivalence between the original finite-horizon MDP (7) and a new MDP with transformed states. Specifically, if $L < \tau - 1$, we set $J_{N+1,\alpha}(z, \boldsymbol{d}) = 0$, where $\boldsymbol{d} = (d_0, ..., d_{2-\tau+L})$ and recursively define for $n = N, N-1, ..., 1$,

$$J_{n,\alpha}(z, \boldsymbol{d}) = \min_{w \ge z} \left\{ kw + \ell(w) + \alpha\mathbb{E}\Big[J_{n+1,\alpha}\left(w - \tilde{D}_n, \boldsymbol{D}_n\right) |\boldsymbol{D}_{n-1} = \boldsymbol{d}\Big] \right\} - kz, \tag{15}$$

where

$$\tilde{D}_n := r(\hat{\boldsymbol{D}}_{n-1}) + D_n - R_{n+1} - r(\hat{\boldsymbol{D}}_n). \tag{16}$$

Note that $J_{n,\alpha}(z, \boldsymbol{d})$ depends on N, and will be denoted as $J_{n,\alpha}^{(N)}(z, \boldsymbol{d})$ when needed. Similarly, if $L \geq \tau - 1$, we set $\hat{J}_{N+1,\alpha}(z) = 0$, and recursively define for $n = N, N-1, ..., 1$,

$$\hat{J}_{n,\alpha}(z) = \min_{w \geq z} \left\{ kw + \ell(w) + \alpha \mathbb{E}\left[\hat{J}_{n+1,\alpha}\left(w - \tilde{D}_n \right) \right] \right\} - kz.$$

Proposition 3 demonstrates the equivalence of optimal value functions between the original MDP and the new MDP with transformed states, under the finite-horizon setting.

Proposition 3. *For $n = 1, 2, ..., N+1$, if $L < \tau - 1$, $V_{n,\alpha}(x, \hat{\boldsymbol{d}}) = J_{n,\alpha}(x + r(\hat{\boldsymbol{d}}), \boldsymbol{d})$; otherwise, $V_{n,\alpha}(x, \hat{\boldsymbol{d}}) = \hat{J}_{n,\alpha}(x + r(\hat{\boldsymbol{d}}))$.*

The above result can be readily extended to the infinite-horizon discounted setting. Note that $J_{n,\alpha}^{(N)}(z, \boldsymbol{d})$ in (15) increases in N because the expected single-period cost is non-negative. Hence, we define $J_\alpha(z, \boldsymbol{d})$ and $\hat{J}_\alpha(z)$ as the limit of $J_{1,\alpha}^{(N)}(z, \boldsymbol{d})$ and $\hat{J}_{n,\alpha}(z)$, respectively, as N goes to infinity:

$$J_\alpha(z, \boldsymbol{d}) = \lim_{N \to \infty} J_{1,\alpha}^{(N)}(z, \boldsymbol{d}), \quad \hat{J}_\alpha(z) = \lim_{N \to \infty} \hat{J}_{1,\alpha}^{(N)}(z).$$

$J_\alpha(z, \boldsymbol{d})$ and $\hat{J}_\alpha(z)$ correspond to the optimal discounted costs of the infinite-horizon problem after the state transformation, in the cases where $L < \tau - 1$ or $L \geq \tau - 1$, respectively.

Proposition 4. *If $L < \tau - 1$, $V_\alpha(x, \hat{\boldsymbol{d}}) = J_\alpha(x + r(\hat{\boldsymbol{d}}), \boldsymbol{d})$; otherwsie, $V_\alpha(x, \hat{\boldsymbol{d}}) = \hat{J}_\alpha(x + r(\hat{\boldsymbol{d}}))$.*

Propositions 3 and 4 demonstrate that the information about past demands, captured by $r(\hat{\boldsymbol{d}})$, can be embedded into the adjusted inventory position through a simple transformation. This is a key insight that we will use when we analyze the average-cost MDP (6) via the vanishing discount factor approach in the next section.

4 Optimal Policies

In this section, we analyze the average-cost MDP (6) defined in Section 2. In particular, we establish the existence of stationary optimal policy and characterize its structure. To facilitate the presentation, we first introduce the definitions of two important policies, based on the new states introduced in Section 3.

Definition 1. *A FABS (Forecast-Adjusted Base-Stock) policy is defined as the policy that raises the adjusted inventory position z_n to a constant S_{F} in each period n.*

That is, the manufacturing quantity in period n under a FABS policy is given by

$$Q_n = \max\left\{ S_{\mathrm{F}} - r(\hat{\boldsymbol{D}}_{n-1}) - IP_n, 0 \right\}. \tag{17}$$

Definition 2. *A FABS-F policy is defined if $L < \tau - 1$ and it is given as follows: there exists a function $S^d_F(\cdot) : \mathbb{R}^{\tau-L-1} \to \mathbb{R}$ such that if the adjusted inventory position $z_n < S^d_F(\boldsymbol{D}_{n-1})$, then manufacture up to $S^d_F(\boldsymbol{D}_{n-1})$; otherwise, not manufacture any.*

That is, the manufacturing quantity in period n under a FABS-F policy is given by

$$Q_n = \max\left\{S^d_{\mathrm{F}}(\boldsymbol{D}_{n-1}) - r(\hat{\boldsymbol{D}}_{n-1}) - IP_n, 0\right\}.$$

In the following, with slight abuse of notation, we refer to the constant S_F in Definition 1 as a base stock level, and the function $S^d_F(\cdot)$ in Definition 2 as a state-dependent base stock level.

Now we analyze the optimal policy that minimizes the average cost $\Phi^\pi(x, \hat{\boldsymbol{d}})$ given in (5). The main result of this section is presented in the following theorem.

Theorem 5. *Suppose Assumption 1 holds.*

(i) There exists a stationary policy π^* that is average-cost optimal and the optimal average cost is independent of the initial state, i.e., $\Phi^{\pi^*}(x, \hat{\boldsymbol{d}}) = \inf_\pi \Phi^\pi(x, \hat{\boldsymbol{d}}) = \inf_{(x,\hat{\boldsymbol{d}})\in\mathcal{S}} \inf_\pi \Phi^\pi(x, \hat{\boldsymbol{d}})$.

(ii) If $L \geq \tau - 1$, then there exists a constant S^*_F such that the FABS policy with the base-stock level S^*_F is average-cost optimal.

(iii) If $L < \tau - 1$, then there exists a function $S^{d*}_F(\cdot) : [\underline{D}, \bar{D}]^{\tau-L-1} \to \mathbb{R}$ such that the FABS-F policy with the state-dependent base-stock level $S^{d*}_F(\cdot)$ is average-cost optimal. Moreover, $S^{d*}_F(\boldsymbol{d})$ is nonincreasing in d_i for $i = 0, -1, \ldots, L - \tau + 2$.

Theorem 5 (i) shows the existence of a stationary optimal policy for the average-cost MDP (6). This result is non-trivial, because problem (6) has an uncountable state space, a non-compact action space, and an unbounded single-period cost function. Theorem 5 (ii) and (iii) characterize the structure of the optimal policy. Under the original state space $\mathcal{S} = \mathbb{R} \times [\underline{D}, \bar{D}]^{\tau-1}$ for (6), one would expect that the optimal stationary policy is a mapping from $\mathcal{S}$ to $\mathcal{A} = \mathbb{R}_+$. If $L \geq \tau - 1$, Theorem 5 (ii) helps reduce the problem to finding an optimal base stock level (which is a scalar) instead of optimal functions on $\mathcal{S}$. Similarly, if $L < \tau - 1$, Theorem 5 (iii) reduces the problem to finding an optimal state-dependent base stock level/function on $[\underline{D}, \bar{D}]^{\tau-L-1}$. Such reductions are essentially based on the state transformation discussed in Section 3 for the finite-horizon and infinite-horizon discounted problems, and the vanishing discount factor method that connects the average-cost problem with the infinite-horizon discounted problem. The structural results obtained here not only provide insights about the optimal policy, but also facilitate computations due to the dimension reduction. We provide a sketch of the proof of Theorem 5 in the following, and relegate the detailed proof to Appendix B.

We first discuss how to prove Theorem 5 (ii) and (iii), because the proof strategy is conceptually easier to explain (but technically it is highly non-trivial). The key ingredient of the proof is the following result, which characterizes the structure of the optimal policy for the infinite-horizon discounted problem (8).

Proposition 6. *Suppose Assumption 1 holds. Consider the infinite-horizon discounted problem (8) with $\alpha \in [0, 1)$.*

(a) If $L \geq \tau - 1$, then there exists a constant $S^*_{F,\alpha}$ such that the FABS policy with the base-stock level $S^*_{F,\alpha}$ is optimal.

(b) If $L < \tau - 1$, then there exists a function $S^{d*}_{F,\alpha}(\cdot) : [\underline{D}, \bar{D}]^{\tau-L-1} \to \mathbb{R}$ such that the FABS-F policy with the state-dependent base-stock level $S^{d*}_{F,\alpha}(\cdot)$ is optimal. Moreover, $S^{d*}_{F,\alpha}(\boldsymbol{d})$ is nonincreasing in d_i for $i = 0, -1, \ldots, L - \tau + 2$.

We relegate the proof of Proposition 6 into the appendix. With Proposition 6, one can then prove Theorem 5 (ii) and (iii) by applying the results in Feinberg and Lewis (2018), which hold for MDPs with non-compact action set and unbounded single-period cost functions. In particular, Theorem 4.3 in Feinberg and Lewis (2018) establishes the convergence of optimal discount-cost actions to optimal average-cost actions for infinite-horizon problems, as the discount factor approaches 1. However, to apply Theorem 4.3 in Feinberg and Lewis (2018), it requires verify Assumption (W*) and Assumption (B) of Feinberg and Lewis (2018) hold for our system. If both assumptions hold, Theorem 5 (i) can also be readily proved by applying Theorem 4.1 in Feinberg and Lewis (2018) (or Theorem 4 in Feinberg et al. (2012)). While verifying Assumption (W*) is relatively straightforward, major difficulties lie in verifying Assumption (B) for our system. In particular, the key challenge is to show the uniform boundedness of the relative discounted cost functions, i.e., for any $(z, \boldsymbol{d}) \in \mathbb{X}$,

$$\sup_{\alpha \in [0,1)} \left[J_\alpha(z, \boldsymbol{d}) - \inf_{(z,\boldsymbol{d}) \in \mathbb{X}} J_\alpha(z, \boldsymbol{d}) \right] < \infty. \tag{18}$$

Different from the inventory system studied in Feinberg and Lewis (2018), the state in our model involves not only the inventory position but also historical demands which are not controllable, which causes difficulties in proving (18).

In the following, we outline the key idea in proving (18) when $L < \tau - 1$. The scenario where $L \geq \tau - 1$ is simpler to analyze (as the transformed state is of dimension one) and can be addressed in a similar manner.

To prove (18), by Proposition 4, it is equivalent to show for any given $(z, \boldsymbol{d}) \in \mathbb{X}$,

$$\sup_{\alpha \in [0,1)} u_\alpha(z, \boldsymbol{d}) < \infty, \tag{19}$$

where

$$u_\alpha(z,\boldsymbol{d}) = J_\alpha(z,\boldsymbol{d}) - m_\alpha, \quad \text{and} \quad m_\alpha = \inf_{(x,\hat{\boldsymbol{d}})\in\mathcal{S}} V_\alpha(x,\hat{\boldsymbol{d}}) = \inf_{(z,\boldsymbol{d})\in\mathbb{X}} J_\alpha(z,\boldsymbol{d}).$$

Denote by $z_{\alpha,\boldsymbol{d}}$ the optimal adjusted inventory position given the past demand vector $\boldsymbol{d}$, i.e., $J_\alpha(z_{\alpha,\boldsymbol{d}},\boldsymbol{d}) = \inf_{z\in\mathbb{R}} J_\alpha(z,\boldsymbol{d})$. The key idea of proving (19) is to construct a feasible policy, denoted by π^σ, such that the system with an initial state $(z,\boldsymbol{d})$ under the policy π^σ can be coupled with the system under the optimal (discounted) policy at some stopping time denoted by $\mathcal{N}(z,\boldsymbol{d})$ (which is designed to be greater than τ). In particular, under policy π^σ, the state $(z_{\alpha,\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}},\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1})$ becomes attainable with an initial state $(z,\boldsymbol{d})$) at time $\mathcal{N}(z,\boldsymbol{d})$, and policy π^σ then follows the optimal discounted policy from this period onward.

Using the optimalilty equation satisfied by J_α and the fact that $\mathcal{N}(z,\boldsymbol{d}) \geq \tau$, one can show that

$$m_\alpha \geq \mathbb{E}\left[\alpha^{\mathcal{N}(z,\boldsymbol{d})-1} \inf_{z\in\mathbb{R}} J_\alpha\Big(z,\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}\Big)\right] = \mathbb{E}\Big[\alpha^{\mathcal{N}(z,\boldsymbol{d})-1} J_\alpha\left(z_{\alpha,\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}},\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}\right)\Big].$$

In addition, because π^σ is a feasible policy, we obtain that $J_\alpha(z,\boldsymbol{d}) \leq J_\alpha^{\pi^\sigma}(z,\boldsymbol{d})$. This implies

$$\begin{aligned} u_\alpha(z,\boldsymbol{d}) &\leq J_\alpha^{\pi^\sigma}(z,\boldsymbol{d}) - \mathbb{E}\Big[\alpha^{\mathcal{N}(z,\boldsymbol{d})-1} J_\alpha\left(z_{\alpha,\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}},\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}\right)\Big], \\ &= \mathbb{E}^{\pi^\sigma}\left[\sum_{n=1}^{\mathcal{N}(z,\boldsymbol{d})-1} \alpha^{n-1}\ell(z_n) + \alpha^{\mathcal{N}(z,\boldsymbol{d})-1} k\Big(z_{\alpha,\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}} - z_{\mathcal{N}(z,\boldsymbol{d})}\Big)\Big| z_1 = z, \boldsymbol{D}_0 = \boldsymbol{d}\right], \end{aligned}$$

which amounts to the total discounted cost for the system under π^σ before the stopping time $\mathcal{N}(z,\boldsymbol{d})$. Using Assumption 1, we can then show it is uniformly bounded for all $\alpha \in [0,1)$, by bounding $z_{\alpha,\boldsymbol{d}}$ and the moments of $\mathcal{N}(z,\boldsymbol{d})$. This concludes the sketch of the proof.

We remark that our approach outlined above is significantly different from the proof of Proposition 6.3 in Feinberg and Lewis (2018), which verifies Assumption (W*) and Assumption (B) and establishes the optimality of (s,S) policies for inventory control problems with setup costs.

5 Stationary Analysis of FABS Policy and Approximations

Theorem 5 establishes the optimality of the FABS policy when $L \geq \tau - 1$, highlighting the importance of determining an appropriate base-stock level S_{F}. If $L < \tau - 1$, the optimal policy is no longer FABS, and computing the optimal state-dependent base-stock function $S_{\text{F}}^{d}(\cdot)$ is computationally challenging, especially when $\tau \gg L$. Hence, the firm might want to simply employ a FABS policy as a heuristic to manage the system. This requires identifying a "good" constant base-stock level S_{F} to approximate the state-dependent base-stock function $S_{\text{F}}^{d}(\cdot)$ when $L < \tau - 1$. In the following, we provide an exact procedure to evaluate the system's long-run average cost under the

FABS policy with a given base-stock level S_{F}. Moreover, we will develop simple approximations that allow for fast computation of a "good" base-stock level S_{F}.

