Joint Inventory and Scheduling Control in a Repair Facility

Citation for published version (APA):
Özkan, E., & van Houtum, G-J. J. A. N. (2023). Joint Inventory and Scheduling Control in a Repair Facility. *Operations Research*, *71*(5), 1498-1514. https://doi.org/10.1287/opre.2023.2459

DOI:
10.1287/opre.2023.2459

Document status and date:
Published: 01/09/2023

Document Version:
Accepted manuscript including changes made at the peer-review stage

Joint Inventory and Scheduling Control in a Repair Facility

Erhun Özkan
Koç University, Istanbul, Turkey, erhozkan@ku.edu.tr

Geert-Jan van Houtum
Eindhoven University of Technology, Eindhoven, the Netherlands, g.j.v.houtum@tue.nl

We study inventory and repair scheduling decisions of a maintenance service provider for repairable capital goods. Due to high downtime costs, the service provider keeps spare parts on stock to replace broken parts quickly. The service provider should determine the inventory level of spare parts for each component and the repair scheduling policy. Furthermore, in case of a stock-out, the service provider should decide whether to backorder the demand or execute an emergency repair, which is an urgent but expensive repair operation for a broken part followed by a fast form of installation. The objective is to minimize the long-run average inventory holding, backorder, and emergency repair costs. We formulate the repairable network as a closed queueing system and consider an asymptotic regime in which the repair facility is in the conventional heavy-traffic regime. Then, we formulate and solve a Brownian control problem (BCP). From the optimal BCP solution, we derive a simple and intuitive decision rule stating if the emergency repairs are necessary to achieve a close-to-optimal system performance. Moreover, we propose a simple, intuitive, and easy-to-implement heuristic control policy and demonstrate its close-to-optimal performance via numerical experiments.

Key words: Spare parts, inventory control, scheduling, asymptotic analysis

---

1. Introduction

Capital goods are machines or products that are used in the production of goods or service deliveries. Some examples are lithography machines (used by semiconductor manufacturers), medical systems, trains, and baggage handling systems in airports. Capital goods can be very expensive. For example, the price of an EUV lithography system is 100+ million Euros, the price of an MRI scanner is 1-2 million Euros, and a baggage handling system at a major airport can cost up to 300-400 million Euros. Due to high maintenance and downtime costs, acquisition costs of some capital goods constitute only a fraction (e.g., one third) of the total life cycle costs (see Öner et al. (2007) and Kim et al. (2015)). Therefore, maintenance of capital goods is important and many capital-good users (customers) outsource the maintenance activities to either the suppliers or third-party service providers. For example, there are cases in which customers pay a fixed service cost per year and the service providers manage everything related to maintenance. There are even cases in which customers buy the function of the system rather than the capital good implying a function-oriented market instead of a product-oriented one (see Kim et al. (2015) and chapter 1 of van Houtum and Kranenburg (2015) for details).

Due to the high downtime costs, maintenance service providers keep spare parts on stock. If a critical component of a capital good breaks down, the broken part is replaced with a spare part to prevent long and costly downtime. If the broken part is repairable, it is sent to a repair facility for future usage. Because spare parts of capital goods are generally expensive, there is an interest in efficient control of spare parts inventory systems.

We study inventory and repair scheduling decisions of a maintenance service provider for capital goods with repairable components, which we denote as stock keeping units (SKUs) when we refer to them as article numbers and as parts when we refer to physical units. There is a fixed number of customers using multiple SKUs. The service provider keeps spare parts on stock so that whenever a part used by a customer breaks down, a demand for a ready-for-use part occurs and the broken part joins the repair queue. If there is on-hand inventory (consisting of ready-for-use parts), the service provider fulfills the demand from the on-hand inventory immediately. Otherwise, if there is no on-hand inventory, the service provider either backorders the demand or uses an emergency repair to fulfill the demand. An emergency repair is an urgent repair operation for a broken part followed by a fast form of installation. Although emergency repairs are expensive in general, they can be useful due to the high downtime costs. The parts of each SKU circulate in the system as ready-for-use, broken, or installed parts. We assume that broken parts are always repairable. Therefore, the total number of parts of each SKU in the network is fixed and thus the repairable network can be formulated as a closed queueing system as depicted in Figure 1.

The objective of the service provider is to minimize the long-run average inventory holding, backorder, and emergency repair costs by determining:

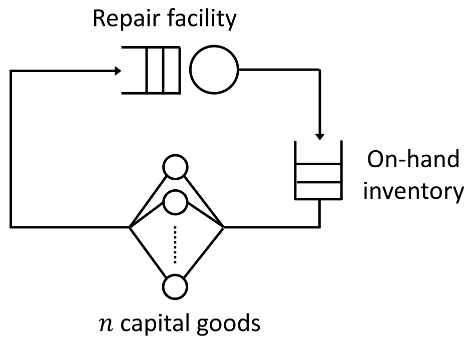


Figure 1 A schematic representation of the repairable network.

i. The initial inventory level for each SKU.

ii. The repair scheduling policy in the repair facility, that is, the service provider should decide which broken part should receive repair priority at any given time. We assume that only the work-conserving and non-preemptive repair scheduling policies are allowed in the repair facility.

iii. Demand fulfillment policy. If there is no on-hand inventory upon a part breakdown, the service provider should decide whether to backorder the demand or execute an emergency repair to fulfill the demand.

Deriving an optimal control policy for the aforementioned problem is challenging due to the curse of dimensionality. Therefore we use asymptotic analysis. Specifically, we construct an asymptotic control problem and solve it optimally. Based on that solution, we formulate a *heuristic control policy* for the original (non-asymptotic) system, to which we refer to as the *pre-limit system*. In the asymptotic regime that we consider, both the number of capital goods and the emergency repair costs tend to infinity and the breakdown rate per part of each SKU tends to zero such that the total breakdown rate of each SKU, which is the product of the number of capital goods using that SKU and the breakdown rate per part of that SKU, converges to a non-degenerate limit. Furthermore, the repair facility is in the conventional heavy-traffic regime, implying a fully utilized repair facility in the limit. The aforementioned asymptotic regime is consistent with the following observations: (i) The reliability of the capital goods has increased significantly over time due to technological advancements such that the breakdown rate of a part can be very small (see chapter

1.2 of van Houtum and Kranenburg (2015)). (ii) Emergency repairs are expensive in general. (iii) It is economically undesirable to have idle resources.

Under the aforementioned asymptotic regime, we formulate a Brownian control problem (BCP) (see Harrison (1988)). Because the BCP is multi-dimensional, we formulate an equivalent workload formulation (EWF) of the BCP (see Harrison and Van Mieghem (1997)). The EWF is single dimensional and solvable. By utilizing the optimal EWF solution, we make the following main contributions.

1. The EWF leads to a simple and intuitive decision rule stating if the emergency repairs are necessary. This rule achieves a close-to-optimal system performance for the pre-limit system. To the best of our knowledge, such a rule is not available in the literature. That rule can be very helpful for repair centers to decide on if the emergency repairs are needed.
2. The optimal EWF solution backorders demands of a single SKU at all times and the index of that SKU can change with the workload level. Therefore, the optimal EWF solution implies a workload-dependent repair prioritization policy for the pre-limit system. The parts of SKUs that are not backordered are repaired and those repaired parts will breakdown again and increase the future workload due to the closed-network structure. Therefore, while making the backordering decisions, the optimal solution takes into account not only the instantaneous backorder costs but also the breakdown rates of the SKUs. As the workload level increases, the optimal solution becomes more forward-looking. Specifically, as the workload level increases, the instantaneous backorder costs become less important and the breakdown rates, that is, the effect of the repaired parts on the future workload levels, become more important for the backordering decisions. Finally, the optimal EWF solution allows emergency repairs for at most one particular SKU (see Theorems 1 and 2).
3. From the optimal EWF solution, we derive a simple, intuitive, and easy-to-implement heuristic control policy for the pre-limit system. This heuristic policy follows the optimal EWF solution for the repair scheduling decisions and the demand fulfillment policy. For the initial

inventory levels of the SKUs, we formulate a simple local search algorithm (LS). By numerical experiments, we show that the heuristic control policy has a reasonably small optimality gap for systems with 80% or more utilization rate, 2 SKUs, and 20 or more capital goods. Furthermore, in the numerical experiments, the LS performs almost as good as the inventory enumeration does and has a reasonable computation time (see Section 6.2).

4. Under a mild assumption on the first-order system parameters (see Assumption 2), we simplify the EWF such that its solution becomes greedy in the sense that only the demands of one particular SKU are allowed to be backordered and a non-zero inventory is kept only for one particular SKU (see Theorems 3 and 4). Consequently, the heuristic control policy for the pre-limit system simplifies as well. For example, under that policy, the repair scheduling decisions become less workload-dependent than before and low initial inventory levels are kept for all SKUs except the one with the "lowest" inventory holding cost.

The main contributions are explained in more detail below.

There are I different SKUs in the system and we let $\mathcal{I} := \{1, 2, \ldots, I\}$ denote the set of SKUs. For all $i \in \mathcal{I}$, b_i denotes the backorder cost per unit time per backordered demand of SKU i, $\tilde{c}_i$ denotes the cost of an emergency repair for a broken part of SKU i, and $1/\tilde{\lambda}_i$ denotes the average time an installed part of SKU i spends until its breakdown. If $\min_{j\in\mathcal{I}} b_j\mu_j/\tilde{\lambda}_j \leq \min_{k\in\mathcal{I}} \tilde{c}_k\mu_k$, the optimal EWF solution does not make any emergency repairs. Otherwise, if $\min_{j\in\mathcal{I}} b_j\mu_j/\tilde{\lambda}_j > \min_{k\in\mathcal{I}} \tilde{c}_k\mu_k$, the optimal EWF solution makes emergency repairs to keep the workload in the repair facility below a threshold. The aforementioned decision rule can be interpreted as follows. For all $i \in \mathcal{I}$, $\tilde{c}_i\mu_i = \tilde{c}_i/(1/\mu_i)$ and thus it represents the cost per unit repair time that is saved by having an emergency repair elsewhere for one part of SKU i. If a broken part of SKU i is repaired and installed, it will take on average $1/\tilde{\lambda}_i$ amount of time for that part to break down again and return back to the repair facility. Therefore, if that part is repaired and it is very likely that there will be backordered parts of SKU i when the part breaks down again, then the associated saved backorder cost is $b_i/\tilde{\lambda}_i$, and the saved backorder cost per unit repair time that is spent is $(b_i/\tilde{\lambda}_i)/(1/\mu_i) = b_i\mu_i/\tilde{\lambda}_i$.

The optimal EWF solution determines if the emergency repairs are necessary by comparing the minimum saved backorder cost per unit of spent repair time with the minimum cost per unit of saved repair time via an emergency repair.

We derive a simple, intuitive, and easy-to-implement heuristic control policy from the optimal EWF solution. If $\min_{j\in\mathcal{I}} b_j\mu_j/\tilde{\lambda}_j \leq \min_{k\in\mathcal{I}} \tilde{c}_k\mu_k$, we propose the *no-emergency-repair (NER) policy* under which emergency repairs are never used. Otherwise, if $\min_{j\in\mathcal{I}} b_j\mu_j/\tilde{\lambda}_j > \min_{k\in\mathcal{I}} \tilde{c}_k\mu_k$, we propose the *barrier policy*, under which emergency repairs are used to keep the workload in the repair facility below a barrier level (or threshold). Emergency repairs are used only when the workload reaches the barrier level and only for the broken parts of a single SKU whose index is in the set $\arg\min_{k\in\mathcal{I}} \tilde{c}_k\mu_k$. Under both the NER and the barrier policies, the SKU whose broken parts receive the least amount of repair priority can change with the workload level due to the non-greedy nature of the optimal EWF solution. Finally, the optimal inventory levels are computed by a simple LS, which is a commonly used technique in the inventory control of spare parts (see van Houtum and Kranenburg (2015)).

Under a mild assumption on the first-order system parameters, the optimal EWF solution and thus our proposed policy simplifies. Under that assumption, we propose two different algorithms to determine the initial inventory levels. The first algorithm is similar to the algorithms in the literature (e.g., Wein (1992) and Ata and Barjesteh (2022)). It keeps a small and equal inventory of spare parts (that is, a small and equal safety stock) for all SKUs except the one with the "lowest" inventory holding cost. The inventory level for the latter SKU is expressed in closed-form under the NER policy and can be computed efficiently by simulation under the barrier policy. The second algorithm is a simple LS. By numerical experiments, we show that the performances of the two algorithms are reasonably well and close to each other (see Section 6.3).

The rest of the paper is organized as follows. We present a literature review in Section 2. We present the model, the BCP, and the EWF in Section 3. We solve the EWF in Section 4. In Section 5, we simplify the EWF and its solution under a mild assumption. Then, we present numerical experiments in Section 6. Finally, we present some future research directions in Section 7. All the proofs are presented in the online appendix (OA).

2. Literature Review

Our paper is closely related to the literature on spare parts inventory control. For recent literature reviews, see Basten and van Houtum (2014) and Driessen et al. (2015). An important feature of spare parts inventory control is the use of lateral (trans)shipments and emergency shipments. If a local warehouse is out-of-stock upon a demand arrival, the demand can be satisfied from a nearby local warehouse by a so-called lateral transshipment (see Axsäter (1990), Alfredsson and Verrijdt (1999), Kranenburg and van Houtum (2009), Paterson et al. (2011)) or from an upstream warehouse or supplier by a so-called emergency shipment (see Muckstadt and Thomas (1980), Alfredsson and Verrijdt (1999), Özkan et al. (2015)). In this way, long downtimes are avoided for the capital goods that are supported. The use of lateral and emergency shipments complicates the analysis and therefore authors often make simplifying assumptions for other aspects. For example, it is often assumed that the repair facility (or the supplier) has infinite capacity, which can lead to significant errors in performance evaluation especially when the capacity of the repair facility is tight in reality (see Sleptchenko et al. (2002)). Therefore, there are papers considering finite repair capacity (see Pyke (1990), Sleptchenko et al. (2002, 2005), Caggiano et al. (2006), Tiemessen and van Houtum (2013)) and expedited repairs (see Arts et al. (2016) and Drent and Arts (2021)). Because capacity allocation decisions are important under finite repair capacity, there are papers studying repair scheduling decisions (see Hausman and Scudder (1982), Pyke (1990), Sleptchenko et al. (2005), Caggiano et al. (2006), Adan et al. (2009), Tiemessen and van Houtum (2013)). There are also papers studying selective emergency repair decisions. For example, Verrijdt et al. (1998) study the effect of executing emergency repairs depending on the system state rather than with respect to a simple rule. van der Heijden et al. (2013) study selecting different repair leadtime options for different SKUs. Bitton et al. (2019) study joint inventory and emergency repair control in aircraft maintenance. They study a system with multiple SKUs and consider simple and SKU-dependent emergency repair policies.

Our paper has major differences from the existing literature on spare parts inventory control. First, we jointly consider finite repair capacity and selective emergency repairs. Second, the existing literature considers an open network with exogenous demand, whereas we consider a closed network with endogenous demand. Third, we are the first who apply an asymptotic analysis and succeed in deriving intuitive structural results.

Our paper is also related to the literature on the control of multi-class make-to-stock manufacturing systems. Early examples of papers in that stream of literature are Zheng and Zipkin (1990) and Wein (1992). Zheng and Zipkin (1990) consider a system with two symmetric products and prove that giving manufacturing priority to the longest queue outperforms the FCFS policy. Wein (1992) studies a multi-class system with the objective of minimizing the long-run average holding and backorder costs. Wein (1992) assumes that the system operates in the heavy-traffic regime and then formulates a BCP and its EWF. The optimal EWF solution implies a barrier type control policy under which the on-hand inventory level is never allowed to exceed a threshold. A recent study in that literature is Ata and Barjesteh (2022), which extends Wein (1992) by considering outsourcing and dynamic pricing. Similar to Wein (1992), Ata and Barjesteh (2022) formulate a BCP and its EWF. The optimal EWF solution implies a barrier type policy under which both the on-hand inventory and the backorder levels are never allowed to exceed threshold values, implying a two-sided barrier. A comprehensive literature review of control of multi-class make-to-stock manufacturing systems can be found in Ata and Barjesteh (2022). A study that is closely related to our work is by Rubino and Ata (2009). They consider control of a multi-class make-to-order manufacturing system with parallel servers, order cancellations, and outsourcing. They formulate a BCP and its EWF. Similar to the case in our paper, their optimal EWF solution is non-greedy in the sense that the resource allocation decisions depend on the workload level.

There are major differences between our study and the papers about control of make-to-stock or make-to-order manufacturing systems. Because the manufactured products are consumable, papers in the aforementioned literatures consider an open system with exogenous demand (unless there is pricing control). Because we consider repairable SKUs, the total number of parts per SKU in the network is fixed, where the state of each part is broken, ready-for-use, or installed. The number of broken parts affects the number of installed parts which, in turn, affects the demand rate for the repair facility. This feature naturally leads to a closed queueing network formulation and affects the structure of the optimal EWF solution. For example, if the backorder costs are sufficiently smaller than the emergency repair costs, there is no need to keep the number of broken parts below a barrier level unlike the cases in Rubino and Ata (2009) and Ata and Barjesteh (2022).

3. Model Description

We formulate the repair facility as a single server queue and assume that emergency repairs are done instantaneously by outsourcing. We present the stochastic primitives, introduce the asymptotic regime, and present the model in Section 3.1. We present the fluid and diffusion scaled processes in Section 3.2. Finally, we present the BCP and the associated EWF in Sections 3.3 and 3.4, respectively.

3.1. Stochastic Primitives and the Asymptotic Regime

We let $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}_{++}$, $\mathbb{N}$, $\mathbb{N}_+$ denote the set of real numbers, nonnegative real numbers, strictly positive real numbers, nonnegative integers, and strictly positive integers, respectively. There are I different SKUs in the system and we let $\mathcal{I} := \{1, 2, \ldots, I\}$ denote the set of SKUs. We consider a sequence of systems indexed by $n \in \mathbb{N}_+$ and we let $n \to \infty$. In the nth system, there are n capital goods in total, a single part can be installed at each capital good, and $n\alpha_i^n$ capital goods use parts of SKU i such that $\alpha_i^n \in \mathbb{R}_+$ and $\sum_{i\in\mathcal{I}} \alpha_i^n = 1$ for all $i \in \mathcal{I}$ and $n \in \mathbb{N}_+$.

An inter-breakdown time denotes the time between the installation of a ready-for-use part and its breakdown. In the nth system, the inter-breakdown times for the parts of SKU i are independent and identically distributed (i.i.d.) and have an exponential distribution with mean $1/\lambda_i^n \in \mathbb{R}_{++}$ for all $i \in \mathcal{I}$ and $n \in \mathbb{N}_+$.

For all $i \in \mathcal{I}$, let $\{v_{ik}, k \in \mathbb{N}_+\}$ be a strictly positive and i.i.d. sequence of random variables with mean $1/\mu_i$ and coefficient of variation $\sigma_i \in \mathbb{R}_+$. We let v_{ik} denote the repair time of the kth broken part of SKU i for all $i \in \mathcal{I}$ and $k \in \mathbb{N}_+$. For all $i \in \mathcal{I}$, $k \in \mathbb{N}_+$, and $t \in \mathbb{R}_+$, let $V_i(0) := 0$ and

$$V_i(k) := \sum_{l=1}^{k} v_{il}, \qquad R_i(t) := \sup\{k \in \mathbb{N} : V_i(k) \le t\}.$$

Then, R_i is a renewal process such that $R_i(t)$ denotes the number of parts of SKU i repaired up to time t if the repair facility spends 100% of its time for the repair of broken parts of SKU i during the time interval $[0,t]$.

We assume that the sequences of random variables associated with the inter-breakdown times and repair times are mutually independent of each other and all other stochastic primitives. Let

$$\rho^n := \sum_{i\in\mathcal{I}} \frac{n\alpha_i^n \lambda_i^n}{\mu_i}$$

denote the load on the repair facility in the nth system. The following assumption sets up the asymptotic regime.

Assumption 1. 1. *$\alpha_i^n \to \alpha_i \in (0,1)$ and $n\lambda_i^n \to \lambda_i \in \mathbb{R}_{++}$ for all $i \in \mathcal{I}$ as $n \to \infty$.*

2. *$\sqrt{n}\,(\rho^n - 1) \to \theta \in (-\infty, 0]$ as $n \to \infty$.*

Assumption 1 implies that $\rho^n \to 1$ as $n \to \infty$. Therefore, the repair facility is fully utilized in the limit. Assumption 1 Part 1 implies that the average inter-breakdown times increase in the order of n as $n \to \infty$. Assumption 1 Part 2 implies that the repair facility operates under the conventional heavy-traffic regime, that is, the capacity of the repair facility is barely enough to repair all incoming broken parts.

One can also consider the Halfin & Whitt (HW) asymptotic regime for the repair facility (see Halfin and Whitt (1981)). In that case, the system resembles the classical "machine repair model" in the HW regime (see Momčilović and Motaei (2018) and the references therein). The two main differences between the conventional and the HW asymptotic regimes are that i.) the number of servers is fixed in the conventional heavy-traffic regime whereas it tends to infinity in the HW regime; ii.) the delay in the repair facility, that is, the amount of time a broken part spends until its repair starts, is in the order of $\sqrt{n}$ in the conventional heavy-traffic regime whereas it is in the order of $1/\sqrt{n}$ in the HW regime (see theorem 4 of de Véricourt and Jennings (2008)). The reason for our choice is that for many repair facilities in practice, the conventional heavy traffic regime is a better fit than the HW regime. Chapter 5 of Driessen (2018) presents a case study about three maintenance service providers for capital goods, namely GVA (the municipal public transport operator for Amsterdam), KLM Engineering & Maintenance, and Royal Netherlands Army. According to the chapter 5 of Driessen (2018), the repair facilities are highly utilized, the numbers of repair men in the repair facilities are not very large (see table 5.1 therein), the repair men are certified professionals that have to complete a long training program and thus they are difficult to recruit, and the repair operations may require expensive, dedicated, and highly utilized equipments whose unavailability can cause long delays.

We have the following notation: For all $i \in \mathcal{I}$ and $t \in \mathbb{R}_+$,

- S_i^n is the initial inventory level for SKU i. Specifically, $\alpha_i^n n + S_i^n$ is the total number of parts of SKU i in the system at all times.
- $U_i^n(t)$ is the number of parts of SKU i installed at the capital goods at time t.
- $Q_i^n(t)$ is the number of broken parts of SKU i at time t.
- $OH_i^n(t)$ is the number of on-hand parts of SKU i at time t, that is, it is the number of ready-for-use parts of SKU i at time t.
- $BO_i^n(t)$ is the number of backordered demands of SKU i at time t.
- $T_i^n(t)$ is the cumulative time that repair facility spends to repair parts of SKU i up to time t.
- $I^n(t)$ is the cumulative time in which the repair facility idles up to time t.
- $E_i^n(t)$ is the cumulative number of emergency repairs done for parts of SKU i up to time t.

Let $A_i(\cdot)$ be a unit rate Poisson process which is independent of all other stochastic primitives for all $i \in \mathcal{I}$. Let $(x)^+ := \max\{0, x\}$ for all $x \in \mathbb{R}$. For all $i \in \mathcal{I}$, $t \in \mathbb{R}_+$, and $n \in \mathbb{N}_+$, we have

$$Q_i^n(t) = A_i\left(\lambda_i^n \int_0^t U_i^n(s)\mathrm{d}s\right) - R_i(T_i^n(t)) - E_i^n(t), \tag{1a}$$

$$BO_i^n(t) = \alpha_i^n n - U_i^n(t) = \left(Q_i^n(t) - S_i^n\right)^+, \tag{1b}$$

$$OH_i^n(t) = \left(S_i^n - Q_i^n(t)\right)^+, \tag{1c}$$

$$I^n(t) + \sum_{i \in \mathcal{I}} T_i^n(t) = t, \tag{1d}$$

$$I^n(0) = E_i^n(0) = 0 \text{ and both } I^n \text{ and } E_i^n \text{ are nondecreasing,} \tag{1e}$$

$$U_i^n(0) = \alpha_i^n n, \qquad 0 \le U_i^n(t) \le \alpha_i^n n, \tag{1f}$$

$$\sum_{i\in\mathcal{I}} \int_0^\infty Q_i^n(t)\mathrm{d}I^n(t) = 0, \tag{1g}$$

a repair of a broken part cannot be interrupted until completion, (1h)

$$S_i^n,\ U_i^n(t),\ Q_i^n(t) \in \mathbb{N}. \tag{1i}$$

Constraint (1g) implies that the repair facility operates in a work-conserving fashion, that is, the repair facility never idles as long as there is a broken part in the repair facility. Constraint (1h) implies that the repair facility uses non-preemptive repair policies. We consider work-conserving and non-preemptive control policies because they have practical appeal and are commonly used in practice. Next, we define the set of admissible control policies for the pre-limit system.

Definition 1. (Admissible policy) Fix an arbitrary $n \in \mathbb{N}_+$. A control policy is admissible if it is non-anticipative and under that policy, the process $\left(S_i^n, I^n, T_i^n, E_i^n, U_i^n, Q_i^n, i \in \mathcal{I}\right)$ satisfies (1).

The set of admissible control processes includes randomized and history-dependent policies but does not include the policies that can use future information.

For all $i \in \mathcal{I}$, $t \in \mathbb{R}_+$, and $n \in \mathbb{N}_+$, (1b) and (1c) imply

$$\alpha_i^n n + S_i^n = OH_i^n(t) + Q_i^n(t) + U_i^n(t),$$

which states that the parts circulate in the system as ready-for-use, broken, or installed parts.

We let h_i denote the holding cost per part of SKU i per unit time, b_i denote the backorder cost per unit time per backordered demand of SKU i, and c_i^n denote the emergency repair cost for a part of SKU i for all $i \in \mathcal{I}$ and $n \in \mathbb{N}_+$. On the one hand, both the inventory holding and the backorder costs are independent on n. On the other hand, we assume that

$$\frac{c_i^n}{n} \to c_i \in \mathbb{R}_+ \text{ as } n \to \infty \text{ for all } i \in \mathcal{I}, \tag{2}$$

that is, the emergency repair costs increase in the order of n as $n \to \infty$. By the assumption in (2), inventory holding, backorder, and emergency repair costs will be non-zero in the asymptotic regime defined in Assumption 1 under an appropriate scaling.

We assume that $\min\{h_i, b_i, c_i\} > 0$ for all $i \in \mathcal{I}$. Let

$$J^n(t) := \sum_{i\in\mathcal{I}} \mathbf{E}\left[h_i S_i^n + \frac{b_i}{t}\int_0^t (Q_i^n(s) - S_i^n)^+ \, \mathrm{d}s + \frac{c_i^n}{t} E_i^n(t) \right], \qquad \forall t \in \mathbb{R}_+,\ n \in \mathbb{N}_+,$$

denote the average cost rate on the interval $[0,t]$ in the nth system. The first term in the definition of $J^n(t)$ is the inventory holding cost per unit time. By (1b), the second term denotes the backorder cost per unit time on the interval $[0,t]$. The last term denotes the emergency repair cost per unit time on the interval $[0,t]$. The objective is to minimize the long-run average cost, that is, the objective is to minimize

$$\limsup_{t\to\infty} J^n(t). \tag{3}$$

3.2. Fluid and Diffusion Scaled Processes

Solving the optimization problem (1) & (3) is very challenging due to the curse of dimensionality. Therefore, we will derive and solve an optimization problem in the asymptotic regime defined in Assumption 1 by considering the diffusion scaled processes. The reason is that the diffusion limits of stochastic processes are more analytically tractable than the processes themselves. The intuition is as follows. Consider a sequence of $M/M/1$ queues indexed by $n \in \mathbb{N}_+$ such that $\rho^n_{M/M/1}$ denotes the load in the nth system. Similar to Assumption 1 Part 2, suppose that $\sqrt{n}\left(1 - \rho^n_{M/M/1}\right) \to \tilde{\theta} \in \mathbb{R}_{++}$ as $n \to \infty$ and thus we consider the conventional heavy-traffic regime. The long-run average number of jobs in the nth $M/M/1$ queue is equal to $\rho^n_{M/M/1}/(1-\rho^n_{M/M/1}) \approx \sqrt{n}/\tilde{\theta}$ and thus we must scale that value with $1/\sqrt{n}$ to obtain a simple and finite limiting value as $n \to \infty$. Under the diffusion scaling, the time is scaled with n and the space is scaled with $1/\sqrt{n}$. Therefore, as $n \to \infty$, due to the time scaling, the stochastic processes of interest reach the steady state quickly, and due to the space scaling, those processes converge to relatively tractable processes.

Let us first define the workload process associated with the broken parts and the weighted cumulative number of emergency repairs as

$$W^n(t) := \sum_{i\in\mathcal{I}} \frac{Q_i^n(t)}{\mu_i}, \qquad E^n(t) := \sum_{i\in\mathcal{I}} \frac{E_i^n(t)}{\mu_i}, \qquad \forall t \in \mathbb{R}_+,\ n \in \mathbb{N}_+, \tag{4}$$

respectively. The workload process $W^n(t)$ can be interpreted as the weighted total number of broken parts in the repair facility at time t. Let $\mathbb{D}$ denote the càdlàg space. For a process $X^n \in \mathbb{D}$, we denote the fluid scaled version of it by $\bar{X}^n$ such that $\bar{X}^n(t) := X^n(nt)/n$ for all $n \in \mathbb{N}_+$ and $t \in \mathbb{R}_+$. For all $i \in \mathcal{I}$, $n \in \mathbb{N}_+$, and $t \in \mathbb{R}_+$, we let $\bar{S}_i^n := S_i^n/n$ and

$$F_i^n(t) := \lambda_i^n \int_0^t U_i^n(s)\mathrm{d}s.$$

After some algebra, we have the following equation for all $i \in \mathcal{I}$, $t \in \mathbb{R}_+$, and $n \in \mathbb{N}_+$:

$$\bar{Q}_i^n(t) = \bar{A}_i^n \circ \bar{F}_i^n(t) - \bar{R}_i^n \circ \bar{T}_i^n(t) - \bar{E}_i^n(t),$$

where "$\circ$" denotes the composition operator.

Let $e, \mathbf{0}, \iota \in \mathbb{D}$ be such that $e(t) := t$, $\mathbf{0}(t) = 0$, and $\iota(t) = 1$ for all $t \in \mathbb{R}_+$. We consider the control policies under which the following convergence result holds:

$$\left(\bar{S}_i^n, \bar{Q}_i^n, \bar{E}_i^n, \bar{U}_i^n, \bar{T}_i^n, \bar{F}_i^n\right) \xrightarrow{a.s.} \left(0, \mathbf{0}, \mathbf{0}, \alpha_i\iota, \frac{\alpha_i\lambda_i}{\mu_i}e, \alpha_i\lambda_i e\right) \quad \text{u.o.c. as } n \to \infty \text{ for all } i \in \mathcal{I}, \tag{5}$$

where "u.o.c." is the abbreviation of uniformly on compact intervals. The convergence in (5) implies that $\bar{I}^n \xrightarrow{a.s.} \mathbf{0}$ u.o.c., that is, the repair facility is fully utilized in the fluid limit. Furthermore, because

$$\bar{J}^n(t) = \sum_{i \in \mathcal{I}} \mathbf{E}\left[h_i \bar{S}_i^n + \frac{1}{t}\left(b_i \int_0^t \left(\bar{Q}_i^n(s) - \bar{S}_i^n\right)^+ \mathrm{d}s + \frac{c_i^n}{n}\bar{E}_i^n(t)\right)\right], \qquad \forall t \in \mathbb{R}_+,\ n \in \mathbb{N}_+,$$

the fluid-scaled cost is zero under the limiting processes in (5).

Next, for all $i \in \mathcal{I}$, $t \in \mathbb{R}_+$, and $n \in \mathbb{N}_+$, let us define the following diffusion scaled processes:

$$\hat{S}_i^n := S_i^n/\sqrt{n}, \qquad \hat{A}_i^n(t) := \left(A_i(nt) - nt\right)/\sqrt{n}, \qquad \hat{R}_i^n(t) := \left(R_i(nt) - n\mu_i t\right)/\sqrt{n}.$$

For any other process $Z^n \in \mathbb{D}$, we denote the diffusion scaled version of it by $\hat{Z}^n$ such that $\hat{Z}^n(t) := Z^n(nt)/\sqrt{n}$ for all $n \in \mathbb{N}_+$ and $t \in \mathbb{R}_+$.

For all $i \in \mathcal{I}$, $t \in \mathbb{R}_+$, and $n \in \mathbb{N}_+$, let

$$\hat{X}_i^n(t) := \hat{A}_i^n \circ \bar{F}_i^n(t) - \hat{R}_i^n \circ \bar{T}_i^n(t),$$

$$\hat{X}^n(t) := \sum_{i \in \mathcal{I}} \frac{\hat{X}_i^n(t)}{\mu_i}.$$

For all $t \in \mathbb{R}_+$ and $n \in \mathbb{N}_+$, the diffusion scaled equations are as follows.

$$\hat{Q}_i^n(t) = \hat{X}_i^n(t) - n\lambda_i^n \int_0^t \left(\hat{Q}_i^n(s) - \hat{S}_i^n\right)^+ \mathrm{d}s + \sqrt{n}\left(n\alpha_i^n \lambda_i^n t - \mu_i \bar{T}_i^n(t)\right) - \hat{E}_i^n(t), \quad \forall i \in \mathcal{I}, \tag{6a}$$

$$\hat{W}^n(t) = \hat{X}^n(t) + \sqrt{n}\left(\rho^n - 1\right)t - \sum_{i \in \mathcal{I}} \frac{n\lambda_i^n}{\mu_i} \int_0^t \left(\hat{Q}_i^n(s) - \hat{S}_i^n\right)^+ \mathrm{d}s + \hat{I}^n(t) - \hat{E}^n(t). \tag{6b}$$

By (3), in the nth system, our objective is to minimize the diffusion-scaled long-run average cost henceforth, that is, our objective is to minimize

$$\limsup_{t\to\infty} \hat{J}^n(t). \tag{7}$$

3.3. The Brownian Control Problem

By considering the limits of the diffusion scaled processes, we formulate the BCP which is a singular stochastic control problem (see Harrison (2013)). By (5), the functional central limit theorem (see theorems 4.3.5 and 7.3.2 of Whitt (2002)), joint convergence when one limit is deterministic (see theorem 11.4.5 of Whitt (2002)), the random time-change theorem (see theorem 13.2.2 of Whitt (2002)), and the continuous mapping theorem (see theorem 3.4.4 of Whitt (2002)), we have

$$\left(\hat{X}_i^n, i \in \mathcal{I}\right) \Rightarrow \left(X_i, i \in \mathcal{I}\right) \quad \text{as } n \to \infty,$$

where $\Rightarrow$ denotes weak convergence and X_i is a Brownian motion starting from origin with drift 0 and variance $\alpha_i\lambda_i\left(1+\sigma_i^2\right)$ for all $i \in \mathcal{I}$, that is, X_i is a $BM\left(0, \alpha_i\lambda_i\left(1+\sigma_i^2\right)\right)$. Furthermore, X_i is independent of X_j for all $j \neq i$. Let

$$X := \sum_{i \in \mathcal{I}} \frac{X_i}{\mu_i}. \tag{8}$$

Then, X is a $BM\left(0, \Sigma\right)$ such that

$$\Sigma := \sum_{i \in \mathcal{I}} \frac{\alpha_i\lambda_i\left(1+\sigma_i^2\right)}{\mu_i^2}.$$

By the continuous mapping theorem, we have $\hat{X}^n \Rightarrow X$ as $n \to \infty$.

Let us assume that as $n \to \infty$,

$$\left(\hat{I}^n, \sqrt{n}\left(\frac{n\alpha_i^n\lambda_i^n}{\mu_i} e - \bar{T}_i^n\right), \hat{S}_i^n, \hat{E}_i^n, i \in \mathcal{I}\right) \Rightarrow \left(I, Y_i, S_i, E_i, i \in \mathcal{I}\right),$$

which implies

$$\sum_{i\in\mathcal{I}} Y_i(t) = \theta t + I(t), \qquad \forall t \in \mathbb{R}_+.$$

The limiting stochastic process Y_i denotes the deviation of the actual time spent for the repair of the broken parts of SKU i from the average time that should be spent for the repair of the broken parts of SKU i. We will control those deviations in the BCP that we will propose.

Let us consider the filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbf{P})$ satisfying the usual conditions (see appendix A.1 of Harrison (2013) for details) such that $\mathbb{F}$ is the filtration generated by $\big(X_i, i \in \mathcal{I}\big)$. The following BCP is the limiting version of the optimization problem (1) & (3) under the diffusion scaling.

$$\min \limsup_{t\to\infty} \sum_{i\in\mathcal{I}} \mathbf{E}\left[h_i S_i + \frac{1}{t}\left(b_i \int_0^t (Q_i(s) - S_i)^+ \,\mathrm{d}s + c_i E_i(t)\right)\right] \tag{9a}$$

$$\text{s.t. } Q_i(t) = X_i(t) - \lambda_i \int_0^t (Q_i(s) - S_i)^+ \,\mathrm{d}s + \mu_i Y_i(t) - E_i(t), \qquad \forall i \in \mathcal{I},\ t \in \mathbb{R}_+, \tag{9b}$$

$$\sum_{i\in\mathcal{I}} Y_i(t) = \theta t + I(t), \qquad \forall t \in \mathbb{R}_+, \tag{9c}$$

$$I(0) = E_i(0) = 0,\ I \text{ and } E_i \text{ are in } \mathbb{D}, \text{ nondecreasing, and } \mathbb{F}\text{-adapted for all } i \in \mathcal{I}, \tag{9d}$$

$$Q_i \text{ is an } \mathbb{F}\text{-adapted stochastic process for all } i \in \mathcal{I}, \tag{9e}$$

$$S_i, Q_i(t) \geq 0, \qquad \forall i \in \mathcal{I},\ t \in \mathbb{R}_+, \tag{9f}$$

where the decision variables are the process $\big(Q_i, S_i, Y_i, I, E_i, i \in \mathcal{I}\big)$, the objective function (9a) follows from (2), (3), and (7), and (9b) follows from (6a). The $\mathbb{F}$-adapted process requirements in (9d) and (9e) enforce a non-anticipative control. Observe that we do not enforce work-conserving control policies in the BCP (9) (recall (1g)) but we will later prove that the optimal BCP (9) solution is indeed a work-conserving policy. Furthermore, we do not enforce any constraints related to non-preemptive control policies (recall (1h)). The reason is that the performance gap between the preemptive and non-preemptive policies generally disappears in the heavy-traffic regime (see for example Atar et al. (2004)).

3.4. The Equivalent Workload Formulation

Because the BCP (9) is multi-dimensional, we will formulate an equivalent, single-dimensional, and analytically tractable stochastic control problem. Specifically, we will consider the limit of the single-dimensional workload evolution constraint (6b) instead of the multi-dimensional queue length evolution constraint (6a). In the EWF, our goal is to dynamically allocate the workload to the individual queue lengths to minimize the long-run average cost. To do this, for any given workload level, we introduce the set of feasible workload allocations in the following way. By considering (4), for all $w \in \mathbb{R}_+$, let

$$\mathcal{A}(w) := \left\{ \boldsymbol{q} \in \mathbb{R}_+^I : \sum_{i\in\mathcal{I}} \frac{q_i}{\mu_i} = w \right\}.$$

Similar to Rubino and Ata (2009), we define the workload allocation process $\boldsymbol{a} = (a_i, i \in \mathcal{I})$ such that $\boldsymbol{a} : \mathbb{R}_+^2 \times \Omega \to \mathbb{R}_+^I$ and $\boldsymbol{a}(t, W(t,\omega), \omega)$ denotes an allocation of the workload W to the queue lengths at time t under the sample path ω. For notational convenience, we will suppress ω from the notation as much as possible.

By recalling (4), let $W(t) := \sum_{i\in\mathcal{I}} Q_i(t)/\mu_i$ and $E(t) := \sum_{i\in\mathcal{I}} E_i(t)/\mu_i$. Then, W is the limiting workload process and $E(t)$ is the limiting cumulative weighted total number of emergency repairs done up to time t. Let $k \in \arg\min_{i\in\mathcal{I}} c_i\mu_i$. By recalling (8), the EWF of the BCP (9) is as follows.

$$\min \limsup_{t\to\infty} \mathbf{E}\left[\sum_{i\in\mathcal{I}} h_i S_i + \frac{1}{t}\left(\sum_{i\in\mathcal{I}} b_i \int_0^t (a_i(s, W(s)) - S_i)^+ \,\mathrm{d}s + c_k\mu_k E(t)\right)\right], \tag{10a}$$

$$\text{s.t. } W(t) = X(t) + \theta t - \sum_{i\in\mathcal{I}} \frac{\lambda_i}{\mu_i} \int_0^t (a_i(s, W(s)) - S_i)^+ \,\mathrm{d}s + I(t) - E(t), \qquad \forall t \in \mathbb{R}_+, \tag{10b}$$

$$I(0) = E(0) = 0,\; I \text{ and } E \text{ are in } \mathbb{D}, \text{ nondecreasing, and } \mathbb{F}\text{-adapted}, \tag{10c}$$

$$\boldsymbol{a}(t, W(t,\omega), \omega) \in \mathcal{A}(W(t,\omega)), \qquad \forall t \in \mathbb{R}_+,\ \omega \in \Omega, \tag{10d}$$

$$S_i, W(t) \geq 0, \qquad \forall i \in \mathcal{I},\ t \in \mathbb{R}_+, \tag{10e}$$

$$\boldsymbol{a}(\cdot, W(\cdot)) \text{ is an } \mathbb{F}\text{-adapted stochastic process}, \tag{10f}$$

where the decision variables are the process $\big(W, I, E, \boldsymbol{a}, \boldsymbol{S}\big)$ such that $\boldsymbol{S} := (S_i, i \in \mathcal{I})$. As seen in the objective function (10a), emergency repairs are used only for SKU k, which is an SKU with the smallest emergency repair cost per unit repair time.

The EWF (10) is much easier to solve than the BCP (9) because the EWF constraint (10b) is one-dimensional whereas the BCP constraint (9b) is multi-dimensional. The following proposition states that the EWF (10) is equivalent to the BCP (9).

Proposition 1. *(i) For any feasible solution to the EWF (10), there exists a feasible solution to the BCP (9) with the same objective function value.*

(ii) For any feasible solution to the BCP (9), there exists a feasible solution to the EWF (10) with a less than or equal objective function value.

Therefore, the optimal objective function values of the BCP (9) and the EWF (10) are the same.

The proof of Proposition 1 is presented in the OA A.1.

4. An Optimal EWF (10) Solution

We will derive an optimal EWF (10) solution. First, we present the associated Bellman equations in Section 4.1. Then, we discuss the structure of the optimal workload allocation policy in Section 4.2. Finally, we present the optimal EWF (10) solution in Sections 4.3 and 4.4.

4.1. The Bellman Equations

Let $j, l \in \mathcal{I}$ be such that

$$j \in \arg\min_{i \in \mathcal{I}} \frac{b_i \mu_i}{\lambda_i}, \qquad l \in \arg\min_{i \in \mathcal{I}} h_i \mu_i. \tag{11}$$

The structure of the optimal EWF (10) solution depends on whether $b_j\mu_j/\lambda_j \leq c_k\mu_k$ or $b_j\mu_j/\lambda_j > c_k\mu_k$, and so does the structure of the Bellman equation. For all $m \in \mathbb{N}_+$, let $\mathbb{C}_m$ denote the set of differentiable functions with domain $\mathbb{R}_+$ and range $\mathbb{R}$ and whose mth derivatives are continuous. For a given $f \in \mathbb{C}_2$, let f' and f'' denote the first and the second derivatives of f, respectively.

Suppose that $b_j\mu_j/\lambda_j \leq c_k\mu_k$. Fix an arbitrary $\boldsymbol{S} \in \mathbb{R}_+^I$. The associated Bellman equation is as follows. Find a pair (f, γ) such that $f \in \mathbb{C}_2$, $\gamma \in \mathbb{R}_+$,

$$\frac{1}{2}\Sigma f''(w) + \theta f'(w) + \min_{\boldsymbol{q} \in \mathcal{A}(w)} \left\{ \sum_{i \in \mathcal{I}} \left(b_i - \frac{\lambda_i}{\mu_i} f'(w) \right) (q_i - S_i)^+ \right\} = \gamma, \qquad \forall w \in \mathbb{R}_+, \tag{12a}$$

$$0 \le f'(w) \le \frac{b_j \mu_j}{\lambda_j}, \qquad \forall w \in \mathbb{R}_+, \tag{12b}$$

$$f'(0) = 0. \tag{12c}$$

Under a given initial inventory level $\boldsymbol{S}$, the function $f \in \mathbb{C}_2$ and the constant $\gamma \in \mathbb{R}_+$ satisfying the condition (12) are the relative cost function and the long-run average backorder and emergency repair cost under an optimal policy, respectively. The intuitive explanation of the constraint (12b) is as follows. Consider an extreme scenario in which the workload is very high, parts of all SKUs are backordered, and the workload will remain high for a long time. If the repair facility repairs a part of SKU i, it will spend $1/\mu_i$ amount of repair time (resource) for that task on average, the repaired part of SKU i will stay installed for $1/\lambda_i$ amount of time on average, and thus the repair facility will save b_i/λ_i total backorder cost on average. Therefore, the average saving per unit resource (repair time) spent is equal to $(b_i/\lambda_i)/(1/\mu_i) = b_i\mu_i/\lambda_i$. Consequently, the repair facility will give the least amount of repair priority to the SKU j where $j \in \arg\min_{i\in\mathcal{I}}(b_i\mu_i/\lambda_i)$ and thus parts of SKU j will be backordered and the cost will increase with the rate $b_j\mu_j/\lambda_j$. Because, the aforementioned scenario is a worst-case scenario, the relative cost function increases in the workload with a rate less than or equal to $b_j\mu_j/\lambda_j$, which is what (12b) states.

Suppose that $b_j\mu_j/\lambda_j > c_k\mu_k$. Fix an arbitrary $\boldsymbol{S} \in \mathbb{R}_+^I$. The associated Bellman equation is as follows. Find a triple (f, B, γ) such that $f \in \mathbb{C}_2$, $B \in \mathbb{R}_{++}$, $\gamma \in \mathbb{R}_+$,

$$\frac{1}{2}\Sigma f''(w) + \theta f'(w) + \min_{\boldsymbol{q}\in\mathcal{A}(w)} \left\{ \sum_{i\in\mathcal{I}} \left(b_i - \frac{\lambda_i}{\mu_i} f'(w) \right) (q_i - S_i)^+ \right\} = \gamma, \qquad \forall w \in [0, B], \tag{13a}$$

$$\frac{1}{2}\Sigma f''(w) + \theta f'(w) + \min_{\boldsymbol{q}\in\mathcal{A}(w)} \left\{ \sum_{i\in\mathcal{I}} \left(b_i - \frac{\lambda_i}{\mu_i} f'(w) \right) (q_i - S_i)^+ \right\} \ge \gamma, \qquad \forall w > B, \tag{13b}$$

$$0 \le f'(w) \le c_k \mu_k, \qquad \forall w \in \mathbb{R}_+, \tag{13c}$$

$$f'(0) = 0 \quad \text{and} \quad f'(B) = c_k \mu_k. \tag{13d}$$

The constant B is the barrier level such that the optimal solution does not allow the workload level to exceed the barrier level B. Whenever the workload level hits the barrier level B, the optimal solution uses emergency repairs to keep the workload below the barrier level as seen in (13d). The intuitive explanation of the constraints (13c) and (13d) is as follows. Recall the extreme scenario discussed after (12). When a part of SKU i breaks down, if an emergency repair is executed, then the associated cost is c_i and that emergency repair prevents the workload to increase by $1/\mu_i$ on average. That increase in the workload would increase the cost by $f'(w)/\mu_i$ approximately. Therefore, applying the emergency repair is attractive for workloads $w \in \mathbb{R}_+$ for which $c_i \leq f'(w)/\mu_i$, that is, $c_i\mu_i \leq f'(w)$. Recall that $k \in \arg\min_{i\in\mathcal{I}} c_i\mu_i$. We will prove in Lemma 2 that $f'(w)$ is strictly increasing, which implies that as the workload increases, the relative cost function increases faster. Therefore, on the one hand, if $c_k\mu_k \geq b_j\mu_j/\lambda_j$, emergency repairs never become attractive because the cost never increases with a rate greater than $b_j\mu_j/\lambda_j$ (recall (12b)). On the other hand, if $c_k\mu_k < b_j\mu_j/\lambda_j$, there exists a workload level $B \in \mathbb{R}_{++}$ at which $f'(B) = c_k\mu_k$. Then, whenever the workload reaches B, the repair facility will execute emergency repairs for the parts of SKU k and thus the workload will never exceed the barrier level B. Consequently, it will never happen that emergency repairs are executed for SKUs other than SKU k.

We will simplify the Bellman equations (12) and (13) by finding a closed-form solution to the minimization problem in (12a), (13a), and (13b). For given $w \in \mathbb{R}_+$ and $f \in \mathbb{C}_2$, consider the optimization problem

$$\min_{\boldsymbol{q}\in\mathcal{A}(w)} \left\{ \sum_{i\in\mathcal{I}} \left(b_i - \frac{\lambda_i}{\mu_i} f'(w) \right) (q_i - S_i)^+ \right\}. \tag{14}$$

For any given workload level $w \in \mathbb{R}_+$, the optimization problem (14) allocates the workload to the individual queues to minimize the cost. For all $w \in \mathbb{R}_+$, let $z(w)$ denote the optimal objective function value and $\boldsymbol{q}^*(w) = (q_i^*(w), i \in \mathcal{I})$ denote an optimal solution to (14). The following lemma presents closed-form expressions for $z(w)$ and $\boldsymbol{q}^*(w)$.

LEMMA 1. *Fix an arbitrary $w \in \mathbb{R}_+$ and $f \in \mathbb{C}_2$ such that $0 \leq f'(w) \leq b_j\mu_j/\lambda_j$. An optimal solution to (14) is as follows.*

i. If $w \leq \sum_{i\in\mathcal{I}} S_i/\mu_i$, then $z(w) = 0$. Furthermore, any feasible solution under which $q_i(w) \leq S_i$ for all $i \in \mathcal{I}$ is optimal and at least one such feasible solution exists.

ii. Suppose that $w > \sum_{i\in\mathcal{I}} S_i/\mu_i$. *Let* $\kappa(w) \in \arg\min_{i\in\mathcal{I}} \{b_i\mu_i - \lambda_i f'(w)\}$. *Then,*

$$q_i^*(w) = S_i, \qquad \forall i \in \mathcal{I} \setminus \{\kappa(w)\},$$
$$q_{\kappa(w)}^*(w) = \mu_{\kappa(w)} \left(w - \sum_{i\in\mathcal{I}\setminus\{\kappa(w)\}} \frac{S_i}{\mu_i} \right),$$
$$z(w) = \min_{i\in\mathcal{I}} \{b_i\mu_i - \lambda_i f'(w)\} \left(w - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i} \right).$$

The proof of Lemma 1 is presented in the OA A.2. Therein, we also present a specific optimal solution under the case $w \le \sum_{i\in\mathcal{I}} S_i/\mu_i$. Next, the following lemma states that both of the Bellman equations (12) and (13) have solutions.

LEMMA 2. *Fix an arbitrary* $\boldsymbol{S} \in \mathbb{R}_+^I$.

i. There exists a pair (f,γ) *such that* $f \in \mathbb{C}_2$, $\gamma \in \mathbb{R}_+$, *and* (f,γ) *satisfies* (12). *Furthermore,* f *is unique up to a constant,* $\gamma > 0$ *and is unique, and* f' *is strictly increasing.*

ii. Suppose that $b_j\mu_j/\lambda_j > c_k\mu_k$. *There exists a triple* (f,B,γ) *such that* $f \in \mathbb{C}_2$, $B \in \mathbb{R}_{++}$, $\gamma \in \mathbb{R}_+$, *and* (f,B,γ) *satisfies* (13). *Furthermore,* f *is unique up to a constant on* $[0,B]$, (B,γ) *is unique,* $B > \sum_{i\in\mathcal{I}} S_i/\mu_i$, $\gamma > 0$, *and* f' *is strictly increasing on* $[0,B]$.

The proof of Lemma 2 is presented in the OA A.3.

DEFINITION 2. If $b_j\mu_j/\lambda_j \le c_k\mu_k$, let $f \in \mathbb{C}_2$ be the function defined in Lemma 2 Part i. Otherwise, if $b_j\mu_j/\lambda_j > c_k\mu_k$, let $f \in \mathbb{C}_2$ be the function defined in Lemma 2 Part ii. We let $\boldsymbol{q}^*(w) = (q_i^*(w), i \in \mathcal{I})$ denote an optimal solution to (14) as defined in Lemma 1 with the function f for all $w \in \mathbb{R}_+$.

4.2. Structure of the Optimal Workload Allocation

Lemma 1 provides important insights about the optimal workload allocation. On the one hand, if the workload level is sufficiently small, that is, if $w \le \sum_{i\in\mathcal{I}} S_i/\mu_i$, no SKU is backordered. On the other hand, if the workload level is sufficiently high, that is, if $w > \sum_{i\in\mathcal{I}} S_i/\mu_i$, then the optimal workload allocation is such that demands of a single SKU in the set $\arg\min_{i\in\mathcal{I}} \{b_i\mu_i - \lambda_i f'(w)\}$ are backordered. Observe that the index of the SKU with backordered demands, $\kappa(w)$, depends on the workload level, that is, different SKUs may be backordered at different workload levels (see also Example 1 below).

Because f' is strictly increasing (recall Lemma 2) and bounded (see (12b) and (13c)), as the workload level w increases on $\mathbb{R}_+$, there are at most $I-1$ thresholds at which the index $\kappa(w) \in \arg\min_{i\in\mathcal{I}} \{b_i\mu_i - \lambda_i f'(w)\}$ changes. Therefore, only demands of some of the SKUs are backordered under the optimal workload allocation policy $\boldsymbol{q}^*$. The following example illustrates how the optimal backordering decisions are made.

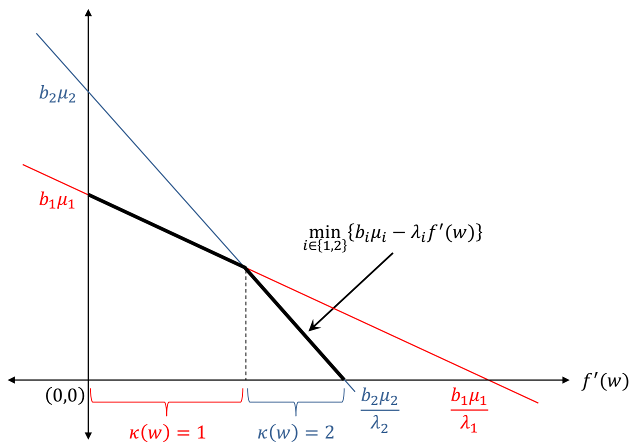


Figure 2 (Color online) Parameters associated with Example 1.

EXAMPLE 1. Suppose that there are two SKUs, that is, $\mathcal{I} = \{1, 2\}$. Suppose that $b_1\mu_1 < b_2\mu_2$, $\lambda_1 < \lambda_2$, and $b_1\mu_1/\lambda_1 > b_2\mu_2/\lambda_2$ such that there exists a threshold below (above) which demands of SKU 1 (2) are backordered (see Figure 2 and recall Lemma 1). Under the optimal workload allocation policy $\boldsymbol{q}^*$, as the workload level exceeds the "total" inventory level, first demands of SKU 1 are backordered. The reason is that the backorder cost per unit repair time for SKU 1 is less than the one for SKU 2 (recall that $b_1\mu_1 < b_2\mu_2$). As the demands of SKU 1 are backordered, broken parts of SKU 2 are given repair priority. However, SKU 2 has a higher breakdown rate than SKU 1 has (recall that $\lambda_1 < \lambda_2$) and thus repaired parts of SKU 2 will breakdown relatively quickly and increase the workload in the near future. Therefore, the initial backordering decision is myopic.

If the workload level exceeds the threshold, then the optimal workload allocation policy takes a drastic measure and backorders demands of SKU 2 and gives repair priority to the broken parts of SKU 1, even though SKU 2 has a higher backorder cost per unit repair time than SKU 1 has. The reason is that SKU 1 has a smaller breakdown rate than SKU 2 has and thus it will take relatively more time on average for the repaired parts of SKU 1 to breakdown and increase the workload in the future. Therefore, as the workload level increases, the optimal policy makes more forward-looking decisions. Finally, if $b_2\mu_2/\lambda_2$ is larger than $\min\{c_1\mu_1, c_2\mu_2\}$ and the workload level increases even more, then the optimal workload allocation policy takes the most drastic measure and implements emergency repairs to keep the workload below a threshold. In summary, as the workload level increases, the backordering decisions become more and more forward looking. □

Next, we will prove the existence of an optimal EWF (10) solution with the following structure:

- The idle time process can increase only if there is no workload, that is, for all $t \in \mathbb{R}_+$, $I(t)$ can increase only if $W(t) = 0$. Therefore, the repair facility operates in a work-conserving fashion.
- When the workload process exceeds the "total" inventory level, demands of a particular SKU are backordered. The index of the SKU whose demands are backordered depends on the workload level (recall Lemma 1).
- If $b_j\mu_j/\lambda_j \le c_k\mu_k$, the optimal solution does not make any emergency repairs, that is, $E = \mathbf{0}$ under the optimal solution. We call that optimal solution the *NER policy*.
- If $b_j\mu_j/\lambda_j > c_k\mu_k$, then a *barrier policy*, under which the workload process is not allowed to exceed an upper barrier level (or threshold), is optimal. The emergency repair process increases only when the workload hits the upper barrier and only broken parts of SKU k are sent into emergency repair.

Next, we will define the NER and the barrier policies rigorously and present the associated optimality results.

4.3. The NER Policy

The following one-sided regulator mapping will help us to define the NER policy rigorously.

Definition 3. (A one-sided regulator mapping) Let $\boldsymbol{\lambda} := (\lambda_i, i \in \mathcal{I}) \in \mathbb{R}_+^I$, $\boldsymbol{\mu} := (\mu_i, i \in \mathcal{I}) \in \mathbb{R}_{++}^I$, $\boldsymbol{s} := (s_i, i \in \mathcal{I}) \in \mathbb{R}_+^I$, and $x \in \mathbb{D}$ be such that $x(0) \geq 0$. The one-sided regulator mapping $\left(\phi^{(1)}, \psi^{(1)}\right)$: $\mathbb{R}_+^{2I} \times \mathbb{R}_{++}^I \times \mathbb{D} \to \mathbb{D}^2$ is defined by $\left(\phi^{(1)}, \psi^{(1)}\right)(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{s}, x) = (w, \ell)$ where

C1. $w(t) = x(t) - \sum_{i \in \mathcal{I}} \frac{\lambda_i}{\mu_i} \int_0^t \left(q_i^*(w(y)) - s_i\right)^+ \mathrm{d}y + \ell(t) \geq 0$ for all $t \in \mathbb{R}_+$,

C2. $\ell(0) = 0$, ℓ is nondecreasing, and $\int_0^\infty w(t) \mathrm{d}\ell(t) = 0$.

Condition C2 states that the pushing process ℓ increases only if w becomes 0. If $\lambda_i = 0$ for all $i \in \mathcal{I}$ in Definition 3, then the regulator mapping becomes the conventional one-sided regulator mapping (see chapter 13.5 of Whitt (2002)). The following lemma proves the existence of the one-sided regulator mapping in Definition 3 .

Lemma 3. *For any given $\boldsymbol{\lambda} \in \mathbb{R}_+^I$, $\boldsymbol{\mu} \in \mathbb{R}_{++}^I$, $\boldsymbol{s} \in \mathbb{R}_+^I$, and $x \in \mathbb{D}$ such that $x(0) \geq 0$, there exists a unique pair $\left(\phi^{(1)}, \psi^{(1)}\right)(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{s}, x)$ satisfying the conditions C1 and C2 in Definition 3. Furthermore, if $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{s})$ is given, $\left(\phi^{(1)}, \psi^{(1)}\right)(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{s}, x)$ is non-anticipative with respect to x.*

The proof of Lemma 3 is presented in the OA A.4.2.

The rigorous definition of the NER policy is as follows.

Definition 4. (NER policy) For given $\boldsymbol{S} \in \mathbb{R}_+^I$, the NER policy is the process $\left(W, I, E, \boldsymbol{q}^*, \boldsymbol{S}\right)$ such that $E = \mathbf{0}$, $\boldsymbol{q}^*$ is defined in Definition 2, and $(W, I) = \left(\phi^{(1)}, \psi^{(1)}\right)(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{S}, X + \theta e)$.

Observe that the NER policy is a feasible solution to the EWF (10) for all $\boldsymbol{S} \in \mathbb{R}_+^I$ by Lemma 3. Under the NER policy, the repair facility never makes an emergency repair by definition and operates in a work-conserving fashion by condition C2 in Definition 3.

The following theorem states the optimality of the NER policy.

Theorem 1. *Suppose that $b_j \mu_j / \lambda_j \leq c_k \mu_k$. Fix an arbitrary $\boldsymbol{S} \in \mathbb{R}_+^I$ in the EWF (10). Then, the NER policy defined in Definition 4 is an optimal EWF (10) solution. Furthermore, the long-run average backorder and emergency repair cost under an optimal solution is equal to γ defined in Lemma 2 Part i.*

The proof of Theorem 1 is presented in the OA A.5.

4.4. The Barrier Policy

Barrier type policies are well known in the context of optimal control of a Brownian motion (see for example chapter 7 of Harrison (2013)). We will define the barrier policy rigorously by the following two-sided regulator mapping.

Definition 5. (A two-sided regulator mapping) Let $\boldsymbol{\lambda} \in \mathbb{R}_+^I$, $\boldsymbol{\mu} \in \mathbb{R}_{++}^I$, $\boldsymbol{s} \in \mathbb{R}_+^I$, $b \in \mathbb{R}_{++}$, and $x \in \mathbb{D}$ be such that $x(0) \in [0,b]$. The two-sided regulator mapping $\left(\phi^{(2)}, \psi_1^{(2)}, \psi_2^{(2)}\right) : \mathbb{R}_+^{2I} \times \mathbb{R}_{++}^{I+1} \times \mathbb{D} \to \mathbb{D}^3$ is such that $\left(\phi^{(2)}, \psi_1^{(2)}, \psi_2^{(2)}\right)(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{s}, b, x) = (w, \ell, u)$ where

C1. $w(t) = x(t) - \sum_{i \in \mathcal{I}} \frac{\lambda_i}{\mu_i} \int_0^t (q_i^*(w(y)) - s_i)^+ \,\mathrm{d}y + \ell(t) - u(t) \in [0,b]$ for all $t \in \mathbb{R}_+$,

C2. $\ell(0) = u(0) = 0$ and both ℓ and u are nondecreasing,

C3. $\int_0^\infty w(t)\mathrm{d}\ell(t) = \int_0^\infty (b - w(t))\mathrm{d}u(t) = 0$.

The two-sided regulator mapping in Definition 5 has the lower barrier 0 and the upper barrier b and thus $w(t) \in [0,b]$ for all $t \in \mathbb{R}_+$. Condition C3 states that the lower and upper pushing processes (that is, ℓ and u) increase only if w hits the corresponding barrier. If $\lambda_i = 0$ or $s_i \geq b$ for all $i \in \mathcal{I}$ in Definition 5, then the regulator mapping becomes the conventional two-sided regulator mapping (see chapter 14.8 of Whitt (2002)). The following lemma proves the existence of the two-sided regulator mapping in Definition 5.

Lemma 4. *For any given $\boldsymbol{\lambda} \in \mathbb{R}_+^I$, $\boldsymbol{\mu} \in \mathbb{R}_{++}^I$, $\boldsymbol{s} \in \mathbb{R}_+^I$, $b \in \mathbb{R}_{++}$, and $x \in \mathbb{D}$ such that $x(0) \in [0,b]$, there exists a unique set of processes $\left(\phi^{(2)}, \psi_1^{(2)}, \psi_2^{(2)}\right)(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{s}, b, x) \in \mathbb{D}^3$ which satisfies the conditions C1-C3 in Definition 5. Furthermore, if $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{s}, b)$ is given, $\left(\phi^{(2)}, \psi_1^{(2)}, \psi_2^{(2)}\right)(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{s}, b, x)$ is non-anticipative with respect to x.*

The proof of Lemma 4 is presented in the OA A.4.3.

The rigorous definition of the barrier policy is as follows.

Definition 6. (Barrier policy) For given $B \in \mathbb{R}_{++}$ and $\boldsymbol{S} \in \mathbb{R}_+^I$, the barrier policy with lower barrier 0 and upper barrier B is the process $(W, I, E, \boldsymbol{q}^*, \boldsymbol{S})$ such that $\boldsymbol{q}^*$ is defined in Definition 2 and $(W, I, E) = \left(\phi^{(2)}, \psi_1^{(2)}, \psi_2^{(2)}\right)(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{S}, B, X + \theta e)$.