5.1 The Overshoot Process

Recall that under a FABS policy, the manufacturing quantity is based on the adjusted inventory position and given by (17). Different from the classic inventory system without product return, under a base-stock policy, the inventory position before manufacturing IP_n in period n may not always be less than or equal to the base-stock level S_{F}. This is mainly due to remanufacturing of the cores. This causes some challenge to derive the steady state inventory position (we will show the existence later) and the resulting system cost. To facilitate our analysis, we introduce the overshoot process $\{O_n : n \geq 1\}$, which is defined by

$$O_n = \max\left[IP_n - (S_{\mathrm{F}} - \sum_{i=1}^{L} r_{n,n+i}), 0\right]. \tag{20}$$

Note that O_n will vanish in a classic inventory system without returns when following a stationary base-stock policy. Using (2) and the fact that $O_n - Q_n = IP_n - (S_{\mathrm{F}} - \sum_{i=1}^{L} r_{n,n+i})$, we can derive the dynamics of the overshoot process under a FABS policy as follows:

$$\begin{aligned} O_{n+1} &= \max\left[IP_n + Q_n - D_n + R_{n+1} - (S_{\mathrm{F}} - \sum_{i=1}^{L} r_{n+1,n+1+i}),\ 0\right] \\ &= \max\left[IP_n + Q_n - (S_{\mathrm{F}} - \sum_{i=1}^{L} r_{n,n+i}) - D_n + R_{n+1} + \sum_{i=1}^{L}(r_{n+1,n+1+i} - r_{n,n+i}),\ 0\right] \\ &= \max\left[O_n - D_n + R_{n+1} + \sum_{i=1}^{L}(r_{n+1,n+1+i} - r_{n,n+i}),\ 0\right]. \end{aligned} \tag{21}$$

By (1) and (13), it can be readily verified that

$$O_{n+1} = \max\{O_n + \theta D_n + \epsilon_{n+1},\ 0\}, \quad L \geq \tau - 1, \tag{22}$$

$$O_{n+1} = \max\{O_n + A_n,\ 0\}, \quad \text{otherwise.} \tag{23}$$

where $\theta := -1 + \sum_{i=1}^{\tau} p_i < 0$, and

$$A_n := -\left(1 - \sum_{i=1}^{L+1} p_i\right) D_n + p_{L+2} D_{n-1} + \ldots + p_\tau D_{n-\tau+L+1} + \epsilon_{n+1}. \tag{24}$$

Hence, the overshoot process $\{O_n : n \geq 1\}$ is a Lindley random walk, also known as a Lindley recursion, in the literature (Asmussen, 2003). When $L \geq \tau - 1$, the increments $\{\theta D_n + \epsilon_{n+1}\}$ are i.i.d., with $\mathbb{E}[\theta D_n + \epsilon_{n+1}] = \theta\lambda$ and $Var(\theta D_n + \epsilon_{n+1}) = \theta^2\sigma^2 + \kappa^2$. On the other hand, when

$L \geq \tau - 1$, the increments $\{A_n\}$ are correlated. Indeed, $\{A_n : n \geq 1\}$ is a strictly stationary sequence with finite-lag dependence, and can be extended to form a stationary processs $\{A_n : -\infty < n < \infty\}$ (Loynes, 1962). Note that $\mathbb{E}[A_n] = \theta\lambda$, and it is also straightforward to compute the asymptotic variance $\lim_{t\to\infty} \frac{1}{t} Var\left(\sum_{i=1}^{t} A_i\right) = \theta^2\sigma^2 + \kappa^2$. Hence, the mean and the asymptotic variance are identical to those of the i.i.d. sequence $\{\theta D_n + \epsilon_{n+1}\}$. This observation will be used later (see Lemma 9).

From the Lindley recursion for $\{O_n\}$, we immediately obtain the following result (see, e.g., (Asmussen, 2003, Chapter III.6) and Loynes (1962)).

Lemma 7. *When $\sum_{i=1}^{\tau} p_i < 1$, O_n converges in distribution to an almost surely finite random variable O_∞ as $n \to \infty$, where*

$$O_\infty \overset{d}{=} \begin{cases} \sup_{k\geq 0} \sum_{i=1}^{k} (\theta D_i + \epsilon_i), & \text{if } L \geq \tau - 1, \\ \sup_{k\geq 0} \sum_{i=1}^{k} A_{-i}, & \text{otherwise,} \end{cases} \tag{25}$$

and $\overset{d}{=}$ denotes equal in distribution.

5.2 Performance Evaluation of the FABS Policy

With the analysis of the overshoot process above, we next discuss the inventory process, which enables us to evaluate the long-run average cost of any given FABS policy. From (3) and (17), we obtain that the inventory level at the end of period $n + L$ is given by

$$\begin{aligned} I_{n+L} &= S_F - \sum_{i=1}^{L} r_{n,n+i} + O_n - \sum_{i=n}^{n+L} D_i + \sum_{i=n+1}^{n+L} R_i \\ &= S_F + O_n - D_{n+L} + \left(\sum_{i=1}^{L} [R_{n+i} - \mathbb{E}[R_{n+i}|\mathcal{F}_{n-1}]] - \sum_{i=0}^{L-1} D_{n+i} \right) \\ &= S_F + O_n - D_{n+L} + Y_{n,L}, \end{aligned} \tag{26}$$

where

$$Y_{n,L} := \left(\sum_{i=1}^{L} [R_{n+i} - \mathbb{E}[R_{n+i}|\mathcal{F}_{n-1}]] - \sum_{i=0}^{L-1} D_{n+i} \right). \tag{27}$$

Let $D_j^c := D_j - \lambda$. It is clear that for $i \geq \tau$, we have

$$R_{n+i} - \mathbb{E}[R_{n+i}|\mathcal{F}_{n-1}] = R_{n+i} - \lambda(1+\theta) = p_1 D_{n+i-1}^c + \ldots + p_\tau D_{n+i-\tau}^c + \epsilon_{n+i}, \tag{28}$$

and for $1 \leq i \leq \tau - 1$, we have

$$R_{n+i} - \mathbb{E}[R_{n+i}|\mathcal{F}_{n-1}] = \sum_{j=1}^{i} p_j D_{n+i-j}^c + \epsilon_{n+i}. \tag{29}$$

Therefore, $Y_{n,L}$ (which depends on demand D_t, $t = n, \ldots, n+L-1$) is independent of O_n and D_{n+L} for each fixed n, and its mean is $-L\lambda$. The variance of $Y_{n,L}$ is given by

$$L\kappa^2 + \left(\sum_{j=1}^{\tau-1} (\sum_{i=1}^{j} p_i - 1)^2 + (L - \tau + 1)(\sum_{i=1}^{\tau} p_i - 1)^2 \right) \sigma^2, \quad \text{if } L \geq \tau - 1, \tag{30}$$

and

$$L\kappa^2 + \left(\sum_{j=1}^{L} (\sum_{i=1}^{j} p_i - 1)^2 \right) \sigma^2, \quad \text{if } L < \tau - 1. \tag{31}$$

Note that the distribution of $-D_{n+L} + Y_{n,L}$ is independent of n. Hence, the sequence of random variables $\{-D_{n+L} + Y_{n,L} : n \geq 1\}$ converges in distribution to a random variable $-D + Y$ where D has the same distribution as demand D_1, and it is independent of Y which has the same distribution as $Y_{1,L}$. Since O_n and $-D_{n+L} + Y_{n,L}$ are independent for each fixed n and O_n converges in distribution to O_∞ when $\sum_{t=1}^{\tau} p_t < 1$, it follows that $\{I_{n+L} : n \geq 0\}$ converges in distribution to an almost surely finite random variable

$$I_\infty := S_F + O_\infty - D + Y, \tag{32}$$

where the three random variables O_∞, D and Y are independent. We summarize this result in the following lemma. Recall that we assume $\sum_{t=1}^{\tau} p_t < 1$.

Lemma 8. *The inventory level $\{I_{n+L} : n \geq 1\}$ under any given FABS policy with base-stock level S_F converges in distribution to the almost surely finite random variable I_∞ in (32).*

We can now evaluate the long-run average cost under the FABS policy with a given base stock level S_F. We denote this policy by π^{S_F}. By the definitions of O_n and Q_n, we have

$$O_{n+1} - Q_{n+1} = O_n + \sum_{i=1}^{L} (r_{n+1,n+1+i} - r_{n,n+i}) - D_n + R_{n+1}.$$

Hence, we obtain

$$\mathbb{E}^{\pi^{S_F}}[Q_{n+1}] = \mathbb{E}^{\pi^{S_F}}[O_{n+1} - O_n] + \mathbb{E}[D_n - R_{n+1}] = \mathbb{E}^{\pi^{S_F}}[O_{n+1} - O_n] + |\theta|\lambda.$$

By Assumption 1, we know that $\{O_n : n \geq 1\}$ is uniformly integrable (see e.g., p.270 of Asmussen (2003) and Loynes (1962)). Moreover, when $\theta < 0$, O_n converges in distribution to O_∞ as $n \to \infty$, regardless of the value of O_1 which depends on the initial state $(IP_1, \hat{\boldsymbol{D}}_0) = (x, \hat{\boldsymbol{d}})$. It follows that the long-run average manufacturing cost is given by

$$\lim_{N \to \infty} \frac{1}{N} \mathbb{E}^{\pi^{S_F}} \left[\sum_{n=1}^{N} kQ_n \Big| IP_1 = x, \hat{\boldsymbol{D}}_0 = \hat{\boldsymbol{d}} \right] = k|\theta|\lambda. \tag{33}$$

Moreover, by (26), $\{I_{n+L} : n \geq 1\}$ is also uniformly integrable. This implies

$$\lim_{N\to\infty} \frac{1}{N}\mathbb{E}^{\pi^{S_F}}\left[\sum_{n=1}^{N}\left(hI_{n+L}^{+} + bI_{n+L}^{-}\right)\Big| IP_1 = x, \hat{\boldsymbol{D}}_0 = \hat{\boldsymbol{d}}\right] = \mathbb{E}^{\pi^{S_F}}[hI_\infty^{+} + bI_\infty^{-}] < \infty.$$

Therefore, the average cost under the FABS policy π^{S_F} is simply given by

$$C(S_{\mathrm{F}}) := k|\theta|\lambda + h\mathbb{E}^{\pi^{S_F}}[I_\infty^{+}] + b\mathbb{E}^{\pi^{S_F}}[I_\infty^{-}], \tag{34}$$

where I_∞ is given by (32), and its distribution depends on S_F.

It is ideal if we could find S_{F}^* that minimizes $C(S_{\mathrm{F}})$. This, however, requires the knowledge of the distribution of I_∞ in (32), which in turn requires the knowledge of the distribution of O_∞. When $L \geq \tau - 1$, it is known that by (22), the distribution function of O_∞ satisfies the so-called Lindley integral equation, which is is difficult to solve in general, both analytically and numerically (see, e.g., Blanchet and Glynn (2006)). Moreover, when $L < \tau - 1$, it is even more challenging to compute the exact distribution of O_∞ due to the correlations in the sequence of $\{A_n\}$ in (24). As a result, we will instead employ approximations for the distribution of O_∞ so that the (approximate) optimal base-stock level S_{F}^* can be easily solved.

5.3 Approximations

To develop approximations for the distribution of O_∞, we consider the heavy traffic regime, i.e., θ approaches zero, or equivalently, $\sum_{i=1}^{\tau} p_i$ approaches one. This means that most of the previous demanded products will be eventually returned, which would occur in practice when the products are leased to customers or the firm provides strong incentives (ask the customers to pay some deposit and will give back the deposit upon return of the product). The heavy-traffic (or diffusion) approximation is well studied in the applied probability literature. Based on the discussion in Section 5.1, we immediately obtain the following result, see, e.g., Kingman (1961) and Jacobs (1980).

Lemma 9. *In the heavy traffic limit, O_∞ has an exponential distribution with parameter $2|\theta|\lambda/[\theta^2\sigma^2+\kappa^2]$. Specifically, $\frac{2|\theta|\lambda}{\theta^2\sigma^2+\kappa^2}\cdot O_\infty$ converges in distribution to an exponential random variable with rate 1 as $\theta := -1 + \sum_{i=1}^{\tau} p_i \to 0$.*

This result holds for both $L \geq \tau - 1$ and $L < \tau - 1$. Based on this result, we propose the following two approximations for O_∞. The first approximation is based on Kingman (1961).

Approximation 1: We approximate O_∞ by a random variable $\widetilde{O}_\infty$ whose distribution is given as follows:

$$\mathbb{P}(\widetilde{O}_\infty = 0) = 1 - \sum_{i=1}^{\tau} p_i = -\theta,$$

$$P(\widetilde{O}_\infty > x) = (1+\theta) \cdot e^{2\theta\lambda x/[\theta^2\sigma^2+\kappa^2]}, \quad \text{for } x \geq 0. \tag{35}$$

The second approximation is based on Siegmund (1979) for a corrected diffusion approximation of random walks, and is conceptually similar to the approximation of the work-in-process distribution in Toktay and Wein (2001) (see Equation (3) there).

Approximation 2: We approximate O_∞ by a random variable $\tilde{O}_\infty$ whose distribution is given as follows:

$$\begin{aligned} &\mathbb{P}(\tilde{O}_\infty = 0) = 1 - \exp(2\theta\lambda\beta/[\theta^2\sigma^2+\kappa^2]) \\ &\mathbb{P}(\tilde{O}_\infty > x) = e^{2\theta\lambda(x+\beta)/[\theta^2\sigma^2+\kappa^2]}, \quad \text{for } x \geq 0, \end{aligned} \tag{36}$$

where $\beta = 0.583\sqrt{\theta^2\sigma^2+\kappa^2}$. This value of β is given in Siegmund (1979) for the special case when the step size of the random walk is i.i.d. normal, and we use it in our approximation similarly as Toktay and Wein (2001).

While some refinements of Siegmund's corrected diffusion approximation has been studied in the literature (see e.g. Blanchet and Glynn (2006)), we will focus on the above two approximations due to their simplicity. With the preceding approximations of the distribution of O_∞, we can derive the distribution (approximately) of I_∞ based on (32). Then, under a given base-stock level S_{F}, we can evaluate the resulting long-run average cost $C(S_{\mathrm{F}})$ in (34). We next discuss how to find the near-optimal base-stock level $\widetilde{S}_F$ under this approximation.

Note that in order to find the optimal S_{F}^*, we need to solve the following problem in view of (34):

$$\min_{S_{\mathrm{F}}} \left\{ h\mathbb{E}[(S_{\mathrm{F}} + O_\infty + Y - D)^+] + b\mathbb{E}[(S_{\mathrm{F}} + O_\infty + Y - D)^-] \right\}.$$

This optimization problem is essentially a newsvendor problem. Applying the approximations for the distribution of O_∞, we immediately have the following result. Denote by F_X^{-1} the quantile function associated with a random variable X, which is defined as $F_X^{-1}(y) = \inf\{x \in \mathbb{R} : F_X(x) \geq y\}$ for $y \in \mathbb{R}$.

Proposition 10. *The optimal base stock level for the FABS Policy S_F^* can be approximated by*

$$\widetilde{S}_F = F^{-1}_{D-Y-\tilde{O}_\infty}\left(\frac{b}{b+h}\right),$$

where D, Y and $\tilde{O}_\infty$ are independent and the distribution of $\tilde{O}_\infty$ is given by either (35) or (36).

In the next section, we will numerically investigate the performance of FABS policy and the approximations developed in Proposition 10 across a wide range of system parameters.

6 Numerical Study

In this section, we first compare the performance of the FABS policy with that of the optimal policy for small-scale problems (Section 6.1) so that exact computations based on dynamic programming is viable. Then, we investigate the effectiveness of our approximations of the optimal base-stock level of the FABS policy (Section 6.2). We further compare the performance of the FABS policy with that of simple base stock policies (defined later) to illustrate the value of return forecast (Section 6.3). Finally, we briefly discuss the impact of misspecification of the return model (Section 6.4). All the experiments are programmed using MATLAB while running on a computer with Intel(R) Core(TM) i5-11500 CPU 2.71 GHz and 16 GB of RAM.

6.1 Performance of the FABS Policy

We compare the performance of the (optimal) FABS policy with that of the optimal policy. We focus on the case where $L < \tau - 1$, because we have shown in Theorem 5 that the FABS policy is optimal when $L \geq \tau - 1$. When $L < \tau - 1$, computing the optimal policy is challenging for high-dimensional problems (that is, when τ is much greater than L), so we focus on small-scale problems here. We measure the performance of the FABS policy using the relative optimality gap defined as

$$\frac{C(S_{\mathrm{F}}^{*}) - OPT}{OPT} \times 100\%,$$

where OPT denotes the long-run average cost of the optimal policy, and $C(S_{\mathrm{F}}^{*})$ denotes the average cost of the optimal FABS policy.

We use binomial distribution $B(10, 0.6)$ for demand D. We assume the noise ϵ_n in the return model has a simple two-point distribution given by $\mathbb{P}(\epsilon_n = -2) = \mathbb{P}(\epsilon_n = 2) = 0.5$. We set the unit manufacturing cost $k = 1.2$ and the unit holding cost $h = 1$. We consider three different scenarios for the return model: (1) $\tau = 3$ with $(p_1, p_2, p_3) = (0.1, 0.1, 0.75)$; (2) $\tau = 4$ with $(p_1, p_2, p_3, p_4) = (0.1, 0.1, 0.1, 0.65)$; and (3) $\tau = 5$ with $(p_1, p_2, p_3, p_4, p_5) = (0.1, 0.1, 0.1, 0.1, 0.55)$. Under each scenario, we choose different values of the unit backlogging cost b and manufacturing lead time L as shown in Table 1. To compute OPT under a given set of system parameters, we apply the relative value iteration algorithm (Bertsekas et al. (2011)). In addition, to compute $C(S_{\mathrm{F}}^{*})$, we simulate the system under a FABS policy with 100 independent replications and compute the average cost for a given S_F. We then find the optimal base stock level S_{F}^{*} via exhaustive search. Each replication consists of 10^5 periods, ensuring the stability and reliability of the results.

Table 1 reports the average costs of the FABS policy and the optimal policy, and the relative optimality gap. We can observe from Table 1 that the cost of FABS policy is very close to the optimal cost. In particular, the relative optimality gap is consistently below 2% across all the instances. We also test some additional small-scale problems under different demand distributions, and the relative optimality gap remains consistently low. One observation from this small scale numerical study is that the FABS would perform worse if $\tau - L$ becomes larger or the backlogging cost b is smaller (the service level is lower), in which case the exact state information is more valuable.

Table 1: Relative optimality gap of FABS policy

τ	L	b	optimal cost	FABS cost	optimality gap (%)
5	2	1.5	6.93	7.02	1.30
	3	1.5	7.26	7.32	0.83
	2	4.0	9.10	9.19	0.99
	3	4.0	9.66	9.72	0.62
4	1	1.5	6.51	6.62	1.66
	2	1.5	6.90	6.94	0.58
	1	4.0	8.38	8.50	1.43
	2	4.0	9.02	9.09	0.82
3	1	1.5	6.42	6.50	1.17
	1	4.0	8.25	8.34	1.05

We illustrate the optimal policy (computed via relative value iteration) when $L < \tau - 1$ in Table 2. Set $\tau = 4, L = 1, h = 1, b = 1.5$, and $(p_1, p_2, p_3, p_4) = (0.1, 0.1, 0.1, 0.65)$. Recall from Theorem 5 that the FABS-F policy with the state-dependent base-stock level $S_{\mathrm{F}}^{d*}(\boldsymbol{d})$ ($\boldsymbol{d} = (d_0, d_{-1})$) is average-cost optimal. Table 2 shows the value of $S_{\mathrm{F}}^{d*}(\boldsymbol{d})$. We can observe that the optimal state-dependent base-stock level $S_{\mathrm{F}}^{d*}(d_0, d_{-1})$ is not in a simple form such as a linear function of (d_0, d_{-1}). Moreover, $S_{\mathrm{F}}^{d*}(d_0, d_{-1})$ decreases as d_0 or d_{-1} increases. Nevertheless, the performance of the FABS policy with a constant base-stock level is remarkably good for small-scale problems, as evidenced in Table 1.

6.2 Effectiveness of Approximations for FABS Policy

To implement the optimal FABS policy, determining the optimal base-stock level S_{F}^* often involves significant search costs. In this section, we investigate the effectiveness of our approximations of S_{F}^* developed in Section 5.3 (see Proposition 10).

In view of Assumption 1, we use the truncated normal distribution for demand $D_n \in [\lambda - 3\sigma, \lambda + 3\sigma]$, with mean $\lambda = 5$, variance σ^2, i.e., $D_n \sim TN(5, \sigma^2, 5 - 3\sigma, 5 + 3\sigma)$. In addition, we assume the noise in the return model $\epsilon_n \sim TN(0, \kappa^2, -3\kappa, 3\kappa)$. We fix the unit holding cost $h = 1$, the unit manufacturing cost $k = 3$, the maximum return lag $\tau = 5$, and choose the values of the other parameters from Table 3. The selected parameters represent different levels of service requirement, manufacturing lead time, and return noise variability. They also capture different return rates

Table 2: Optimal state-dependent base-stock level $S_{\mathrm{F}}^{d*}(\boldsymbol{d})$

d_0 \ d_{-1}	0	1	2	3	4	5	6	7	8	9	10
0	12.90	12.90	12.88	12.62	12.18	11.88	11.48	11.08	10.78	10.22	9.88
1	12.90	12.90	12.64	12.16	11.98	11.58	11.10	10.84	10.18	9.88	9.36
2	12.90	12.72	12.28	11.98	11.68	11.24	10.88	10.24	9.88	9.38	8.98
3	12.78	12.26	12.08	11.72	11.22	10.88	10.28	9.98	9.46	9.02	8.48
4	12.38	12.02	11.78	11.22	10.88	10.38	9.98	9.38	8.98	8.38	7.98
5	12.08	11.74	11.22	10.88	10.42	10.08	9.48	9.08	8.46	7.98	7.48
6	11.88	11.24	10.98	10.40	9.98	9.58	9.08	8.54	7.98	7.38	6.78
7	11.28	10.88	10.38	10.02	9.58	9.08	8.48	8.08	7.38	6.72	6.18
8	10.98	10.38	9.98	9.58	9.08	8.58	7.98	7.40	6.78	6.22	5.58
9	10.48	10.08	9.48	9.08	8.48	8.04	7.38	6.74	6.18	5.58	4.88
10	9.98	9.52	8.98	8.38	7.98	7.38	6.78	6.24	5.58	4.90	4.28

Table 3: Set of Parameters

b	1.5 5 10
L	1 2 5
κ	0.5 1 2
σ	1 2
$(p_1, p_2, p_3, p_4, p_5)$	(0.10, 0.10, 0.10, 0.10, 0.10) (0.14, 0.14, 0.14, 0.14, 0.14) (0.19, 0.19, 0.19, 0.19, 0.19) (0.45, 0.00, 0.00, 0.00, 0.45) (0.48, 0.00, 0.00, 0.00, 0.48)

as well as return patterns, for example, $(p_1, p_2, p_3, p_4, p_5) = (0.14, 0.14, 0.14, 0.14, 0.14)$ means for demand in any period, 14% of such demand will return in each of the subsequent 5 periods and the total return rate is 70%; $(p_1, p_2, p_3, p_4, p_5) = (0.45, 0.00, 0.00, 0.00, 0.45)$ means for demand in any period, 45% of such demand will return in the immediate next period while another 45% will return 5 periods later and the total return rate is 90%. There are $3 \times 3 \times 3 \times 2 \times 5 = 270$ problem instances in total.

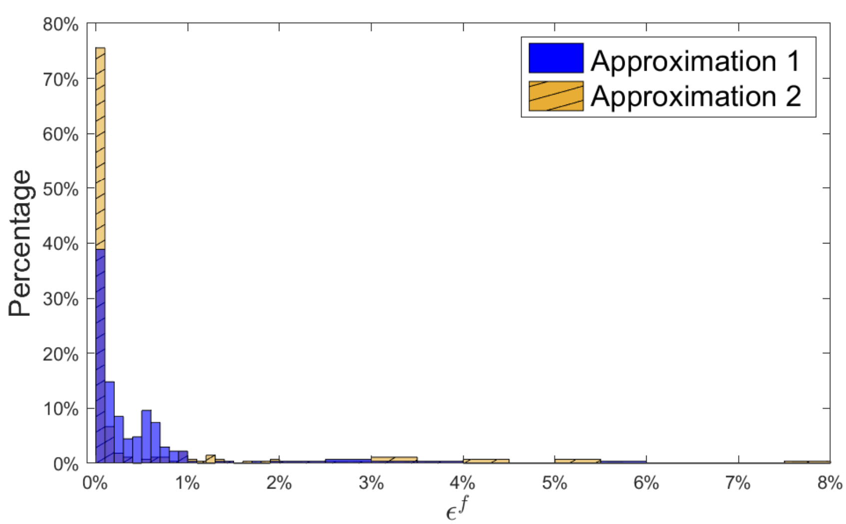
To evaluate the effectiveness of approximations of S^*_{F} developed in Section 5.3, we denote by ϵ^{f} the (relative) cost gap:

$$\epsilon^{\mathrm{f}} = \frac{C(\widetilde{S}_F) - C(S^*_{\mathrm{F}})}{C(S^*_{\mathrm{F}})} \times 100\%,$$

where $\widetilde{S}_F$ denotes the approximated optimal base-stock level of the FABS policy in Proposition 10. To compute $C(\widetilde{S}_F)$, we simulate the system under the FABS policy with $\widetilde{S}_F$. The computation of $C(S^*_{\mathrm{F}})$ is similar, but the optimal base stock level S^*_{F} is computed via exhaustive search as described in Section 6.1. In our experiments, we find that searching an optimal base stock level S^*_{F} takes about 140 seconds on average for each instance, while solving $\widetilde{S}_F$ (using Proposition 10) only requires less than 0.001 seconds on average.

Figure 1 illustrates the cost gap ϵ^{f} for each of the two approximation methods discussed in Section 5.3. The x-axis represents the cost gap introduced by each approximation method, while the y-axis shows the percentage of problem instances (out of 270) of each cost gap. From Figure 1, we can observe that Approximation 1 results in cost gaps that are very close to 0 in approximately 40% of the instances, with all cost gaps smaller than 5.5%. In contrast, Approximation 2 results in cost gaps that are very close to 0 in approximately 76% of the instances, but the maximum cost gap can reach up to 8%. The average cost gap ϵ^{f} for Approximation 1 is 0.4%, with a maximum of 5.5% while for Approximation 2, the average ϵ^{f} is 0.3%, with a maximum of 7.9%. Overall, these cost gaps are close to 0, indicating that both approximations are very effective. This can be explained as when $\sum_i p_i$ is close to 1 (heavy traffic), the approximations are known to be effective; when $\sum_i p_i$ is considerably less than 1 (light traffic), overshoot is unlikely to occur and so its approximation has minimum effect.

Figure 1: Comparison of cost gaps of two approximation methods



6.3 Value of Return Forecast

In this section, we compare the performance of the FABS policy with that of simple base stock policies for illustrating the value of return forecast. Under a simple base stock policy with base-stock level S, the firm manufactures up to S if $IP_n < S$ in period n; otherwise, it does not manufacture. Hence, unlike the FABS policy, the firm does not adjust the inventory position based on the return forecast when implementing the simple base stock policy.

Intuitively, because the FABS policy utilizes return forecast, it will reduce the uncertainty faced by the firm and should outperform the simple base stock policy. We define the relative saving of the FABS policy as

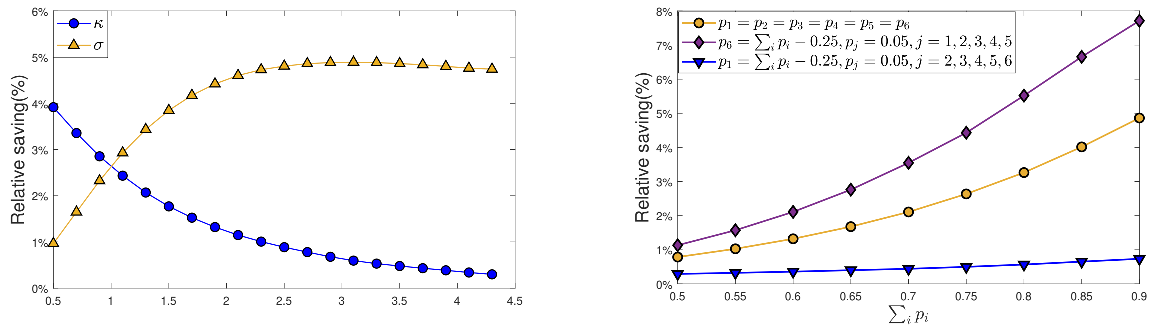
$$\frac{C_B(S^*_{\mathrm{B}}) - C(S^*_{\mathrm{F}})}{C(S^*_{\mathrm{F}})} \times 100\%,$$

where $C_B(S^*_{\mathrm{B}})$ is the optimal cost under the simple base-stock policy and S^*_{B} is the corresponding optimal base-stock level. It is interesting to see when the relative saving from the FABS policy tends to be large, i.e., the value of incorporating return forecast is high. We assume demand $D_n \sim TN(5, \sigma^2, 5-3\sigma, 5+3\sigma)$ and the return noise $\epsilon_n \sim TN(0, \kappa^2, -3\kappa, 3\kappa)$, as in Section 6.2. We consider the unit manufacturing cost $k = 3$, the unit holding cost $h = 1$, and the unit backlogging cost $b = 10$.

First, we examine the effect of variability, including both the noise of return and demand. We fix $L = 4$, $\tau = 6$, $p_1 = p_2 = ... = p_6 = 0.125$, and set one of the two parameters, either σ (demand variance) or κ (return noise variance), to 1, while varying the other within a range of 0.5 to 4.3. The results are plotted in Figure 2(a). It can be observed that the value of return forecast decreases as κ increases whereas it increases and then decreases (slightly) as σ becomes larger. Both demand variability and noise variability contribute to the overall variability of returns. It is intuitive that the value of return forecast becomes smaller as the return noise variability increases. On the other hand, when demand variability is not high, the return forecast is more accurate, making the FABS policy outperforms the simple base-stock more. However, when demand variability becomes large, the (relative) benefit of return forecasting is smaller, as demand variability starts to considerably increase the system's cost.

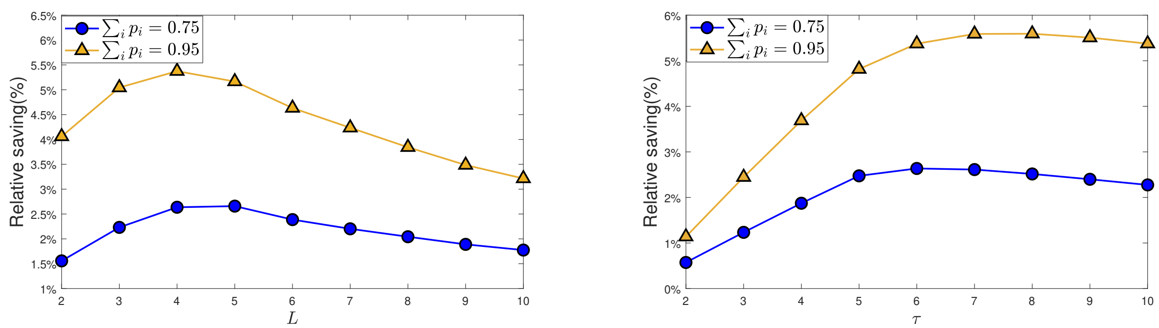
Next, we examine the impact of return rates on the value of return forecast. We plot the results in Figure 2(b), where we fix $L = 4$, $\tau = 6$, $\kappa = \sigma = 1$, and vary $\sum_{i=1}^{6} p_i$ from 0.5 to 0.9. We consider three different scenarios for the return rates (p_i) to examine their effects on the value of return forecast. The results show that the relative saving increases with the total return rate. As $\sum_{i=1}^{6} p_i$ increases, more of the future return uncertainty is attributed to demand variability. Consequently, utilizing past demand information for return forecasting becomes increasingly beneficial for the firm. Moreover, it can be observed that if the return rate is concentrated on p_6 ($p_1 = p_2 = \cdots = p_5 = 0.05$), the value of return forecast is significantly larger than the scenario where the return rate is distributed equally across six periods $((p_1 = p_2 = \cdots = p_5 = p_6)$. However, when the return rate is concentrated on p_1 $(p_2 = p_3 = \cdots = p_6 = 0.05)$, the value of return forecast becomes smaller than the other scenarios.