Observe that the barrier policy is a feasible solution to the EWF (10) for all $B \in \mathbb{R}_{++}$ and $\boldsymbol{S} \in \mathbb{R}_+^I$ by Lemma 4. Under the barrier policy, the workload process always stays between the lower and upper barrier levels, that is, $W(t) \in [0, B]$ for all $t \in \mathbb{R}_+$. Furthermore, by condition C3 in Definition 5, the repair facility operates in a work-conserving fashion and the emergency repairs are used only when the workload level hits the upper barrier level.

The following theorem states the optimality of the barrier policy.

THEOREM 2. *Suppose that $b_j\mu_j/\lambda_j > c_k\mu_k$. Fix an arbitrary $\boldsymbol{S} \in \mathbb{R}_+^I$ in the EWF (10). The barrier policy defined in Definition 6 with the upper barrier level B defined in Lemma 2 Part ii is an optimal EWF (10) solution. Furthermore, the long-run average backorder and emergency repair cost under an optimal solution is equal to γ defined in Lemma 2 Part ii.*

The proof of Theorem 2 is presented in the OA A.6.

5. Further Simplification of the EWF (10)

In Theorems 1 and 2, we show that if the inventory vector $\boldsymbol{S} \in \mathbb{R}_+^I$ is given, the NER and the barrier policies are optimal EWF (10) solutions depending on the comparison $b_j\mu_j/\lambda_j$ vs. $c_k\mu_k$. Furthermore, for a given inventory vector $\boldsymbol{S} \in \mathbb{R}_+^I$, we can numerically compute the optimal long-run average backorder and emergency repair cost, denoted by $\gamma(\boldsymbol{S})$, by solving the Bellman equation (12) or (13). Therefore, the following optimization problem gives us the optimal inventory levels.

$$\min_{\boldsymbol{S} \in \mathbb{R}_+^I} \left\{ \gamma(\boldsymbol{S}) + \sum_{i \in \mathcal{I}} h_i S_i \right\}. \tag{15}$$

Solving (15) is not trivial because we can compute $\gamma(\boldsymbol{S})$ only numerically. If there are many SKUs in the system, that is, if I is large, the optimization problem (15) becomes challenging. Nevertheless, we can simplify the EWF (10) together with its optimal solution under the following assumption.

ASSUMPTION 2. *Let the SKU j in (11) be such that $j \in \arg\min_{i \in \mathcal{I}} b_i\mu_i$ and $j \in \arg\max_{i \in \mathcal{I}} \lambda_i$, that is, SKU j has the smallest backorder cost per unit repair time and the highest breakdown rate.*

Under Assumption 2, the optimal workload allocation policy $\boldsymbol{q}^*$ (see Definition 2) simplifies such that the SKU $j \in \arg\min_{i \in \mathcal{I}} \{b_i\mu_i - \lambda_i f'(w)\}$ for all $w \in \mathbb{R}_+$. In other words, it is enough to backorder only demands of SKU j at all workload levels such that $w > \sum_{i\in\mathcal{I}} S_i/\mu_i$. Therefore, the structure of both the NER and the barrier policies simplify. Furthermore, under Assumption 2, the state-dependent drift term in (10b) simplifies under $\boldsymbol{q}^*$. Specifically, under Assumption 2,

$$\sum_{i\in\mathcal{I}} \frac{\lambda_i}{\mu_i} \int_0^t (q_i^*(W(s)) - S_i)^+ \,\mathrm{d}s = \lambda_j \int_0^t \left(W(s) - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i} \right)^+ \mathrm{d}s, \qquad \forall t \in \mathbb{R}_+.$$

Finally, under Assumption 2, it is enough to consider the inventory vectors $\boldsymbol{S} \in \mathbb{R}_+^I$ such that $S_i = 0$ for all $i \in \mathcal{I} \setminus \{l\}$, that is, by recalling (11), a nonzero initial inventory is allowed only for the SKU with the cheapest inventory holding cost per unit repair time. Therefore, the I-dimensional inventory optimization problem in (15) simplifies into a single-dimensional one. Let $S := \sum_{i\in\mathcal{I}} S_i/\mu_i$ so that S is the weighted total initial inventory level. Then, under Assumption 2, the EWF (10) simplifies into the following one:

$$\min \limsup_{t\to\infty} \mathbf{E}\left[h_l\mu_l S + \frac{1}{t}\left(b_j\mu_j \int_0^t (W(s) - S)^+ \,\mathrm{d}s + c_k\mu_k E(t) \right) \right], \tag{16a}$$

$$\text{s.t. } W(t) = X(t) + \theta t - \lambda_j \int_0^t (W(s) - S)^+ \,\mathrm{d}s + I(t) - E(t), \qquad \forall t \in \mathbb{R}_+, \tag{16b}$$

$$I(0) = E(0) = 0,\ I \text{ and } E \text{ are in } \mathbb{D}, \text{ nondecreasing, and } \mathbb{F}\text{-adapted}, \tag{16c}$$

$$S, W(t) \in \mathbb{R}_+, \qquad \forall t \in \mathbb{R}_+, \tag{16d}$$

where the decision variables are the process $\big(W, I, E, S\big)$. By recalling (11), in the EWF (16), a nonzero initial inventory is allowed only for SKU l, backordering is allowed only for SKU j, and emergency repairs are allowed only for SKU k. The following proposition shows that the EWF (16) is equivalent to the BCP (9) and thus it is also equivalent to the EWF (10) by Proposition 1.

PROPOSITION 2. *(i) For any feasible solution to the EWF (16), there exists a feasible solution to the BCP (9) with the same objective function value. Therefore, the optimal objective function value of the BCP (9) is less than or equal to the one of the EWF (16).*

(ii) Suppose that Assumption 2 holds. For any feasible solution to the BCP (9), there exists a feasible solution to the EWF (16) with less than or equal objective function value.

Consequently, under Assumption 2, the optimal objective function values of the BCP (9) and the EWF (16) are the same.

The proof of Proposition 2 is presented in the OA A.7.

5.1. An Optimal EWF (16) Solution

Let z^* denote the optimal EWF (16) objective function value. The following theorem states the optimality of the NER policy.

THEOREM 3. *Suppose that $b_j\mu_j/\lambda_j \le c_k\mu_k$ and Assumption 2 holds. Let $\boldsymbol{S}^* := (S_i^*, i\in\mathcal{I}) \in \mathbb{R}_+^I$ be such that $S_i^* := 0$ for all $i \in \mathcal{I}\setminus\{l\}$ and $S_l^* := \mu_l S^*$ where $S^* \in \mathbb{R}_+$ is defined below. Consider the NER policy $\left(W^*, I^*, \mathbf{0}, \boldsymbol{q}^*, \boldsymbol{S}^*\right)$ in Definition 4. Then, $(W^*, I^*, \mathbf{0}, S^*)$ is an optimal EWF (16) solution.*

- *If $\theta = 0$,*

$$S^* = \sqrt{\frac{\Sigma}{2\lambda_j}}\left(\sqrt{\frac{b_j\mu_j}{h_l\mu_l}} - \sqrt{\frac{\pi}{2}}\right)^+, \tag{17}$$
$$z^* = h_l\mu_l S^* + \frac{\frac{b_j\mu_j}{\lambda_j}}{\frac{2}{\Sigma}S^* + \sqrt{\frac{\pi}{\lambda_j\Sigma}}}.$$

- *If $\theta < 0$,*

$$S^* = \frac{\Sigma}{-\theta}\left[\ln\left(-\theta\sqrt{\frac{2b_j\mu_j}{h_l\mu_l\lambda_j\Sigma}(1+\theta C)} + \sqrt{\left(4 + \frac{2b_j\mu_j\theta^2}{h_l\mu_l\lambda_j\Sigma}\right)(1+\theta C)}\right) - \ln(2)\right]^+, \tag{18}$$
$$z^* = h_l\mu_l S^* + \frac{\frac{b_j\mu_j}{\lambda_j}(1+\theta C)}{\frac{1}{\theta}\left(1 - \mathrm{e}^{-\frac{2\theta}{\Sigma}S^*}\right) + C},$$

where

$$C := \sqrt{\frac{\pi}{\lambda_j\Sigma}}\mathrm{e}^{\frac{\theta^2}{\lambda_j\Sigma}}\left(1 + \mathrm{Erf}\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right]\right)$$

and $\mathrm{Erf}[\cdot]$ *is the Gauss error function.*

The proof of Theorem 3 is presented in the OA A.8.

Next, we will introduce some notation. If $\theta = 0$, let for all $w, S \in \mathbb{R}_+$

$$G(w,S) := \frac{b_j\mu_j}{\lambda_j} + \mathrm{e}^{\frac{\lambda_j}{\Sigma}(w-S)^2}\left(\frac{2}{\Sigma}S\left(b_j\mu_j - \lambda_j c_k\mu_k\right)(w-S) - \frac{b_j\mu_j}{\lambda_j} + \left(b_j\mu_j - \lambda_j c_k\mu_k\right)(w-S)\sqrt{\frac{\pi}{\lambda_j\Sigma}}\mathrm{Erf}\left[\sqrt{\frac{\lambda_j}{\Sigma}}(w-S)\right]\right). \tag{19}$$

If $\theta < 0$, let for all $w, S \in \mathbb{R}_+$

$$G(w,S) := \frac{b_j\mu_j}{\lambda_j} + \mathrm{e}^{\frac{1}{\Sigma}\left(\lambda_j(w-S)^2 - 2\theta(w-S)\right)}\Bigg[\left(\theta c_k\mu_k + (b_j\mu_j - \lambda_j c_k\mu_k)(w-S)\right) \times \left(\frac{1}{\theta}\left(1-\mathrm{e}^{-\frac{2\theta}{\Sigma}S}\right) + \sqrt{\frac{\pi}{\lambda_j\Sigma}}\mathrm{e}^{\frac{\theta^2}{\lambda_j\Sigma}}\left(\mathrm{Erf}\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right] + \mathrm{Erf}\left[\frac{-\theta+\lambda_j(w-S)}{\sqrt{\lambda_j\Sigma}}\right]\right)\right) - \frac{b_j\mu_j}{\lambda_j}\left(1 + \frac{\theta}{\sqrt{\lambda_j\Sigma}}\sqrt{\pi}\mathrm{e}^{\frac{\theta^2}{\lambda_j\Sigma}}\left(\mathrm{Erf}\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right] + \mathrm{Erf}\left[\frac{-\theta+\lambda_j(w-S)}{\sqrt{\lambda_j\Sigma}}\right]\right)\right)\Bigg]. \tag{20}$$

The following theorem states the optimality of the barrier policy.

THEOREM 4. *Suppose that $b_j\mu_j/\lambda_j > c_k\mu_k$ and Assumption 2 holds. Let $\boldsymbol{S}^* \in \mathbb{R}_+^I$ be such that $S_i^* := 0$ for all $i \in \mathcal{I}\setminus\{l\}$ and $S_l^* := \mu_l S^*$. Consider the barrier policy $\left(W^*, I^*, E^*, \boldsymbol{q}^*, \boldsymbol{S}^*\right)$ in Definition 6 with the barrier level $B^* \in (S^*, \infty)$ such that (S^*, B^*) is an optimal solution to the optimization problem (21) presented below. Then, (W^*, I^*, E^*, S^*) is an optimal EWF (16) solution.*

$$\min_{S,B} \quad h_l\mu_l S + \theta c_k\mu_k + (b_j\mu_j - \lambda_j c_k\mu_k)(B-S), \tag{21a}$$

$$B = \inf\{w > S : G(w,S) = c_k\mu_k\}, \tag{21b}$$

$$S \geq 0. \tag{21c}$$

Finally, there exists an optimal solution to (21) and the optimal objective function value of (21) is equal to the one of EWF (16).

The proof of Theorem 4 is presented in the OA A.9. In that proof, we show that for any given $S \in \mathbb{R}_+$, there exists a unique $B \in (S, \infty)$ satisfying the constraint (21b). Furthermore, the constraint (21b) implies that the derivative of the relative value function is continuous at the barrier level and thus that constraint is an application of the "*principle of smooth fit*" (see chapter 5 of Harrison (2013)).

6. Numerical Experiments

We present control policies for the pre-limit system in Section 6.1. Then, we present numerical results for problem instances for which Assumption 2 does not hold in Section 6.2. Finally, we present numerical results for problem instances for which Assumption 2 holds in Section 6.3.

6.1. Policies for the pre-limit system

Let us consider a pre-limit system, that is, let us fix an arbitrary $n \in \mathbb{N}_+$. For notational convenience, we let $n_i := n\alpha_i^n$ denote the number of capital goods using SKU i for all $i \in \mathcal{I}$ in this section and thus $\sum_{i \in \mathcal{I}} n_i = n$. Let $\boldsymbol{S}^n := (S_i^n, i \in \mathcal{I})$ denote an initial inventory vector in the pre-limit system. Let e_i denote the I-dimensional vector whose elements are all equal to 0 except the ith element which is equal to 1. For all $x \in \mathbb{R}$, we let $[x]$ and $\lceil x \rceil$ denote the closest integer to x and the smallest integer greater than or equal to x, respectively. We let $o(\cdot)$ denote the little-o notation. We assume that the repair times are exponentially distributed in the numerical experiments.

First, we present an optimal policy and a policy that does not use emergency repairs.

An optimal policy. For a given inventory vector $\boldsymbol{S}^n$, we use a Markov decision process (MDP) to compute the optimal long-run average cost. We compute the optimal inventory vector by enumeration. The details about the MDP formulation are presented in the OA A.10.1. In the MDP model, the set of admissible control policies are the ones defined in Definition 1.

The NER policy.* This policy is the best performing policy among the ones that do not use emergency repairs. For a given inventory vector, the NER* policy is derived with the same technique used to derive the optimal policy except that emergency repairs are deleted from the action space of the MDP model. Finally, we compute the optimal inventory vector by enumeration.

Next, we describe the pre-limit interpretations of the NER and the barrier policies. We use the following approximations for the parameters that appear in the limit:

$$\lambda_i \approx n\lambda_i^n, \quad c_i \approx c_i^n/n, \quad S_i \approx S_i^n/\sqrt{n}, \quad \forall i \in \mathcal{I}, \tag{22a}$$

$$\theta \approx \sqrt{n}\,(\rho^n - 1)\,, \qquad \Sigma \approx \sum_{i \in \mathcal{I}} \frac{n_i \lambda_i^n\,(1+\sigma_i^2)}{\mu_i^2}. \tag{22b}$$

By recalling (11), let $j \in \arg\min_{i \in \mathcal{I}} b_i\mu_i/\lambda_i^n$, $k \in \arg\min_{i \in \mathcal{I}} c_i^n\mu_i$, and $l \in \arg\min_{i \in \mathcal{I}} h_i\mu_i$. By recalling Lemma 2, for any given inventory vector $\boldsymbol{S}^n$, if $b_j\mu_j/\lambda_j^n \le c_k^n\mu_k$, then let f be the function defined in Lemma 2 Part i that solves the Bellman equation (12) with the limiting parameters defined in

(22). Similarly, for any given $\boldsymbol{S}^n$, if $b_j\mu_j/\lambda_j^n > c_k^n\mu_k$, then let (f, B) be the pair defined in Lemma 2 Part ii that solves the Bellman equation (13) with the limiting parameters defined in (22).

If $b_j\mu_j/\lambda_j^n \le c_k^n\mu_k$, our proposed policy is the NER policy; and if $b_j\mu_j/\lambda_j^n > c_k^n\mu_k$, our proposed policy is the barrier policy. The formal definitions of the NER and the barrier policies are as follows.

The NER and the barrier policies for the general case. Suppose that the inventory vector $\boldsymbol{S}^n$ is given. By recalling Lemma 1, for all $t \in \mathbb{R}_+$ and $i \in \mathcal{I}$, let us define the index $\chi_i(t) := b_i\mu_i - n\lambda_i^n f'(W^n(t)/\sqrt{n})$. The repair facility operates in a work-conserving fashion and never preempts the repair of a broken part.

- At any given time $t \in \mathbb{R}_+$, if the set $\mathcal{Q}^<(t) := \{i \in \mathcal{I} : Q_i^n(t) < S_i^n\}$ is empty, then the repair facility gives the repair priority to the broken parts of SKUs with respect to the index $\chi_i(t)$ such that the higher the index of an SKU, the higher the repair priority that the broken parts of that SKU receive (ties can be broken arbitrarily).
- If $\mathcal{Q}^<(t) \neq \emptyset$, then the repair facility gives higher repair priority to the broken parts of SKUs in the set $\mathcal{I} \setminus \mathcal{Q}^<(t)$ over the ones in the set $\mathcal{Q}^<(t)$. The prioritization within the sets $\mathcal{Q}^<(t)$ and $\mathcal{I} \setminus \mathcal{Q}^<(t)$ is again done with respect to the index $\chi_i(t)$.
- Under the barrier policy, emergency repairs are used only for SKU k and only if a part of SKU k breaks down when $W^n(t)/\sqrt{n} \ge B$.

Because both the NER and the barrier policies are stationary Markov policies, they are admissible, and we compute the long-run average cost associated with those policies by policy evaluation for given inventory and barrier levels. In order to determine the inventory vector $\boldsymbol{S}^n$, we use the following local-search algorithm (LS):

1. (*Initialization*) Set $S_i^n = 0$ for all $i \in \mathcal{I}$ and compute the long-run average cost. Go to Step 2.
2. Compute the long-run average cost under the inventory vectors $\boldsymbol{S}^n + e_i$ for all $i \in \mathcal{I}$ and $\boldsymbol{S}^n - e_i$ with $S_i^n - 1 \ge 0$ for all $i \in \mathcal{I}$ (so that $\boldsymbol{S}^n - e_i$ is nonnegative componentwise). If the cost under one of those inventory vectors is less than the one under $\boldsymbol{S}^n$, update $\boldsymbol{S}^n$ with the inventory vector with the minimum cost and repeat step 2. Otherwise, end the algorithm.

The NER and the barrier policies under Assumption 2. The repair facility operates in a work-conserving fashion and never preempts the repair of a broken part. The NER and the barrier policies simplify in the following way.

- By the EWF (16) and the proof of Proposition 2, if $Q_l^n(t) < S_l^n$, the repair facility gives the least amount of repair priority to the parts of SKU l at time $t \in \mathbb{R}_+$.
- Otherwise, if $Q_l^n(t) \geq S_l^n$, the repair facility gives the least amount of repair priority to the parts of SKU j at time $t \in \mathbb{R}_+$.
- For the remaining SKUs, the repair facility gives higher repair priority to the broken parts of SKUs with $Q_i^n(t) \geq S_i^n$ over the ones with $Q_i^n(t) < S_i^n$. Within the aforementioned two sets, the repair facility prioritizes the broken parts of the remaining SKUs in the descending order of the index $b_i\mu_i$ (ties can be broken arbitrarily).

We present two different algorithms to determine the inventory vector $\boldsymbol{S}^n$. The first algorithm is similar to the ones in the literature. Specifically, recall from Theorems 3 and 4 that, under an optimal EWF (16) solution, $S_i = 0$ for all $i \in \mathcal{I} \setminus \{l\}$ which implies $S_i^n = o(\sqrt{n})$ for all $i \in \mathcal{I} \setminus \{l\}$. Therefore, we keep a small inventory of spare parts for the SKUs in the set $\mathcal{I} \setminus \{l\}$ in the pre-limit. Specifically, we assume that $S_i^n = s^n$ for all $i \in \mathcal{I} \setminus \{l\}$. This is exactly what Wein (1992) and Ata and Barjesteh (2022) do. Because $s^n = o(\sqrt{n})$, we enumerate s^n values in the set $\{0, 1, \ldots, \lceil \sqrt{n} \rceil\}$. Under the NER policy, we let $S_l^n = \lceil \mu_l S^* \sqrt{n} \rceil$ where S^* is computed by (17) and (18). Under the barrier policy, by considering (21b), for a given $S_l^n \in \mathbb{N}$, the associated barrier level B is computed by the following formula.

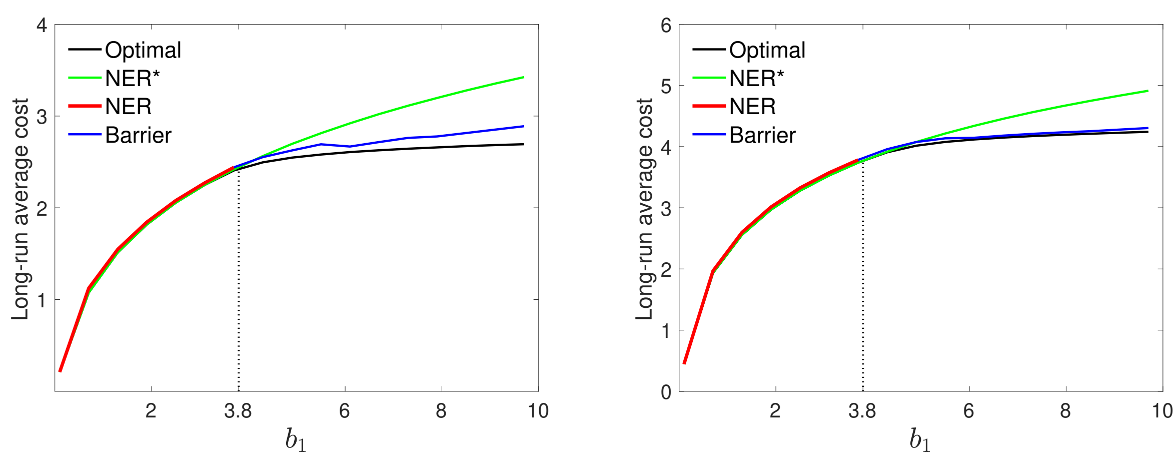
$$B := \frac{1}{\sqrt{n}} \times \underset{w \in \left\{ \frac{S_l^n}{\mu_l}, \ldots, \frac{S_l^n}{\mu_l} + \frac{n_j + \mathbb{I}\{j \neq l\} S_j^n}{\mu_j} \right\}}{\arg\min} \left| G\left(\frac{w}{\sqrt{n}}, \frac{S_l^n}{\mu_l \sqrt{n}} \right) - \frac{c_k^n \mu_k}{n} \right|,$$

where G is defined in (19) and (20) and $\mathbb{I}\{\cdot\}$ denotes the indicator function. Finally, we enumerate S_l^n. We denote the proposed policy with the inventory computation algorithm described above by PP1.

Under the barrier policy, the second inventory computation algorithm is exactly the LS described above. Under the NER policy, the second algorithm is the LS with the following modification for the initialization step: Set $S_i^n = 0$ for all $i \in \mathcal{I} \setminus \{l\}$ and $S_l^n = \left[\mu_l S^* \sqrt{n}\right]$ where S^* is computed by (17) and (18). Under Assumption 2, we denote the proposed policy with LS by PP2.

6.2. Numerical Results

We assume that there are two SKUs, $n_1 = n_2$ (recall that $n_1 + n_2 = n$), $\mu_1 = 1$, $h_1 = 0.1$, $c_1^n = 2n$, $\lambda_1^n \in \{0.8, 0.9, 0.95, 1\}/n$, $b_1 \in \{0.1, 0.7, 1.3, \ldots, 9.7\}$. For SKU 2, we choose $\mu_2 \in \{3, 4, 5, 6\}$, $h_2 = 0.2/\mu_2$, $c_2^n = 4n/\mu_2$, $\lambda_2^n = \lambda_1^n \mu_2$, and $b_2 = 2b_1/\mu_2$. Therefore, $h_2\mu_2 = 2h_1\mu_1$, $b_2\mu_2 = 2b_1\mu_1$, and $c_2^n\mu_2 = 2c_1^n\mu_1$ implying that SKU 2 is more "expensive" than SKU 1. Furthermore, $\rho^n = n_1\lambda_1^n/\mu_1 + n_2\lambda_2^n/\mu_2 = n\lambda_1^n \in \{0.8, 0.9, 0.95, 1\}$. Because $\lambda_1^n/\lambda_2^n \in \{1/3, 1/4, 1/5, 1/6\}$ and $(b_1\mu_1/\lambda_1^n)/(b_2\mu_2/\lambda_2^n) \in \{1.5, 2, 2.5, 3\}$, Assumption 2 does not hold. By recalling Lemma 1 and Figure 2, under the proposed policy, as the workload level increases, demands of SKU 1 are backordered first, and then if the workload level exceeds a threshold, demands of SKU 2 are backordered.



(a) $n_1 = n_2 = 10$ (b) $n_1 = n_2 = 50$

Figure 3 (Color online) Performances of the optimal, the proposed, and the NER* policies when $n_1 = n_2 \in \{10, 50\}$, $\rho^n = 0.95$, $\mu_2 = 4$, and $b_1 \in \{0.1, 0.7, 0.13, \ldots, 9.7\}$.

Recall the comparison $b_2\mu_2/\lambda_2^n$ vs. $c_1^n\mu_1$. Because $b_2\mu_2 = 2b_1\mu_1$, $\lambda_2^n = \lambda_1^n\mu_2$, $\rho^n = n\lambda_1^n$, and $c_1^n = 2n$, the aforementioned comparison is equivalent to b_1 vs. $\rho^n\mu_2$. We call the number $\rho^n\mu_2$ as the *switch point* such that if b_1 is less than or equal to the switch point, our proposed policy is the NER policy, otherwise, it is the barrier policy. Figure 3 presents numerical experiments in which $\rho^n = 0.95$, $\mu_2 =$

4, and the switch point is 3.8. According to the results, the proposed policy performs reasonably well even when $n_1 = n_2 = 10$. An interesting observation is that the performance of the NER* policy significantly deteriorates as b_1 exceeds the switch point. This is because as the backorder cost increases, the optimal policy backorders less and uses emergency repairs more. However, as the backorder cost increases, the NER* policy keeps more and more inventory to prevent backordering which increases the cost. Consequently, if $b_2\mu_2/\lambda_2^n > c_1^n\mu_1$, emergency repairs must be used to obtain a reasonable performance. Otherwise, there is no need for emergency repairs.

Table 1 presents the results of the experiments in which $n_1 = n_2 \in \{10, 25, 50\}$, $\mu_2 \in \{3, 4, 5, 6\}$, $\rho^n \in \{0.8, 0.9, 0.95, 1\}$, and $b_1 \in \{0.1, 0.7, 1.3, \ldots, 9.7\}$, that is, Table 1 presents the results of $3 \times 4 \times 4 \times 17 = 816$ parameter instances. According to the results, the proposed policy performs reasonably well and its performance improves as the number of capital goods increases.

Table 1 Average and maximum % deviations of the costs under the proposed policy from the optimal costs when $n_1 = n_2 \in \{10, 25, 50\}$, $\rho^n \in \{0.8, 0.9, 0.95, 1\}$, $\mu_2 \in \{3, 4, 5, 6\}$, and $b_1 \in \{0.1, 0.7, 1.3, \ldots, 9.7\}$.

		$n_1 = n_2 = 10$		$n_1 = n_2 = 25$		$n_1 = n_2 = 50$	
	ρ^n	Avg.	Max.	Avg.	Max.	Avg.	Max.
$\mu_2 = 3$	0.8	1.82	3.48	1.00	1.68	0.86	2.19
	0.9	2.77	6.52	1.18	2.62	0.65	1.26
	0.95	3.73	8.93	1.64	3.60	0.84	1.53
	1	5.49	14.19	2.57	6.70	1.35	3.78
$\mu_2 = 4$	0.8	1.93	3.27	1.00	1.49	0.79	1.51
	0.9	2.85	4.84	1.58	2.22	0.76	1.07
	0.95	3.24	7.29	1.72	2.91	1.18	1.76
	1	4.40	11.36	2.20	6.03	1.20	2.97
$\mu_2 = 5$	0.8	2.70	4.98	1.15	1.67	0.80	1.10
	0.9	2.76	4.91	1.98	3.86	1.19	1.73
	0.95	3.03	5.97	1.67	3.53	1.29	2.53
	1	4.05	10.22	2.36	7.12	1.46	5.12
$\mu_2 = 6$	0.8	3.13	6.19	1.50	2.31	0.87	1.67
	0.9	2.82	5.06	2.19	4.13	1.68	2.69
	0.95	2.97	7.27	1.78	4.05	1.38	2.85
	1	3.75	12.09	2.52	8.57	1.74	6.26

We also consider a modified version of the proposed policy which computes the inventory levels by enumeration instead of the LS. Among the 816 parameter instances, only for 3 instances, enumeration performs different (better) than the LS does. The average and the maximum deviations of the cost under the LS compared to the cost under enumeration are 0.0022% and 1.56%, respectively. We present a detailed analysis of the instance with the maximum deviation (1.56%) in the

OA A.10.3. The success of the LS is not surprising because local search algorithms are commonly used in the inventory control of spare parts due to their effective performance and computational efficiency (see for example van Houtum and Kranenburg (2015)).

6.3. Numerical Results under Assumption 2

We assume that there are two SKUs, $n_1 = n_2$, $\lambda_1^n = \lambda_2^n$, $\mu_1 = \mu_2 = 1$, and thus $\rho^n = n\lambda_1^n$. We assume that the costs associated with SKU 2 are twice the costs associated with SKU 1 and thus Assumption 2 holds. Specifically, we assume that $h_2 = 2h_1 = 0.2$, $c_2^n = 2c_1^n = 4n$, $b_2 = 2b_1$, and $b_1 \in \{0.1, 0.7, 1.3, \ldots, 9.7\}$. In these experiments, the switch point is $\lambda_1^n c_1^n$. Figure 4 presents numerical experiments in which $\rho^n = 0.95$ and the switch point is 1.9.

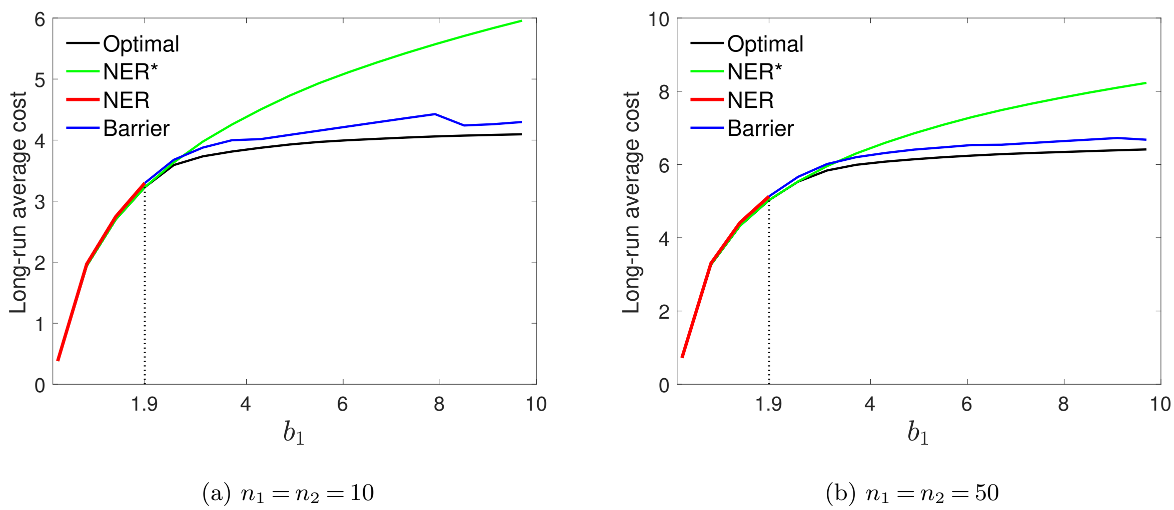


Figure 4 (Color online) Performances of the optimal and the NER* policies and the PP1 when $n_1 = n_2 \in \{10, 50\}$, $\rho^n = 0.95$, and $b_1 \in \{0.1, 0.7, 0.13, \ldots, 9.7\}$.

Table 2 presents the results of the experiments in which $n_1 = n_2 \in \{10, 25, 50\}$ and $\rho^n \in \{0.8, 0.9, 0.95, 1\}$, that is, Table 2 presents the results of $3 \times 4 \times 17 = 204$ parameter instances. Our observations from Table 2 are similar to the ones from Table 1.

Because the performance of the PP1 is very close to the one of PP2, we present the numerical results associated with the PP2 in the OA A.10.4. On the one hand, the PP2 performs slightly (0.25% on average) better than the PP1 does. On the other hand, because the PP1 keeps the same

Table 2 Average and maximum % deviations of the costs under the PP1 from the optimal costs when $n_1 = n_2 \in \{10, 25, 50\}$, $\rho^n \in \{0.8, 0.9, 0.95, 1\}$, and $b_1 \in \{0.1, 0.7, 0.13, \dots, 9.7\}$.