Figure 2: Impact of variability and return rate



(a) Impact of σ and κ (b) Impact of $\sum_i p_i$

Figure 3: Impact of the maximum distributed lag τ and manufacturing lead time$L$



(a) $\tau = 6$ (b) $L = 4$

We next study the effect of manufacturing lead time L. We fix $\tau = 6, \sigma = \kappa = 1, p_1 = p_2 = \ldots = p_6$, and vary L from 2 to 10. The results are presented in Figure 3(a). We can see that as the lead time L increases, the value of return forecast first increases and then decreases, peaked roughly at $L = \tau - 1$. When $L > \tau - 1$, with a longer lead time, the total uncertainty faced by the firm increases, where as the information on returns remains unchanged. Consequently, the value return forecast decreases. On the other hand, when the lead time L is shorter than $\tau - 1$, a longer lead time implies more effective utilization of return forecast, which increases its value.

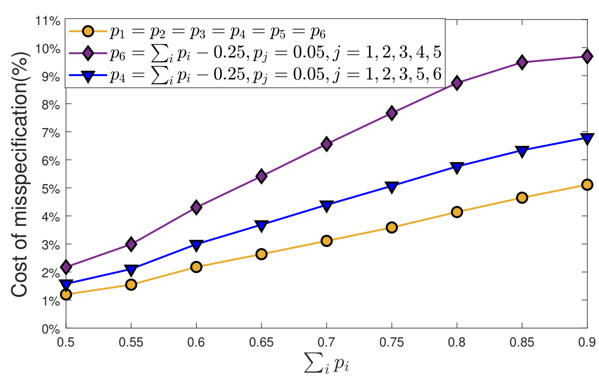
Finally, we study the effect of the maximum return lag τ. We fix $L = 4$, $\sigma = \kappa = 1$, and $p_1 = p_2 = \ldots = p_\tau$, and vary τ from 2 to 10. We plot the value of return forecast as a function of τ in Figure 3(b). We can observe that as the maximum return lag τ increases, the value of return forecast first increases and then decreases (slowly). On one hand, as τ increases, it is more important to utilize the past demand information to forecast return and decide manufacturing quantity. On the other, as p distributed thinner across τ periods, the most valuable part of information (within $L+1$) becomes smaller.

6.4 Return Model Misspecification

In this section we briefly discuss the issue of model misspecification. For illustration, we consider a simplified setting where the true return model is given by (1) with $\tau > 1$, but the firm may not correctly specify it. In particular, we assume the firm considers a "false" return model with $\hat{R}_n = \hat{p}_1 D_{n-1} + \epsilon_n$ where $\hat{p}_1 = \sum_{i=1}^{\tau} p_i$ so that the average return rate is the same as the one under the true return model. Under this misspecified return model, a simple base-stock policy can be shown optimal (following the same analysis as in Section 4). In the following, we compare the system costs under the correct and misspecified return model. We compute the relative cost increase, i.e., [system cost under the policy assuming $\hat{R}_n$ − optimal system cost]/optimal system cost $\times 100\%$.

Figure 4 present the cost increase under three different patterns of return rates which represent distinct distributions of return rates over time. We consider the unit manufacturing cost $k = 3$, the unit holding cost $h = 1$, and the unit penalty cost $b = 10$. We set $L = 5$, $\tau = 6$, which indicates that the FABS policy is optimal. We fix $\sigma = \kappa = 1$ and vary $\sum_i p_i$ from 0.5 to 0.9. We observe that the cost resulted from model misspecification increases as $\sum_i p_i$ increases and can be as high as 9%. Moreover, a less even distributed return over the previous demands results in a higher model misspecification cost.

Figure 4: Impact of model misspecification



7 Conclusion

We study a periodic-review remanufacturing inventory system, where the quantity of returned products each period is random and dependent on the demands of the previous periods, following the distributed lag model. A firm operating the system remanufactures all the cores returned each period while deciding additional manufacturing quantity to fulfill random customer demand. The firm aims to minimize the expected long-run average manufacturing, inventory holding, and demand backlogging cost. We establish the existence of optimal stationary policy via the vanishing discount factor approach. Moreover, we conduct a state-space reduction that facilitates the characterization of the optimal policy. In particular, when the manufacturing lead time is longer than the maximum return lag, we show that a FABS policy that incorporates return forecast with constant base-stock level is optimal; otherwise, a state-dependent base-stock policy is optimal with the base-stock level decreasing in the state. As the optimal state-dependent policy is hard to compute and implement, we apply the FABS policy in both scenarios and provide a procedure to evaluate the FABS policy. We further develop simple approximations for the optimal base-stock level. Our numerical results demonstrate that the FABS policy performs very close to optimal in general and our approximations are very effective. We further numerically examine the value of return forecast and the impact of return model misspecification.

References

Asmussen, S. (2003). *Applied probability and queues*. Springer.

Ata, B. and Barjesteh, N. (2023). An approximate analysis of dynamic pricing, outsourcing, and scheduling policies for a multiclass make-to-stock queue in the heavy traffic regime. *Operations Research*, 71(1):341–357.

Avci, H., Gokbayrak, K., and Nadar, E. (2020). Structural results for average-cost inventory models with markov-modulated demand and partial information. *Production and Operations Management*, 29(1):156–173.

Bai, X., Chen, X., and Stolyar, A. L. (2023). Average cost optimality in partially observable lost-sales inventory systems. *Operations Research*, 71(6):2390–2396.

Bayiz, M. and Tang, C. (2004). An integrated planning system for managing the refurbishment of thermoluminescent badges. *Interfaces*, 34(5):383–393.

Bertsekas, D. P. et al. (2011). Dynamic programming and optimal control 3rd edition, volume ii. *Belmont, MA: Athena Scientific*, 1.

Beyer, D. and Sethi, S. P. (1997). Average cost optimality in inventory models with markovian demands. *Journal of Optimization Theory and Applications*, 92(3):497–526.

Blanchet, J. and Glynn, P. (2006). Complete corrected diffusion approximations for the maximum of a random walk. *The Annals of Applied Probability*, 16(2):951–983.

Bu, J., Zhang, H., and Jasin, S. (2023). Asymptotic optimality of simple replenishment policies for a lost-sales inventory system with delivery lead time and purchase returns. *Available at SSRN*.

Clottey, T., Benton, W., and Srivastava, R. (2012). Forecasting product returns for remanufacturing operations. *Decision Sciences*, 43(4):589–614.

Decroix, G. (2006). Optimal policy for a multiechelon inventory system with remanufacturing. *Operations Research*, 54(3):532–543.

Feinberg, E. A., Kasyanov, P. O., and Zadoianchuk, N. V. (2012). Average cost markov decision processes with weakly continuous transition probabilities. *Mathematics of Operations Research*, 37(4):591–607.

Feinberg, E. A. and Lewis, M. E. (2018). On the convergence of optimal actions for markov decision processes and the optimality of (s, s) inventory policies. *Naval Research Logistics (NRL)*, 65(8):619–637.

Gao, X. and Huang, J. (2024). Asymptotically optimal control of make-to-stock systems. *Mathematics of Operations Research*, 49(2):948–985.

Goh, T. and Varaprasad, N. (1986). A statistical methodology for the analysis of the life-cycle of reusable containers. *IIE Transactions*, 18(1):42–47.

Gong, X. and Chao, X. (2013). Technical note—optimal control policy for capacitated inventory systems with remanufacturing. *Operations Research*, 61(3):603–611.

Gong, X. and Liu, S. (2024). Managing hybrid manufacturing/remanufacturing inventory systems with random production capacities. *Production and Operations Management*, 33(7):1518–1534.

Gut, A. (1974). On the moments and limit distributions of some first passage times. *The Annals of Probability*, 2(2):277–308.

Huh, W. T., Janakiraman, G., and Nagarajan, M. (2011). Average cost single-stage inventory models: An analysis using a vanishing discount approach. *Operations Research*, 59(1):143–155.

Jacobs, P. A. (1980). Heavy traffic results for single-server queues with dependent (earma) service and interarrival times. *Advances in Applied Probability*, 12(2):517–529.

Kelle, P. and Silver, E. (1987). Purchasing policy of new containers considering the random returns of previously issued containers. *IIE Transactions*, 21(4):349–354.

Kiesmüller, G. and Van Der Laan, E. (2001). An inventory model with dependent product demands and returns. *International Journal of Production Economics*, 72(1):73–87.

Kingman, J. (1961). The single server queue in heavy traffic. In *Mathematical proceedings of the cambridge philosophical society*, volume 57, pages 902–904. Cambridge University Press.

Loynes, R. M. (1962). The stability of a queue with non-independent inter-arrival and service times. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 58, pages 497–520. Cambridge University Press.

Miao, S. (2023). Managing the inventory with product return: An approximation algorithm. *Available at SSRN*.

Parker, D., Riley, K., Robinson, S., Symington, H., Tewson, J., Jansson, K., Ramkumar, S., and Peck, D. (2015). *Remanufacturing market study*. European Remanufacturing Network (ERN), United Kingdom.

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

Rosen, E. (2020). What's old can be new again for businesses.

Schal, M. (1993). Average optimality in dynamic programming with general state space. *Mathematics of Operations Research*, 18(1):163–172.

Shang, G., McKie, E. C., Ferguson, M. E., and Galbreth, M. R. (2020). Using transactions data to improve consumer returns forecasting. *Journal of Operations Management*, 66(3):326–348.

Siegmund, D. (1979). Corrected diffusion approximations in certain random walk problems. *Advances in Applied Probability*, 11(4):701–719.

Simpson, V. (1978). Optimum solution structure for a repairable inventory problem. *Operations Research*, 26(2):270–281.

Tao, Z. and Zhou, S. X. (2014). Approximation balancing policies for inventory systems with remanufacturing. *Mathematics of Operations Research*, 39(4):1179–1197.

Tao, Z., Zhou, S. X., and Tang, C. S. (2012). Managing a remanufacturing system with random yield: Properties, observations, and heuristics. *Production and Operations Management*, 21(5):797–813.

Toktay, L. and Wein, L. (2001). Analysis of a forecasting-production-inventory system with stationary demand. *Management Science*, 47(9):1268–1281.

Toktay, L. B., Wein, L. M., and Zenios, S. A. (2000). Inventory management of remanufacturable products. *Management Science*, 46(11):1412–1426.

Topkis, D. M. (1978). Minimizing a submodular function on a lattice. *Operations research*, 26(2):305–321.

Xin, L. and Yang, J. (2022). Asymptotic analysis of a remanufacturing system with non-identical lead times. *Available at SSRN: https://ssrn.com/abstract=3760906*.

Zhou, S., Tao, Z., and Chao, X. (2011). Optimal control of inventory systems with multiple types of remanufacturable products. *Manufacturing & Service Operations Management*, 13(1):20–34.

Appendix: Proofs

A Proofs of Propositions 2, 3, 4 and 6

A.1 Proof of Proposition 2

We can readily infer from (14) and (2) that the dynamics of z_n is given by:

$$z_{n+1} = z_n + Q_n - \tilde{D}_n, \tag{A.1}$$

where $\tilde{D}_n$ is given in (16). Using the definition of $r(\hat{\boldsymbol{d}})$ in (9), we can obtain the explicit form of $\tilde{D}_n$:

$$\tilde{D}_n = \begin{cases} \left(1 - \sum_{i=1}^{\tau} p_i\right) D_n - \epsilon_{n+1}, & \text{if } L \geq \tau - 1, \\ \left(1 - \sum_{i=1}^{L+1} p_i\right) D_n - \epsilon_{n+1} - \sum_{j=1}^{\tau-L-1} p_{L+1+j} D_{n-j}, & \text{if } L < \tau - 1. \end{cases} \tag{A.2}$$

Hence, if $L \geq \tau - 1$, we obtain that $\tilde{D}_n$ is independent of $\mathcal{F}_{n-1}$ (the history of demand) and z_n. It follows that $\{z_n\}$ is a controlled Markov process and the state $(IP_n, \hat{\boldsymbol{D}}_{n-1})$ in the original MDP can be reduced to the one-dimensional state z_n. On the other hand, if $L < \tau - 1$, we observe that $\tilde{D}_n$ is correlated with past demands via $\boldsymbol{D}_{n-1} = (D_{n-1}, ..., D_{n-\tau+L+1})$. It follows that the state $(IP_n, \hat{\boldsymbol{D}}_{n-1})$ in the orignal MDP can be reduced to a lower-dimensional state $(z_n, \boldsymbol{D}_{n-1})$.

A.2 Proof of Proposition 3

We prove the first part (when $L < \tau - 1$) of Proposition 3 through induction. The proof for $L \geq \tau - 1$ follows a similar approach and is therefore omitted.

For $n = N+1$, by the definition of $J_{N+1,\alpha}(\cdot)$ and $V_{N+1,\alpha}(\cdot)$, we have,

$$V_{N+1,\alpha}(x, \hat{\boldsymbol{d}}) = J_{N+1,\alpha}(x + r(\hat{\boldsymbol{d}}), \boldsymbol{d}) = 0.$$

Assume that for $n = k+1$, we have

$$V_{k+1,\alpha}(x, \hat{\boldsymbol{d}}) = J_{k+1,\alpha}(x + r(\hat{\boldsymbol{d}}), \boldsymbol{d}), \quad \text{for any } (x, \hat{\boldsymbol{d}}) \in \mathcal{S}.$$

It follows that

$$\begin{aligned} V_{k+1,\alpha}(w - r(\hat{\boldsymbol{D}}_{k-1}) - D_k + R_{k+1}, \hat{\boldsymbol{D}}_k) &= J_{k+1,\alpha}(w - r(\hat{\boldsymbol{D}}_{k-1}) - D_k + R_{k+1} + r(\hat{\boldsymbol{D}}_k), \boldsymbol{D}_k) \\ &= J_{k+1,\alpha}(w - \tilde{D}_k, \boldsymbol{D}_k), \end{aligned} \tag{A.3}$$

where the second equality is due to the definition of $\tilde{D}_k$. From the optimality equation in (12), letting $z = x + r(\hat{\boldsymbol{d}})$ and $w = y + r(\hat{\boldsymbol{d}})$, we have,

$$V_{k,\alpha}(x, \hat{\boldsymbol{d}}) = \min_{w \geq z} \Big\{ k(w - z) + \ell(w) \\ + \alpha \mathbb{E}\Big[V_{k+1,\alpha}\Big(w - r(\hat{\boldsymbol{D}}_{k-1}) - D_k + R_{k+1}, \hat{\boldsymbol{D}}_k \Big) \Big| \hat{\boldsymbol{D}}_{k-1} = \hat{\boldsymbol{d}} \Big] \Big\}.$$

By (A.3), we have,

$$V_{k,\alpha}(x, \hat{\boldsymbol{d}}) = \min_{w \geq z} \Big\{ k(w - z) + \ell(w) + \alpha \mathbb{E}\Big[J_{k+1,\alpha}\Big(w - \tilde{D}_k, \boldsymbol{D}_k \Big) | \hat{\boldsymbol{D}}_{k-1} = \hat{\boldsymbol{d}} \Big] \Big\}.$$

From (A.2), we know that if $L < \tau - 1$, $J_{k,\alpha}\Big(w - \tilde{D}_k, \boldsymbol{D}_k \Big)$ is independent of $(D_{k-\tau+L}, .., D_{k-\tau+1})$. Hence, we obtain

$$V_{k,\alpha}(x, \hat{\boldsymbol{d}}) = \min_{w \geq z} \Big\{ k(w - z) + \ell(w) + \alpha \mathbb{E}\Big[J_{k+1,\alpha}\Big(w - \tilde{D}_k, \boldsymbol{D}_k \Big) \Big| \boldsymbol{D}_{k-1} = \boldsymbol{d} \Big] \Big\}.$$

By the optimality equation in (15), we get $V_{k,\alpha}(x, \hat{\boldsymbol{d}}) = J_{k,\alpha}(x + r(\hat{\boldsymbol{d}}), \boldsymbol{d})$. The proof is therefore complete.

A.3 Proof of Proposition 4

We prove the first part (when $L < \tau - 1$) of Proposition 4. The second part when $L \geq \tau - 1$ can be established with a similar argument.

From Proposition 3, we have,

$$\lim_{N \to \infty} V_{n,\alpha}^{(N)}(x, \hat{\boldsymbol{d}}) = \lim_{N \to \infty} J_{n,\alpha}^{(N)}(x + r(\hat{\boldsymbol{d}}), \boldsymbol{d}).$$

Theorem 3.4 in Feinberg and Lewis (2018) shows that if Assumption (W*) holds, then $V_{1,\alpha}^{(N)}(x, \hat{\boldsymbol{d}}) \uparrow V_\alpha(x, \hat{\boldsymbol{d}})$ as $N \to \infty$. For our system, Assumption (W*) indeed holds, and we defer the verification of this assumption to Appendix B.1. Hence, we have,

$$V_\alpha(x, \hat{\boldsymbol{d}}) = \lim_{N \to \infty} V_{1,\alpha}^{(N)}(x, \hat{\boldsymbol{d}}) = \lim_{N \to \infty} J_{1,\alpha}^{(N)}(x + r(\hat{\boldsymbol{d}}), \boldsymbol{d}) = J_\alpha(x + r(\hat{\boldsymbol{d}}), \boldsymbol{d}).$$

The proof is therefore complete.

A.4 Proof of Proposition 6

We prove part (b) of Proposition 6. Part (a) can be shown by a similar (and slightly simpler) argument.

For an N-horizon discounted problem, define for $n = 1, 2, ..., N$,

$$G_{n,\alpha}(w, \boldsymbol{d}) = kw + \ell(w) + \alpha\mathbb{E}[J_{n+1,\alpha}(w - \tilde{D}_n, \boldsymbol{D}_n)|\boldsymbol{D}_{n-1} = \boldsymbol{d}].$$

Thus, by (15) we obtain $J_{n,\alpha}(z, \boldsymbol{d}) = \min_{w\geq z} \; G_{n,\alpha}(w, \boldsymbol{d}) - kz$. In addition, by a similar argument as in the proof of Proposition 4, we obtain that $J_\alpha(z, \boldsymbol{d})$ defined before is the optimal cost of the infinite-horizon discounted problem after the state transformation. Note that one can readily verify $J_\alpha(z, \boldsymbol{d}) < \infty$, for given $(z, \boldsymbol{d}) \in \mathbb{X}$ and $\alpha \in [0, 1)$. This is relatively straightforward because one can bound $J_\alpha(z, \boldsymbol{d})$ by the discounted costs of using a policy that orders nothing, which can be shown to be finite. Hence, we have the following optimality equation:

$$J_\alpha(z, \boldsymbol{d}) = \min_{w\geq z}\{kw + \ell(w) + \mathbb{E}[J_\alpha(w - \tilde{D}_+, \boldsymbol{D}_+)|\boldsymbol{D}_0 = \boldsymbol{d}]\} - kz, \tag{A.4}$$

where $D \overset{d}{=} D_1$, $\epsilon \overset{d}{=} \epsilon_1$, $\boldsymbol{D}_+ = (D, d_0, ..., d_{-\tau+L+3})$, and $\tilde{D}_+ = \Big(1 - \sum_{i=1}^{L+1} p_i\Big)D - \epsilon - \sum_{j=1}^{\tau-1-L} p_{L+1+j} d_{1-j}$ following from the definition of $\tilde{D}_n$ in (16). Letting

$$G_\alpha(w, \boldsymbol{d}) = kw + \ell(w) + \alpha\mathbb{E}[J_\alpha(w - \tilde{D}_+, \boldsymbol{D}_+)|\boldsymbol{D}_0 = \boldsymbol{d}],$$

we obtain $J_\alpha(z, \boldsymbol{d}) = \min_{w\geq z} G_\alpha(w, \boldsymbol{d}) - kz$. We have the following result, the proof of which is deferred to the end of this section.

Lemma A.1. *For any demand vector $\boldsymbol{d} \in [\underline{D}, \bar{D}]^{\tau-L-1}$, $G_{n,\alpha}(w, \boldsymbol{d})$ and $J_{n,\alpha}(w, \boldsymbol{d})$ are convex in w, for $n = 1, 2, ..., N$. As a result, $G_\alpha(w, \boldsymbol{d})$ and $J_\alpha(w, \boldsymbol{d})$ are convex in w for fixed $\boldsymbol{d}$.*

Let

$$S^{d*}_{\mathrm{F},\alpha}(\boldsymbol{d}) = \mathrm{argmin}_{w\in\mathbb{R}} G_\alpha(w, \boldsymbol{d}),$$

where by convention, if there are multiple minimizers of $G_\alpha(w, \boldsymbol{d})$, $S^{d*}_{\mathrm{F},\alpha}(\boldsymbol{d})$ is the smallest one. By Lemma A.1, $G_\alpha(w, \boldsymbol{d})$ and $J_\alpha(w, \boldsymbol{d})$ are convex in w. Then we have

$$J_\alpha(z, \boldsymbol{d}) = \min_{w\geq z} \; G_\alpha(w, \boldsymbol{d}) - kz = G_\alpha(\max\{S^{d*}_{\mathrm{F},\alpha}(\boldsymbol{d}), z\}, \boldsymbol{d}) - kz. \tag{A.5}$$

Thus, given $(z, \boldsymbol{d}) \in \mathbb{X}$, the optimal action at this state is $Q^*_\alpha(z, \boldsymbol{d}) = (S^{d*}_{\mathrm{F},\alpha}(\boldsymbol{d}) - z)^+$. It follows that the FABS-F policy with state-dependent base-stock level $S^{d*}_{\mathrm{F},\alpha}(\boldsymbol{d})$ is optimal for the infinite-horizon discounted MDP with state space $\mathbb{X}$.

It remains to show the FABS-F policy is also optimal for the infinite-horizon discounted MDP (8) with state space $\mathcal{S}$. By (8), we have the following optimality equations, for $(x, \hat{\boldsymbol{d}}) \in \mathcal{S}$,

$$V_\alpha(x, \hat{\boldsymbol{d}}) = \min_{y\geq x}\left\{k(y - x) + \ell(y + r(\hat{\boldsymbol{d}})) + \alpha\mathbb{E}[V_\alpha(y - D + R_+, \hat{\boldsymbol{D}}_+)|\hat{\boldsymbol{D}}_0 = \hat{\boldsymbol{d}}]\right\}, \tag{A.6}$$

where $\hat{\boldsymbol{D}}_+ = (D, d_0, ..., d_{-\tau+3})$, and $R_+ = p_1 D + p_2 d_0 + ... + p_\tau d_{-\tau+2} + \epsilon$. Using Proposition 4 and (A.7), we obtain

$$V_\alpha(x, \hat{\boldsymbol{d}}) = J_\alpha(x + r(\hat{\boldsymbol{d}}), \boldsymbol{d}) = \min_{w \geq x + r(\hat{\boldsymbol{d}})} \left\{ kw + \ell(w) + \mathbb{E}[J_\alpha(w - \tilde{D}_+, \boldsymbol{D}_+) | \boldsymbol{D}_0 = \boldsymbol{d}] \right\} - k(x + r(\hat{\boldsymbol{d}}))$$
$$= kw^* + \ell(w^*) + \mathbb{E}[J_\alpha(w^* - \tilde{D}_+, \boldsymbol{D}_+) | \boldsymbol{D}_0 = \boldsymbol{d}] - k(x + r(\hat{\boldsymbol{d}})),$$

where $w^* = \max\{S^{d*}_{\mathrm{F},\alpha}(\boldsymbol{d}), x + r(\hat{\boldsymbol{d}})\}$ by (A.5). From the definition of $\tilde{D}_n$ in (16), we get $\tilde{D}_+ = r(\hat{\boldsymbol{D}}_0) + D - R_+ - r(\hat{\boldsymbol{D}}_+)$. We can then infer from Proposition 4 that

$$J_\alpha(w^* - \tilde{D}_+, \boldsymbol{D}_+) = V_\alpha(w^* - r(\hat{\boldsymbol{D}}_+) - \tilde{D}_+, \boldsymbol{D}_+) = V_\alpha(w^* - r(\hat{\boldsymbol{D}}_0) - D + R_+, \boldsymbol{D}_+).$$

It follows that

$$V_\alpha(x, \hat{\boldsymbol{d}}) = kw^* + \ell(w^*) + \mathbb{E}[V_\alpha(w^* - r(\hat{\boldsymbol{d}}) - D + R_+, \hat{\boldsymbol{D}}_+) | \hat{\boldsymbol{D}}_0 = \hat{\boldsymbol{d}}] - k(x + r(\hat{\boldsymbol{d}})).$$

This implies $y^* = w^* - r(\hat{\boldsymbol{d}})$ achieves the minimum in (A.6), and hence the optimal manufacturing quantity at state $(x, \hat{\boldsymbol{d}})$ is given by $Q(x, \hat{\boldsymbol{d}}) = (S^{d*}_{\mathrm{F},\alpha}(\boldsymbol{d}) - x - r(\hat{\boldsymbol{d}}))^+$. Therefore, the FABS-F policy with state-dependent base-stock level $S^{d*}_{\mathrm{F},\alpha}(\boldsymbol{d})$ is also optimal for the original infinite-horizon discounted problem (8).

To prove the second claim that $S^{d*}_{\mathrm{F},\alpha}(\boldsymbol{d})$ is nonincreasing in d_i, we need the following result. The proof is deferred to the end of this section.

Lemma A.2. *$J^N_{n,\alpha}(w, \boldsymbol{d})$ is supermodular in (w, d_i), for any $i = 0, -1, \ldots, L-\tau+2$, $n = 1, 2, ..., N$. As a result, $J_\alpha(w, \boldsymbol{d}) = \lim_{N\to\infty} J^N_{n,\alpha}(w, \boldsymbol{d})$ is supermodular in (w, d_i), for any $i = 0, -1, \ldots, L - \tau + 2$.*

Now let $\tilde{G}_{m,\alpha}(w, \boldsymbol{d}') = G_{m,\alpha}(w, -\boldsymbol{d}')$, where $\boldsymbol{d}' = -\boldsymbol{d}$. By Theorem 6.2 in Topkis (1978) and Lemma A.2, since $\tilde{G}_{m,\alpha}(w, \boldsymbol{d}')$ is submodular in (w, d'_i) for $i \in 0, -1, \ldots, 2 - \tau$, which implies that $S^{d*}_{\mathrm{F},\alpha}(\boldsymbol{d})$ is nonincreasing in d_i. The proof is complete.

Proof of Lemma A.1. We prove the result by induction. First, $J_{N+1,\alpha}(w, \boldsymbol{d})$ is clearly convex in w by definition. Suppose $J_{n+1,\alpha}(w, \boldsymbol{d})$ is convex in w for $n = N, N-1 \ldots, k$. Then $G_{k,\alpha}(w, \boldsymbol{d})$ is convex in w because $\ell(w)$ is convex in w. Let $S_n(\boldsymbol{d}) = \mathrm{argmin}_{w\in\mathbb{R}} G_{n,\alpha}(w, \boldsymbol{d})$. By minimizing $G_{k,\alpha}(w, \boldsymbol{d})$ over $\{w : w \geq z\}$, we have $J_{k,\alpha}(w, \boldsymbol{d}) = G_{k,\alpha}(\max\{w, S_n(\boldsymbol{d})\}, \boldsymbol{d})$, which is convex in w. Hence, we have proved the first part of the lemma.

By the definition of $J_\alpha(z, \boldsymbol{d})$, we have $J^{(N)}_{1,\alpha} \uparrow J_\alpha(z, \boldsymbol{d})$ as $N \to \infty$. By Monotone Convergence Theorem, we obtain $G^{(N)}_{1,\alpha}(w, \boldsymbol{d}) \uparrow G_\alpha(w, \boldsymbol{d})$ as $N \to \infty$. It then follows from the first part of the lemma that $G_\alpha(w, \boldsymbol{d})$ and $J_\alpha(w, \boldsymbol{d})$ are also convex in w. □

Proof of Lemma A.2. We prove the result by induction. $J_{N+1,\alpha}(z, \boldsymbol{d})$ is trivially supermodular in (z, d_i) for all i. Suppose $J_{m+1,\alpha}(z, \boldsymbol{d})$ is supermodular in (z, d_i) for all i. Let $\tilde{J}_{m,\alpha}(z, \boldsymbol{d}) =$

$J_{m,\alpha}(z, d_0, d_{-1}, \ldots - d_i, d_{i-1}, d_{L-\tau+2})$. Note that

$$\begin{aligned}\tilde{J}_{m,\alpha}(z, \boldsymbol{d}) &= \min_{w \geq z} \Bigg\{ kw + \ell(w) + \mathbb{E}\Bigg[J_{m+1,\alpha}\Bigg(\Big(w - \Big(1 - \sum_{i=1}^{L+1} p_i\Big) D_m + \epsilon_{m+1} \\ &\quad + \sum_{j=1}^{\tau-L-1, j\neq 1-i} p_{L+1+j} d_{1-j} - p_{L+2-i} d_i, (D_m, d_0, d_{-1}, \ldots - d_i, \ldots, d_{3-\tau+L})\Big)\Bigg)\Bigg]\Bigg\} - kz,\end{aligned}$$

where the bracketed terms are submodular in (w, d_i) from the inductive assumption and the convexity of $J_{m,\alpha}(w, \boldsymbol{d})$ in w. Hence, By Theorem 4.3 of Topkis (1978), i.e., the submodularity preserves under minimization over a sublattice, $\tilde{J}_{m,\alpha}(z, \boldsymbol{d})$ is submodular in (z, d_i). This holds true for any pair of (z, d_i) and so $J_{m,\alpha}(z, \boldsymbol{d})$ is supermodular in (z, d_i) for all i. This completes the proof.

B Proof of Theorem 5

In this section, we prove Theorem 5. Our proof relies on the results established in Feinberg and Lewis (2018), but it is substantially different from the proof in Feinberg and Lewis (2018) due to the high-dimensional state space in our problem with correlated returns over time. The proof is lengthy, so we divide it into several steps for clarity. In Appendix B.1, we verify Assumption (W*) in Feinberg and Lewis (2018). In Appendix B.2, we verify that Assumption (B) in Feinberg and Lewis (2018) holds for our system, highlighting significant challenges encountered during this verification process. In Appendix B.3, we prove Theorem 5 based on the analysis in Appendices B.1 and B.2. Auxiliary lemmas are proved in Appendix B.4. Throughout this section, we focus on analyzing the case when $L < \tau - 1$. The scenario where $L \geq \tau - 1$ is simpler to analyze (as the transformed state of the MDP is of dimension one) and can be addressed in a similar manner.

B.1 Verification of Assumption (W*) in Feinberg and Lewis (2018)

Assumption (W*) of Feinberg and Lewis (2018) holds, if (i) the single-period cost function c is bounded below and inf-compact (which implies $\mathbb{K}$-inf-compact by Proposition 3.2 in Feinberg and Lewis (2018)); (ii) the transition probability is weakly continuous. We verify these two conditions below.