	$n_1 = n_2 = 10$		$n_1 = n_2 = 25$		$n_1 = n_2 = 50$	
ρ^n	Avg.	Max.	Avg.	Max.	Avg.	Max.
0.8	5.61	8.17	5.86	7.97	6.07	7.89
0.9	5.20	9.98	4.62	7.13	4.78	6.67
0.95	4.19	9.01	3.56	5.57	3.45	5.24
1	3.28	7.3	2.36	5.3	1.98	3.5

small initial inventory levels for all SKUs in the set $\mathcal{I} \setminus \{l\}$, we expect it to be computationally more efficient than the PP2 if there are many SKUs.

We repeat the numerical experiments depicted in Figure 4b by changing the scaling degree of the emergency repair cost from n (recall the assumption in (2)) to $\sqrt{n}$ and $n^{1.5}$ in the OA A.10.5. According to the results, PP1 performs reasonably well. Finally, in the OA A.10.6, we repeat the numerical experiments of Table 1 with the PP1 even though Assumption 2 does not hold in those experiments. According to the results, the PP1 performs poorly. Therefore, if Assumption 2 does not hold, it is crucial to implement the general version of the proposed policy.

7. Future Research Directions

We assume that the repair facility is in the conventional heavy-traffic regime. Yet another valuable topic for future research is considering the HW regime for the repair facility. There are studies considering performance evaluation in the HW regime (e.g., de Véricourt and Jennings (2008), Momčilović and Motaei (2018)). Studying control in that regime is an interesting future research topic. Finally, relaxing the exponential inter-breakdown times assumption is a relevant area for future research (see for example Momčilović and Motaei (2018)).

Acknowledgments

The authors thank the area editor, the associate editor, and two anonymous referees, whose comments significantly improved the paper.

References

Adan IJBF, Sleptchenko A, van Houtum GJ (2009) Reducing costs of spare parts supply systems via static priorities. *Asia-Pacific Journal of Operational Research* 26:559–585.

Alfredsson P, Verrijdt J (1999) Modeling emergency supply flexibility in a two-echelon inventory system. *Management Science* 45:1416–1431.

Arts J, Basten R, van Houtum GJ (2016) Repairable stocking and expediting in a fluctuating demand environment: Optimal policy and heuristics. *Operations Research* 64:1285–1301.

Ata B, Barjesteh N (2022) An approximate analysis of dynamic pricing, outsourcing, and scheduling policies for a multiclass make-to-stock queue in the heavy traffic regime. *Operations Research* URL http://dx.doi.org/10.1287/opre.2022.2361, published online.

Atar R, Mandelbaum A, Reiman MI (2004) Scheduling a multi class queue with many exponential servers: asymptotic optimality in heavy traffic. *The Annals of Applied Probability* 14:1084–1134.

Axsäter S (1990) Modelling emergency lateral transshipments in inventory systems. *Management Science* 36:1329–1338.

Bainov D, Simeonov P (1992) *Integral Inequalities and Applications*, volume 57 of *Mathematics and Its Applications* (Dordrecth: Kluwer Academic Publishers), first edition.

Basten RJI, van Houtum GJ (2014) System-oriented inventory models for spare parts. *Surveys in Operations Research and Management Science* 19:34–55.

Bitton S, Cohen I, Cohen M (2019) Joint repair sourcing and stocking policies for repairables using Erlang-A and Erlang-B queueing models. *IISE Transactions* 51:1151–1166.

Caggiano KE, Muckstadt JA, Rappold JA (2006) Integrated real-time capacity and inventory allocation for reparable service parts in a two-echelon supply system. *Manufacturing & Service Operations Management* 8:292–319.

de Véricourt F, Jennings OB (2008) Dimensioning large-scale membership services. *Operations Research* 56:173–187.

Drent M, Arts J (2021) Expediting in two-echelon spare parts inventory systems. *Manufacturing & Service Operations Management* 23:1431–1448.

Driessen M (2018) *Integrated Capacity Planning and Inventory Control for Repairable Spare Parts*. Ph.D. thesis, Technische Universiteit Eindhoven.

Driessen M, Arts J, van Houtum GJ, Rustenburg JW, Huisman B (2015) Maintenance spare parts planning and control: a framework for control and agenda for future research. *Production Planning & Control* 26:407–426.

Halfin S, Whitt W (1981) Heavy-traffic limits for queues with many exponential servers. *Operations Research* 29:567–588.

Harrison JM (1988) Brownian models of queueing networks with heterogeneous customer populations. Fleming W, Lions PL, eds., *Stochastic Differential Systems, Stochastic Control Theory and Their Applications*, volume 10 of *The IMA Volumes in Mathematics and Its Applications*, 147–186 (New York: Springer).

Harrison JM (2013) *Brownian Models of Performance and Control* (New York: Cambridge University Press).

Harrison JM, Van Mieghem JA (1997) Dynamic control of brownian networks: State space collapse and equivalent workload formulations. *The Annals of Applied Probability* 7:747–771.

Hausman WH, Scudder GD (1982) Priority scheduling rules for repairable inventory systems. *Management Science* 28:1215–1232.

Kim SH, Cohen MA, Netessine S (2015) Reliability or inventory? An analysis of performance-based contracts for product support services. Ha AY, Tang CS, eds., *Handbook of Information Exchange in Supply Chain Management*, volume 5 of *Springer Series in Supply Chain Management*, 65–88 (Springer, Cham).

Korotkov NE, Korotkov AN (2020) *Integrals Related to the Error Function* (Boca Raton: CRC Press), first edition.

Kranenburg AA, van Houtum GJ (2009) A new partial pooling structure for spare parts networks. *European Journal of Operational Research* 199:908–921.

Momčilović P, Motaei A (2018) An analysis of a large-scale machine repair model. *Stochastic Systems* 8:91–125.

Muckstadt JA, Thomas LJ (1980) Are multi-echelon inventory methods worth implementing in systems with low-demand-rate items? *Management Science* 26:483–494.

Öner KB, Franssen R, Kiesmüller GP, van Houtum GJ (2007) Life cycle costs measurement of complex systems manufactured by an engineer-to-order company. Qui RG, Russell DW, Sullivan WG, Ahmad

M, eds., *The 17th International Conference on Flexible Automation and Intelligent Manufacturing*, 589–596 (Philadephia).

Özkan E, van Houtum GJ, Serin Y (2015) A new approximate evaluation method for two-echelon inventory systems with emergency shipments. *Annals of Operations Research* 224:147–169.

Paterson C, Kiesmüller G, Teunter R, Glazebrook K (2011) Inventory models with lateral transshipments: A review. *European Journal of Operational Research* 210:125–136.

Polyanin AD, Zaitsev VF (2018) *Handbook of Ordinary Differential Equations: Exact Solutions, Methods, and Problems* (Boca Raton: Chapman and Hall/CRC), first edition.

Puterman ML (2005) *Markov Decision Processes: Discrete Stochastic Dynamic Programming* (Hoboken, New Jersey: Wiley).

Pyke DF (1990) Priority repair and dispatch policies for reparable-item logistics systems. *Naval Research Logistics* 37:1–30.

Reed JE, Ward AR (2004) A diffusion approximation for a generalized Jackson network with reneging. *Proceedings of the 42nd Annual Allerton Conference on Communication, Control, and Computing* (Monticello, IL.).

Rubino M, Ata B (2009) Dynamic control of a make-to-order, parallel-server system with cancellations. *Operations Research* 57:94–108.

Sleptchenko A, van der Heijden MC, van Harten A (2002) Effects of finite repair capacity in multi-echelon, multi-indenture service part supply systems. *International Journal of Production Economics* 79:209–230.

Sleptchenko A, van der Heijden MC, van Harten A (2005) Using repair priorities to reduce stock investment in spare part networks. *European Journal of Operational Research* 163:733–750.

Tiemessen HGH, van Houtum GJ (2013) Reducing costs of repairable inventory supply systems via dynamic scheduling. *International Journal of Production Economics* 143:478–488.

van der Heijden M, Alvarez E, Schutten JMJ (2013) Inventory reduction in spare part networks by selective throughput time reduction. *International Journal of Production Economics* 143:509–517.

van Houtum GJ, Kranenburg B (2015) *Spare Parts Inventory Control under System Availability Constraints* (London: Springer).

Verrijdt J, Adan I, Dekok T (1998) A trade off between emergency repair and inventory investment. *IIE Transactions* 30:119–132.

Wein LM (1992) Dynamic scheduling of a multiclass make-to-stock queue. *Operations Research* 40:724–735.

Whitt W (2002) *Stochastic-Process Limits: An Introduction to Stochastic-Process Limits and Their Application to Queues* (New York: Springer).

Zheng YS, Zipkin P (1990) A queueing model to analyze the value of centralized inventory information. *Operations Research* 38:296–307.

ONLINE APPENDIX

This online appendix is associated with the manuscript titled "Joint Inventory and Scheduling Control in a Repair Facility". We present the proof of Proposition 1 in Section A.1. We present the proofs of Lemmas 1 and 2 in Sections A.2 and A.3, respectively. Section A.4 presents the proofs and results related to the regulator mappings defined in Definitions 3 and 5. We present the proofs of Theorems 1 and 2 in Sections A.5 and A.6, respectively. We present the proof of Proposition 2 in Section A.7. We present the proofs of Theorems 3 and 4 in Sections A.8 and A.9, respectively. Finally, we present some additional numerical experiments in Section A.10.

A.1. Proof of Proposition 1

(i) Let $\left(W, I, E, \boldsymbol{a}, \boldsymbol{S}\right)$ be a feasible solution to the EWF (10). Let

$$
\begin{aligned}
&Q_i(t,\omega) := a_i(t, W(t,\omega), \omega), \qquad \forall i \in \mathcal{I},\ t \in \mathbb{R}_+,\ \omega \in \Omega,\\
&E_k := \mu_k E, \qquad\qquad E_i := \mathbf{0}, \quad \forall i \in \mathcal{I}\backslash\{k\},\\
&Y_i(t) := \frac{1}{\mu_i}\left(Q_i(t) + \lambda_i \int_0^t \left(Q_i(s) - S_i\right)^+ \mathrm{d}s + E_i(t) - X_i(t)\right), \qquad \forall i \in \mathcal{I},\ t \in \mathbb{R}_+.
\end{aligned}
$$

One can see that $\left(Q_i, S_i, Y_i, I, E_i, i \in \mathcal{I}\right)$ defined above is a feasible solution to the BCP (9) with the objective function value equal to the one of $\left(W, I, E, \boldsymbol{a}, \boldsymbol{S}\right)$.

(ii) Let $\left(Q_i, S_i, Y_i, I, E_i, i \in \mathcal{I}\right)$ be a feasible solution to the BCP (9). Let $W := \sum_{i\in\mathcal{I}} Q_i/\mu_i$, $E := \sum_{i\in\mathcal{I}} E_i/\mu_i$, and

$$
a_i\left(t, W(t,\omega), \omega\right) := Q_i(t,\omega), \qquad \forall i \in \mathcal{I},\ t \in \mathbb{R}_+,\ \omega \in \Omega.
$$

Then, one can see that $\left(W, I, E, \boldsymbol{a}, \boldsymbol{S}\right)$ defined above is a feasible solution to the EWF (10) with the objective function value less than or equal to the one of $\left(Q_i, S_i, Y_i, I, E_i, i \in \mathcal{I}\right)$.

A.2. Proof of Lemma 1

Because $f'(w) \le b_j\mu_j/\lambda_j$ and by (11), the objective function of the optimization problem (14) is nonnegative and thus $z(w) \ge 0$.

Case i. Suppose that $\sum_{i\in\mathcal{I}} S_i/\mu_i \ge w$. If there exist a feasible solution to (14) under which $q_i(w) \le S_i$, then the associated objective function value is equal to 0 and thus that feasible solution is optimal. Next, we will prove that such a feasible solution always exists. There exists $m \in \mathcal{I}$ such that

$$\sum_{i<m} \frac{S_i}{\mu_i} \leq w, \qquad \sum_{i\leq m} \frac{S_i}{\mu_i} \geq w,$$

where $\sum_{i<1} S_i/\mu_i := 0$ for completeness. Then, the following feasible solution is optimal.

$$q_i(w) = S_i, \qquad \forall i < m,$$

$$q_m(w) = \mu_m \left(w - \sum_{i<m} \frac{S_i}{\mu_i} \right),$$

$$q_i(w) = 0, \qquad \forall i > m.$$

Observe that there can be multiple optimal solutions to (14) in this case.

Case ii. Suppose that $\sum_{i\in\mathcal{I}} S_i/\mu_i < w$. Because $f'(w) \leq b_j\mu_j/\lambda_j$ and by (11), we have $\min_{i\in\mathcal{I}} \{b_i\mu_i - \lambda_i f'(w)\} \geq 0$. Let $\boldsymbol{q}$ be an arbitrary feasible solution to (14). Consider the objective function value of $\boldsymbol{q}$.

$$\begin{aligned} \sum_{i\in\mathcal{I}} \left(b_i - \frac{\lambda_i}{\mu_i} f'(w) \right) (q_i - S_i)^+ &= \sum_{i\in\mathcal{I}} (b_i\mu_i - \lambda_i f'(w)) \left(\frac{q_i}{\mu_i} - \frac{S_i}{\mu_i} \right)^+ \\ &\geq \min_{i\in\mathcal{I}} \{b_i\mu_i - \lambda_i f'(w)\} \sum_{i\in\mathcal{I}} \left(\frac{q_i}{\mu_i} - \frac{S_i}{\mu_i} \right)^+ \\ &\geq \min_{i\in\mathcal{I}} \{b_i\mu_i - \lambda_i f'(w)\} \left(\sum_{i\in\mathcal{I}} \frac{q_i}{\mu_i} - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i} \right)^+ \\ &= \min_{i\in\mathcal{I}} \{b_i\mu_i - \lambda_i f'(w)\} \left(w - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i} \right)^+, \end{aligned} \tag{A.1}$$

where the last equality is by the fact that $\boldsymbol{q} \in \mathcal{A}(w)$. Therefore, (A.1) provides a lower bound on the objective function value of all feasible solutions. Recall the feasible solution $\boldsymbol{q}^*(w)$ defined in the statement of Lemma 1. Because the objective function value of $\boldsymbol{q}^*(w)$ achieves the lower bound in (A.1), $\boldsymbol{q}^*(w)$ is optimal.

A.3. Proof of Lemma 2

Because the Bellman equations (12) and (13) contain f' and f'' but do not contain the function f itself, they are first-order ODEs. Therefore, we start by solving and studying a related initial value problem (IVP) in Section A.3.1. Then, we present the proof of Lemma 2 in Section A.3.2.

A.3.1. An initial value problem

By (12b), (13c), and Lemma 1, we can replace the minimization problems in (12a), (13a), and (13b) with the optimal objective function value of (14). Consequently, we consider the following IVP for all $\gamma \in \mathbb{R}$:

$$g'_\gamma(w) = \frac{2}{\Sigma}\left[\gamma - \theta g_\gamma(w) - \min_{i\in\mathcal{I}}\{b_i\mu_i - \lambda_i g_\gamma(w)\}\left(w - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)^+\right], \qquad \forall w \in \mathbb{R}_+, \tag{A.2a}$$

$$g_\gamma(0) = 0. \tag{A.2b}$$

Let $F_0 : \mathbb{R} \to \mathbb{R}$ and $F : \mathbb{R}^3 \to \mathbb{R}$ be such that

$$F_0(g) := \min_{i\in\mathcal{I}}\{b_i\mu_i - \lambda_i g\}, \qquad \forall g \in \mathbb{R}, \tag{A.3}$$
$$F(w,g,\gamma) := \frac{2}{\Sigma}\left[\gamma - \theta g - F_0(g)\left(w - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)^+\right], \qquad \forall w,g,\gamma \in \mathbb{R}.$$

Observe that the IVP (A.2) is equivalent to the following IVP:

$$g'_\gamma(w) = F(w, g_\gamma(w), \gamma), \quad \forall w \in \mathbb{R}_+, \qquad\qquad g_\gamma(0) = 0.$$

Observe also that F_0 is continuous, piecewise linear with finitely many break points, concave, and strictly decreasing. Without loss of generality, let us assume that F_0 has $m-1$ break points on $\mathbb{R}_+$ where $m \in \{1,2,\ldots,I\}$ so that F_0 is piecewise combination of m different lines on $\mathbb{R}_+$. Let the index of those m lines be denoted by the set $\{i_1, i_2, \ldots, i_m\} \subset \mathcal{I}$. Without loss of generality, we assume that

$$b_{i_1}\mu_{i_1} < b_{i_2}\mu_{i_2} < \ldots < b_{i_m}\mu_{i_m}, \tag{A.4a}$$

$$\lambda_{i_1} < \lambda_{i_2} < \ldots < \lambda_{i_m}, \tag{A.4b}$$

$$\frac{b_{i_1}\mu_{i_1}}{\lambda_{i_1}} > \frac{b_{i_2}\mu_{i_2}}{\lambda_{i_2}} > \ldots > \frac{b_{i_m}\mu_{i_m}}{\lambda_{i_m}}. \tag{A.4c}$$

By (11), we must have $i_m = j$. Figure A.3.1 illustrates the function F_0 when $m = 3$.

Let

$$\bar{\lambda} := \max_{i\in\mathcal{I}}\lambda_i, \qquad \overline{b\mu} := \max_{i\in\mathcal{I}}\{b_i\mu_i\}.$$

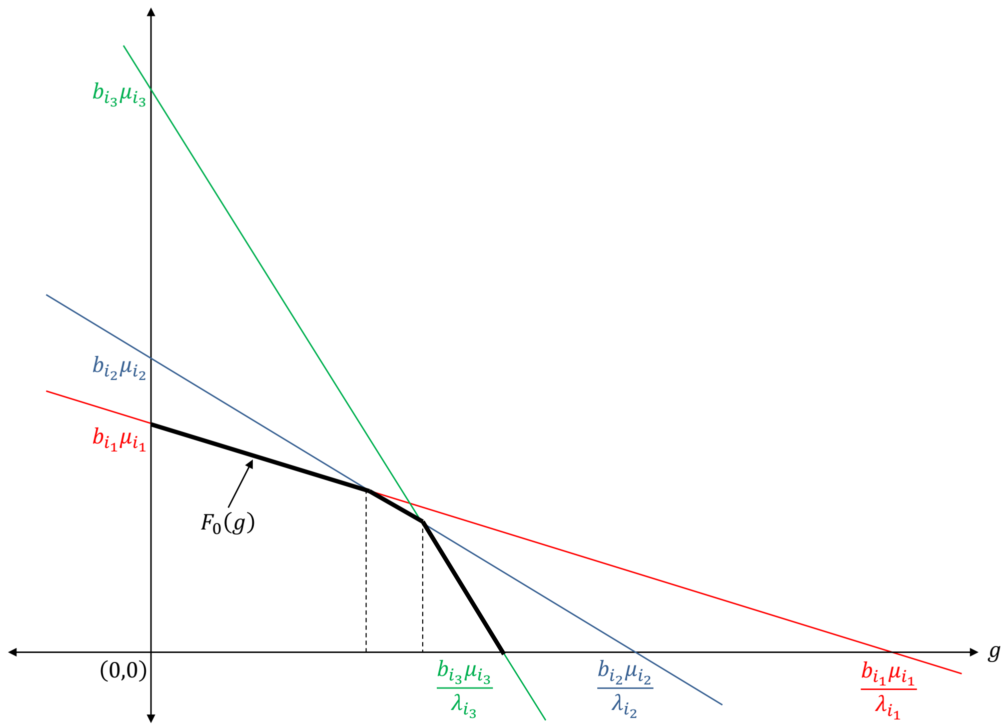


Figure A.3.1 (Color online) An illustration of the function F_0 with two break points.

Fix arbitrary $(w_1, w_2, g_1, g_2, \gamma_1, \gamma_2) \in \mathbb{R}^6$. Then,

$$\begin{aligned}
&|F(w_1, g_1, \gamma_1) - F(w_2, g_2, \gamma_2)| \\
&\quad = \frac{2}{\Sigma} \left| (\gamma_1 - \gamma_2) - \theta(g_1 - g_2) - F_0(g_1) \left(w_1 - \sum_{i \in \mathcal{I}} \frac{S_i}{\mu_i} \right)^+ + F_0(g_2) \left(w_2 - \sum_{i \in \mathcal{I}} \frac{S_i}{\mu_i} \right)^+ \right| \\
&\quad \le \frac{2}{\Sigma} \left[|\gamma_1 - \gamma_2| + |\theta|\,|g_1 - g_2| + |F_0(g_1)|\,|w_1 - w_2| + \left(w_2 - \sum_{i \in \mathcal{I}} \frac{S_i}{\mu_i} \right)^+ |F_0(g_1) - F_0(g_2)| \right] \\
&\quad \le \frac{2}{\Sigma} \left[|\gamma_1 - \gamma_2| + \left(\overline{b\mu} + \bar{\lambda}\,|g_1| \right) |w_1 - w_2| + \left(|\theta| + \bar{\lambda} \left(w_2 - \sum_{i \in \mathcal{I}} \frac{S_i}{\mu_i} \right)^+ \right) |g_1 - g_2| \right]. \qquad \text{(A.5)}
\end{aligned}$$

The following observations are straightforward.

- F is continuous in $(w, g, \gamma) \in \mathbb{R}^3$.
- If two variables among $(w, g, \gamma) \in \mathbb{R}^3$ are fixed, F is Lipschitz continuous in the remaining variable.
- Fix an arbitrary $w \in \mathbb{R}$. By (A.5), for all $(g_1, g_2, \gamma_1, \gamma_2) \in \mathbb{R}^4$,

$$|F(w, g_1, \gamma_1) - F(w, g_2, \gamma_2)| \le \frac{2}{\Sigma} |\gamma_1 - \gamma_2| + L(w)\,|g_1 - g_2|, \qquad \text{(A.6)}$$

where $L:\mathbb{R}\to\mathbb{R}_+$ is a nonnegative and continuous function such that

$$L(w) := \frac{2}{\Sigma}\left[|\theta| + \bar{\lambda}\left(w - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)^+\right], \qquad \forall w\in\mathbb{R}.$$

We have the following existence and uniqueness results for the IVP (A.2).

Lemma A.3.1. *For all $\gamma\in\mathbb{R}$, there exists a unique solution to the IVP (A.2), which is in $\mathbb{C}_1$.*

Proof: Fix an arbitrary $\gamma\in\mathbb{R}$. Because $F(\cdot,\cdot,\gamma)$ is continuous in $(w,g)\in\mathbb{R}^2$, by Peano existence theorem (see for example page 4 of Polyanin and Zaitsev (2018)), there exists an open interval around $w=0$ on which the IVP (A.2) has a solution.

Let $F_1:\mathbb{R}\to\mathbb{R}_+$ and $F_2:\mathbb{R}\to\mathbb{R}_+$ be such that

$$F_1(w) := 1 + \sqrt{|\gamma|} + \left(w - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)^+,$$

$$F_2(g) := \frac{2}{\Sigma}\left(\sqrt{|\gamma|} + \overline{b\mu} + \left(|\theta| + \bar{\lambda}\right)|g|\right).$$

Both F_1 and F_2 are strictly positive, continuous, and nondecreasing. Furthermore,

$$|F(w,g,\gamma)| \le F_1(w)F_2(g), \qquad \forall (w,g)\in\mathbb{R}^2,$$

$$\int_0^\infty \frac{\mathrm{d}g}{F_2(g)} = \infty.$$

By theorem 9.3 of Bainov and Simeonov (1992), all solutions to the IVP (A.2) on the open interval around $w=0$ can be extended to $\mathbb{R}_+$. Therefore, there exists a solution to the IVP (A.2).

Next, we will prove the uniqueness of the solution. Let $g_\gamma^{(1)}$ and $g_\gamma^{(2)}$ be two different solutions to the IVP (A.2). For all $w\in\mathbb{R}_+$,

$$\begin{aligned}\left|g_\gamma^{(1)}(w) - g_\gamma^{(2)}(w)\right| &\le \left|g_\gamma^{(1)}(0) - g_\gamma^{(2)}(0)\right| + \int_0^w \left|F\left(u, g_\gamma^{(1)}(u),\gamma\right) - F\left(u, g_\gamma^{(2)}(u),\gamma\right)\right|\mathrm{d}u \\ &\le \left|g_\gamma^{(1)}(0) - g_\gamma^{(2)}(0)\right| + \int_0^w L(u)\left|g_\gamma^{(1)}(u) - g_\gamma^{(2)}(u)\right|\mathrm{d}u \\ &\le \left|g_\gamma^{(1)}(0) - g_\gamma^{(2)}(0)\right|\exp\left\{\int_0^w L(u)\mathrm{d}u\right\} \\ &= 0,\end{aligned}$$

where the second inequality follows from (A.6), the last inequality is by the Grönwall's inequality (see theorem 1.1 of Bainov and Simeonov (1992)), and the equality follows from the fact that $g_\gamma^{(1)}(0) = g_\gamma^{(2)}(0) = 0$. Therefore, $g_\gamma^{(1)}(w) = g_\gamma^{(2)}(w)$ for all $w \in \mathbb{R}_+$ implying the uniqueness of the IVP (A.2) solution.

By (A.2) and because F_0 is continuous, g'_γ is also continuous implying that $g_\gamma \in \mathbb{C}_1$. □

By algebra, one can see that

$$g_\gamma(w) = \begin{cases} \frac{2\gamma}{\Sigma} w, & \text{if } \theta = 0 \text{ and } w \in \left[0, \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right], \\ \frac{\gamma}{\theta}\left(1 - \mathrm{e}^{-\frac{2\theta}{\Sigma} w}\right), & \text{if } \theta < 0 \text{ and } w \in \left[0, \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right], \end{cases} \tag{A.7}$$

that is, (A.7) presents the unique solution to the IVP (A.2) on the interval $\left[0, \sum_{i\in\mathcal{I}} S_i/\mu_i\right]$. Furthermore,

$$g'_\gamma(w) = \begin{cases} \frac{2\gamma}{\Sigma}, & \text{if } \theta = 0 \text{ and } w \in \left[0, \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right], \\ \frac{2\gamma}{\Sigma}\mathrm{e}^{-\frac{2\theta}{\Sigma} w}, & \text{if } \theta < 0 \text{ and } w \in \left[0, \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right]. \end{cases} \tag{A.8}$$

Next, we will derive some properties of the unique solution to the IVP (A.2).

LEMMA A.3.2. 1. *For all fixed $w \in \mathbb{R}_+$, $g_\gamma(w)$ is continuous in $\gamma \in \mathbb{R}$.*

2. *For all fixed $w \in \mathbb{R}_{++}$, $g_\gamma(w)$ is strictly increasing in $\gamma \in \mathbb{R}$.*

Proof: We start with proving the first part of the lemma. Fix arbitrary $w \in \mathbb{R}_+$ and $(\gamma_1, \gamma_2) \in \mathbb{R}$.

$$\begin{aligned} |g_{\gamma_1}(w) - g_{\gamma_2}(w)| &\le |g_{\gamma_1}(0) - g_{\gamma_2}(0)| + \int_0^w |F(u, g_{\gamma_1}(u), \gamma_1) - F(u, g_{\gamma_2}(u), \gamma_2)|\,\mathrm{d}u \\ &\le \frac{2}{\Sigma}|\gamma_1 - \gamma_2|\,w + \int_0^w L(u)\,|g_{\gamma_1}(u) - g_{\gamma_2}(u)|\,\mathrm{d}u \\ &\le \frac{2}{\Sigma} w \exp\left\{\int_0^w L(u)\mathrm{d}u\right\} |\gamma_1 - \gamma_2|, \end{aligned} \tag{A.9}$$

where the second inequality follows from (A.2b) and (A.6) and the last inequality follows from a Grönwall type inequality (see for example corollary 1.2 of Bainov and Simeonov (1992)). Consequently, for all fixed $w \in \mathbb{R}_+$, $g_\gamma(w)$ is Lipschitz continuous in γ.

Next, we will prove the second part of the lemma by the proof by contradiction technique. Let us fix $\gamma_1, \gamma_2 \in \mathbb{R}$ such that $\gamma_1 > \gamma_2$. Suppose that there exists a $w \in \mathbb{R}_{++}$ at which $g_{\gamma_1}(w) \le g_{\gamma_2}(w)$. By

(A.2), $g'_{\gamma_1}(0) = 2\gamma_1/\Sigma > 2\gamma_2/\Sigma = g'_{\gamma_2}(0)$. Therefore, there must exist $w^* \in \mathbb{R}_{++}$ such that $g_{\gamma_1}(w) > g_{\gamma_2}(w)$ for all $w \in (0, w^*)$ and $g_{\gamma_1}(w^*) = g_{\gamma_2}(w^*)$ due to the continuity of g_{γ_1} and g_{γ_2}. By (A.2),

$$\begin{aligned}0 = g_{\gamma_1}(w^*) - g_{\gamma_2}(w^*) = \frac{2}{\Sigma}\Bigg[&(\gamma_1 - \gamma_2)\, w^* - \theta \int_0^{w^*} (g_{\gamma_1}(w) - g_{\gamma_2}(w))\, \mathrm{d}w \\ &- \int_0^{w^*} (F_0(g_{\gamma_1}(w)) - F_0(g_{\gamma_2}(w))) \left(w - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i} \right)^+ \mathrm{d}w \Bigg] \\ > 0, \end{aligned} \tag{A.10}$$

where the strict inequality follows from the fact that $\gamma_1 > \gamma_2$, $\theta \le 0$, $g_{\gamma_1}(w) > g_{\gamma_2}(w)$ for all $w \in (0, w^*)$, and F_0 is a strictly decreasing function (see for example Figure A.3.1). The strict inequality in (A.10) is a contradiction, implying that w^* does not exist. □

LEMMA A.3.3. *We have* $\max_{w\in\mathbb{R}_+} g_0(w) = 0$ *and* $\inf_{w\in\mathbb{R}_+} g_0(w) < 0$.

Proof: We will use the proof by contradiction technique. Suppose that there exists a $w \in \mathbb{R}_{++}$ at which $g_0(w) > 0$. Then there must exist $w^* \in \mathbb{R}_+$ such that

$$w^* = \inf\{w \in \mathbb{R}_+ : g_0(w) > 0\}.$$

Because g_0 is continuous, we must have $g_0(w^*) = 0$. By (A.7), $g_0(w) = 0$ for all $w \in \left[0, \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right]$. Therefore, it must be the case that $w^* \ge \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}$. There are two cases to consider.

First, suppose that $w^* > \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}$. Because $g_0(w^*) = 0$ and by (A.2),

$$g'_0(w^*) = -\frac{2}{\Sigma} b_{i_1}\mu_{i_1} \left(w^* - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i} \right) < 0,$$

which creates a contradiction because we cannot have $g_0(w^*+) > 0$ if $g'_0(w^*) < 0$.

Second, suppose that $w^* = \sum_{i\in\mathcal{I}} S_i/\mu_i$. By (A.2), we have $g'_0(w^*) = 0$. Therefore, it must be the case that $g_0(w^*+) > 0$ and $g'_0(w^*+) > 0$. There are two sub-cases to consider. In the first sub-case, suppose that $\theta = 0$. Because g_0 is continuous, there exists an $\epsilon > 0$ such that

$$0 \le g_0(u) \le \frac{b_{i_m}\mu_{i_m}}{\lambda_{i_m}}, \qquad \forall u \in [w^*, w^* + \epsilon],$$

which implies that $F_0(g_0(u)) \geq 0$ for all $u \in [w^*, w^* + \epsilon]$ (recall (A.4) and Figure A.3.1). Then, by (A.2),

$$g_0'(u) = -\frac{2}{\Sigma} F_0\left(g_0(u)\right) \left(u - \sum_{i \in \mathcal{I}} \frac{S_i}{\mu_i} \right) \leq 0, \qquad \forall u \in [w^*, w^* + \epsilon],$$

which is a contradiction because $g_0'(w^*+) > 0$ by assumption.

In the second sub-case, suppose that $\theta < 0$. Because both F_0 and g_0' are continuous, $g_0'(w^*) = 0$, and $F_0(g_0(w^*)) = F_0(0) = b_{i_1}\mu_{i_1}$, there exists an $\epsilon > 0$ such that

$$0 \leq g_0'(w^* + v) \leq \frac{b_{i_1}\mu_{i_1}}{2|\theta|}, \qquad \forall v \in [0, \epsilon], \tag{A.11}$$

$$\frac{b_{i_1}\mu_{i_1}}{2} < F_0(x) \leq b_{i_1}\mu_{i_1}, \qquad \forall x \in \left[0, \frac{b_{i_1}\mu_{i_1}}{2|\theta|}\epsilon\right]. \tag{A.12}$$

Because $g_0(w^*) = 0$, (A.11) implies that

$$0 \leq g_0(w^* + v) \leq \frac{b_{i_1}\mu_{i_1}}{2|\theta|} v, \qquad \forall v \in [0, \epsilon]. \tag{A.13}$$

For all $v \in [0, \epsilon]$, we have

$$\begin{aligned} g_0'(w^* + v) &= \frac{2}{\Sigma}\left[-\theta g_0(w^* + v) - F_0\left(g_0(w^* + v)\right) v\right] \\ &\leq \frac{2}{\Sigma}\left[|\theta| \frac{b_{i_1}\mu_{i_1}}{2|\theta|} v - F_0\left(\frac{b_{i_1}\mu_{i_1}}{2|\theta|} v\right) v\right] \\ &= \frac{2v}{\Sigma}\left[\frac{b_{i_1}\mu_{i_1}}{2} - F_0\left(\frac{b_{i_1}\mu_{i_1}}{2|\theta|} v\right)\right] \\ &< 0, \end{aligned} \tag{A.14}$$

where the first equality follows from (A.2) and the fact that $w^* = \sum_{i \in \mathcal{I}} S_i/\mu_i$, the first inequality follows from (A.13) and the fact that F_0 is a strictly decreasing function, and the strict inequality follows from (A.12). However, (A.14) is a contradiction because $g_0'(w^*+) > 0$ by assumption.