(1) Firstly, we show that c is inf-compact on state space $\mathcal{S}$. Consider the level sets of the function c, $\mathscr{D}_c(\lambda) = \{(x, \hat{\boldsymbol{d}}, a) \in \mathcal{S} \times \mathbb{R}_+ : c(x, \hat{\boldsymbol{d}}, a) \leq \lambda\}$. The single-period cost function $c(x, \hat{\boldsymbol{d}}, a)$ is inf-compact if all the level sets $\mathscr{D}_c(\lambda)$ are compact, for $\lambda \in \mathbb{R}$. Given a λ, it is not hard to see that there exists a constant $M_\lambda > 0$ such that $\ell(x + a + r(\hat{\boldsymbol{d}})) > \lambda$ if $|x + a + r(\hat{\boldsymbol{d}})| > M_\lambda$ because $\ell(y) \to \infty$ as $|y| \to \infty$. Moreover, the ordering cost is less than λ implying $0 \leq a < \lambda/k$. For

$\lambda \in \mathbb{R}$, if $(x, \hat{\boldsymbol{d}}, a) \in \mathscr{D}_c(\lambda)$, then it satisfies that $|x + a + r(\hat{\boldsymbol{d}})| \leq M_\lambda$, $|a| \leq \lambda/k$ and $\hat{\boldsymbol{d}} \in [\underline{D}, \bar{D}]^{\tau-1}$, which implies that $\mathscr{D}_c(\lambda)$ is bounded. Since $c(x, \hat{\boldsymbol{d}}, a)$ is continuous with respect to $(x, \hat{\boldsymbol{d}}, a)$, $\mathscr{D}_c(\lambda)$ is closed. Hence, the one-period cost function $c(x, \hat{\boldsymbol{d}}, a)$ is inf-compact.

(2) The transition probability is weakly continuous if for any sequence $\{(x^i, \hat{\boldsymbol{d}}^i, a^i), i \geq 0\}$ converges to $(x, \hat{\boldsymbol{d}}, q)$, where $(x, \hat{\boldsymbol{d}}, a), (x^i, \hat{\boldsymbol{d}}^i, a^i) \in \mathcal{S} \times \mathbb{R}_+$, then

$$\mathbb{E}[f(IP_{n+1}, \hat{\boldsymbol{D}}_n) | IP_n = x^i, \hat{\boldsymbol{D}}_{n-1} = \hat{\boldsymbol{d}}^i, Q_n = a^i] \to \mathbb{E}[f(IP_{n+1}, \hat{\boldsymbol{D}}_n) | IP_n = x, \hat{\boldsymbol{D}}_{n-1} = \hat{\boldsymbol{d}}, Q_n = a],$$

as $i \to \infty$, for any bounded and continuous function f. Recall the state dynamics:

$$\begin{aligned} IP_{n+1} &= IP_n + Q_n - D_n + R_{n+1} = IP_n + \sum_{i=1}^{\tau-1} p_{i+1} D_{n-i} + Q_n - (1 - p_1) D_n + \epsilon_{n+1}, \\ \hat{\boldsymbol{D}}_n &= (D_n, D_{n-1}, ..., D_{n-\tau+2}). \end{aligned}$$

Then by the Dominated Convergence Theorem, we have,

$$\begin{aligned} &\lim_{i\to\infty} \mathbb{E}[f(IP_{n+1}, \hat{\boldsymbol{D}}_n) | IP_n = x^i, \hat{\boldsymbol{D}}_{n-1} = \hat{\boldsymbol{d}}^i, Q_n = a^i] \\ &\quad = \lim_{i\to\infty} \mathbb{E}[f(x^i + a^i + \sum_{i=1}^{\tau-1} p_{i+1} d^i_{1-i} - (1 - p_1) D_n + \epsilon_{n+1}, (D_n, d^i_0, ..., d^i_{2-\tau}))] \\ &\quad = \mathbb{E}[f(x + a + \sum_{i=1}^{\tau-1} p_{i+1} d_{1-i} - (1 - p_1) D_n + \epsilon_{n+1}, (D_n, d_0, ..., d_{2-\tau}))] \\ &\quad = \mathbb{E}[f(IP_{n+1}, \hat{\boldsymbol{D}}_n) | IP_n = x, \hat{\boldsymbol{D}}_{n-1} = \hat{\boldsymbol{d}}, Q_n = a]. \end{aligned}$$

Hence, the transition probability is weakly continuous. The proof is therefore complete. □

B.2 Verification of Assumption (B) in Feinberg and Lewis (2018)

Recall Assumption (B) in Feinberg and Lewis (2018):

(i) There exists a policy π and an initial state $(x, \hat{\boldsymbol{d}}) \in \mathcal{S}$ such that the corresponding long-run average cost is finite;

(ii) For any $(x, \hat{\boldsymbol{d}}) \in \mathcal{S}$, $\sup_{\alpha\in[0,1)} \left[V_\alpha(x, \hat{\boldsymbol{d}}) - \inf_{(x,\hat{\boldsymbol{d}})\in\mathcal{S}} V_\alpha(x, \hat{\boldsymbol{d}})\right] < \infty$.

In Section 5, we have already argued that the long-run average cost of the FABS policy is finite. Hence Assumption (B)(i) holds.

In this section, we focus on validating Assumption (B) (ii), which poses a significantly greater challenge. When $L < \tau - 1$, it is equivalent to show for any given $(z, \boldsymbol{d}) \in \mathbb{X}$,

$$\sup_{\alpha\in[0,1)} u_\alpha(z, \boldsymbol{d}) < \infty, \tag{B.1}$$

where

$$u_\alpha(z,\boldsymbol{d}) = J_\alpha(z,\boldsymbol{d}) - m_\alpha, \quad \text{and} \quad m_\alpha = \inf_{(z,\boldsymbol{d})\in\mathbb{X}} J_\alpha(z,\boldsymbol{d}). \tag{B.2}$$

To establish (B.1), we first bound m_α from below and then construct a feasible policy to upper bound the optimal value function $J_\alpha(z,\boldsymbol{d})$ for the infinite-horizon discounted problem. To facilitate the presentation, we introduce some notation. Define

$$\begin{aligned} &\mathcal{X}_\alpha = \{(z,\boldsymbol{d}) \in \mathbb{X} : J_\alpha(z,\boldsymbol{d}) = m_\alpha\}, \\ &\mathcal{X}_{\alpha,\boldsymbol{d}} = \{z \in \mathbb{R} : J_\alpha(z,\boldsymbol{d}) = \inf_{z\in\mathbb{R}} J_\alpha(z,\boldsymbol{d})\},\ \boldsymbol{d} \in [\underline{D}, \bar{D}]^{\tau-L-1}. \end{aligned} \tag{B.3}$$

Now we show how to obtain a lower bound of m_α. Let π_α be the optimal stationary policy for the discounted infinite-horizon problem $J_\alpha(z,\boldsymbol{d})$, whose existence is implied by Theorem 3.4 (vii) of Feinberg and Lewis (2018). Consider any state $(z^1,\boldsymbol{d}^0) \in \mathbb{X}$ and any time $t \geq \tau$. Let z_t^1 be the adjusted inventory position at period t of the inventory system following policy π_α, with the initial state $(z^1,\boldsymbol{d}^0)$. Denote by $(z_t^1,\boldsymbol{D}_{t-1})$ the state of the above inventory system at period t. Then we have

$$J_\alpha(z^1,\boldsymbol{d}^0) \geq \mathbb{E}\left[\alpha^{t-1} J_\alpha(z_t^1,\boldsymbol{D}_{t-1}) | z_1 = z^1, \boldsymbol{D}_0 = \boldsymbol{d}^0\right] \geq \mathbb{E}\left[\alpha^{t-1} \inf_{z\in\mathbb{R}} J_\alpha(z,\boldsymbol{D}_{t-1})\right],$$

where the first inequality follows from the fact that the single-period cost is non-negative, and the second inequality holds because $\boldsymbol{D}_{t-1}$ is independent of the initial state when $t > \tau$. Then, for a stopping time $\mathcal{T}$ (with respect to the natural filtration generated by historical demand and forecast noise) satisfying $\mathcal{T} \geq \tau$, we have

$$J_\alpha(z^1,\boldsymbol{d}^0) \geq \mathbb{E}\left[\alpha^{\mathcal{T}-1} \inf_{z\in\mathbb{R}} J_\alpha(z,\boldsymbol{D}_{\mathcal{T}-1})\right]. \tag{B.4}$$

With the above inequality, for any state $(z^\alpha,\boldsymbol{d}^\alpha) \in \mathcal{X}_\alpha$ and $(z,\boldsymbol{d}) \in \mathbb{X}$, letting $(z^\alpha,\boldsymbol{d}^\alpha)$ replace $(z^1,\boldsymbol{d}^0)$ in (B.4), we can obtain

$$m_\alpha = J_\alpha(z^\alpha,\boldsymbol{d}^\alpha) \geq \mathbb{E}\left[\alpha^{\mathcal{T}-1} \cdot \inf_{z\in\mathbb{R}} J_\alpha(z,\boldsymbol{D}_{\mathcal{T}-1})\right]. \tag{B.5}$$

Next, we construct a feasible policy, its resulting cost is then the upper bound of the optimal value function $J_\alpha(z,\boldsymbol{d})$. Let π^σ be the policy defined by the following rules depending on the initial state $(z,\boldsymbol{d})$: (i) the policy does not order before period τ; (ii) after period τ, the policy does not order until the adjusted inventory position z_n is lower than $z_{\alpha,\boldsymbol{D}_{n-1}}$, for any $z_{\alpha,\boldsymbol{D}_{n-1}} \in \mathcal{X}_{\alpha,\boldsymbol{D}_{n-1}}$, for $n = \tau, \tau+1, ...$ If $z_n < z_{\alpha,\boldsymbol{D}_{n-1}}$, the policy orders up to level $z_{\alpha,\boldsymbol{D}_{n-1}}$ and then switch to the optimal policy π_α. Given an initial state $(z,\boldsymbol{d})$, denote by

$$\mathcal{N}(z,\boldsymbol{d}) := \inf\left\{n \geq \tau : z_n < z_{\alpha,\boldsymbol{D}_{n-1}}\right\} = \inf\left\{n \geq \tau : z - \tilde{D}[1,n-1] < z_{\alpha,\boldsymbol{D}_{n-1}}\right\}, \tag{B.6}$$

which is the first time that the policy π^σ places an order. Here, the second equation holds due to (A.1). Define $J_\alpha^{\pi^\sigma}(z, \boldsymbol{d})$ as the value function of the policy π^σ for the infinite-horizon discounted problem. Since π^σ is a feasible policy, we have

$$J_\alpha(z, \boldsymbol{d}) \leq J_\alpha^{\pi^\sigma}(z, \boldsymbol{d}). \tag{B.7}$$

We next analyze $J_\alpha^{\pi^\sigma}(z, \boldsymbol{d})$. It is not difficult to show that

$$\begin{aligned} J_\alpha^{\pi^\sigma}(z, \boldsymbol{d}) = \mathbb{E}^{\pi^\sigma} \Bigg[& \sum_{n=1}^{\mathcal{N}(z,\boldsymbol{d})-1} \alpha^{n-1} \ell(z_n) + \alpha^{\mathcal{N}(z,\boldsymbol{d})-1} k \left(z_{\alpha, \boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}} - z_{\mathcal{N}(z,\boldsymbol{d})} \right) \\ & + \alpha^{\mathcal{N}(z,\boldsymbol{d})-1} J_\alpha \left(z_{\alpha, \boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}}, \boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1} \right) \Big| \, z_1 = z, \boldsymbol{D}_0 = \boldsymbol{d} \Bigg], \end{aligned} \tag{B.8}$$

where the first term on the right-hand-side of the above question represents the holding/backlogging costs (the policy π^σ does not order before period $\mathcal{N}(z, \boldsymbol{d})$), the second term represents the ordering cost at period $\mathcal{N}(z, \boldsymbol{d})$, and the third term represents the discounted future cost after period $\mathcal{N}(z, \boldsymbol{d})$ when following/switching to the discounted optimal policy π_α.

We can now proceed to bound $u_\alpha(z, \boldsymbol{d})$ for $\alpha \in [0, 1)$. Letting $\mathcal{N}(z, \boldsymbol{d})$ replace $\mathcal{T}$ in (B.5), we have,

$$m_\alpha = J_\alpha(z^\alpha, \boldsymbol{d}^\alpha) \geq \mathbb{E} \left[\alpha^{\mathcal{N}(z,\boldsymbol{d})-1} \cdot \inf_{z \in \mathbb{R}} J_\alpha(z, \boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}) \right]. \tag{B.9}$$

Using (B.7), (B.8) and (B.9), we obtain

$$\begin{aligned} u_\alpha(z, \boldsymbol{d}) &= J_\alpha(z, \boldsymbol{d}) - m_\alpha \\ &\leq J_\alpha^{\pi^\sigma}(z, \boldsymbol{d}) - \mathbb{E} \left[\alpha^{\mathcal{N}(z,\boldsymbol{d})-1} \cdot \inf_{z \in \mathbb{R}} J_\alpha\big(z, \boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}\big) \right] \\ &\leq \mathbb{E}^{\pi^\sigma} \left[\sum_{n=1}^{\mathcal{N}(z,\boldsymbol{d})-1} \alpha^{n-1} \ell(z_n) + \alpha^{\mathcal{N}(z,\boldsymbol{d})-1} k \left(z_{\alpha, \boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}} - z_{\mathcal{N}(z,\boldsymbol{d})} \right) \Big| z_1 = z, \boldsymbol{D}_0 = \boldsymbol{d} \right] \\ &\quad + \mathbb{E}\left[\alpha^{\mathcal{N}(z,\boldsymbol{d})-1} J_\alpha \left(z_{\alpha, \boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}}, \boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1} \right) \right] - \mathbb{E}\left[\alpha^{\mathcal{N}(z,\boldsymbol{d})-1} \cdot \inf_{z \in \mathbb{R}} J_\alpha \left(z, \boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1} \right) \right]. \end{aligned}$$

By the definition of $\mathcal{X}_{\alpha,\boldsymbol{d}}$ in (B.3), we have,

$$J_\alpha \left(z_{\alpha,\boldsymbol{d}}, \boldsymbol{d} \right) = \inf_{z \in \mathbb{R}} J_\alpha \left(z, \boldsymbol{d} \right), \quad \text{for } \boldsymbol{d} \in [\underline{D}, \bar{D}]^{\tau - L - 1}.$$

Then, we have,

$$\mathbb{E}\left[\alpha^{\mathcal{N}(z,\boldsymbol{d})-1} J_\alpha \left(z_{\alpha, \boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}}, \boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1} \right) \right] - \mathbb{E}\left[\alpha^{\mathcal{N}(z,\boldsymbol{d})-1} \cdot \inf_{z \in \mathbb{R}} J_\alpha \left(z, \boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1} \right) \right] = 0.$$

Therefore, we can get the following upper bound of $u_\alpha(z, \boldsymbol{d})$:

$$u_\alpha(z, \boldsymbol{d}) \leq \mathbb{E}^{\pi^\sigma} \left[\sum_{n=1}^{\mathcal{N}(z,\boldsymbol{d})-1} \alpha^{n-1} \ell(z_n) + \alpha^{\mathcal{N}(z,\boldsymbol{d})-1} k \Big(z_{\alpha, \boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}} - z_{\mathcal{N}(z,\boldsymbol{d})} \Big) \Big| z_1 = z, \boldsymbol{D}_0 = \boldsymbol{d} \right]. \tag{B.10}$$

It remains to show the right hand side of (B.10) is uniformly bounded with respect to $\alpha \in [0,1)$. We need the following lemma, the proof of which is deferred to Appendix B.4.

Lemma B.1. *There exists a compact set $[y_L, y_U] \subseteq \mathbb{R}$ such that $\mathcal{X}_{\alpha,\boldsymbol{d}} \subseteq [y_L, y_U]$ for all $\alpha \in [0,1)$ and $\boldsymbol{d} \in [\underline{D}, \bar{D}]^{\tau-L-1}$.*

Lemma B.1 implies that $z_{\alpha,\boldsymbol{d}} \in [y_L, y_U]$ for all $\alpha \in [0,1)$ and $\boldsymbol{d} \in [\underline{D}, \bar{D}]^{\tau-L-1}$. With this result, we can first bound the second term in the right hand side of (B.10). We consider two cases: (i) $\mathcal{N}(z,\boldsymbol{d}) \geq \tau + 1$ and (ii) $\mathcal{N}(z,\boldsymbol{d}) = \tau$. For Case (i), we note that by the dynamics of z_n in (A.1) and the definition of $\mathcal{N}(z,\boldsymbol{d})$ in (B.6), we have

$$z_{\mathcal{N}(z,\boldsymbol{d})} = z_{\mathcal{N}(z,\boldsymbol{d})-1} - \tilde{D}_{\mathcal{N}(z,\boldsymbol{d})-1} \geq z_{\alpha,\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-2}} - \tilde{D}_{\mathcal{N}(z,\boldsymbol{d})-1}.$$

By Assumption 1, we infer that there exists a constant M such that $|\tilde{D}_n| \leq M$ for any n. It then follows that $z_{\alpha,\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}} - z_{\mathcal{N}(z,\boldsymbol{d})} \leq z_{\alpha,\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}} - z_{\alpha,\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-2}} + \tilde{D}_{\mathcal{N}(z,\boldsymbol{d})-1} \leq y_U - y_L + M$. For Case (ii), we need a separate analysis because it might occur that $z_{\tau-1} < z_{\alpha,\boldsymbol{D}_{\tau-2}}$. In this case, from the dynamic of z_n in (A.1) with $z_1 = z$, we have

$$z_\tau = z_1 - \tilde{D}[1, \tau-1] \geq z - (\tau-1)M.$$

It then follows from Lemma B.1 that

$$z_{\alpha,\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}} - z_{\mathcal{N}(z,\boldsymbol{d})} = z_{\alpha,\boldsymbol{D}_{\tau-1}} - z_\tau \leq y_U + (\tau-1)M - z.$$

Because $\alpha \leq 1$, We then conclude that the term

$$\mathbb{E}^{\pi^\sigma}\left[\alpha^{\mathcal{N}(z,\boldsymbol{d})-1} k\Big(z_{\alpha,\boldsymbol{D}_{\mathcal{N}(z,\boldsymbol{d})-1}} - z_{\mathcal{N}(z,\boldsymbol{d})}\Big)\Big| z_1 = z, \boldsymbol{D}_0 = \boldsymbol{d}\right]$$

is uniformly bounded for all $\alpha \in [0,1)$.

We next bound the first term in the right hand side of (B.10). Because the stopping time $\mathcal{N}(z,\boldsymbol{d})$ depends on α and is difficult to analyze directly, we introduce another stopping time that is independent of α and that bounds $\mathcal{N}(z,\boldsymbol{d})$. Specifically, define for a given $(z,\boldsymbol{d}) \in \mathbb{X}$,

$$\tilde{\mathcal{N}}(z,\boldsymbol{d}) = \inf\left\{n \geq \tau : z - \tilde{D}[1, n-1] \leq y_L\right\}. \tag{B.11}$$

By Lemma B.1, we immediately have $\mathcal{N}(z,\boldsymbol{d}) \leq \tilde{\mathcal{N}}(z,\boldsymbol{d})$. Since the holding/backlogging cost is non-negative and $\alpha \leq 1$, we have,

$$\mathbb{E}^{\pi^\sigma}\left[\sum_{n=1}^{\mathcal{N}(z,\boldsymbol{d})-1} \alpha^{n-1}\ell(z_n)\ \Big| z_1 = z, \boldsymbol{D}_0 = \boldsymbol{d}\right] \leq \mathbb{E}^{\pi^\circ}\left[\sum_{n=1}^{\tilde{\mathcal{N}}(z,\boldsymbol{d})-1} \ell(z_n)\ \Big| z_1 = z, \boldsymbol{D}_0 = \boldsymbol{d}\right], \tag{B.12}$$

where π° denotes the policy that does nothing. Note that the right-hand-side of the above inequality does not depend on α, and it is indeed finite as shown in the following result. The proof of Lemma B.2 is also deferred to Appendix B.4.

Lemma B.2. *Suppose Assumption 1 holds. For a given* $(z, \boldsymbol{d}) \in \mathbb{X}$*, we have,*

$$\mathbb{E}^{\pi^\circ}\Big[\sum_{n=1}^{\tilde{\mathcal{N}}(z,\boldsymbol{d})-1} \ell(z_n)\ \Big| z_1 = z, \boldsymbol{D}_0 = \boldsymbol{d}\Big] < \infty, \tag{B.13}$$

where π° *denotes the policy that does nothing.*

Therefore, we can conclude from (B.10) that given $(z, \boldsymbol{d}) \in \mathbb{X}$, $u_\alpha(z, \boldsymbol{d})$ is uniformly bounded for all $\alpha \in [0, 1)$, that is, (B.1) holds. The proof is hence complete. $\square$

B.3 Proof of Theorem 5

In this section we prove Theorem 5. Because we have shown Assumption (W*) and Assumption (B) in Feinberg and Lewis (2018) hold for our system, Part (i) of Theorem 5 immediately follows from Theorem 4.1 in Feinberg and Lewis (2018) (or Theorem 4 in Feinberg et al. (2012))

In the following, we provide a proof of Theorem 5 (iii). Theorem 5 (ii) can be proved by a similar (and slightly simpler) argument.

Let $\{\alpha_n\}_{n=1,2,\ldots}$ be a sequence of discount factors with $\alpha_n \in [0, 1)$ and $\alpha_n \uparrow 1$. According to Proposition 6, we consider a function $S^{d*}_{\mathrm{F},\alpha_n}(\boldsymbol{d})$ that is nonincreasing in $\boldsymbol{d}$. The optimal action at the state $(x, \hat{\boldsymbol{d}})$ for the infinite-horizon discounted problem (8) with discount factor α_n is $Q^*_{\alpha_n}(x, \hat{\boldsymbol{d}}) = (S^{d*}_{\mathrm{F},\alpha_n}(\boldsymbol{d}) - x - r(\hat{\boldsymbol{d}}))^+$. By Theorem 4.5 of Feinberg and Lewis (2018), if Assumption (W*) and Assumption (B) hold, the sequence of $\{Q^*_{\alpha_n}(x, \hat{\boldsymbol{d}}) : n \geq 1\}$ is bounded. It follows that the sequence $\{S^{d*}_{\mathrm{F},\alpha_n}(\boldsymbol{d}) : n \geq 1\}$ is bounded above.

Next we show that the sequence $\{S^{d*}_{\mathrm{F},\alpha_n}(\boldsymbol{d}) : n \geq 1\}$ is also bounded below. We prove it by contradiction. Suppose there exists a subsequence $\{\alpha_{n_m}\}_{m=1,2,\ldots}$ of discount factors such that $\lim_{m\to\infty} S^{d*}_{\mathrm{F},\alpha_{n_m}}(\boldsymbol{d}) = -\infty$. Then we have $Q^*_{\alpha_{n_m}}(x, \hat{\boldsymbol{d}}) \to 0$ as $m \to \infty$. By Theorem 4.3(ii) in Feinberg and Lewis (2018), this suggests that the optimal ordering policy for the long-run average problem is ordering nothing. We denote this policy by π°. By (4), (10) and the definition of $\ell(\cdot)$, we have,

$$\ell(IP_n + r(\hat{\boldsymbol{D}}_{n-1})) = \mathbb{E}[g(IP_n - D[n, n+L] + R[n+1, n+L]|\hat{\boldsymbol{D}}_{n-1}].$$

Hence, we can infer from the dynamics of IP_n in (2) and obtain for $n \geq \tau$,

$$\begin{aligned}
&\mathbb{E}^{\pi^\circ}\Big[\ell(IP_n + r(\hat{\boldsymbol{D}}_{n-1}))\Big| IP_1 = x, \hat{\boldsymbol{D}}_0 = \hat{\boldsymbol{d}}\Big] \\
&= \mathbb{E}\Big[g(x - D[1, n+L] + R[2, n+L])\Big|\hat{\boldsymbol{D}}_0 = \hat{\boldsymbol{d}}\Big] \\
&\geq g\left(\mathbb{E}\Big[x - D[1, \tau] + R[2, \tau]|\hat{\boldsymbol{D}}_0 = \hat{\boldsymbol{d}}\Big] - (n + L - \tau)\mathbb{E}[D_1 - R_1]\right),
\end{aligned}$$

where we have used Jensen's inequality. Then, from Equation (5), we have

$$\begin{aligned}
\Phi^{\pi^\circ}(x, \hat{\boldsymbol{d}}) &= \liminf_{N\to\infty} \frac{1}{N}\mathbb{E}^{\pi^\circ}\Big[\sum_{t=1}^{N} c(IP_t, \hat{\boldsymbol{D}}_{t-1}, Q_t)\Big| IP_1 = x, \hat{\boldsymbol{D}}_0 = \hat{\boldsymbol{d}}\Big] \\
&= \liminf_{N\to\infty} \frac{1}{N}\mathbb{E}^{\pi^\circ}\left[\sum_{t=1}^{N} \ell(IP_t + r(\hat{\boldsymbol{D}}_{t-1}))\Big| IP_1 = x, \hat{\boldsymbol{D}}_0 = \hat{\boldsymbol{d}}\right] \\
&\geq \liminf_{N\to\infty} \frac{1}{N}\mathbb{E}^{\pi^\circ}\left[\sum_{n=1}^{N} g\left(\mathbb{E}\Big[x - D[1,\tau] + R[2,\tau] | \hat{\boldsymbol{D}}_0 = \hat{\boldsymbol{d}}\Big] - (n + L - \tau)\mathbb{E}[D_1 - R_1]\right)\right],
\end{aligned}$$

where the first equality follows from the definition in (5), and the second equality holds because policy π° orders nothing. Note that $g(y) = hy^+ + by^-$ and $\mathbb{E}[D_1 - R_1] \neq 0$. It follows that $\Phi(x, \hat{\boldsymbol{d}})^{\pi^\circ} = \infty$, which contradicts with Assumption (B)(i) verified in Section B.2. Therefore, the sequence $\{S^{d*}_{\mathrm{F},\alpha_n}(\boldsymbol{d}) : n \geq 1\}$ is bounded below.

Because $\{S^{d*}_{\mathrm{F},\alpha_n}(\boldsymbol{d}) : n \geq 1\}$ is bounded, we let $S^{d*}_{\mathrm{F}}(\boldsymbol{d})$ be a limit point of this sequence. Recall that $Q^*_{\alpha_n}(x, \hat{\boldsymbol{d}}) = (S^{d*}_{\mathrm{F},\alpha_n}(\boldsymbol{d}) - x - r(\hat{\boldsymbol{d}}))^+$. We infer that the sequence $\{Q^*_{\alpha_n}(x, \hat{\boldsymbol{d}}) : n \geq 1\}$ has a limit point given by $(S^{d*}_{\mathrm{F}}(\boldsymbol{d}) - x - r(\hat{\boldsymbol{d}}))^+$. By Theorem 4.3(ii) in Feinberg and Lewis (2018), we deduce that the FABS-F policy with base-stock level $S^{d*}_{\mathrm{F}}(\boldsymbol{d})$ is optimal for the long-run average problem. Furthermore, the function $S^{d*}_{\mathrm{F}}(\boldsymbol{d})$ is nonincreasing in $\boldsymbol{d}$. The proof is hence complete.

B.4 Proofs of Auxililary Lemmas (Lemmas B.1 and B.2)

In this section, we prove Lemma B.1 and Lemma B.2, which have been used in the verification of Assumption (B)(ii) in Feinberg and Lewis (2018) in Appendix B.2.

Proof of Lemma B.1. We use an approach inspired by the proofs of Theorem 6 and Lemma 6 in Feinberg et al. (2012), but our analysis is more sophisticated due to the presence of the (uncontrolled) demand vector encoded in the state of the MDP when $L < \tau - 1$. We first introduce some notations. We define for $\boldsymbol{d} \in [\underline{D}, \bar{D}]^{\tau-L-1}$ and $\alpha \in [0, 1)$,

$$\tilde{m}_\alpha(\boldsymbol{d}) = \inf\nolimits_{z\in\mathbb{R}} J_\alpha(z, \boldsymbol{d}), \qquad u_{\alpha,\boldsymbol{d}}(z) = J_\alpha(z, \boldsymbol{d}) - \tilde{m}_\alpha(\boldsymbol{d}). \tag{B.14}$$

By the definition of set $\mathcal{X}_{\alpha,\boldsymbol{d}}$, we have

$$\mathcal{X}_{\alpha,\boldsymbol{d}} = \{z \in \mathbb{R} : J_\alpha(z, \boldsymbol{d}) - \inf_{z\in\mathbb{R}} J_\alpha(z, \boldsymbol{d}) \leq 0\} = \{z \in \mathbb{R} : u_{\alpha,\boldsymbol{d}}(z) \leq 0\}.$$

The main idea of proving Lemma B.1 is to find a function $f : \mathbb{R} \to \mathbb{R}$ that is inf-compact (i.e., all level sets are compact), independent of α and $\boldsymbol{d}$, such that $u_{\alpha,\boldsymbol{d}}(z) \geq f(z)$ for all α and $\boldsymbol{d}$. It then implies that

$$\mathcal{X}_{\alpha,\boldsymbol{d}} = \{z \in \mathbb{R} : u_{\alpha,\boldsymbol{d}}(z) \leq 0\} \subseteq \{z \in \mathbb{R} : f(z) \leq 0\}, \quad \text{for all } \alpha \in [0, 1), \boldsymbol{d} \in [\underline{D}, \bar{D}]^{\tau-L-1}, \tag{B.15}$$

where the set $\{z \in \mathbb{R} : f(z) \leq 0\}$ is compact because f is inf-compact.

We next construct such a function $f(\cdot)$. By definition, we have

$$
\begin{aligned}
u_{\alpha,\boldsymbol{d}}(z) + \tilde{m}_\alpha(\boldsymbol{d}) - \alpha m_\alpha &= J_\alpha(z, \boldsymbol{d}) - \alpha m_\alpha \\
&= \inf_{w \geq z} \left\{ k(w - z) + \ell(w) + \alpha \mathbb{E}[J_\alpha(w - \tilde{D}_+, \boldsymbol{D}_+) | \boldsymbol{D}_0 = \boldsymbol{d}] \right\} - \alpha m_\alpha \\
&= \inf_{w \geq z} \left\{ k(w - z) + \ell(w) + \alpha \mathbb{E}[J_\alpha(w - \tilde{D}_+, \boldsymbol{D}_+) - m_\alpha | \boldsymbol{D}_0 = \boldsymbol{d}] \right\} \\
&\geq \inf_{w \geq z} \left\{ k(w - z) + \ell(w) \right\} \\
&:= J_0(z),
\end{aligned}
$$

where the second equation follows from the optimality equation (A.7), the inequality is due to the the definition of m_α in (B.2), and $J_0(z)$ is simply the optimal value function for the infinite-horizon problem with discount factor $\alpha = 0$ and it is independent of $\boldsymbol{d}$. It follows that

$$
u_{\alpha,\boldsymbol{d}}(z) \geq J_0(z) - \sup_{\boldsymbol{d} \in [\underline{D}, \bar{D}]^{\tau - L - 1},\ \alpha \in [0,1)} (\tilde{m}_\alpha(\boldsymbol{d}) - \alpha m_\alpha) := f(z). \tag{B.16}
$$

It remains to show f in (B.16) is well defined (i.e. finite) and it is inf-compact. It is easy to see that J_0 is well defined on $\mathbb{R}$ and it is an inf-compact function by using a similar argument as the one used in Appendix B.1. The more challenging part is to show f is finite and well defined, i.e.,

$$
\lambda^* := \sup_{\boldsymbol{d} \in [\underline{D}, \bar{D}]^{\tau - L - 1}, \alpha \in [0,1)} (\tilde{m}_\alpha(\boldsymbol{d}) - \alpha \cdot m_\alpha) < \infty. \tag{B.17}
$$