Consequently, $g_0(w) \leq 0$ for all $w \in \mathbb{R}_+$. Because $g_0(0) = 0$ by (A.2b), we have $\max_{w \in \mathbb{R}_+} g_0(w) = 0$. Finally, one can easily see that $g_0(w) = 0$ for all $w \in \mathbb{R}_+$ is not a solution to the IVP (A.2). Therefore, there must exist a $w \in \mathbb{R}_+$ at which $g_0(w) < 0$ implying that $\inf_{w \in \mathbb{R}_+} g_0(w) < 0$. □

LEMMA A.3.4. *Fix an arbitrary $\gamma \in \mathbb{R}_{++}$. If $g_\gamma(w^*) = b_j\mu_j/\lambda_j$ for some $w^* \in \mathbb{R}_+$, then $g'_\gamma(w) > 0$ for all $w \in [w^*, \infty)$ and thus $g_\gamma(w)$ is strictly increasing in w on the interval $[w^*, \infty)$ and $\lim_{w\to\infty} g_\gamma(w) = \infty$. In other words, if g_γ hits the point $b_j\mu_j/\lambda_j$, henceforth, it strictly increases to infinity.*

Proof: If $g_\gamma(w) \geq b_j\mu_j/\lambda_j$ for some $w \in \mathbb{R}_+$, then by (A.2a)

$$g'_\gamma(w) = \frac{2}{\Sigma}\left[\gamma - \theta g_\gamma(w) - F_0\left(g_\gamma(w)\right)\left(w - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right)^+\right] \geq \frac{2}{\Sigma}\left[\gamma - \theta\frac{b_j\mu_j}{\lambda_j}\right] \geq \frac{2}{\Sigma}\gamma,$$

where the first inequality follows from the fact that $F_0(x) \leq 0$ for all $x \geq b_j\mu_j/\lambda_j$ by (A.4) and the fact that $j = i_m$ (see also Figure A.3.1). Therefore, if $g_\gamma(w^*) = b_j\mu_j/\lambda_j$ for some $w^* \in \mathbb{R}_+$, then $g'_\gamma(w) > 0$ for all $w \in [w^*, \infty)$ and $g_\gamma(w)$ is strictly increasing in w on the interval $[w^*, \infty)$. Finally,

$$g_\gamma(w^* + w) = g_\gamma(w^*) + \int_{w^*}^{w^*+w} g'_\gamma(u)\mathrm{d}u \geq \frac{b_j\mu_j}{\lambda_j} + \frac{2}{\Sigma}\gamma w, \qquad \forall w \in \mathbb{R}_+,$$

implying that $\lim_{w\to\infty} g_\gamma(w) = \infty$. □

LEMMA A.3.5. *There exists a $\gamma \in \mathbb{R}_{++}$ at which $\sup_{w\in\mathbb{R}_+} g_\gamma(w) = \infty$.*

Proof: We will make the proof case by case. First, suppose that $\sum_{i\in\mathcal{I}} S_i/\mu_i > 0$. Let

$$\gamma := \begin{cases} \frac{b_j\mu_j\Sigma}{2\lambda_j}\left(\sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)^{-1}, & \text{if } \theta = 0, \\ -\theta\frac{b_j\mu_j}{\lambda_j}\left(\mathrm{e}^{-\frac{2\theta}{\Sigma}\left(\sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)} - 1\right)^{-1}, & \text{if } \theta < 0. \end{cases}$$

Observe that $\gamma > 0$. Furthermore, by (A.7), we have $g_\gamma\left(\sum_{i\in\mathcal{I}} S_i/\mu_i\right) = b_j\mu_j/\lambda_j$. By Lemma A.3.4, $\lim_{w\to\infty} g_\gamma(w) = \infty$.

Next, suppose that $\sum_{i\in\mathcal{I}} S_i/\mu_i = 0$. By (A.2), we have

$$g'_\gamma(w) = \frac{2}{\Sigma}\left[\gamma - \theta g_\gamma(w) - F_0\left(g_\gamma(w)\right)w\right], \qquad \forall w \in \mathbb{R}_+,$$

$$g_\gamma(0) = 0.$$

By recalling (A.4), let

$$\gamma := \sqrt{2b_{i_1}\mu_{i_1}\Sigma\frac{b_j\mu_j}{\lambda_j}}, \qquad w^* := \frac{\gamma}{2b_{i_1}\mu_{i_1}}.$$

We will prove that $g_\gamma(w) \geq 0$ for all $w \in [0, w^*]$ by the proof by contradiction technique. Suppose that there exists a $w \in [0, w^*]$ such that $g_\gamma(w) < 0$. Let

$$\hat{w} := \inf\{w \in [0, w^*] : g_\gamma(w) < 0\}.$$

Because both g_γ and g'_γ are continuous in w and $g_\gamma(\hat{w}+) < 0$, we must have $g_\gamma(\hat{w}) = 0$ and $g'_\gamma(\hat{w}) \leq 0$. However, we have the following contradiction:

$$g'_\gamma(\hat{w}) = \frac{2}{\Sigma}\left[\gamma - \theta g_\gamma(\hat{w}) - F_0\left(g_\gamma(\hat{w})\right)\hat{w}\right] = \frac{2}{\Sigma}\left[\gamma - b_{i_1}\mu_{i_1}\hat{w}\right] \geq \frac{2}{\Sigma}\left[\gamma - b_{i_1}\mu_{i_1}w^*\right] = \frac{\gamma}{\Sigma} > 0,$$

where the second equality follows from the fact that $g_\gamma(\hat{w}) = 0$ and $F_0(0) = b_{i_1}\mu_{i_1}$ (recall Figure A.3.1). Therefore, we prove that $g_\gamma(w) \geq 0$ for all $w \in [0, w^*]$. By utilizing this fact, we have

$$g'_\gamma(w) = \frac{2}{\Sigma}\left[\gamma - \theta g_\gamma(w) - F_0\left(g_\gamma(w)\right)w\right] \geq \frac{2}{\Sigma}\left[\gamma - b_{i_1}\mu_{i_1}w\right] \geq \frac{2}{\Sigma}\left[\gamma - b_{i_1}\mu_{i_1}w^*\right] = \frac{\gamma}{\Sigma}, \quad \forall w \in [0, w^*].$$

Therefore,

$$g_\gamma(w^*) = \int_0^{w^*} g'_\gamma(w)\mathrm{d}w \geq \frac{\gamma}{\Sigma}w^* = \frac{b_j\mu_j}{\lambda_j}.$$

By Lemma A.3.4, $\lim_{w\to\infty} g_\gamma(w) = \infty$. □

LEMMA A.3.6. 1. *If* $\sup_{w\in\mathbb{R}_+} g_\gamma(w) \geq b_j\mu_j/\lambda_j$, *then* $g'_\gamma(w) > 0$ *for all* $w \in \mathbb{R}_+$ *and thus* $g_\gamma(w)$ *is strictly increasing in* w *and* $\min_{w\in\mathbb{R}_+} g_\gamma(w) = 0$.

2. *Suppose that* $\sup_{w\in\mathbb{R}_+} g_\gamma(w) < b_j\mu_j/\lambda_j$.

   i. $\lim_{w\to\infty} g_\gamma(w) = -\infty$ *and* g_γ *attains its maximum, that is,* $\max_{w\in\mathbb{R}_+} g_\gamma(w) = \sup_{w\in\mathbb{R}_+} g_\gamma(w)$.

   ii. *If* $\gamma > 0$, *the maximum points are strictly greater than* $\sum_{i\in\mathcal{I}} S_i/\mu_i$ *and the derivative of* $g_\gamma(\cdot)$ *at a maximum point is equal to 0.*

   iii. *If* $\gamma > 0$, $g_\gamma(w)$ *is strictly increasing in* w *until the first time the maximum is attained.*

   iv. *If* $\gamma > 0$, *there exists a unique maximum point. After the maximum point,* $g'_\gamma(\cdot)$ *is strictly negative and thus* $g_\gamma(\cdot)$ *is strictly decreasing in* w.

   v. *If* $\gamma > 0$, *the unique maximum point is strictly increasing in* γ.

Proof of Part 1. Because $\sup_{w\in\mathbb{R}_+} g_\gamma(w) \geq b_j\mu_j/\lambda_j$ and by Lemma A.3.2 Part 2 and Lemma A.3.3, we must have $\gamma > 0$. By (A.8) and because $\gamma > 0$, $g'_\gamma(w) > 0$ for all $w \in \left[0, \sum_{i\in\mathcal{I}} S_i/\mu_i\right]$, that is, $g_\gamma(w)$ is strictly increasing on the interval $\left[0, \sum_{i\in\mathcal{I}} S_i/\mu_i\right]$.

Suppose that $g_\gamma(w)$ is not strictly increasing in w. Because $g_\gamma(\cdot) \in \mathbb{C}_1$ and by Lemma A.3.4, there must exist $(w_1, w_2) \in \mathbb{R}_+^2$ such that

$$\sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i} < w_1 < w_2, \qquad 0 = g'_\gamma(w_1) \leq g'_\gamma(w_2), \qquad g_\gamma(w_1) = g_\gamma(w_2) < \frac{b_j\mu_j}{\lambda_j}.$$

By (A.2a), we obtain the following contradiction.

$$\begin{aligned} g'_\gamma(w_1) &= \frac{2}{\Sigma}\left[\gamma - \theta g_\gamma(w_1) - F_0\left(g_\gamma(w_1)\right)\left(w_1 - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right)\right] \\ &> \frac{2}{\Sigma}\left[\gamma - \theta g_\gamma(w_2) - F_0\left(g_\gamma(w_2)\right)\left(w_2 - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right)\right] \\ &= g'_\gamma(w_2), \end{aligned}$$

where the strict inequality follows from the fact that $w_1 < w_2$ and $F_0(x) > 0$ for all $x < b_j\mu_j/\lambda_j$. Therefore, $g_\gamma(w)$ is strictly increasing in w and thus $\min_{w\in\mathbb{R}_+} g_\gamma(w) = g_\gamma(0) = 0$ by (A.2b).

Finally, we will prove that $g'_\gamma(w) > 0$ for all $w \in \mathbb{R}_+$. Observe that the fact that $g_\gamma(w)$ is strictly increasing in w does not necessarily imply that $g'_\gamma(w) > 0$ for all $w \in \mathbb{R}_+$. We will use the proof by contradiction technique. Suppose that $g'_\gamma(w) \leq 0$ for some $w \in \mathbb{R}_+$. Because g'_γ is continuous, there must exist $w^* \in \mathbb{R}_+$ at which $g'_\gamma(w^*) = 0$. Furthermore, by (A.8), we must have $w^* > \sum_{i\in\mathcal{I}} S_i/\mu_i$, and by Lemma A.3.4, $g_\gamma(w^*) < b_j\mu_j/\lambda_j$ and thus $F_0\left(g_\gamma(w^*)\right) > 0$ (recall Figure A.3.1). Because $g_\gamma \in \mathbb{C}_1$ and F_0 is continuous (see (A.3)), there exist $\delta > 0$ and $\epsilon > 0$ such that

$$g'_\gamma(w^* + x) < \epsilon \quad \text{and} \quad \left(|\theta| + \bar{\lambda}\left(w^* - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right)\right)\epsilon + \bar{\lambda}\epsilon x < F_0\left(g_\gamma(w^*)\right), \qquad \forall x \in (0, \delta). \tag{A.15}$$

Recall that g_γ is strictly increasing and F_0 is strictly decreasing. Then,

$$g_\gamma(w^* + x) - g_\gamma(w^*) \leq \epsilon x, \qquad \forall x \in (0, \delta), \tag{A.16a}$$

$$F_0\left(g_\gamma(w^*)\right) - F_0\left(g_\gamma(w^* + x)\right) \leq \bar{\lambda}\left(g_\gamma(w^* + x) - g_\gamma(w^*)\right) \leq \bar{\lambda}\epsilon x, \qquad \forall x \in (0, \delta), \tag{A.16b}$$

where (A.16a) follows from (A.15) and (A.16b) follows from (A.3) (recall Figure A.3.1) and (A.15).

Fix an arbitrary $x \in (0, \delta)$.

$$\begin{aligned}
g_\gamma'(w^*+x) &= g_\gamma'(w^*+x) - g_\gamma'(w^*) \\
&= \frac{2}{\Sigma}\Bigg[|\theta|\left(g_\gamma(w^*+x) - g_\gamma(w^*)\right) + \left(F_0\left(g_\gamma(w^*)\right) - F_0\left(g_\gamma(w^*+x)\right)\right)\left(w^* - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right) \\
&\qquad - F_0\left(g_\gamma(w^*+x)\right)x\Bigg] \\
&\le \frac{2}{\Sigma}\left[|\theta|\epsilon x + \bar{\lambda}\left(w^* - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)\epsilon x - F_0\left(g_\gamma(w^*+x)\right)x\right] \\
&\le \frac{2}{\Sigma}\left[|\theta|\epsilon x + \bar{\lambda}\left(w^* - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)\epsilon x + \bar{\lambda}\epsilon x^2 - F_0\left(g_\gamma(w^*)\right)x\right] \\
&= \frac{2}{\Sigma}x\left[\left(|\theta| + \bar{\lambda}\left(w^* - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)\right)\epsilon + \bar{\lambda}\epsilon x - F_0\left(g_\gamma(w^*)\right)\right] \\
&< 0, \qquad\qquad (A.17)
\end{aligned}$$

where the first equality follows from the fact that $g_\gamma'(w^*) = 0$, the second equality follows from (A.2a) and (A.3), the first inequality follows from (A.16), the second inequality follows from (A.16b), and the strict inequality follows from (A.15).

Because $g_\gamma(w)$ is strictly increasing in w and $g_\gamma \in \mathbb{C}_1$, (A.17) creates a contradiction. Therefore, w^* cannot exist and thus $g_\gamma'(w) > 0$ for all $w \in \mathbb{R}_+$. □

Proof of Part 2. **i.)** By (A.2a),

$$\begin{aligned}
g_\gamma'(w) &= \frac{2}{\Sigma}\left[\gamma - \theta g_\gamma(w) - F_0\left(g_\gamma(w)\right)\left(w - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)^+\right] \\
&\le \frac{2}{\Sigma}\left[\gamma - \theta\frac{b_j\mu_j}{\lambda_j} - F_0\left(\sup_{\tilde{w}\in\mathbb{R}_+} g_\gamma(\tilde{w})\right)\left(w - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)^+\right] \\
&\to -\infty \quad \text{as } w \to \infty,
\end{aligned}$$

where the inequality follows from the fact that $\theta \le 0$, $g_\gamma(w) < b_j\mu_j/\lambda_j$ for all $w \in \mathbb{R}_+$, F_0 is strictly decreasing, and $F_0\big(\sup_{\tilde{w}\in\mathbb{R}_+} g_\gamma(\tilde{w})\big) > F_0\big(b_j\mu_j/\lambda_j\big) = 0$ (see (A.4) and Figure A.3.1). Therefore, $\lim_{w\to\infty} g_\gamma'(w) = -\infty$. There exist $w_1 \in \mathbb{R}_+$ and $K_1 > 0$ such that if $w \ge w_1$, $g_\gamma'(w) \le -K_1$. Then,

$$g_\gamma(w_1 + w) = g_\gamma(w_1) + \int_{w_1}^{w_1+w} g_\gamma'(u)\mathrm{d}u \le \frac{b_j\mu_j}{\lambda_j} - K_1 w \to -\infty, \quad \text{as } w \to \infty,$$

implying that $\lim_{w\to\infty} g_\gamma(w) = -\infty$. Therefore, there exist $w_2 \in \mathbb{R}_+$ and $K_2 > 0$ such that if $w \geq w_2$, $g_\gamma(w) \leq -K_2$ and thus $\sup_{w\in\mathbb{R}_+} g_\gamma(w) = \sup_{w\in[0,w_2]} g_\gamma(w)$ by (A.2b). Because g_γ is continuous, it attains its maximum in compact intervals and thus

$$\sup_{w\in\mathbb{R}_+} g_\gamma(w) = \sup_{w\in[0,w_2]} g_\gamma(w) = \max_{w\in[0,w_2]} g_\gamma(w) = \max_{w\in\mathbb{R}_+} g_\gamma(w).$$

ii.) By (A.8), if $\gamma > 0$, then $g'_\gamma(w) > 0$ for all $w \in \left[0, \sum_{i\in\mathcal{I}} S_i/\mu_i\right]$. This and the continuity of $g'_\gamma(\cdot)$ in w imply that the maximum points of $g_\gamma(\cdot)$ must be strictly greater than $\sum_{i\in\mathcal{I}} S_i/\mu_i$. Let $w^* \in \mathbb{R}_+$ be a maximum point of g_γ, that is, $g_\gamma(w^*) = \max_{w\in\mathbb{R}_+} g_\gamma(w)$. Suppose that, $g'_\gamma(w^*) < 0$. Because $w^* > 0$ and $g'_\gamma(w)$ is continuous in w, there exists an $\epsilon > 0$ such that $g'_\gamma(w^* - u) < 0$ for all $u \in [0, \epsilon]$. Then,

$$g_\gamma(w^* - \epsilon) = g_\gamma(w^*) - \int_{w^*-\epsilon}^{w^*} g'_\gamma(u)\mathrm{d}u > g_\gamma(w^*),$$

which is a contradiction because w^* is a maximum point. Therefore, $g'_\gamma(w^*) \not< 0$. Similarly, we can show that $g'_\gamma(w^*) \not> 0$ implying that $g'_\gamma(w^*) = 0$.

iii.) The fact that $g_\gamma(w)$ is strictly increasing in w until the first time the maximum is attained follows from exactly the same argument that we use in the proof of Part 1 of this lemma.

iv.) Suppose that $w_1, w_2 \in \mathbb{R}_+$ be two different maximum points such that $w_2 > w_1$. By Part 2.ii of this lemma, $w_2 > w_1 > \sum_{i\in\mathcal{I}} S_i/\mu_i$. Because g'_γ is equal to zero at the maximum points, by (A.2a), we obtain the following contradiction.

$$0 = \gamma - \theta g_\gamma(w_1) - F_0\left(g_\gamma(w_1)\right)\left(w_1 - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right) > \gamma - \theta g_\gamma(w_2) - F_0\left(g_\gamma(w_2)\right)\left(w_2 - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right) = 0,$$

where the strict inequality is by the fact that $w_1 < w_2$, $g_\gamma(w_1) = g_\gamma(w_2) < b_j\mu_j/\lambda_j$, and F_0 is strictly decreasing such that $F_0\left(b_j\mu_j/\lambda_j\right) = 0$. Therefore, there exists a unique maximum point if $\gamma > 0$. Let w^* denote the unique maximum point. Consider an arbitrary $w > w^*$. By (A.2a),

$$0 = \gamma - \theta g_\gamma(w^*) - F_0\left(g_\gamma(w^*)\right)\left(w^* - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right) > \gamma - \theta g_\gamma(w) - F_0\left(g_\gamma(w)\right)\left(w - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right) = g'_\gamma(w),$$

where the strict inequality follows from the fact that $w > w^*$, $\theta \leq 0$, $g_\gamma(w) < g_\gamma(w^*) < b_j\mu_j/\lambda_j$, F_0 is strictly decreasing, and $F_0\left(b_j\mu_j/\lambda_j\right) = 0$. Because $g'_\gamma(w) < 0$ for all $w \in [w^*, \infty)$, $g_\gamma(\cdot)$ is strictly decreasing on $[w^*, \infty)$.

v.) Let $w(\gamma)$ denote the unique maximum point for all $\gamma > 0$ such that $\sup_{w\in\mathbb{R}_+} g_\gamma(w) < b_j\mu_j/\lambda_j$. Let $(\gamma_1,\gamma_2)\in\mathbb{R}_+^2$ be such that $0<\gamma_1<\gamma_2$ and $w(\gamma_1)\geq w(\gamma_2)$. By Part 2.ii of this lemma and (A.2a),

$$\begin{aligned} 0 = g'_{\gamma_2}(w(\gamma_2)) &= \frac{2}{\Sigma}\left[\gamma_2 - \theta g_{\gamma_2}(w(\gamma_2)) - F_0\left(g_{\gamma_2}(w(\gamma_2))\right)\left(w(\gamma_2) - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)\right] \\ &> \frac{2}{\Sigma}\left[\gamma_1 - \theta g_{\gamma_1}(w(\gamma_2)) - F_0\left(g_{\gamma_1}(w(\gamma_2))\right)\left(w(\gamma_2) - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)\right] \\ &= g'_{\gamma_1}(w(\gamma_2)), \end{aligned}$$

where the strict inequality follows from the fact that $\gamma_1<\gamma_2$, $\theta\leq 0$, $g_{\gamma_2}(w(\gamma_2)) > g_{\gamma_1}(w(\gamma_2))$ (recall Lemma A.3.2 Part 2), and F_0 is strictly decreasing. The fact that $g'_{\gamma_1}(w(\gamma_2))<0$ is a contradiction because $w(\gamma_1)\geq w(\gamma_2)$ and $g'_{\gamma_1}(w)\geq 0$ for all $w\in[0,w(\gamma_1)]$ by Part 2.iii of this lemma. □

Lemma A.3.7. *There exists a $\gamma\in\mathbb{R}_{++}$ such that $0\leq g_\gamma(w) < b_j\mu_j/\lambda_j$ for all $w\in\mathbb{R}_+$.*

Proof: First, we will prove that there exists a $\bar\gamma\in\mathbb{R}_+$ such that if $\gamma\leq\bar\gamma$, then $g_\gamma(w)<b_j\mu_j/\lambda_j$ for all $w\in\mathbb{R}_+$; otherwise, if $\gamma>\bar\gamma$, then $\sup_{w\in\mathbb{R}_+} g_\gamma(w)=\infty$. By Lemma A.3.2 Part 2 and Lemmas A.3.3, A.3.4, and A.3.5, there exists a $\bar\gamma\in\mathbb{R}_+$ such that if $\gamma<\bar\gamma$, then $g_\gamma(w)<b_j\mu_j/\lambda_j$ for all $w\in\mathbb{R}_+$; and if $\gamma>\bar\gamma$, then $\sup_{w\in\mathbb{R}_+} g_\gamma(w)=\infty$. Let $\{\gamma_n, n\in\mathbb{N}\}$ be a convergent sequence in $\mathbb{R}_+$ such that $\gamma_n\to\bar\gamma$ as $n\to\infty$ and $g_{\gamma_n}(w)<b_j\mu_j/\lambda_j$ for all $w\in\mathbb{R}_+$ and $n\in\mathbb{N}$. Fix an arbitrary $w\in\mathbb{R}_+$. By Lemma A.3.2 Part 1, $g_{\gamma_n}(w)\to g_{\bar\gamma}(w)$ as $n\to\infty$. Because $g_{\gamma_n}(w)<b_j\mu_j/\lambda_j$ for all $n\in\mathbb{N}$, we must have $g_{\bar\gamma}(w)\leq b_j\mu_j/\lambda_j$. Because $w\in\mathbb{R}_+$ is arbitrarily chosen, we have $g_{\bar\gamma}(w)\leq b_j\mu_j/\lambda_j$ for all $w\in\mathbb{R}_+$. By Lemma A.3.4, $g_{\bar\gamma}(w)<b_j\mu_j/\lambda_j$ for all $w\in\mathbb{R}_+$ (otherwise if $g_{\bar\gamma}(w)=b_j\mu_j/\lambda_j$ for some $w\in\mathbb{R}_+$, then $g_{\bar\gamma}(\cdot)$ tends to infinity).

Next, we will prove that there exists a $\underline{\gamma}\in\mathbb{R}_{++}$ such that if $\gamma\geq\underline{\gamma}$, then $\min_{w\in\mathbb{R}_+} g_\gamma(w)=0$; otherwise, if $\gamma<\underline{\gamma}$, then $\inf_{w\in\mathbb{R}_+} g_\gamma(w)<0$. By Lemma A.3.2 Part 2, Lemmas A.3.3 and A.3.5, and Lemma A.3.6 Part 1, there exists a $\underline{\gamma}\in\mathbb{R}_+$ such that if $\gamma>\underline{\gamma}$, then $\min_{w\in\mathbb{R}_+} g_\gamma(w)=0$; otherwise, if $\gamma<\underline{\gamma}$, then $\inf_{w\in\mathbb{R}_+} g_\gamma(w)<0$. Let $\{\gamma_n, n\in\mathbb{N}\}$ be a convergent sequence in $\mathbb{R}_+$ such that $\gamma_n\to\underline{\gamma}$ as $n\to\infty$ and $\min_{w\in\mathbb{R}_+} g_{\gamma_n}(w)=0$ for all $n\in\mathbb{N}$. Fix an arbitrary $w\in\mathbb{R}_+$. By Lemma A.3.2 Part 1, $g_{\gamma_n}(w)\to g_{\underline{\gamma}}(w)$ as $n\to\infty$. Because $g_{\gamma_n}(w)\geq 0$ for all $n\in\mathbb{N}$, we must have $g_{\underline{\gamma}}(w)\geq 0$.

Because $w \in \mathbb{R}_+$ is arbitrarily chosen, we have $\inf_{w\in\mathbb{R}_+} g_{\underline{\gamma}}(w) \geq 0$. By (A.2b), $g_{\underline{\gamma}}(0) = 0$ and thus $\min_{w\in\mathbb{R}_+} g_{\underline{\gamma}}(w) = 0$. By Lemma A.3.3, we have $\underline{\gamma} > 0$.

Suppose that $\underline{\gamma} > \bar{\gamma}$. For all $\gamma \in (\bar{\gamma}, \underline{\gamma})$, we have $\sup_{w\in\mathbb{R}_+} g_\gamma(w) = \infty$ and $\inf_{w\in\mathbb{R}_+} g_\gamma(w) < 0$. However, by Lemma A.3.6 Part 1, we have $\min_{w\in\mathbb{R}_+} g_\gamma(w) = 0$, which is a contradiction. Therefore, $\underline{\gamma} \not> \bar{\gamma}$ implying that $\underline{\gamma} \leq \bar{\gamma}$. Consequently, for all $\gamma \in [\underline{\gamma}, \bar{\gamma}]$, $0 \leq g_\gamma(w) < b_j\mu_j/\lambda_j$ for all $w \in \mathbb{R}_+$. □

LEMMA A.3.8. *There exists a unique $\gamma \in \mathbb{R}_{++}$ such that $0 \leq g_\gamma(w) < b_j\mu_j/\lambda_j$ for all $w \in \mathbb{R}_+$. Furthermore, $g'_\gamma(w) > 0$ for all $w \in \mathbb{R}_+$ and thus $g_\gamma(w)$ is strictly increasing in w, $\lim_{w\to\infty} g_\gamma(w) = b_j\mu_j/\lambda_j$, and $\sup_{w\in\mathbb{R}_+} g_\gamma(w) = b_j\mu_j/\lambda_j$*

Proof: Let $\gamma \in \mathbb{R}_{++}$ be such that $0 \leq g_\gamma(w) < b_j\mu_j/\lambda_j$ for all $w \in \mathbb{R}_+$. Such a γ exists by Lemma A.3.7. Because $g_\gamma(w) \geq 0$ for all $w \in \mathbb{R}_+$, we must have $\sup_{w\in\mathbb{R}_+} g_\gamma(w) = b_j\mu_j/\lambda_j$ by Lemma A.3.6 Parts 1 and 2.i. Furthermore, $g'_\gamma(w) > 0$ for all $w \in \mathbb{R}_+$ and thus $g_\gamma(w)$ is strictly increasing in w by Lemma A.3.6 Part 1. Because $\sup_{w\in\mathbb{R}_+} g_\gamma(w) = b_j\mu_j/\lambda_j$ and $g_\gamma(w)$ is strictly increasing in w, we must have $0 \leq g_\gamma(w) < b_j\mu_j/\lambda_j$ for all $w \in \mathbb{R}_+$ and $\lim_{w\to\infty} g_\gamma(w) = b_j\mu_j/\lambda_j$.

Finally, we will prove uniqueness. Suppose that $0 \leq g_\gamma(w) < b_j\mu_j/\lambda_j$ for all $w \in \mathbb{R}_+$ and $\gamma \in \{\gamma_1, \gamma_2\}$ such that $\gamma_1 < \gamma_2$. By Lemma A.3.2 Part 2, we have $g_{\gamma_1}(w) < g_{\gamma_2}(w)$ for all $w \in \mathbb{R}_{++}$. By (A.2a), for all $w \in \mathbb{R}_+$,

$$\begin{aligned} g'_{\gamma_1}(w) &= \frac{2}{\Sigma}\left[\gamma_1 - \theta g_{\gamma_1}(w) - F_0\left(g_{\gamma_1}(w)\right)\left(w - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)^+\right] \\ &< \frac{2}{\Sigma}\left[\gamma_2 - \theta g_{\gamma_2}(w) - F_0\left(g_{\gamma_2}(w)\right)\left(w - \sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)^+\right] \\ &= g'_{\gamma_2}(w), \end{aligned}$$

where the strict inequality follows from the fact that $\gamma_1 < \gamma_2$, $\theta \leq 0$, $g_{\gamma_1}(w) \leq g_{\gamma_2}(w)$ for all $w \in \mathbb{R}_+$, and F_0 is strictly decreasing. Recall that $\lim_{w\to\infty} g_\gamma(w) = b_j\mu_j/\lambda_j$ for all $\gamma \in \{\gamma_1, \gamma_2\}$. Then, we have the following contradiction

$$0 = \lim_{w\to\infty}\left(g_{\gamma_2}(w) - g_{\gamma_1}(w)\right) = \lim_{w\to\infty}\int_0^w \left(g'_{\gamma_2}(u) - g'_{\gamma_1}(u)\right)\mathrm{d}u > 0,$$

where the strict inequality follows from the fact that $g'_{\gamma_1}(w) < g'_{\gamma_2}(w)$ for all $w \in \mathbb{R}_+$ and both g'_{γ_1} and g'_{γ_2} are continuous in w. □

Let γ^{NER} be the unique γ described in Lemma A.3.8.

Lemma A.3.9. 1. *On the interval* $[0, \gamma^{NER})$, $\max_{w\in\mathbb{R}_+} g_\gamma(w)$ *exists and is continuous and strictly increasing in* γ.

2. *The function* $\max_{w\in\mathbb{R}_+} g_{(\cdot)}(w)$ *is a bijection from* $[0, \gamma^{NER}) \to [0, b_j\mu_j/\lambda_j)$.

Proof of Part 1: By Lemma A.3.2 Part 2 and Lemma A.3.8, we have $g_\gamma(w) < b_j\mu_j/\lambda_j$ for all $w \in \mathbb{R}_+$ and $\sup_{w\in\mathbb{R}_+} g_\gamma(w) \leq b_j\mu_j/\lambda_j$ for all $\gamma \in [0, \gamma^{NER}]$. We will show that if $\gamma < \gamma^{NER}$, then $\sup_{w\in\mathbb{R}_+} g_\gamma(w) < b_j\mu_j/\lambda_j$. To see this, suppose that $\gamma < \gamma^{NER}$ but $\sup_{w\in\mathbb{R}_+} g_\gamma(w) = b_j\mu_j/\lambda_j$. By Lemma A.3.6 Part 1, $g_\gamma(w)$ is strictly increasing in w and $\min_{w\in\mathbb{R}_+} g_\gamma(w) = 0$. Then, by the uniqueness of γ^{NER} (see Lemma A.3.8), we must have $\gamma = \gamma^{NER}$, which is a contradiction. Therefore, if $\gamma < \gamma^{NER}$, then $\sup_{w\in\mathbb{R}_+} g_\gamma(w) < b_j\mu_j/\lambda_j$. Furthermore, by Lemma A.3.6 Part 2.i, $\max_{w\in\mathbb{R}_+} g_\gamma(w)$ exists for all $\gamma \in [0, \gamma^{NER})$. However, the aforementioned maximum does not exist when $\gamma = \gamma^{NER}$ by Lemma A.3.8.

By recalling Lemma A.3.6 Part 2.iv, for given $\gamma \in (0, \gamma^{NER})$, let $w(\gamma)$ denote the unique maximum of $g_\gamma(\cdot)$, that is, $w(\gamma) := \arg\max_{w\in\mathbb{R}_+} g_\gamma(w)$. By Lemma A.3.6 Part 2.ii, $w(\gamma) > 0$. Let $\gamma_1, \gamma_2 \in (0, \gamma^{NER})$ such that $\gamma_1 < \gamma_2$. Then,

$$\max_{w\in\mathbb{R}_+} g_{\gamma_1}(w) = g_{\gamma_1}(w(\gamma_1)) < g_{\gamma_2}(w(\gamma_1)) \leq \max_{w\in\mathbb{R}_+} g_{\gamma_2}(w),$$

where the strict inequality follows from Lemma A.3.2 Part 2 and the fact that $w(\gamma_1) > 0$. Therefore, $\max_{w\in\mathbb{R}_+} g_\gamma(w)$ is strictly increasing in γ on the interval $(0, \gamma^{NER})$. By (A.8), $g'_\gamma(0) > 0$ for all $\gamma > 0$. Because $g'_\gamma(\cdot)$ is continuous in w, if $\gamma > 0$, then $g_\gamma(w) > 0$ for some $w \in \mathbb{R}_+$. Hence, $\max_{w\in\mathbb{R}_+} g_\gamma(w) > 0$ for all $\gamma \in (0, \gamma^{NER})$. By Lemma A.3.3, we have $\max_{w\in\mathbb{R}_+} g_0(w) = 0$. Consequently, $\max_{w\in\mathbb{R}_+} g_\gamma(w)$ is strictly increasing in γ on the interval $[0, \gamma^{NER})$.

Finally, we will prove that $\max_{w\in\mathbb{R}_+} g_\gamma(w)$ is continuous in γ on the interval $[0, \gamma^{NER})$. Consider an arbitrary $\gamma \in [0, \gamma^{NER})$. We will consider two cases. First, let $\{\gamma_n, n \in \mathbb{N}\}$ be a sequence in $\mathbb{R}_+$ such that $\gamma_n \uparrow \gamma$ as $n \to \infty$. If $\gamma_n = 0$ for some $n \in \mathbb{N}$, we let $w(\gamma_n) := 0$ for completeness. By Lemma A.3.6 Part 2.v, we have $w(\gamma_n) < w(\gamma)$ for all $n \in \mathbb{N}$. Then,

$$\left|\max_{w\in\mathbb{R}_+} g_\gamma(w) - \max_{w\in\mathbb{R}_+} g_{\gamma_n}(w)\right| = g_\gamma(w(\gamma)) - g_{\gamma_n}(w(\gamma_n))$$

$$
\begin{aligned}
&\leq g_\gamma(w(\gamma)) - g_{\gamma_n}(w(\gamma)) \\
&\leq \sup_{w\in[0,w(\gamma)]} |g_\gamma(w) - g_{\gamma_n}(w)| \\
&\leq \frac{2}{\Sigma} w(\gamma) \exp\left\{\int_0^{w(\gamma)} L(u)\mathrm{d}u\right\} |\gamma - \gamma_n| \\
&\to 0 \quad \text{as } n \to \infty,
\end{aligned}
$$

where the last inequality follows from (A.9). Therefore, $\max_{w\in\mathbb{R}_+} g_\gamma(w)$ is left continuous in γ on the interval $[0, \gamma^{NER})$.

Second, let $\{\gamma_n, n \in \mathbb{N}\}$ be a sequence in $\mathbb{R}_+$ such that $\gamma_n \downarrow \gamma$ as $n \to \infty$. Let $\epsilon > 0$ be an arbitrary constant such that $\gamma + \epsilon < \gamma^{NER}$. There exists an $n_0 \in \mathbb{N}$ such that if $n \geq n_0$, then $\gamma_n - \gamma < \epsilon$ and thus $\gamma_n < \gamma^{NER}$. By Lemma A.3.6 Part 2.v, if $n \geq n_0$, $w(\gamma) < w(\gamma_n) < w(\gamma + \epsilon) < \infty$. If $n \geq n_0$,

$$
\begin{aligned}
\left|\max_{w\in\mathbb{R}_+} g_{\gamma_n}(w) - \max_{w\in\mathbb{R}_+} g_\gamma(w)\right| &= g_{\gamma_n}(w(\gamma_n)) - g_\gamma(w(\gamma)) \\
&\leq g_{\gamma_n}(w(\gamma_n)) - g_\gamma(w(\gamma_n)) \\
&\leq \sup_{w\in[0,w(\gamma+\epsilon)]} |g_{\gamma_n}(w) - g_\gamma(w)| \\
&\leq \frac{2}{\Sigma} w(\gamma+\epsilon) \exp\left\{\int_0^{w(\gamma+\epsilon)} L(u)\mathrm{d}u\right\} |\gamma_n - \gamma| \\
&\to 0 \quad \text{as } n \to \infty,
\end{aligned}
$$

where the last inequality follows from (A.9). Therefore, $\max_{w\in\mathbb{R}_+} g_\gamma(w)$ is right continuous in γ on the interval $[0, \gamma^{NER})$.

Consequently, $\max_{w\in\mathbb{R}_+} g_\gamma(w)$ is continuous in γ on the interval $[0, \gamma^{NER})$. □

Proof of Part 2: Recall that $\max_{w\in\mathbb{R}_+} g_0(w) = 0$ by Lemma A.3.3. By Part 1 of this lemma $\max_{w\in\mathbb{R}_+} g_\gamma(w)$ is continuous and strictly increasing in γ on the interval $[0, \gamma^{NER})$. Furthermore, we prove that $\max_{w\in\mathbb{R}_+} g_\gamma(w) = \sup_{w\in\mathbb{R}_+} g_\gamma(w) < b_j\mu_j/\lambda_j$ for all $\gamma < \gamma^{NER}$ in Part 1 of this lemma. Therefore, $\max_{w\in\mathbb{R}_+} g_\gamma(w)$ converges to a finite number as $\gamma \uparrow \gamma^{NER}$ and thus the following convergence result is enough to prove the second part of this lemma:

$$
\max_{w\in\mathbb{R}_+} g_\gamma(w) \to \frac{b_j\mu_j}{\lambda_j} \qquad \text{as } \gamma \uparrow \gamma^{NER}. \tag{A.18}
$$

First, we will prove that the unique maximum $w(\gamma)\uparrow\infty$ as $\gamma\uparrow\gamma^{NER}$. Recall from Lemma A.3.6 Part 2.v, $w(\gamma)$ is strictly increasing in γ on $(0,\gamma^{NER})$. Therefore, $\lim_{\gamma\uparrow\gamma^{NER}} w(\gamma)$ exists. Suppose that $\lim_{\gamma\uparrow\gamma^{NER}} w(\gamma)=\bar{w}<\infty$. Then, $w(\gamma)<\bar{w}$ for all $\gamma\in[0,\gamma^{NER})$. By Lemma A.3.6 Part 2.ii, $\bar{w}>\sum_{i\in\mathcal{I}} S_i/\mu_i$. By (A.2a) and Lemma A.3.6 Part 2.iv,

$$g'_\gamma(\bar{w})=\frac{2}{\Sigma}\left[\gamma-\theta g_\gamma(\bar{w})-F_0\left(g_\gamma(\bar{w})\right)\left(\bar{w}-\sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)\right]<0, \qquad \forall\gamma\in\left(0,\gamma^{NER}\right). \tag{A.19}$$

Recall that F_0 is continuous and recall from Lemma A.3.2 Part 1 that $g_\gamma(\bar{w})$ is continuous in γ. Hence, let us take the limit $\gamma\uparrow\gamma^{NER}$ in the left-hand-side (LHS) of the strict inequality in (A.19). Then, we obtain the following contradiction:

$$\begin{aligned}0\geq\lim_{\gamma\uparrow\gamma^{NER}} g'_\gamma(\bar{w})&=\frac{2}{\Sigma}\left[\gamma^{NER}-\theta g_{\gamma^{NER}}(\bar{w})-F_0\left(g_{\gamma^{NER}}(\bar{w})\right)\left(\bar{w}-\sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)\right]\\ &=g'_{\gamma^{NER}}(\bar{w})\\ &>0,\end{aligned}$$

where the first inequality and the first equality follow from (A.19), the second equality follows from (A.2a), and the strict inequality follows from Lemma A.3.8. Therefore, we must have $w(\gamma)\uparrow\infty$ as $\gamma\uparrow\gamma^{NER}$.

By (A.2a) and Lemma A.3.6 Part 2.iv,

$$\begin{aligned}0=g'_\gamma(w(\gamma))=\frac{2}{\Sigma}\left[\gamma-\theta g_\gamma(w(\gamma))-F_0\left(g_\gamma(w(\gamma))\right)\left(w(\gamma)-\sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)\right],\quad\forall\gamma\in\left(0,\gamma^{NER}\right),\\ \Longrightarrow\ \gamma-\theta g_\gamma(w(\gamma))=F_0\left(g_\gamma(w(\gamma))\right)\left(w(\gamma)-\sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right),\qquad\forall\gamma\in\left(0,\gamma^{NER}\right).\end{aligned} \tag{A.20}$$

Recall that $g_\gamma(w(\gamma))=\max_{w\in\mathbb{R}_+} g_\gamma(w)$ converges to some finite number as $\gamma\uparrow\gamma^{NER}$. Hence, let us take the limit $\gamma\uparrow\gamma^{NER}$ in both sides of the equality in (A.20). Because the LHS of (A.20) converges to a finite number, the RHS of (A.20) must also converge to a finite number as $\gamma\uparrow\gamma^{NER}$. Because $w(\gamma)\uparrow\infty$ as $\gamma\uparrow\gamma^{NER}$, the only way for the RHS of (A.20) to converge to a finite number is that $F_0\left(g_\gamma(w(\gamma))\right)\downarrow 0$ as $\gamma\uparrow\gamma^{NER}$. Because $F_0(x)>F_0(b_j\mu_j/\lambda_j)=0$ for all $x<b_j\mu_j/\lambda_j$, F_0 is strictly decreasing, and $g_\gamma(w(\gamma))$ is strictly increasing in γ, $F_0\left(g_\gamma(w(\gamma))\right)\downarrow 0$ as $\gamma\uparrow\gamma^{NER}$ if only if (A.18) holds. □

A.3.2. The main proof of Lemma 2

Part i. By Lemma 1 and (12b), the Bellman equation (12) is equivalent to the IVP (A.2) with the additional constraint $0 \le g_\gamma(w) \le b_j\mu_j/\lambda_j$ for all $w \in \mathbb{R}_+$. By Lemma A.3.4, the aforementioned additional constraint is equivalent to $0 \le g_\gamma(w) < b_j\mu_j/\lambda_j$ for all $w \in \mathbb{R}_+$. By Lemmas A.3.1 and A.3.8, there exists a unique pair (f', γ) such that $f' \in \mathbb{C}_1$, $\gamma \in \mathbb{R}_+$, and (f, γ) satisfies (12). Therefore, $f \in \mathbb{C}_2$ and is unique up to a constant. Finally, by Lemmas A.3.3 and A.3.8, f' is strictly increasing and $\gamma > 0$.

Part ii. Suppose that $b_j\mu_j/\lambda_j > c_k\mu_k$. By Lemmas A.3.3 and A.3.9 and because $c_k\mu_k > 0$, there exists a unique $\gamma \in (0, \gamma^{NER})$ such that $\max_{w\in\mathbb{R}_+} g_\gamma(w) = c_k\mu_k$. Let γ^{bar} denote the aforementioned unique γ. By Lemma A.3.6 Parts 2.iii and 2.iv, $g_{\gamma^{bar}}$ has a unique maximum, which is strictly greater than $\sum_{i\in\mathcal{I}} S_i/\mu_i$. We let

$$B := \arg\max_{w\in\mathbb{R}_+} g_{\gamma^{bar}}(w).$$

By construction, $g_{\gamma^{bar}}(B) = c_k\mu_k$. By Lemma A.3.6 Part 2.ii, $B > \sum_{i\in\mathcal{I}} S_i/\mu_i$. Let

$$f(w) := \begin{cases} C + \int_0^w g_{\gamma^{bar}}(x)\mathrm{d}x, & \text{if } w \in [0, B], \\ C + \int_0^B g_{\gamma^{bar}}(x)\mathrm{d}x + c_k\mu_k(w - B), & \text{if } w > B, \end{cases}$$

where $C \in \mathbb{R}_+$ is an arbitrary constant. Observe that

$$\lim_{w\uparrow B} f'(w) = \lim_{w\uparrow B} g_{\gamma^{bar}}(w) = c_k\mu_k = \lim_{w\downarrow B} f'(w).$$

Therefore, f' is continuous. Furthermore,

$$\lim_{w\uparrow B} f''(w) = \lim_{w\uparrow B} g'_{\gamma^{bar}}(w) = 0 = \lim_{w\downarrow B} f''(w),$$

where the second equality follows from Lemma A.3.6 Part 2.ii. Therefore, f'' is continuous and thus $f \in \mathbb{C}_2$. By construction and Lemma 1, (f, B, γ^{bar}) satisfies (13a) and (13d). By (A.2b) and Lemma A.3.6 Part 2.iii, (f, B, γ^{bar}) satisfies (13c). We have

$$\begin{aligned} \gamma^{bar} &= \frac{1}{2}\Sigma g'_{\gamma^{bar}}(B) + \theta g_{\gamma^{bar}}(B) + \min_{i\in\mathcal{I}} \{b_i\mu_i - \lambda_i g_{\gamma^{bar}}(B)\} \left(B - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right) \\ &= \theta c_k\mu_k + \min_{i\in\mathcal{I}} \{b_i\mu_i - \lambda_i c_k\mu_k\} \left(B - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}\right), \end{aligned} \tag{A.21}$$

where the first equality is by (A.2a) and the second equality follows from the fact that $g_{\gamma^{bar}}(B) = c_k\mu_k$ and $g'_{\gamma^{bar}}(B) = 0$. Next, for all $w > B$,

$$\begin{aligned}
&\frac{1}{2}\Sigma f''(w) + \theta f'(w) + \min_{\boldsymbol{q}\in\mathcal{A}(w)} \left\{ \sum_{i\in\mathcal{I}} \left(b_i - \frac{\lambda_i}{\mu_i} f'(w) \right) (q_i - S_i)^+ \right\} \\
&\qquad = \frac{1}{2}\Sigma f''(w) + \theta f'(w) + \min_{i\in\mathcal{I}} \{ b_i\mu_i - \lambda_i f'(w) \} \left(w - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i} \right)^+ \\
&\qquad = \theta c_k\mu_k + \min_{i\in\mathcal{I}} \{ b_i\mu_i - \lambda_i c_k\mu_k \} \left(w - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i} \right) \\
&\qquad > \theta c_k\mu_k + \min_{i\in\mathcal{I}} \{ b_i\mu_i - \lambda_i c_k\mu_k \} \left(B - \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i} \right) \\
&\qquad = \gamma^{bar},
\end{aligned}$$

where the first equality follows from Lemma 1 and the fact that f satisfies (13c), the strict inequality follows from the fact that $b_j\mu_j/\lambda_j > c_k\mu_k$ and $w > B$, and the last equality follows from (A.21). Therefore, (f, B, γ^{bar}) satisfies (13b).

By Lemmas A.3.1 and A.3.9 and because $f \in \mathbb{C}_2$, (B, γ^{bar}) is unique and f' is unique on $[0, B]$. Hence, f is unique up to a constant on $[0, B]$. Finally, by Lemma A.3.6 Part 2.iii, f' is strictly increasing on $[0, B]$.

A.4. Regulator Mapping Proofs

We introduce a new one-sided regulator mapping that has a simpler structure than the one in Definition 3 and a new two-sided regulator mapping that has a simpler structure than the one in Definition 5 in Section A.4.1. By using the aforementioned relatively simple regulator mappings, we prove Lemmas 3 and 4 in Sections A.4.2 and A.4.3, respectively. We also present an auxiliary limiting result associated with the one-sided regulator mapping in Definition 3 in Section A.4.4, which will help us in the later proofs.

A.4.1. Two new regulator mappings

We introduce two new regulator mappings that will help us to prove Lemmas 3 and 4.

A.4.1.1. A new one-sided regulator mapping We introduce the following one-sided regulator mapping that has a simpler structure than the one in Definition 3.

DEFINITION A.4.1. Let $(\lambda, s) \in \mathbb{R}_+^2$ and $x \in \mathbb{D}$ be such that $x(0) \geq 0$. The one-sided regulator mapping $(\phi^{\mathcal{M}}, \psi^{\mathcal{M}}) : \mathbb{R}_+^2 \times \mathbb{D} \to \mathbb{D}^2$ is defined by $(\phi^{\mathcal{M}}, \psi^{\mathcal{M}})(\lambda, s, x) = (w, \ell)$ where

C1. $w(t) = x(t) - \lambda \int_0^t (w(y) - s)^+ \,\mathrm{d}y + \ell(t) \geq 0$ for all $t \in \mathbb{R}_+$,

C2. $\ell(0) = 0$, ℓ is nondecreasing, and $\int_0^\infty w(t)\mathrm{d}\ell(t) = 0$.

The following lemma proves the existence of the regulator mapping in Definition A.4.1.

LEMMA A.4.1. *For any given $(\lambda, s) \in \mathbb{R}_+^2$ and $x \in \mathbb{D}$ such that $x(0) \geq 0$, there exists a unique pair of functions $(\phi^{\mathcal{M}}, \psi^{\mathcal{M}})(\lambda, s, x)$ satisfying the conditions C1 and C2 in Definition A.4.1. Furthermore, if (λ, s) is given, $(\phi^{\mathcal{M}}, \psi^{\mathcal{M}})(\lambda, s, x)$ is non-anticipative with respect to x.*

Before presenting the proof of Lemma A.4.1, we will present an auxiliary result. Let $\phi, \psi : \mathbb{D} \to \mathbb{D}$ be such that for all $x \in \mathbb{D}$ and $t \in \mathbb{R}_+$,

$$\psi(x)(t) := \sup_{0 \leq s \leq t} (-x(s))^+, \qquad \phi(x)(t) := x(t) + \psi(x)(t). \tag{A.22}$$

Then, (ϕ, ψ) is the conventional one-sided regulator mapping (see chapter 13.5 of Whitt (2002)). The following auxiliary lemma will help us to connect the one-sided regulator mapping in Definition A.4.1 with the conventional one-sided regulator mapping in (A.22).

LEMMA A.4.2. *Let $(\lambda, s) \in \mathbb{R}_+^2$ and $x \in \mathbb{D}$. There exists a unique $\nu \in \mathbb{D}$ which solves the equation*

$$\nu(t) = x(t) - \lambda \int_0^t (\phi(\nu)(y) - s)^+ \,\mathrm{d}y, \quad \forall t \in \mathbb{R}_+. \tag{A.23}$$

We let $\mathcal{M} : \mathbb{R}_+^2 \times \mathbb{D} \to \mathbb{D}$ denote the unique solution to (A.23), that is, $\nu = \mathcal{M}(\lambda, s, x)$. If (λ, s) is given, $\mathcal{M}(\lambda, s, x)$ is non-anticipative with respect to x.

Proof: Let us fix arbitrary $(\lambda, s) \in \mathbb{R}_+^2$ and $x \in \mathbb{D}$. Let $\eta^{\lambda,s} : \mathbb{D} \to \mathbb{D}$ be such that for all $\nu \in \mathbb{D}$ and $t \in \mathbb{R}_+$, $\eta^{\lambda,s}(\nu)(t) := \lambda (\phi(\nu)(t) - s)^+$. Let us define the uniform norm such that for all $z \in \mathbb{D}$ and $t \in \mathbb{R}_+$, $\|z\|_t := \sup_{0 \leq y \leq t} |z(y)|$. Then for all $\nu_1, \nu_2 \in \mathbb{D}$ and $t \in \mathbb{R}_+$,

$$\left\|\eta^{\lambda,s}(\nu_1) - \eta^{\lambda,s}(\nu_2)\right\|_t \leq \lambda \left\|\phi(\nu_1) - \phi(\nu_2)\right\|_t \leq 2\lambda \left\|\nu_1 - \nu_2\right\|_t,$$

where the last inequality is by the fact that the mapping ϕ is Lipschitz continuous with respect to the uniform norm with Lipschitz constant 2 (see lemma 13.5.1 of Whitt (2002)). Thus, $\eta^{\lambda,s}$ is Lipschitz continuous with respect to the uniform norm. Because (A.23) is equivalent to

$$\nu(t) = x(t) - \int_0^t \eta^{\lambda,s}(\nu)(y)\mathrm{d}y, \quad \forall t \in \mathbb{R}_+,$$

there exists a unique $\nu \in \mathbb{D}$ which solves (A.23) by lemma 1 of Reed and Ward (2004). Finally, we see in (A.23) that when (λ, s) is given, $\nu(t)$ is determined by $\{x(y), 0 \le y \le t\}$ for all $t \in \mathbb{R}_+$, implying that ν is non-anticipative with respect to x when (λ, s) is given. □

Proof of Lemma A.4.1. Let us fix arbitrary $(\lambda, s) \in \mathbb{R}_+^2$ and $x \in \mathbb{D}$ such that $x(0) \ge 0$. Let $(w, \ell) := (\phi, \psi)(\nu)$ where $\nu = \mathcal{M}(\lambda, s, x)$. By (A.22) and the fact that $\nu(0) = x(0) \ge 0$, we have $w(t) \ge 0$ for all $t \in \mathbb{R}_+$. Furthermore, by (A.23),

$$\begin{aligned} w(t) = \nu(t) + \ell(t) &= x(t) - \lambda \int_0^t (\phi(\nu)(y) - s)^+ \,\mathrm{d}y + \ell(t) \\ &= x(t) - \lambda \int_0^t (w(y) - s)^+ \,\mathrm{d}y + \ell(t), \quad \forall t \in \mathbb{R}_+. \end{aligned}$$

Therefore, condition C1 in Definition A.4.1 is satisfied by (w, ℓ). Because $\nu(0) = x(0) \ge 0$, $\ell(0) = \psi(\nu)(0) = 0$ and ℓ is nondecreasing and in $\mathbb{D}$ by the definition of the mapping ψ (see (A.22)). Finally,

$$\int_0^\infty w(t)\mathrm{d}\ell(t) = \int_0^\infty \phi(\nu)(t)\mathrm{d}\psi(\nu)(t) = 0$$

by the complementarity property of the conventional regulator mapping (see theorem 14.2.3 of Whitt (2002)). Therefore, $(w, \ell) = (\phi, \psi)(\nu)$ satisfies the conditions C1 and C2 in Definition A.4.1.

Next, we will prove the uniqueness of (w, ℓ). Let (w_1, ℓ_1) be another pair satisfying the conditions C1 and C2 and $g \in \mathbb{D}$ be such that

$$g(t) := x(t) - \lambda \int_0^t (w_1(y) - s)^+ \,\mathrm{d}y, \qquad \forall t \in \mathbb{R}_+.$$

Then, $w_1 = g + \ell_1$ by condition C1. By condition C2 and the uniqueness of the conventional regulator mapping (see theorem 14.2.2 of Whitt (2002)), $(w_1, \ell_1) = (\phi, \psi)(g)$, and thus

$$g(t) = x(t) - \lambda \int_0^t (\phi(g)(y) - s)^+ \,\mathrm{d}y, \qquad \forall t \in \mathbb{R}_+,$$

that is, $g = \mathcal{M}(\lambda, s, x)$. Because the mapping $\mathcal{M}$ is unique (see Lemma A.4.2), we have $g = \nu$. Therefore, $(w_1, \ell_1) = (\phi, \psi)(g) = (\phi, \psi)(\nu) = (w, \ell)$, which proves uniqueness. Furthermore, $(\phi^{\mathcal{M}}, \psi^{\mathcal{M}})(\lambda, s, x) = (w, \ell) = (\phi, \psi)(\nu) = (\phi, \psi)(\mathcal{M}(\lambda, s, x))$.

Finally, we will prove that $(\phi^{\mathcal{M}}, \psi^{\mathcal{M}})(\lambda, s, x)$ is non-anticipative with respect to x when (λ, s) is given. Observe that $(\phi^{\mathcal{M}}, \psi^{\mathcal{M}})(\lambda, s, x) = (\phi, \psi)(\mathcal{M}(\lambda, s, x))$ and $\mathcal{M}(\lambda, s, x)$ is non-anticipative with respect to x when (λ, s) is given by Lemma A.4.2. Because the conventional one-sided regulator mapping is non-anticipative by definition (see (A.22)), the desired result follows. □

A.4.1.2. A new two-sided regulator mapping We introduce the following two-sided regulator mapping that has a simpler structure than the one in Definition 5.

Definition A.4.2. (A two-sided regulator mapping) Let $b \in \mathbb{R}_{++}$, $(\lambda, s) \in \mathbb{R}^2_+$, and $x \in \mathbb{D}$ be such that $x(0) \in [0, b]$. The two-sided regulator mapping $(\phi^{\mathcal{N}}, \psi_1^{\mathcal{N}}, \psi_2^{\mathcal{N}}) : \mathbb{R}_{++} \times \mathbb{R}^2_+ \times \mathbb{D} \to \mathbb{D}^3$ is such that $(\phi^{\mathcal{N}}, \psi_1^{\mathcal{N}}, \psi_2^{\mathcal{N}})(b, \lambda, s, x) = (w, \ell, u)$ where

C1. $w(t) = x(t) - \lambda \int_0^t (w(y) - s)^+ \, \mathrm{d}y + \ell(t) - u(t) \in [0, b]$ for all $t \in \mathbb{R}_+$,

C2. $\ell(0) = u(0) = 0$ and both ℓ and u are nondecreasing,

C3. $\int_0^\infty w(t) \mathrm{d}\ell(t) = \int_0^\infty (b - w(t)) \mathrm{d}u(t) = 0$.

The following lemma proves the existence of the two-sided regulator mapping in Definition A.4.2.

Lemma A.4.3. *For any given $b \in \mathbb{R}_{++}$, $(\lambda, s) \in \mathbb{R}^2_+$, and $x \in \mathbb{D}$ such that $x(0) \in [0, b]$, there exists a unique triple $(\phi^{\mathcal{N}}, \psi_1^{\mathcal{N}}, \psi_2^{\mathcal{N}})(b, \lambda, s, x) \in \mathbb{D}^3$ which satisfies the conditions C1-C3 in Definition A.4.2. Furthermore, if (b, λ, s) is given, $(\phi^{\mathcal{N}}, \psi_1^{\mathcal{N}}, \psi_2^{\mathcal{N}})(b, \lambda, s, x)$ is non-anticipative with respect to x.*

Before presenting the proof of Lemma A.4.3, we will present an auxiliary result. Let $\left(\phi^{(3)}, \psi_1^{(3)}, \psi_2^{(3)}\right) : \mathbb{R}_{++} \times \mathbb{D} \to \mathbb{D}^3$ be such that for given $b \in \mathbb{R}_{++}$ and $x \in \mathbb{D}$, $\left(\phi^{(3)}, \psi_1^{(3)}, \psi_2^{(3)}\right)(b, x) = (w, \ell, u)$ where

$$l = \psi(x - u), \qquad u = \psi(b - x - l), \qquad w = x + l - u, \tag{A.24}$$

where the mapping ψ is defined in (A.22). Then, $\left(\phi^{(3)}, \psi_1^{(3)}, \psi_2^{(3)}\right)$ is the conventional two-sided regulator mapping (see chapter 2.4 of Harrison (2013) and chapter 14.8 of Whitt (2002)). The following lemma will help us to connect the two-sided regulator mapping in Definition A.4.2 with the conventional two-sided regulator mapping.

Lemma A.4.4. *Let $b \in \mathbb{R}_{++}$, $(\lambda, s) \in \mathbb{R}_{+}^{2}$, and $x \in \mathbb{D}$. There exists a unique $\nu \in \mathbb{D}$ which solves the equation*

$$\nu(t) = x(t) - \lambda \int_0^t \left(\phi^{(3)}(\nu)(y) - s\right)^+ \mathrm{d}y, \quad \forall t \in \mathbb{R}_+. \tag{A.25}$$

We let $\mathcal{N} : \mathbb{R}_{++} \times \mathbb{R}_{+}^{2} \times \mathbb{D} \to \mathbb{D}$ denote the unique solution to (A.25), that is, $\nu = \mathcal{N}(b, \lambda, s, x)$. If (b, λ, s) is given, $\mathcal{N}(b, \lambda, s, x)$ is non-anticipative with respect to x.

Proof: Because the mapping $\phi^{(3)}$ is Lipschitz continuous with respect to the uniform norm with Lipschitz constant 2 (see theorem 14.8.1 of Whitt (2002)), the proof of Lemma A.4.4 is very similar to the one of Lemma A.4.2 and thus we skip the details. □

Proof of Lemma A.4.3. Let us fix arbitrary $b \in \mathbb{R}_{++}$, $(\lambda, s) \in \mathbb{R}_{+}^{2}$, and $x \in \mathbb{D}$ such that $x(0) \in [0, b]$. Let $(w, \ell, u) := \left(\phi^{(3)}, \psi_1^{(3)}, \psi_2^{(3)}\right)(\nu)$ where $\nu = \mathcal{N}(b, \lambda, s, x)$. By (A.24) and the fact that $\nu(0) = x(0) \in [0, b]$, we have $w(t) \in [0, b]$ for all $t \in \mathbb{R}_+$. Furthermore, by (A.24) and (A.25),

$$\begin{aligned} w(t) = \nu(t) + \ell(t) - u(t) &= x(t) - \lambda \int_0^t \left(\phi^{(3)}(\nu)(y) - s\right)^+ \mathrm{d}y + \ell(t) - u(t) \\ &= x(t) - \lambda \int_0^t (w(y) - s)^+ \mathrm{d}y + \ell(t) - u(t), \quad \forall t \in \mathbb{R}_+. \end{aligned}$$

Therefore, condition C1 in Definition A.4.2 is satisfied by (w, ℓ, u). Because $\nu(0) = x(0) \in [0, b]$, $\ell(0) = u(0) = 0$, both ℓ and u are nondecreasing and in $\mathbb{D}$ by (A.24). Finally,

$$\begin{aligned} \int_0^\infty w(t) \mathrm{d}\ell(t) &= \int_0^\infty \phi^{(3)}(\nu)(t) \mathrm{d}\psi_1^{(3)}(\nu)(t) = 0, \\ \int_0^\infty (b - w(t)) \mathrm{d}u(t) &= \int_0^\infty \left(b - \phi^{(3)}(\nu)(t)\right) \mathrm{d}\psi_2^{(3)}(\nu)(t) = 0, \end{aligned}$$

by the complementarity property of the conventional two-sided regulator mapping (see section 14.8.1 of Whitt (2002)). Therefore, $(w, \ell, u) = \left(\phi^{(3)}, \psi_1^{(3)}, \psi_2^{(3)}\right)(\nu)$ satisfies the conditions C1-C3 in Definition A.4.2.