To prove (B.17), we first note that

$$
\begin{aligned}
\lambda^* &= \sup_{\alpha \in [0,1),\ \boldsymbol{d} \in [\underline{D}, \bar{D}]^{\tau - L - 1}} (\tilde{m}_\alpha(\boldsymbol{d}) - m_\alpha + (1 - \alpha) \cdot m_\alpha) \\
&\leq \sup_{\alpha \in [0,1),\ \boldsymbol{d} \in [\underline{D}, \bar{D}]^{\tau - L - 1}} (\tilde{m}_\alpha(\boldsymbol{d}) - m_\alpha) + \sup_{\alpha \in [0,1)} (1 - \alpha) \cdot m_\alpha
\end{aligned} \tag{B.18}
$$

In the proof of Lemma 6 of Feinberg et al. (2012), it has been that $\sup_{\alpha \in [0,1)} (1 - \alpha) \cdot m_\alpha < +\infty$. Hence, it suffices to prove

$$
\sup_{\alpha \in [0,1),\ \boldsymbol{d} \in [\underline{D}, \bar{D}]^{\tau - L - 1}} (\tilde{m}_\alpha(\boldsymbol{d}) - m_\alpha) < \infty. \tag{B.19}
$$

To prove (B.19), we note from the definition of $\tilde{m}_\alpha(\boldsymbol{d})$ in (B.14) that $\tilde{m}_\alpha(\boldsymbol{d}) \leq J_\alpha(z, \boldsymbol{d})$ for any z. In addition, there exists $(z^\alpha, \boldsymbol{d}^\alpha) \in \mathcal{X}_\alpha$ such that $m_\alpha = J_\alpha(z^\alpha, \boldsymbol{d}^\alpha)$ by its definition. It then follows that for any z,

$$
\tilde{m}_\alpha(\boldsymbol{d}) - m_\alpha \leq J_\alpha(z, \boldsymbol{d}) - J_\alpha(z^\alpha, \boldsymbol{d}^\alpha). \tag{B.20}
$$

The key idea to bound the right-hand-side of the above equation is to construct a feasible policy, denoted by π^δ, and select a proper initial adjusted inventory position z_* so that the inventory system with an initial state $(z_*, \boldsymbol{d})$ under the policy π^δ will be coupled at some future time with another system with an initial state $(z^\alpha, \boldsymbol{d}^\alpha)$ under the infinite-horizon discounted optimal policy denoted by π_α. Let z_n^δ (z_n^α) and Q_n^δ (Q_n^α) denote the adjusted inventory position and manufacturing quantity in the system under policy π^δ (policy π_α respectively), and the dependency of these quantities on the initial state has been made implicit for notational simplicity. For simplicity, the latter system will be referred to the optimal system in the following, and we denote by z_n^α the adjusted inventory position of this optimal system at period n with $z_1^\alpha = z^\alpha$. Now we specify the feasible policy π^δ as follows: Do not manufacture before period $\tau - L$; At period $\tau - L$, we manufacture up to level $z_{\tau-L}^\alpha + Q_{\tau-L}^\alpha$ and then switch to policy π_α onward. We will choose the initial adjusted inventory position z_* to be very low (to be specified later), so that the inventory system under policy π^δ will have an adjusted inventory position lower than $z_{\tau-L}^\alpha$. Then this system will be coupled with the optimal system at period $\tau - L$, and behaves the same as the optimal system afterwards. This follows by the design of the policy π^δ and the fact that $\boldsymbol{D}_{n-1} = (D_{n-1}, ..., D_{n-\tau+L+1})$ is independent of the initial demand state $\boldsymbol{D}_0$ after period $\tau - L$. Based on this discussion, we infer that

$$
\begin{aligned}
J_\alpha(z_*, \boldsymbol{d}) - J_\alpha(z^\alpha, \boldsymbol{d}^\alpha) &= \mathbb{E}\Big[\sum_{n=1}^{\tau-L} \alpha^{n-1} k Q_n^\delta + \alpha^{n-1}\ell(z_n^\delta + Q_n^\delta)\Big] - \mathbb{E}\Big[\sum_{n=1}^{\tau-L} \alpha^{n-1} k Q_n^\alpha + \alpha^{n-1}\ell(z_n^\alpha + Q_n^\alpha)\Big], \\
&\le \mathbb{E}\Big[\sum_{n=1}^{\tau-L} \alpha^{n-1} k (Q_n^\delta - Q_n^\alpha)\Big] + \mathbb{E}\Big[\sum_{n=1}^{\tau-L-1} \alpha^{n-1}\ell(z_n^\delta + Q_n^\delta)\Big],
\end{aligned} \tag{B.21}
$$

It remains to specify z_* and bound the right-hand-side of (B.21). By Assumption 1, there exists a constant $M \ge 0$ such that $|\tilde{D}_n| \le M$. Set $z_* = z^\alpha - 2(\tau - L - 1)M$. Then one can readily verify $z_{\tau-L}^\delta \le z_{\tau-L}^\alpha$. To see this, note that the policy π^δ does not manufacture before period $\tau - L$, which implies that $z_{\tau-L}^\delta = z_* - \sum_{n=1}^{\tau-L-1} \tilde{D}_n \le z_* + (\tau - L - 1)M$ by (A.1). Moreover, it is clear that for the optimal system we have $z_{\tau-L}^\alpha \ge z^\alpha - \sum_{n=1}^{\tau-L-1} \tilde{D}_n \ge z^\alpha - (\tau - L - 1)M$. Hence by the choice of z_* we obtain $z_{\tau-L}^\delta \le z_{\tau-L}^\alpha$. Next we bound the right-hand-side of (B.21). By the design of the policy π^δ, we obtain

$$
\begin{aligned}
\mathbb{E}\Big[\sum_{n=1}^{\tau-L} \alpha^{n-1}(Q_n^\delta - Q_n^\alpha)\Big] &= \alpha^{\tau-L-1} \cdot \mathbb{E}\Big[Q_{\tau-L}^\delta - Q_{\tau-L}^\alpha\Big] - \mathbb{E}\Big[\sum_{n=1}^{\tau-L-1} \alpha^{n-1} Q_n^\alpha\Big] \\
&= \alpha^{\tau-L-1} \cdot \mathbb{E}\Big[z_{\tau-L}^\alpha - z_{\tau-L}^\delta\Big] - \mathbb{E}\Big[\sum_{n=1}^{\tau-L-1} \alpha^{n-1} Q_n^\alpha\Big].
\end{aligned}
$$

It is clear that $z_{\tau-L}^\delta = z_* - \sum_{n=1}^{\tau-L-1} \tilde{D}_n \ge z^\alpha - 3(\tau - L - 1)M$. In addition, $z_{\tau-L}^\alpha = z^\alpha + \sum_{n=1}^{\tau-L-1} Q_n^\alpha - \sum_{n=1}^{\tau-L-1} \tilde{D}_n \le z^\alpha + \sum_{n=1}^{\tau-L-1} Q_n^\alpha + (\tau - L - 1)M$. This suggests that for any $\alpha \in [0, 1)$,

$$
\mathbb{E}\Big[\sum_{n=1}^{\tau-L} \alpha^{n-1}(Q_n^\delta - Q_n^\alpha)\Big] \le \alpha^{\tau-L-1} \cdot \mathbb{E}\Big[\sum_{n=1}^{\tau-L-1} Q_n^\alpha + 4(\tau - L - 1)M\Big] - \mathbb{E}\Big[\sum_{n=1}^{\tau-L-1} \alpha^{n-1} Q_n^\alpha\Big]
$$

$$\leq 4(\tau - L - 1)M.$$

Finally, we bound the term $\mathbb{E}\Big[\sum_{n=1}^{\tau-L-1} \alpha^{n-1}\ell(z_n^\delta + Q_n^\delta)\Big]$ in (B.21), which is clearly bounded by $\mathbb{E}\Big[\sum_{n=1}^{\tau-L-1} \ell(z_n^\delta + Q_n^\delta)\Big]$. From the definition of $\ell(\cdot)$ and the fact that the policy π^δ does not manufacture before period $\tau - L$, we have

$$\mathbb{E}\Big[\sum_{n=1}^{\tau-L-1} \alpha^{n-1}\ell(z_n^\delta + Q_n^\delta)\Big] = \mathbb{E}\Big[\sum_{n=1}^{\tau-L-1} \alpha^{n-1} g\Big(z_n^\delta - \sum_{i=1}^{L} r_{n,n+i} - \sum_{i=0}^{L} D_{n+i} + \sum_{i=1}^{L} R_{n+i}\Big)\Big].$$

By the definition of function $g(\cdot)$, we know that $g(y) \leq (h \vee b)|y|$ for any y. In addition, by Assumption 1, there exists a constant U that is independent of n such that for $n = 1, 2, ...,$

$$\Big| - \sum_{i=1}^{L} r_{n,n+i} - \sum_{i=0}^{L} D_{n+i} + \sum_{i=1}^{L} R_{n+i}\Big| \leq U.$$

Moreover, for $n < \tau - L$, we have $|z_n^\delta| = |z_* - \sum_{i=1}^{n-1} \tilde{D}_i| \leq |z^\alpha| + 3(\tau - L - 1)M$ by the definition of z_*. Because there is a constant C such that $|z^\alpha| \leq C$ uniformly in $\alpha \in [0, 1)$ (see Theorem 4.2 in Feinberg and Lewis (2018)), we can then infer that

$$\mathbb{E}\Big[\sum_{n=1}^{\tau-L-1} \alpha^{n-1}\ell(z_n^\delta + Q_n^\delta)\Big] \leq (\tau - L - 1) \cdot (U + C + 3(\tau - L - 1)M).$$

Hence we can obtain from (B.21) that

$$\tilde{m}_\alpha(\boldsymbol{d}) - m_\alpha \leq J_\alpha(z_*, \boldsymbol{d}) - J_\alpha(z^\alpha, \boldsymbol{d}^\alpha) \leq (\tau - L - 1) \cdot (4kM + U + C + 3(\tau - L - 1)M). \quad \text{(B.22)}$$

The constant on the right-hand-side of the above equation is independent of α and $\boldsymbol{d}$, and it is finite. Hence, we obtain (B.19). The proof is therefore complete. $\square$

Proof of Lemma B.2.

By the definition of $\ell(\cdot)$ in (11) and the definition of $r_{n,n+i}$ in (13), we obtain

$$\begin{aligned}
&\mathbb{E}^{\pi^\circ}\Big[\sum_{n=1}^{\tilde{\mathcal{N}}(z,\boldsymbol{d})-1} \ell(z_n) \;\Big| z_1 = z, \boldsymbol{D}_0 = \boldsymbol{d}\Big] \\
&= \mathbb{E}^{\pi^\circ}\Big[\sum_{n=1}^{\tilde{\mathcal{N}}(z,\boldsymbol{d})-1} g\Big(z_n - \sum_{i=1}^{L} r_{n,n+i} - \sum_{i=0}^{L} D_{n+i} + \sum_{i=1}^{L} R_{n+i}\Big)\Big| z_1 = z, \boldsymbol{D}_0 = \boldsymbol{d}\Big].
\end{aligned}$$

Because the policy π° does not order, we get

$$z_n = z_1 - \tilde{D}[1, n-1] = z - \tilde{D}[1, n-1], \qquad \text{for } n < \tilde{\mathcal{N}}(z, \boldsymbol{d}).$$

By Assumption 1, there exist positive constants M and U such that $|\tilde{D}_n| \le M$ and $|-\sum_{i=1}^{L} r_{n,n+i} - \sum_{i=0}^{L} D_{n+i} + \sum_{i=1}^{L} R_{n+i}| \le \mathrm{U}$ for all n. It follows that for $n < \tilde{\mathcal{N}}(z, \boldsymbol{d})$,

$$\left| z_n - \sum_{i=1}^{L} r_{n,n+i} - \sum_{i=0}^{L} D_{n+i} + \sum_{i=1}^{L} R_{n+i} \right| \le |z| + \tilde{\mathcal{N}}(z, \boldsymbol{d}) M + \mathrm{U}.$$

Thus, we have

$$\mathbb{E}^{\pi^\circ}\Big[\sum_{n=1}^{\tilde{\mathcal{N}}(z,\boldsymbol{d})-1} \ell(z_n) \Big| z_1 = z, \boldsymbol{D}_0 = \boldsymbol{d} \Big] \le \mathbb{E}\Big[(h \vee b)(\tilde{\mathcal{N}}(z, \boldsymbol{d}) - 1) \Big(|z| + \tilde{\mathcal{N}}(z, \boldsymbol{d}) M + \mathrm{U} \Big) \Big]. \tag{B.23}$$

Therefore, if we can prove that the first two moments of $\tilde{\mathcal{N}}(z, \boldsymbol{d})$ are finite, then inequality (B.13) holds. It is difficult to analyze the moments of the stopping time $\tilde{\mathcal{N}}(z, \boldsymbol{d})$ in (B.11) directly, because the the random variables $\{\tilde{D}_n\}_{n=1,2,\dots}$ defined in (A.2) are not independent when $L < \tau - 1$. Hence, we introduce a new random variable that bounds $\tilde{\mathcal{N}}(z, \boldsymbol{d})$. Define

$$\tilde{\mathcal{T}}(z, \boldsymbol{d}) = \inf \left\{ n \ge \tau : \sum_{j=1}^{n-\tau+L} \left(\sum_{i=1}^{\tau} p_i - 1 \right) D_j + \sum_{j=1}^{n-\tau+L} \epsilon_j \le y_L - y(z, \boldsymbol{d}) \right\},$$

where

$$y(z, \boldsymbol{d}) = z + \sum_{j=0}^{\tau-L-2} \sum_{i=j+L+2}^{\tau} p_i d_{-j} + \sum_{j=0}^{\tau-L-2} \left(\sum_{i=1}^{L+1+j} p_i - 1 \right) \underline{D} + (\tau - L - 1)\bar{\epsilon}.$$

Note that the sequence of random variables $\{(\sum_{i=1}^{\tau} p_i - 1) D_j + \epsilon_{j+1} : j \ge 1\}$ is i.i.d. Hence, $\tilde{\mathcal{T}}(z, \boldsymbol{d})$ is essentially the first passage time of a random walk. By the definition of $\tilde{D}_n$ in (16), we have for $(z_1, \boldsymbol{D}_0) = (z, \boldsymbol{d}) \in \mathbb{X}$ and $n \ge \tau$,

$$\begin{aligned} z - \tilde{D}[1, n-1] = z &+ \sum_{j=1}^{n-\tau+L} \left(\sum_{i=1}^{\tau} p_i - 1 \right) D_j + \sum_{j=1}^{n-1} \epsilon_{j+1} \\ &+ \sum_{j=0}^{\tau-L-2} \sum_{i=j+L+2}^{\tau} p_i d_{-j} + \sum_{j=0}^{\tau-L-2} \left(\sum_{i=1}^{L+1+j} p_i - 1 \right) D_{n-1-j} \\ &\le \sum_{j=1}^{n-\tau+L} \left(\sum_{i=1}^{\tau} p_i - 1 \right) D_j + \sum_{j=1}^{n-\tau+L} \epsilon_{j+1} + y(z, \boldsymbol{d}). \end{aligned}$$

This directly implies $\tilde{\mathcal{N}}(z, \boldsymbol{d}) \le \tilde{\mathcal{T}}(z, \boldsymbol{d})$. Because we assume $\sum_{t=1}^{\tau} p_t < 1$, we can then infer from Theorem 2.1 of Gut (1974) that for a given $(z, \boldsymbol{d}) \in \mathbb{X}$,

$$\mathbb{E}[\tilde{\mathcal{N}}(z, \boldsymbol{d})] \le \mathbb{E}[\tilde{\mathcal{T}}(z, \boldsymbol{d})] < \infty, \quad \mathbb{E}[\tilde{\mathcal{N}}^2(z, \boldsymbol{d})] \le \mathbb{E}[\tilde{\mathcal{T}}^2(z, \boldsymbol{d})] < \infty.$$

Hence, by (B.23), we have

$$\mathbb{E}^{\pi^\circ}\Big[\sum_{n=1}^{\tilde{\mathcal{N}}(z,\boldsymbol{d})-1} \ell(z_n) \ \Big| z_1 = z, \boldsymbol{D}_0 = \boldsymbol{d} \Big] < \infty.$$

This concludes the proof of Lemma B.2. ∎