Next, we will prove the uniqueness of (w,ℓ,u). Let (w_1,ℓ_1,u_1) be another triple satisfying the conditions C1-C3 and $g\in\mathbb{D}$ be such that

$$g(t):=x(t)-\lambda\int_0^t (w_1(y)-s)^+\,\mathrm{d}y,\qquad \forall t\in\mathbb{R}_+.$$

Then, $w_1=g+\ell_1-u_1$ by condition C1. By conditions C2-C3 and the uniqueness of the conventional two-sided regulator mapping (see theorem 14.8.1 of Whitt (2002)), $(w_1,\ell_1,u_1)=\left(\phi^{(3)},\psi_1^{(3)},\psi_2^{(3)}\right)(g)$, and thus

$$g(t)=x(t)-\lambda\int_0^t \left(\phi^{(3)}(g)(y)-s\right)^+\mathrm{d}y,\qquad \forall t\in\mathbb{R}_+,$$

that is, $g=\mathcal{N}(b,\lambda,s,x)$. Because the mapping $\mathcal{N}$ is unique (see Lemma A.4.4), we have $g=\nu$. Therefore, $(w_1,\ell_1,u_1)=\left(\phi^{(3)},\psi_1^{(3)},\psi_2^{(3)}\right)(g)=\left(\phi^{(3)},\psi_1^{(3)},\psi_2^{(3)}\right)(\nu)=(w,\ell,u)$, which proves uniqueness. Furthermore,

$$(\phi^{\mathcal{N}},\psi_1^{\mathcal{N}},\psi_2^{\mathcal{N}})(b,\lambda,s,x)=(w,\ell,u)=\left(\phi^{(3)},\psi_1^{(3)},\psi_2^{(3)}\right)(\nu)=\left(\phi^{(3)},\psi_1^{(3)},\psi_2^{(3)}\right)(\mathcal{N}(b,\lambda,s,x)).$$

Finally, we will prove that $(\phi^{\mathcal{N}},\psi_1^{\mathcal{N}},\psi_2^{\mathcal{N}})(b,\lambda,s,x)$ is non-anticipative with respect to x when (b,λ,s) is given. Observe that

$$\left(\phi^{\mathcal{N}},\psi_1^{\mathcal{N}},\psi_2^{\mathcal{N}}\right)(b,\lambda,s,x)=\left(\phi^{(3)},\psi_1^{(3)},\psi_2^{(3)}\right)(\mathcal{N}(b,\lambda,s,x))$$

and $\mathcal{N}(b,\lambda,s,x)$ is non-anticipative with respect to x when (b,λ,s) is given by Lemma A.4.4. Because the conventional two-sided regulator mapping is non-anticipative by definition (see (A.24)), the desired result follows. □

A.4.2. Proof of Lemma 3

For given $\boldsymbol{\lambda}\in\mathbb{R}_+^I$, $\boldsymbol{\mu}\in\mathbb{R}_{++}^I$, and $\boldsymbol{s}\in\mathbb{R}_+^I$, let $\beta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}:\mathbb{R}_+\to\mathbb{R}_+$ be such that

$$\beta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}(w):=\sum_{i\in\mathcal{I}}\frac{\lambda_i}{\mu_i}\left(q_i^*(w)-s_i\right)^+,\qquad \forall w\in\mathbb{R}_+.$$

By recalling Lemma 1, for all $w\in\mathbb{R}_+$, let $\kappa(w)\in\arg\min_{i\in\mathcal{I}}\{b_i\mu_i-\lambda_i f'(w)\}$. By Lemma 1 and Definition 2,

$$\beta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}(w)=\lambda_{\kappa(w)}\left(w-\sum_{i\in\mathcal{I}}\frac{s_i}{\mu_i}\right)^+,\qquad \forall w\in\mathbb{R}_+. \tag{A.26}$$

Recall that the function f' is nonnegative and strictly increasing (see (12), (13), and Lemma 2). By (A.3) and (A.4) (and recall Figure A.3.1), the function $\beta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}$ is nondecreasing and piecewise linear with finitely many break points. Furthermore, we can choose $\kappa(w)$ such that $\beta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}$ becomes right continuous. Let $\eta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}:\mathbb{D}\to\mathbb{D}$ be such that

$$\eta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}(w)(t):=\beta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}(w(t)), \qquad \forall w\in\mathbb{D},\, t\in\mathbb{R}_+,$$

that is, $\eta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}$ is the process version of $\beta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}$. In order to satisfy condition C1 in Definition 3, we need

$$w(t)=x(t)-\int_0^t \eta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}(w)(y)\mathrm{d}y+\ell(t), \qquad \forall t\in\mathbb{R}_+.$$

Due to the potential discontinuity of $\beta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}$, the mapping $\eta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}$ is not necessarily Lipschitz continuous with respect to the uniform norm and thus the proof of Lemma A.4.1 cannot be used to prove Lemma 3. Specifically, we cannot extend Lemma A.4.2 with the mapping $\eta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}$. Therefore, we will use a different proof technique to prove Lemma 3.

Recall that $\beta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}$ is nondecreasing and piecewise linear with finitely many break points. Therefore, $\mathbb{R}_+$ can be partitioned into finite number of intervals such that whenever the workload level $w(t)$ enters an interval, $\beta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}$ is a linear function in that interval. Specifically, let $\mathbb{R}_+$ be partitioned into $\xi\in\{1,2,\ldots,I,I+1\}$ workload intervals with break points $(w^*_{i_1},w^*_{i_2},\ldots,w^*_{i_{\xi-1}})\in\mathbb{R}_+^{\xi-1}$ such that

$$0=:w^*_{i_0}\le\sum_{i\in\mathcal{I}}\frac{s_i}{\mu_i}=:w^*_{i_1}<w^*_{i_2}<\ldots<w^*_{i_{\xi-1}}<\infty=:w^*_{i_\xi},$$

$$0=:\lambda_{i_0}<\lambda_{i_1}<\lambda_{i_2}<\ldots<\lambda_{i_{\xi-1}},$$

where $\lambda_{i_m}=\lambda_{\kappa(w)}$ in the $(m+1)$th interval for all $m\in\{0,1,\ldots,\xi-1\}$. Figure A.4.1 illustrates the function $\beta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}$ with three break points.

By (A.26), on each workload interval $[w^*_{i_m},w^*_{i_{m+1}}]$ where $m\in\{0,1,2,\ldots,\xi-1\}$, the one-sided regulator mapping defined in Definition 3 behaves like the one-sided regulator mapping defined in Definition A.4.1 with the initial condition $x(0)\in[w^*_{i_m},w^*_{i_{m+1}}]$ and the parameters $\left(\lambda_{i_m},\sum_{i\in\mathcal{I}}s_i/\mu_i\right)$.

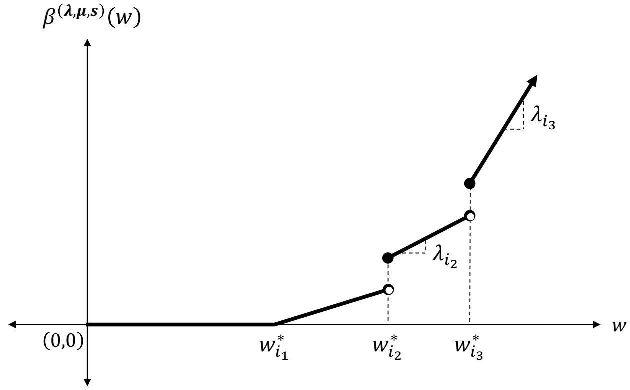


Figure A.4.1 (Color online) An illustration of the function $\beta^{(\boldsymbol{\lambda},\boldsymbol{\mu},\boldsymbol{s})}$ with three break points.

By Lemma A.4.1, there exists a unique pair in $\mathbb{D}^2$ satisfying the conditions C1 and C2 in Definition A.4.1 on the workload interval $[w^*_{i_m}, w^*_{i_{m+1}}]$ for all $m \in \{0, 1, \ldots, \xi - 1\}$. Therefore, there exists a unique pair in $\mathbb{D}^2$ satisfying the conditions C1 and C2 in Definition 3. Finally, by Lemma A.4.1, if $\left(\lambda_{i_m}, \sum_{i\in\mathcal{I}} s_i/\mu_i\right)$ is given, the unique pair defined in Definition A.4.1 is non-anticipative with respect to x on the workload interval $[w^*_{i_m}, w^*_{i_{m+1}}]$ for all $m \in \{0, 1, \ldots, \xi - 1\}$. Consequently, if $(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{s})$ is given, the unique pair defined in Definition 3 is non-anticipative with respect to x.

A.4.3. Proof of Lemma 4

Because the proof of Lemma 4 is very similar to the one of Lemma 3, we skip it. The only difference is that we refer to Lemma A.4.3 instead of Lemma A.4.1 to show the existence of the two-sided regulator mapping with the desired properties at each workload interval.

A.4.4. An auxiliary result associated with the one-sided regulator mapping in Definition 3

We will derive a limiting result for the workload process under the one-sided regulator mapping in Definition 3. In order to accomplish that result, first, we will derive an upper bound on that workload process.

LEMMA A.4.5. *Let us fix arbitrary* $\boldsymbol{\lambda} \in \mathbb{R}^I_+$, $\boldsymbol{\mu} \in \mathbb{R}^I_{++}$, $\boldsymbol{s} \in \mathbb{R}^I_+$, *and* $x \in \mathbb{D}$ *such that* $x(0) \geq 0$. *Let* (w, ℓ) *be the one-sided regulator mapping in Definition 3 and* $(w^*, \ell^*) := (\phi, \psi)(x)$ *be the conventional one-sided regulator mapping defined in* (A.22). *Then,* $w(t) \leq w^*(t)$ *for all* $t \in \mathbb{R}_+$.

Proof: For notational convenience, let us define

$$\beta(t) := \sum_{i\in\mathcal{I}} \frac{\lambda_i}{\mu_i} \left(q_i^*(w(t)) - s_i\right)^+, \qquad \forall t \in \mathbb{R}_+.$$

Then, by condition C1 in Definition 3, we have

$$w(t) = x(t) - \int_0^t \beta(u)\mathrm{d}u + \ell(t), \qquad \forall t \in \mathbb{R}_+. \tag{A.27}$$

Suppose that the statement of Lemma A.4.5 is not correct. Let

$$\tau := \inf\{t \in \mathbb{R}_+ : w(t) > w^*(t)\}.$$

Then, it must be the case that τ exists and $\tau < \infty$. There are four cases to consider.

Case 1 Suppose that both w and w^* are continuous at τ. Then, it must be the case that $w(\tau) = w^*(\tau)$ and $w(\tau+) > w^*(\tau+)$, which implies that there exists $\epsilon > 0$ such that

$$w(\tau + t) > w^*(\tau + t), \qquad \forall t \in (0, \epsilon). \tag{A.28}$$

By (A.28) and condition C2 in Definition 3 , we have

$$\ell(\tau + t) = \ell(\tau), \qquad \forall t \in (0, \epsilon). \tag{A.29}$$

Let us fix an arbitrary $t \in (0, \epsilon)$. Then,

$$\begin{aligned} w(\tau+t) - w^*(\tau+t) &= x(\tau+t) - \int_0^{\tau+t} \beta(u)\mathrm{d}u + \ell(\tau+t) - x(\tau+t) - \ell^*(\tau+t) \\ &= -\int_0^{\tau} \beta(u)\mathrm{d}u + \ell(\tau) - \int_\tau^{\tau+t} \beta(u)\mathrm{d}u + \ell(\tau+t) - \ell(\tau) - \ell^*(\tau+t) \\ &= w(\tau) - x(\tau) - \int_\tau^{\tau+t} \beta(u)\mathrm{d}u - \ell^*(\tau+t) \\ &= w^*(\tau) - x(\tau) - \int_\tau^{\tau+t} \beta(u)\mathrm{d}u - \ell^*(\tau+t) \\ &= \ell^*(\tau) - \ell^*(\tau+t) - \int_\tau^{\tau+t} \beta(u)\mathrm{d}u \\ &\le 0, \end{aligned} \tag{A.30}$$

where the first equality follows from (A.22) and (A.27), the third equality follows from (A.27) and (A.29), the fourth equality follows from the fact that $w(\tau) = w^*(\tau)$, the fifth equality follows from

(A.22), and the inequality follows from the fact that ℓ^* is nondecreasing and β is a nonnegative function. Observe that the inequality in (A.30) is a contradiction because of (A.28).

Case 2 Suppose that w has a jump at τ and w^* is continuous at τ. Then, it must be the case that

$$w(\tau-) \leq w^*(\tau-), \qquad w(\tau) > w^*(\tau), \qquad \Delta w(\tau) > 0, \tag{A.31}$$

where $\Delta a(t) := a(t) - a(t-)$ for all $a \in \mathbb{D}$ and $t \in \mathbb{R}_+$. By condition C2 in Definition 3 , $\Delta\ell(\tau) = 0$, and because ℓ^* is nondecresing, $\Delta\ell^*(\tau) \geq 0$. By condition C1 in Definition 3

$$\Delta w(\tau) = \Delta x(\tau), \tag{A.32}$$

$$\Delta w^*(\tau) = \Delta x(\tau) + \Delta\ell^*(\tau),$$

$$\implies \quad \Delta w(\tau) \leq \Delta w^*(\tau),$$

$$\implies \quad w(\tau) - w^*(\tau) \leq w(\tau-) - w^*(\tau-). \tag{A.33}$$

Because $w(\tau-) \leq w^*(\tau-)$ by (A.31), (A.33) implies $w(\tau) \leq w^*(\tau)$, which is a contradiction by (A.31).

Case 3 Suppose that w is continuous at τ and w^* has a jump at τ. Then, it must be the case that

$$w(\tau-) \leq w^*(\tau-), \qquad w(\tau) > w^*(\tau), \qquad \Delta w^*(\tau) < 0. \tag{A.34}$$

By (A.34), $w(\tau) > 0$ and thus $\Delta\ell(\tau) = 0$ by condition C2 in Definition 3 and the continuity of w at τ. The rest of the proof is exactly the same as the one of Case 2 starting from (A.32).

Case 4 Suppose that both w and w^* have a jump at τ. Then, it must be the case that

$$w(\tau-) \leq w^*(\tau-), \qquad w(\tau) > w^*(\tau),$$

which implies $\Delta\ell(\tau) = 0$ by condition C2 in Definition 3 . The rest of the proof is exactly the same as the one of Case 2 starting from (A.32). □

The following limiting result associated with the one-sided regulator mapping in Definition 3 will be useful in the optimality proofs of the NER policy.

LEMMA A.4.6. *Let us fix arbitrary* $\boldsymbol{\lambda} \in \mathbb{R}_+^I$, $\boldsymbol{\mu} \in \mathbb{R}_{++}^I$, $\boldsymbol{S} \in \mathbb{R}_+^I$, *and* $\theta \in (-\infty, 0]$. *Let* $W := \phi^{(1)}(\boldsymbol{\lambda}, \boldsymbol{\mu}, \boldsymbol{S}, X + \theta e)$, *where* X *is a* $BM(0, \Sigma)$. *By Definition 4,* W *is the workload process under the one-sided regulator mapping in Definition 3. Then,*

$$\lim_{t\to\infty} \frac{\mathbf{E}[W(t)]}{t} = 0.$$

Proof. By condition C1 in Definition 3 and Lemma A.4.5, we have $\mathbf{0} \le W \le W^*$ where $W^* := \phi(X + \theta e)$ is the conventional one-sided regulator mapping defined in (A.22). We have

$$\mathbf{P}(W^*(t) \le w) = \Phi\left(\frac{w - \theta t}{\sqrt{\Sigma t}}\right) - \mathrm{e}^{\frac{2\theta}{\Sigma} w} \Phi\left(\frac{-w - \theta t}{\sqrt{\Sigma t}}\right), \qquad \forall t, w \in \mathbb{R}_+, \tag{A.35}$$

where Φ denotes the cumulative distribution function of a standard normal random variable (see section 1.10 of Harrison (2013) for details). By (A.35), if $\theta = 0$,

$$\mathbf{E}[W^*(t)] = \int_0^\infty \mathbf{P}(W^*(t) > w)\,\mathrm{d}w = \int_0^\infty \left(1 - \mathrm{Erf}\left[\frac{w}{\sqrt{2\Sigma t}}\right]\right) \mathrm{d}w = \sqrt{\frac{2\Sigma t}{\pi}}, \tag{A.36}$$

where the last equality follows from Korotkov and Korotkov (2020) (see page 138 therein). By (A.35), if $\theta \neq 0$,

$$\begin{aligned} \mathbf{E}[W^*(t)] &= \int_0^\infty \mathbf{P}(W^*(t) > w)\,\mathrm{d}w \\ &= \frac{1}{2}\int_0^\infty \left(1 - \mathrm{Erf}\left[\frac{w - \theta t}{\sqrt{2\Sigma t}}\right]\right) \mathrm{d}w + \frac{1}{2}\int_0^\infty \mathrm{e}^{\frac{2\theta}{\Sigma} w} \left(1 - \mathrm{Erf}\left[\frac{w + \theta t}{\sqrt{2\Sigma t}}\right]\right) \mathrm{d}w \\ &= \sqrt{\frac{\Sigma t}{2\pi}} \mathrm{e}^{-\frac{\theta^2}{2\Sigma} t} + \frac{\theta t}{2}\left(1 - \mathrm{Erf}\left[-\theta\sqrt{\frac{t}{2\Sigma}}\right]\right) + \frac{\Sigma}{2\theta}\mathrm{Erf}\left[\theta\sqrt{\frac{t}{2\Sigma}}\right], \end{aligned} \tag{A.37}$$

where the last equality follows from Korotkov and Korotkov (2020) (see pages 138 and 154 therein).

Therefore, if $\theta \le 0$, by (A.36) and (A.37), we have

$$\lim_{t\to\infty} \frac{\mathbf{E}[W^*(t)]}{t} = 0,$$

which gives the desired result. □

A.5. Proof of Theorem 1

Let us fix an arbitrary $\boldsymbol{S} \in \mathbb{R}_+^I$ and consider the function f defined in Lemma 2 Part i. Because f is unique up to a constant, let us choose $f(0)=0$. By (12a),

$$\frac{1}{2}\Sigma f''(w)+\theta f'(w)+\sum_{i\in\mathcal{I}}\left(b_i-\frac{\lambda_i}{\mu_i}f'(w)\right)(q_i-S_i)^+\geq\gamma,\qquad\forall w\in\mathbb{R}_+,\ \boldsymbol{q}\in\mathcal{A}(w).\tag{A.38}$$

By (10b) and Ito's lemma, we have

$$\begin{aligned}
&\mathbf{E}\left[f(W(t))\right]-\mathbf{E}\left[f(W(0))\right]\\
&=\frac{1}{2}\Sigma\,\mathbf{E}\left[\int_0^t f''(W(s))\mathrm{d}s\right]+\mathbf{E}\left[\int_0^t\left(\theta-\sum_{i\in\mathcal{I}}\frac{\lambda_i}{\mu_i}\left(a_i(s,W(s))-S_i\right)^+\right)f'(W(s))\mathrm{d}s\right]\\
&\qquad\qquad+\mathbf{E}\left[\int_0^t f'(W(s))\left(\mathrm{d}I(s)-\mathrm{d}E(s)\right)\right]\\
&\geq\gamma t-\mathbf{E}\left[\sum_{i\in\mathcal{I}}b_i\int_0^t\left(a_i(s,W(s))-S_i\right)^+\mathrm{d}s\right]+\mathbf{E}\left[\int_0^t f'(W(s))\left(\mathrm{d}I(s)-\mathrm{d}E(s)\right)\right],
\end{aligned}\tag{A.39}$$

where the inequality in (A.39) is by (A.38). Observe that the inequality in (A.39) becomes an equality under the NER policy by (12a) and Definitions 2 and 4.

The NER policy. We will prove that γ defined in Lemma 2 Part i is the long-run average backorder cost under the NER policy. Because $E=\mathbf{0}$ under the NER policy,

$$\mathbf{E}\left[\int_0^t f'\big(W(s)\big)\mathrm{d}E(s)\right]=0.\tag{A.40}$$

By (12c) and because the repair facility works in a work-conserving fashion under the NER policy (see condition C2 in Definition 3), we have

$$\mathbf{E}\left[\int_0^t f'\big(W(s)\big)\mathrm{d}I(s)\right]=\mathbf{E}\left[\int_0^t f'\big(0\big)\mathrm{d}I(s)\right]=0.\tag{A.41}$$

Next, recall that $W(0)=0$, $W(t)\geq 0$ for all $t\in\mathbb{R}_+$ (see (10e)), and $f(0)=0$. Furthermore, by (12b) and because $f(0)=0$, $f(w)\geq 0$ for all $w\in\mathbb{R}_+$. Therefore,

$$0\leq\frac{1}{t}\left(\mathbf{E}\left[f(W(t))\right]-\mathbf{E}\left[f(W(0))\right]\right)=\frac{1}{t}\mathbf{E}\left[f(W(t))\right]\leq\frac{b_j\mu_j}{\lambda_j t}\mathbf{E}\left[W(t)\right]\to 0\quad\text{as } t\to\infty,\tag{A.42}$$

where the last inequality follows from (12b) and the convergence result follows from Lemma A.4.6. Therefore, by (A.42),

$$\lim_{t\to\infty}\frac{1}{t}\left(\mathbf{E}\left[f(W(t))\right]-\mathbf{E}\left[f(W(0))\right]\right)=0. \tag{A.43}$$

By (A.40), (A.41), and (A.43), and because (A.39) holds with equality and $E=\mathbf{0}$ under the NER policy, we have

$$\lim_{t\to\infty}\frac{1}{t}\mathbf{E}\left[\sum_{i\in\mathcal{I}}b_i\int_0^t\left(a_i(s,W(s))-S_i\right)^+\mathrm{d}s+c_k\mu_kE(t)\right]=\gamma. \tag{A.44}$$

An arbitrary feasible solution to the EWF (10). We will prove that γ defined in Lemma 2 Part i is a lower-bound on the long-run average backorder and emergency repair cost under any feasible solution to the EWF (10). Let $\left(W,I,E,\boldsymbol{a},\boldsymbol{S}\right)$ be an arbitrary feasible solution to the EWF (10). We will consider two cases. First, suppose that

$$\liminf_{t\to\infty}\frac{\mathbf{E}\left[W(t)\right]}{t}=\epsilon \tag{A.45}$$

for some arbitrary $\epsilon>0$. Let $\underline{b\mu}:=\min_{i\in\mathcal{I}}b_i\mu_i$. Then,

$$\begin{aligned}
\liminf_{t\to\infty}\frac{1}{t}\mathbf{E}\left[\sum_{i\in\mathcal{I}}b_i\left(a_i(t,W(t))-S_i\right)^+\right]&\geq\liminf_{t\to\infty}\frac{1}{t}\mathbf{E}\left[\underline{b\mu}\sum_{i\in\mathcal{I}}\left(\frac{a_i(t,W(t))}{\mu_i}-\frac{S_i}{\mu_i}\right)^+\right]\\
&\geq\liminf_{t\to\infty}\frac{1}{t}\mathbf{E}\left[\underline{b\mu}\left(\sum_{i\in\mathcal{I}}\frac{a_i(t,W(t))}{\mu_i}-\sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)^+\right]\\
&=\liminf_{t\to\infty}\frac{1}{t}\mathbf{E}\left[\underline{b\mu}\left(W(t)-\sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right)^+\right]\\
&\geq\underline{b\mu}\,\liminf_{t\to\infty}\frac{1}{t}\mathbf{E}\left[W(t)-\sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\right]\\
&=\underline{b\mu}\,\liminf_{t\to\infty}\frac{1}{t}\mathbf{E}\left[W(t)\right]-\underline{b\mu}\,\lim_{t\to\infty}\frac{1}{t}\sum_{i\in\mathcal{I}}\frac{S_i}{\mu_i}\\
&=\underline{b\mu}\epsilon,
\end{aligned}$$

where the first equality follows from (10d) and the last equality follows from (A.45). Therefore, there exists a $t_0\in\mathbb{R}_+$ such that

$$\frac{1}{t}\mathbf{E}\left[\sum_{i\in\mathcal{I}}b_i\left(a_i(t,W(t))-S_i\right)^+\right]>0.5\underline{b\mu}\epsilon,\qquad\forall t\geq t_0. \tag{A.46}$$

Next, let us consider the long-run average backorder cost.

$$\begin{aligned}\liminf_{t\to\infty}\frac{1}{t}\mathbf{E}\left[\sum_{i\in\mathcal{I}} b_i\int_0^t (a_i(s,W(s))-S_i)^+\,\mathrm{d}s\right] &= \liminf_{t\to\infty}\frac{1}{t}\int_0^t \mathbf{E}\left[\sum_{i\in\mathcal{I}} b_i\,(a_i(s,W(s))-S_i)^+\right]\mathrm{d}s\\ &\geq \liminf_{t\to\infty}\frac{1}{t}\int_{t_0}^t \mathbf{E}\left[\sum_{i\in\mathcal{I}} b_i\,(a_i(s,W(s))-S_i)^+\right]\mathrm{d}s\\ &\geq \liminf_{t\to\infty}\frac{1}{t}\int_{t_0}^t \left(0.5b\underline{\mu}\epsilon s\right)\mathrm{d}s\\ &= \liminf_{t\to\infty}\frac{0.5b\underline{\mu}\epsilon}{t}\left(\frac{t^2-t_0^2}{2}\right)\\ &= \infty, \end{aligned}\tag{A.47}$$

where the first equality follows from Tonelli's theorem and the second inequality is by (A.46). Therefore, if (A.45) holds under a feasible solution to the EWF (10), the associated objective function value is infinity.

Second, suppose that

$$\liminf_{t\to\infty}\frac{\mathbf{E}\big[W(t)\big]}{t}=0. \tag{A.48}$$

Similar to how we derive (A.43), by (A.48), we can prove that

$$\liminf_{t\to\infty}\frac{1}{t}\left(\mathbf{E}\left[f\big(W(t)\big)\right]-\mathbf{E}\left[f\big(W(0)\big)\right]\right)=0. \tag{A.49}$$

Let $\Delta E(t) := E(t)-E(t-)$ denote the jump of E at time t for all $t\in\mathbb{R}_+$. By (10c), we let

$$E^c(t) := E(t)-\sum_{0<s\leq t}\Delta E(s), \qquad \forall t\in\mathbb{R}_+,$$

where the sum is over the countable set $s\in(0,t]$ at which $\Delta E(s)>0$ and E^c denotes the continuous part of E (see section 4.9 of Harrison (2013) for details). Then,

$$\begin{aligned}\mathbf{E}\left[\int_0^t f'\big(W(s)\big)\mathrm{d}\left(-E(s)\right)\right] &= -\mathbf{E}\left[\int_0^t f'\big(W(s)\big)\mathrm{d}E^c(s)\right]+\mathbf{E}\left[\sum_{0<s\leq t}\left(f\big(W(s)\big)-f\big(W(s-)\big)\right)\right]\\ &\geq -c_k\mu_k\mathbf{E}\left[E^c(t)\right]-\mathbf{E}\left[c_k\mu_k\sum_{0<s\leq t}\Delta E(s)\right]\\ &= -c_k\mu_k\mathbf{E}\left[E(t)\right], \end{aligned}\tag{A.50}$$

where the inequality follows from (10c), (12b), and the fact that $b_j\mu_j/\lambda_j \le c_k\mu_k$. Similarly, by (10c) and (12b), we can show that

$$\mathbf{E}\left[\int_0^t f'\big(W(s)\big)\mathrm{d}I(s)\right] \ge 0. \tag{A.51}$$

By (A.39), (A.50), and (A.51) we have

$$\mathbf{E}\left[f(W(t))\right] - \mathbf{E}\left[f(W(0))\right] \ge \gamma t - \mathbf{E}\left[\sum_{i\in\mathcal{I}} b_i \int_0^t \left(a_i(s,W(s)) - S_i\right)^+ \mathrm{d}s\right] - c_k\mu_k \mathbf{E}\left[E(t)\right]. \tag{A.52}$$

By (A.49) and (A.52), we have

$$\limsup_{t\to\infty} \frac{1}{t}\mathbf{E}\left[\sum_{i\in\mathcal{I}} b_i \int_0^t \left(a_i(s,W(s)) - S_i\right)^+ \mathrm{d}s + c_k\mu_k E(t)\right] \ge \gamma. \tag{A.53}$$

Consequently, for all fixed $\boldsymbol{S} \in \mathbb{R}_+^I$ in the EWF (10), the NER policy is optimal by (A.44) and (A.53).

A.6. Proof of Theorem 2

Let us fix an arbitrary $\boldsymbol{S} \in \mathbb{R}_+^I$ and consider the function f defined in Lemma 2 Part ii. Because f is unique up to a constant, let us choose $f(0) = 0$. By (13a) and (13b), the inequality in (A.38) holds under the function f defined in Lemma 2 Part ii. By (10b) and Ito's lemma, we have

$$\begin{aligned}
&\mathbf{E}\left[f(W(t))\right] - \mathbf{E}\left[f(W(0))\right] \\
&= \frac{1}{2}\Sigma\,\mathbf{E}\left[\int_0^t f''(W(s))\mathrm{d}s\right] + \mathbf{E}\left[\int_0^t \left(\theta - \sum_{i\in\mathcal{I}} \frac{\lambda_i}{\mu_i}\left(a_i(s,W(s)) - S_i\right)^+\right) f'(W(s))\mathrm{d}s\right] \\
&\qquad + \mathbf{E}\left[\int_0^t f'(W(s))\left(\mathrm{d}I(s) - \mathrm{d}E(s)\right)\right] \\
&\ge \gamma t - \mathbf{E}\left[\sum_{i\in\mathcal{I}} b_i \int_0^t \left(a_i(s,W(s)) - S_i\right)^+ \mathrm{d}s\right] + \mathbf{E}\left[\int_0^t f'(W(s))\left(\mathrm{d}I(s) - \mathrm{d}E(s)\right)\right],
\end{aligned} \tag{A.54}$$

where the inequality in (A.54) is by (A.38). Observe that the inequality in (A.54) becomes an equality under the barrier policy by (13a) and Definitions 2 and 6.

The barrier policy. We will prove that γ defined in Lemma 2 Part ii is the long-run average backorder and emergency repair cost under the barrier policy. By (13d) and because the barrier policy uses emergency repairs only when the workload level hits the upper barrier level B (see condition C3 in Definition 5), we have

$$\mathbf{E}\left[\int_0^t f'\big(W(s)\big)\mathrm{d}E(s)\right] = \mathbf{E}\left[\int_0^t f'(B)\mathrm{d}E(s)\right] = c_k\mu_k\mathbf{E}\left[E(t)\right]. \tag{A.55}$$

By (13d) and because the repair facility works in a work-conserving fashion under the barrier policy (see condition C3 in Definition 5), we have

$$\mathbf{E}\left[\int_0^t f'\big(W(s)\big)\mathrm{d}I(s)\right] = \mathbf{E}\left[\int_0^t f'\big(0\big)\mathrm{d}I(s)\right] = 0. \tag{A.56}$$

Next,

$$\frac{1}{t}\left(\mathbf{E}\left[f(W(t))\right] - \mathbf{E}\left[f(W(0))\right]\right) = \frac{1}{t}\mathbf{E}\left[f(W(t))\right] \to 0 \qquad \text{as } t\to\infty, \tag{A.57}$$

where the equality follows from the fact that $W(0) = 0$ and $f(0) = 0$ and the convergence result follows from the fact that $W(t) \in [0,B]$ for all $t\in\mathbb{R}_+$ under the barrier policy and $\sup_{0\le w\le B}|f(w)| < \infty$ because f is continuous as stated in Lemma 2 Part ii.

Therefore, by (A.55), (A.56), (A.57), and because (A.54) holds with equality under the barrier policy, we have

$$\lim_{t\to\infty}\frac{1}{t}\mathbf{E}\left[\sum_{i\in\mathcal{I}} b_i\int_0^t \left(a_i(s,W(s)) - S_i\right)^+ \mathrm{d}s + c_k\mu_k E(t)\right] = \gamma. \tag{A.58}$$

An arbitrary feasible solution to the EWF (10)*.* We will prove that γ defined in Lemma 2 Part ii is a lower-bound on the long-run average backorder and emergency repair cost under any feasible solution to the EWF (10). Let $\big(W,I,E,\boldsymbol{a},\boldsymbol{S}\big)$ be an arbitrary feasible solution to the EWF (10). We will consider two cases. First, if

$$\liminf_{t\to\infty}\frac{\mathbf{E}\big[W(t)\big]}{t} = \epsilon$$

for some arbitrary $\epsilon > 0$, then the associated objective function value is infinity by the same argument leading to (A.47). Second, suppose that

$$\liminf_{t\to\infty}\frac{\mathbf{E}\big[W(t)\big]}{t} = 0. \tag{A.59}$$

Recall that $W(0)=0$, $W(t)\geq 0$ for all $t\in\mathbb{R}_+$ (see (10e)), and $f(0)=0$. Furthermore, by (13c) and because $f(0)=0$, $f(w)\geq 0$ for all $w\in\mathbb{R}_+$. Therefore,

$$0\leq \liminf_{t\to\infty}\frac{1}{t}\left(\mathbf{E}\left[f\big(W(t)\big)\right]-\mathbf{E}\left[f\big(W(0)\big)\right]\right)=\liminf_{t\to\infty}\frac{1}{t}\mathbf{E}\left[f\big(W(t)\big)\right]\leq \liminf_{t\to\infty}\frac{c_k\mu_k}{t}\mathbf{E}\big[W(t)\big]=0,$$

where the last inequality follows from (13c) and the last equality follows from (A.59). Therefore,

$$\liminf_{t\to\infty}\frac{1}{t}\left(\mathbf{E}\left[f\big(W(t)\big)\right]-\mathbf{E}\left[f\big(W(0)\big)\right]\right)=0. \tag{A.60}$$

Similar to how we derive (A.50) and (A.51), by (13c), we can derive that

$$\mathbf{E}\left[\int_0^t f'\big(W(s)\big)\mathrm{d}E(s)\right]\leq c_k\mu_k\mathbf{E}\left[E(t)\right],\qquad \mathbf{E}\left[\int_0^t f'\big(W(s)\big)\mathrm{d}I(s)\right]\geq 0. \tag{A.61}$$

By (A.54) and (A.61), we have

$$\mathbf{E}\left[f(W(t))\right]-\mathbf{E}\left[f(W(0))\right]\geq \gamma t-\mathbf{E}\left[\sum_{i\in\mathcal{I}}b_i\int_0^t\left(a_i(s,W(s))-S_i\right)^+\mathrm{d}s\right]-c_k\mu_k\mathbf{E}\left[E(t)\right]. \tag{A.62}$$

By (A.60) and (A.62), we have

$$\limsup_{t\to\infty}\frac{1}{t}\mathbf{E}\left[\sum_{i\in\mathcal{I}}b_i\int_0^t\left(a_i(s,W(s))-S_i\right)^+\mathrm{d}s+c_k\mu_kE(t)\right]\geq\gamma. \tag{A.63}$$

Consequently, for any fixed $\boldsymbol{S}\in\mathbb{R}_+^I$ in the EWF (10), the barrier policy is optimal by (A.58) and (A.63).

A.7. Proof of Proposition 2

(i) Let $\big(W,I,E,S\big)$ be a feasible solution to the EWF (16). Recall (11) and that $k\in\arg\min_{i\in\mathcal{I}}c_i\mu_i$. For $(x,y)\in\mathbb{D}^2$, let $(x\wedge y)(t):=\min\{x(t),y(t)\}$ for all $t\in\mathbb{R}_+$. There are five cases to consider.

Case i.1. Suppose that $j\neq k\neq l$. Let

$$\begin{aligned}
&Q_l:=\mu_l\left(W\wedge S\right), && Q_j:=\mu_j\left(W-S\right)^+, && Q_i:=\mathbf{0}\qquad \forall i\in\mathcal{I}\backslash\{j,l\},\\
&S_l:=\mu_lS, && S_i:=0\qquad \forall i\in\mathcal{I}\backslash\{l\}, &&\\
&E_k:=\mu_kE, && E_i:=\mathbf{0}\qquad \forall i\in\mathcal{I}\backslash\{k\}, &&\\
&Y_k:=E-\frac{X_k}{\mu_k}, && Y_l:=W\wedge S-\frac{X_l}{\mu_l}, && Y_i:=-\frac{X_i}{\mu_i}\qquad \forall i\in\mathcal{I}\backslash\{j,k,l\},
\end{aligned}$$

$$Y_j(t) := (W(t)-S)^+ - \frac{X_j(t)}{\mu_j} + \lambda_j \int_0^t (W(s)-S)^+ \,\mathrm{d}s, \qquad \forall t \in \mathbb{R}_+.$$

Case i.2. Suppose that $j = k \neq l$. Let

$$Q_l := \mu_l (W \wedge S), \qquad Q_j := \mu_j (W - S)^+, \qquad Q_i := \mathbf{0} \qquad \forall i \in \mathcal{I} \backslash \{j, l\},$$

$$S_l := \mu_l S, \qquad S_i := 0 \qquad \forall i \in \mathcal{I} \backslash \{l\},$$

$$E_j := \mu_j E, \qquad E_i := \mathbf{0} \qquad \forall i \in \mathcal{I} \backslash \{j\},$$

$$Y_l := W \wedge S - \frac{X_l}{\mu_l}, \qquad Y_i := -\frac{X_i}{\mu_i} \qquad \forall i \in \mathcal{I} \backslash \{j, l\},$$

$$Y_j(t) := (W(t)-S)^+ - \frac{X_j(t)}{\mu_j} + \lambda_j \int_0^t (W(s)-S)^+ \,\mathrm{d}s + E(t), \qquad \forall t \in \mathbb{R}_+.$$

Case i.3. Suppose that $j \neq k = l$. Let

$$Q_k := \mu_k (W \wedge S), \qquad Q_j := \mu_j (W - S)^+, \qquad Q_i := \mathbf{0} \qquad \forall i \in \mathcal{I} \backslash \{j, k\},$$

$$S_k := \mu_k S, \qquad S_i := 0 \qquad \forall i \in \mathcal{I} \backslash \{k\},$$

$$E_k := \mu_k E, \qquad E_i := \mathbf{0} \qquad \forall i \in \mathcal{I} \backslash \{k\},$$

$$Y_k := W \wedge S - \frac{X_k}{\mu_k} + E, \qquad Y_i := -\frac{X_i}{\mu_i} \qquad \forall i \in \mathcal{I} \backslash \{j, k\},$$

$$Y_j(t) := (W(t)-S)^+ - \frac{X_j(t)}{\mu_j} + \lambda_j \int_0^t (W(s)-S)^+ \,\mathrm{d}s, \qquad \forall t \in \mathbb{R}_+.$$

Case i.4. Suppose that $j = l \neq k$. Let

$$Q_j := \mu_j W, \qquad Q_i := \mathbf{0} \qquad \forall i \in \mathcal{I} \backslash \{j\},$$

$$S_j := \mu_j S, \qquad S_i := 0 \qquad \forall i \in \mathcal{I} \backslash \{j\},$$

$$E_k := \mu_k E, \qquad E_i := \mathbf{0} \qquad \forall i \in \mathcal{I} \backslash \{k\},$$

$$Y_k := E - \frac{X_k}{\mu_k}, \qquad Y_i := -\frac{X_i}{\mu_i} \qquad \forall i \in \mathcal{I} \backslash \{j, k\},$$

$$Y_j(t) := W(t) - \frac{X_j(t)}{\mu_j} + \lambda_j \int_0^t (W(s)-S)^+ \,\mathrm{d}s, \qquad \forall t \in \mathbb{R}_+.$$

Case i.5. Suppose that $j = k = l$. Let

$$
\begin{aligned}
&Q_j := \mu_j W, && Q_i := \mathbf{0} \quad \forall i \in \mathcal{I}\backslash\{j\},\\
&S_j := \mu_j S, && S_i := 0 \quad \forall i \in \mathcal{I}\backslash\{j\},\\
&E_j := \mu_j E, && E_i := \mathbf{0} \quad \forall i \in \mathcal{I}\backslash\{j\},\\
&Y_i := -\frac{X_i}{\mu_i} \quad \forall i \in \mathcal{I}\backslash\{j\}, && Y_j(t) := W(t) - \frac{X_j(t)}{\mu_j} + \lambda_j \int_0^t (W(s) - S)^+ \,\mathrm{d}s + E(t), \qquad \forall t \in \mathbb{R}_+.
\end{aligned}
$$

Observe that $\big(Q_i, S_i, Y_i, I, E_i, i \in \mathcal{I}\big)$ defined above is a feasible solution to the BCP (9) with the objective function value equal to the one of $\big(W, I, E, S\big)$.

(ii) Let $\big(Q_i, S_i, Y_i, I, E_i, i \in \mathcal{I}\big)$ be a feasible solution to the BCP (9) and let

$$S := \sum_{i\in\mathcal{I}} \frac{S_i}{\mu_i}, \qquad E := \sum_{i\in\mathcal{I}} \frac{E_i}{\mu_i}, \qquad \tilde{X} := X + \theta e - E.$$

By scaling and then summing (9b) over i, we obtain

$$\sum_{i\in\mathcal{I}} \frac{Q_i(t)}{\mu_i} = \tilde{X}(t) - \sum_{i\in\mathcal{I}} \lambda_i \int_0^t \left(\frac{Q_i(s)}{\mu_i} - \frac{S_i}{\mu_i}\right)^+ \mathrm{d}s + I(t), \qquad \forall t \in \mathbb{R}_+. \tag{A.64}$$

Let $(W^{\lambda_j}, I^{\lambda_j}) \in \mathbb{D}^2$ denote the unique process pair that satisfies the conditions C1 and C2 in Definition A.4.1 under $(\lambda_j, S, \tilde{X})$. Then,

$$W^{\lambda_j}(t) = \tilde{X}(t) - \lambda_j \int_0^t \big(W^{\lambda_j}(s) - S\big)^+ \mathrm{d}s + I^{\lambda_j}(t), \qquad \forall t \in \mathbb{R}_+. \tag{A.65}$$

By Definition A.4.1, $\big(W^{\lambda_j}, S, I^{\lambda_j}, E\big)$ is a feasible solution to the EWF (16). We will show that the objective function value of $\big(W^{\lambda_j}, S, I^{\lambda_j}, E\big)$ is less than or equal to the one of $\big(Q_i, S_i, Y_i, I, E_i, i \in \mathcal{I}\big)$ by the following result.

Lemma A.7.1. *Fix an arbitrary sample path. Under Assumption 2, for all* $t \in \mathbb{R}_+$,

$$\int_0^t \big(W^{\lambda_j}(s) - S\big)^+ \mathrm{d}s \le \sum_{i\in\mathcal{I}} \int_0^t \left(\frac{Q_i(s)}{\mu_i} - \frac{S_i}{\mu_i}\right)^+ \mathrm{d}s.$$

Proof: We will use the proof by contradiction technique. Let

$$\tau_3 := \inf\left\{ t \in \mathbb{R}_+ : \int_0^t \big(W^{\lambda_j}(s) - S\big)^+ \mathrm{d}s > \sum_{i\in\mathcal{I}} \int_0^t \left(\frac{Q_i(s)}{\mu_i} - \frac{S_i}{\mu_i}\right)^+ \mathrm{d}s \right\}$$

and suppose that $\tau_3 < \infty$. Then

$$\int_0^t \left(W^{\lambda_j}(s) - S\right)^+ \mathrm{d}s \le \sum_{i\in\mathcal{I}} \int_0^t \left(\frac{Q_i(s)}{\mu_i} - \frac{S_i}{\mu_i}\right)^+ \mathrm{d}s, \qquad \forall t \in [0, \tau_3), \tag{A.66a}$$

$$\int_0^{\tau_3} \left(W^{\lambda_j}(s) - S\right)^+ \mathrm{d}s = \sum_{i\in\mathcal{I}} \int_0^{\tau_3} \left(\frac{Q_i(s)}{\mu_i} - \frac{S_i}{\mu_i}\right)^+ \mathrm{d}s, \tag{A.66b}$$

$$\int_0^{\tau_3+} \left(W^{\lambda_j}(s) - S\right)^+ \mathrm{d}s > \sum_{i\in\mathcal{I}} \int_0^{\tau_3+} \left(\frac{Q_i(s)}{\mu_i} - \frac{S_i}{\mu_i}\right)^+ \mathrm{d}s. \tag{A.66c}$$

By (A.66b) and (A.66c), we have

$$\begin{aligned} &\left(W^{\lambda_j}(\tau_3) - S\right)^+ > \sum_{i\in\mathcal{I}} \left(\frac{Q_i(\tau_3)}{\mu_i} - \frac{S_i}{\mu_i}\right)^+ \ge \left(\sum_{i\in\mathcal{I}} \frac{Q_i(\tau_3)}{\mu_i} - S\right)^+, \\ \implies \quad & W^{\lambda_j}(\tau_3) > \sum_{i\in\mathcal{I}} \frac{Q_i(\tau_3)}{\mu_i}. \end{aligned} \tag{A.67}$$

By (A.64), (A.65), (A.66b), (A.67), Assumption 2, and the definition of τ_3, we have

$$I^{\lambda_j}(\tau_3) > I(\tau_3). \tag{A.68}$$

Let

$$\tau_1 := \sup\left\{t \in [0, \tau_3] : dI^{\lambda_j}(t) > 0\right\},$$

that is, τ_1 is the last time I^{λ_j} increases before τ_3. Observe that τ_1 is well defined (that is, τ_1 exists) by (A.68). By definition of τ_1, we have $I^{\lambda_j}(\tau_1) = I^{\lambda_j}(\tau_3)$. By condition C2 in Definition A.4.1 and because $\tilde{X}$ does not have any upward jumps, we have $W^{\lambda_j}(\tau_1) = 0$. By (A.66a) and (A.66b) and because $\tau_1 \le \tau_3$, we have

$$\int_0^{\tau_1} \left(W^{\lambda_j}(s) - S\right)^+ \mathrm{d}s \le \sum_{i\in\mathcal{I}} \int_0^{\tau_1} \left(\frac{Q_i(s)}{\mu_i} - \frac{S_i}{\mu_i}\right)^+ \mathrm{d}s. \tag{A.69}$$

Let

$$\tau_2 := \inf\left\{t \in [\tau_1, \tau_3] : \int_{\tau_1}^t \left(W^{\lambda_j}(s) - S\right)^+ \mathrm{d}s > \sum_{i\in\mathcal{I}} \int_{\tau_1}^t \left(\frac{Q_i(s)}{\mu_i} - \frac{S_i}{\mu_i}\right)^+ \mathrm{d}s\right\}.$$

Observe that τ_2 is well defined by (A.66c) and (A.69). Furthermore,

$$\begin{aligned} &\int_{\tau_1}^t \left(W^{\lambda_j}(s) - S\right)^+ \mathrm{d}s \le \sum_{i\in\mathcal{I}} \int_{\tau_1}^t \left(\frac{Q_i(s)}{\mu_i} - \frac{S_i}{\mu_i}\right)^+ \mathrm{d}s, \qquad \forall t \in [\tau_1, \tau_2), \\ &\int_{\tau_1}^{\tau_2} \left(W^{\lambda_j}(s) - S\right)^+ \mathrm{d}s = \sum_{i\in\mathcal{I}} \int_{\tau_1}^{\tau_2} \left(\frac{Q_i(s)}{\mu_i} - \frac{S_i}{\mu_i}\right)^+ \mathrm{d}s, \\ &\int_{\tau_1}^{\tau_2+} \left(W^{\lambda_j}(s) - S\right)^+ \mathrm{d}s > \sum_{i\in\mathcal{I}} \int_{\tau_1}^{\tau_2+} \left(\frac{Q_i(s)}{\mu_i} - \frac{S_i}{\mu_i}\right)^+ \mathrm{d}s, \end{aligned} \tag{A.70}$$

which implies

$$\left(W^{\lambda_j}(\tau_2)-S\right)^+ > \sum_{i\in\mathcal{I}}\left(\frac{Q_i(\tau_2)}{\mu_i}-\frac{S_i}{\mu_i}\right)^+ \geq \left(\sum_{i\in\mathcal{I}}\frac{Q_i(\tau_2)}{\mu_i}-S\right)^+,$$

$$\implies \quad W^{\lambda_j}(\tau_2) > \sum_{i\in\mathcal{I}}\frac{Q_i(\tau_2)}{\mu_i}. \tag{A.71}$$

Because $W^{\lambda_j}(\tau_1)=0$ and $I^{\lambda_j}(\tau_1)=I^{\lambda_j}(\tau_2)=I^{\lambda_j}(\tau_3)>I(\tau_3)\geq I(\tau_2)\geq I(\tau_1)$ (recall (A.68)), and by (A.64) and (A.65), we have

$$\sum_{i\in\mathcal{I}}\frac{Q_i(\tau_2)}{\mu_i}-\sum_{i\in\mathcal{I}}\frac{Q_i(\tau_1)}{\mu_i}=\tilde{X}(\tau_2)-\tilde{X}(\tau_1)-\sum_{i\in\mathcal{I}}\lambda_i\int_{\tau_1}^{\tau_2}\left(\frac{Q_i(s)}{\mu_i}-\frac{S_i}{\mu_i}\right)^+\mathrm{d}s+I(\tau_2)-I(\tau_1), \tag{A.72}$$

$$W^{\lambda_j}(\tau_2)=\tilde{X}(\tau_2)-\tilde{X}(\tau_1)-\lambda_j\int_{\tau_1}^{\tau_2}\left(W^{\lambda_j}(s)-S\right)^+\mathrm{d}s. \tag{A.73}$$

By (A.70), (A.72), (A.73), and Assumption 2, we have

$$W^{\lambda_j}(\tau_2)\leq\sum_{i\in\mathcal{I}}\frac{Q_i(\tau_2)}{\mu_i},$$

which is a contradiction by (A.71). □

Next we will prove that the objective function value of $\left(W^{\lambda_j},S,I^{\lambda_j},E\right)$ is less than or equal to the one of $\left(Q_i,S_i,Y_i,I,E_i,i\in\mathcal{I}\right)$ under all sample paths. Let us compare the objective functions (9a) and (16a) at a fixed time point t and under an arbitrary sample path:

$$\begin{aligned}
&\sum_{i\in\mathcal{I}}\left(h_iS_i+\frac{1}{t}\left(b_i\int_0^t\left(Q_i(s)-S_i\right)^+\mathrm{d}s+c_iE_i(t)\right)\right)\\
&\qquad=\sum_{i\in\mathcal{I}}\left(h_i\mu_i\frac{S_i}{\mu_i}+\frac{1}{t}\left(b_i\mu_i\int_0^t\left(\frac{Q_i(s)}{\mu_i}-\frac{S_i}{\mu_i}\right)^+\mathrm{d}s+c_i\mu_i\frac{E_i(t)}{\mu_i}\right)\right)\\
&\qquad\geq\sum_{i\in\mathcal{I}}\left(h_l\mu_l\frac{S_i}{\mu_i}+\frac{1}{t}\left(b_j\mu_j\int_0^t\left(\frac{Q_i(s)}{\mu_i}-\frac{S_i}{\mu_i}\right)^+\mathrm{d}s+c_k\mu_k\frac{E_i(t)}{\mu_i}\right)\right)\\
&\qquad=h_l\mu_lS+\frac{1}{t}\left(b_j\mu_j\sum_{i\in\mathcal{I}}\int_0^t\left(\frac{Q_i(s)}{\mu_i}-\frac{S_i}{\mu_i}\right)^+\mathrm{d}s+c_k\mu_kE(t)\right)\\
&\qquad\geq h_l\mu_lS+\frac{1}{t}\left(b_j\mu_j\int_0^t\left(W^{\lambda_j}(s)-S\right)^+\mathrm{d}s+c_k\mu_kE(t)\right),
\end{aligned}$$

where the first inequality follows from (11), Assumption 2, and the fact that $k\in\arg\min_{i\in\mathcal{I}}c_i\mu_i$, and the second the inequality follows from Lemma A.7.1. Therefore, the objective function value associated with $\left(Q_i,S_i,Y_i,I,E_i,i\in\mathcal{I}\right)$ is greater than or equal to the one associated with $\left(W^{\lambda_j},S,I^{\lambda_j},E\right)$ under all sample paths.

A.8. Proof of Theorem 3

For given $\boldsymbol{S} \in \mathbb{R}_+^I$, let $S := \sum_{i\in\mathcal{I}} S_i/\mu_i$. Under Assumption 2, by Lemma 1, the equality in (12a) becomes equivalent to

$$\frac{1}{2}\Sigma f''(w) + \left(\theta - \lambda_j (w-S)^+\right) f'(w) + b_j\mu_j(w-S)^+ = \gamma, \qquad \forall w \in \mathbb{R}_+. \tag{A.74}$$

By (A.74) and (12c), let us consider the following IVP with the initial condition $g(0)=0$ and

$$\frac{1}{2}\Sigma g'(w) + \left(\theta - \lambda_j (w-S)^+\right) g(w) + b_j\mu_j(w-S)^+ = \gamma, \qquad \forall w \in \mathbb{R}_+. \tag{A.75}$$

Observe that (A.75) is a first-order linear ODE and its unique solution is as follows.

$$\begin{aligned}
g(w) &= \mathrm{e}^{-\int_0^w g_1(x)\mathrm{d}x} \int_0^w \mathrm{e}^{\int_0^x g_1(s)\mathrm{d}s} g_0(x)\mathrm{d}x, \qquad \forall w \in \mathbb{R}_+,\\
g_0(w) &:= \frac{2}{\Sigma}\left(\gamma - b_j\mu_j(w-S)^+\right), \qquad \forall w \in \mathbb{R}_+,\\
g_1(w) &:= \frac{2}{\Sigma}\left(\theta - \lambda_j (w-S)^+\right), \qquad \forall w \in \mathbb{R}_+.
\end{aligned}$$

By algebra, if $\theta = 0$,

$$g(w) = \begin{cases} \frac{2\gamma}{\Sigma} w, & \text{if } w \in [0,S],\\ \frac{b_j\mu_j}{\lambda_j} + \mathrm{e}^{\frac{\lambda_j}{\Sigma}(w-S)^2}\left(\frac{2\gamma}{\Sigma}S - \frac{b_j\mu_j}{\lambda_j} + \gamma\sqrt{\frac{\pi}{\lambda_j\Sigma}}\mathrm{Erf}\left[\sqrt{\frac{\lambda_j}{\Sigma}}(w-S)\right]\right), & \text{if } w > S, \end{cases} \tag{A.76}$$

where $\mathrm{Erf}[\cdot]$ is the Gauss error function, that is,

$$\mathrm{Erf}[x] = \frac{2}{\sqrt{\pi}}\int_0^x \mathrm{e}^{-s^2}\mathrm{d}s, \qquad \forall x \in \mathbb{R}.$$

If $\theta \neq 0$,

$$g(w) = \begin{cases} \frac{\gamma}{\theta}\left(1 - \mathrm{e}^{-\frac{2\theta}{\Sigma}w}\right), & \text{if } w \in [0,S],\\ \frac{b_j\mu_j}{\lambda_j} + \mathrm{e}^{\frac{1}{\Sigma}\left(\lambda_j(w-S)^2 - 2\theta(w-S)\right)} & \\ \quad \times\left(\gamma\left(\frac{1}{\theta}\left(1-\mathrm{e}^{-\frac{2\theta}{\Sigma}S}\right) + \sqrt{\frac{\pi}{\lambda_j\Sigma}}\mathrm{e}^{\frac{\theta^2}{\lambda_j\Sigma}}\left(\mathrm{Erf}\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right] + \mathrm{Erf}\left[\frac{-\theta+\lambda_j(w-S)}{\sqrt{\lambda_j\Sigma}}\right]\right)\right)\right. & \\ \quad\quad \left. - \frac{b_j\mu_j}{\lambda_j}\left(1 + \frac{\theta}{\sqrt{\lambda_j\Sigma}}\sqrt{\pi}\mathrm{e}^{\frac{\theta^2}{\lambda_j\Sigma}}\left(\mathrm{Erf}\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right] + \mathrm{Erf}\left[\frac{-\theta+\lambda_j(w-S)}{\sqrt{\lambda_j\Sigma}}\right]\right)\right)\right), & \text{if } w > S. \end{cases} \tag{A.77}$$

Next, we will derive the closed-form expression for the pair (f, γ) defined in Lemma 2 Part i in Section A.8.1. Then, we will optimize the inventory level S in Section A.8.2 and complete the proof of Theorem 3.

The following auxiliary result will be helpful in the upcoming proofs. Let

$$\Upsilon(x) := \sqrt{\pi}\mathrm{e}^{x^2}\left(1 + \mathrm{Erf}[x]\right), \quad \forall x \in \mathbb{R}. \tag{A.78}$$

Lemma A.8.1. *We have $\Upsilon(x) > 0$ and $1 + x\Upsilon(x) > 0$ for all $x \in \mathbb{R}$.*

Proof. Because $\mathrm{Erf}[x] \in (-1, 1)$ for all $x \in \mathbb{R}$, we have $\Upsilon(x) > 0$ for all $x \in \mathbb{R}$. Clearly $x\Upsilon(x) \geq 0$ for all $x \in \mathbb{R}_+$. Let

$$\tilde{\Upsilon}(x) := \sqrt{\pi}\mathrm{e}^{x^2}\left(1 - \mathrm{Erf}[x]\right), \qquad \forall x \in \mathbb{R}.$$

Because $\mathrm{Erf}[x] = -\mathrm{Erf}[-x]$ for all $x \in \mathbb{R}$, we have $\Upsilon(x) = \tilde{\Upsilon}(-x)$ for all $x \in \mathbb{R}$. Therefore, if we prove $1 - x\tilde{\Upsilon}(x) > 0$ for all $x \in \mathbb{R}_+$, then it will imply $1 + x\Upsilon(x) > 0$ for all $x \in (-\infty, 0)$, which will complete the proof.

By L'Hôpital's rule, $\lim_{x\to\infty} x\tilde{\Upsilon}(x) = 1$. Because $\tilde{\Upsilon}(x) > 0$ for all $x \in \mathbb{R}_+$, if $1 - x\tilde{\Upsilon}(x) \leq 0$ for some $x \in \mathbb{R}_+$, then

$$\frac{\mathrm{d}(1 - x\tilde{\Upsilon}(x))}{\mathrm{d}x} = 2x\left(1 - x\tilde{\Upsilon}(x)\right) - \tilde{\Upsilon}(x) < 0,$$

which in turn implies $1 - x\tilde{\Upsilon}(x)$ will be strictly negative and will keep decreasing forever. Because $\lim_{x\to\infty}\left(1 - x\tilde{\Upsilon}(x)\right) = 0$, it must be the case that $1 - x\tilde{\Upsilon}(x) > 0$ for all $x \in \mathbb{R}_+$. □

A.8.1. Solution to the Bellman equation (12)

Let

$$f(w) := \int_0^w g(x)\mathrm{d}x, \qquad \forall w \in \mathbb{R}_+, \tag{A.79}$$

where g is defined in (A.76) and (A.77) depending on θ. We will derive $\gamma \in \mathbb{R}_+$ such that (f, γ) solves the Bellman equation (12). Observe that f satisfies (A.74) (and thus (12a)) and (12c) by definition and the fact that g is the solution to the ODE (A.75) with the initial condition $g(0) = 0$.

There are two cases to consider.

Case 1: $\theta = 0$. Let

$$\gamma := \frac{\frac{b_j \mu_j}{\lambda_j}}{\frac{2}{\Sigma} S + \sqrt{\frac{\pi}{\lambda_j \Sigma}}}. \tag{A.80}$$

Observe that γ defined in (A.80) is strictly positive and nonincreasing and convex in S.

We have $f'(w) = g(w)$ for all $w \in \mathbb{R}_+$ by (A.79). By (A.76) and (A.80), for all $w \in [0, S]$,

$$f'(w) = g(w) = \frac{2}{\Sigma} \gamma w \leq \frac{2}{\Sigma} \gamma S = \frac{2}{\Sigma} S \frac{\frac{b_j \mu_j}{\lambda_j}}{\frac{2}{\Sigma} S + \sqrt{\frac{\pi}{\lambda_j \Sigma}}} < \frac{b_j \mu_j}{\lambda_j}. \tag{A.81}$$

By (A.76), (A.80), and the fact that $\mathrm{Erf}[w] < 1$ for all $w \in \mathbb{R}_+$, for all $w > S$,

$$f'(w) = g(w) < \frac{b_j \mu_j}{\lambda_j} + \mathrm{e}^{\frac{\lambda_j}{\Sigma}(w-S)^2} \left(\left(\frac{2}{\Sigma} S + \sqrt{\frac{\pi}{\lambda_j \Sigma}} \right) \gamma - \frac{b_j \mu_j}{\lambda_j} \right) = \frac{b_j \mu_j}{\lambda_j}. \tag{A.82}$$

By (A.76), (A.79), (A.80), and some algebra,

$$f''(w) = \frac{2\lambda_j}{\Sigma}(w - S) \mathrm{e}^{\frac{\lambda_j}{\Sigma}(w-S)^2} \gamma \sqrt{\frac{\pi}{\lambda_j \Sigma}} \left(\mathrm{Erf}\left[\sqrt{\frac{\lambda_j}{\Sigma}} (w - S) \right] - 1 \right) + \frac{2\gamma}{\Sigma}, \qquad \forall w > S.$$

By L'Hôpital's rule, $\lim_{w \to \infty} f''(w) = 0$. Furthermore,

$$f'''(w) = \frac{2\lambda_j}{\Sigma}(w - s) f''(w) + \left[\frac{2\lambda_j}{\Sigma} \mathrm{e}^{\frac{\lambda_j}{\Sigma}(w-S)^2} \gamma \sqrt{\frac{\pi}{\lambda_j \Sigma}} \left(\mathrm{Erf}\left[\sqrt{\frac{\lambda_j}{\Sigma}} (w - S) \right] - 1 \right) \right]. \tag{A.83}$$

Because $\mathrm{Erf}[w] < 1$ for all $w \in \mathbb{R}_+$, the term in the square brackets in (A.83) is strictly negative. Hence, if $f''(w) < 0$ for some $w > S$, then $f'''(w) < 0$ for that w. Therefore, if f'' becomes strictly negative at some point on the interval (S, ∞), it keeps decreasing henceforth. Because $\lim_{w \to \infty} f''(w) = 0$, it must be the case that $f''(w) \geq 0$ for all $w > S$, which in turn implies $f''(w) \geq 0$ for all $w \in \mathbb{R}_+$ because $f''(w) = 2\gamma/\Sigma$ for all $w \in [0, S]$ by (A.76). Finally, by (12c), (A.81), (A.82), and the fact that $f''(w) \geq 0$ for all $w \in \mathbb{R}_+$, the pair (f, γ) defined in (A.79) and (A.80) satisfies (12b).

Case 2: $\theta < 0$. Let

$$\gamma := \frac{\frac{b_j \mu_j}{\lambda_j} \left(1 + \frac{\theta}{\sqrt{\lambda_j \Sigma}} \Upsilon \left[\frac{\theta}{\sqrt{\lambda_j \Sigma}} \right] \right)}{\frac{1}{\theta} \left(1 - \mathrm{e}^{-\frac{2\theta}{\Sigma} S} \right) + \frac{1}{\sqrt{\lambda_j \Sigma}} \Upsilon \left[\frac{\theta}{\sqrt{\lambda_j \Sigma}} \right]}, \tag{A.84}$$

where Υ is defined in (A.78). By Lemma A.8.1, γ defined in (A.84) is strictly positive for all $\theta < 0$. By Lemma A.8.1 and algebra, one can see that γ is nonincreasing and convex in S.

We have $f'(w) = g(w)$ for all $w \in \mathbb{R}_+$ by (A.79). Observe that $g(0) = 0$ and $g(w)$ is nondecreasing in w for all $w \in [0, S]$ by (A.77). Therefore, for all $w \in [0, S]$,

$$\begin{aligned}
0 \le f'(w) = g(w) &= \frac{\gamma}{\theta}\left(1 - \mathrm{e}^{-\frac{2\theta}{\Sigma}w}\right) \le \frac{\gamma}{\theta}\left(1 - \mathrm{e}^{-\frac{2\theta}{\Sigma}S}\right) \\
&= \frac{1}{\theta}\left(1 - \mathrm{e}^{-\frac{2\theta}{\Sigma}S}\right) \frac{\frac{b_j\mu_j}{\lambda_j}\left(1 + \frac{\theta}{\sqrt{\lambda_j\Sigma}}\Upsilon\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right]\right)}{\frac{1}{\theta}\left(1 - \mathrm{e}^{-\frac{2\theta}{\Sigma}S}\right) + \frac{1}{\sqrt{\lambda_j\Sigma}}\Upsilon\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right]} \\
&= \frac{b_j\mu_j}{\lambda_j}\left(1 - \mathrm{e}^{-\frac{2\theta}{\Sigma}S} \frac{\frac{1}{\sqrt{\lambda_j\Sigma}}\Upsilon\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right]}{\frac{1}{\theta}\left(1 - \mathrm{e}^{-\frac{2\theta}{\Sigma}S}\right) + \frac{1}{\sqrt{\lambda_j\Sigma}}\Upsilon\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right]}\right) \\
&< \frac{b_j\mu_j}{\lambda_j},
\end{aligned} \tag{A.85}$$

where the last inequality follows from Lemma A.8.1.

Second, for all $w > S$,

$$\begin{aligned}
f'(w) &= g(w) \\
&< \frac{b_j\mu_j}{\lambda_j} + \mathrm{e}^{\frac{\lambda_j(w-S)^2 - 2\theta(w-S)}{\Sigma}}\Bigg(\gamma\left(\frac{1}{\theta}\left(1 - \mathrm{e}^{-\frac{2\theta}{\Sigma}S}\right) + \sqrt{\frac{\pi}{\lambda_j\Sigma}}\mathrm{e}^{\frac{\theta^2}{\lambda_j\Sigma}}\left(1 + \mathrm{Erf}\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right]\right)\right) \\
&\qquad - \frac{b_j\mu_j}{\lambda_j}\left(1 + \frac{\theta}{\sqrt{\lambda_j\Sigma}}\sqrt{\pi}\mathrm{e}^{\frac{\theta^2}{\lambda_j\Sigma}}\left(1 + \mathrm{Erf}\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right]\right)\right)\Bigg) \\
&= \frac{b_j\mu_j}{\lambda_j} + \mathrm{e}^{\frac{\lambda_j(w-S)^2 - 2\theta(w-S)}{\Sigma}}\Bigg(\gamma\left(\frac{1}{\theta}\left(1 - \mathrm{e}^{-\frac{2\theta}{\Sigma}S}\right) + \frac{1}{\sqrt{\lambda_j\Sigma}}\Upsilon\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right]\right) \\
&\qquad - \frac{b_j\mu_j}{\lambda_j}\left(1 + \frac{\theta}{\sqrt{\lambda_j\Sigma}}\Upsilon\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right]\right)\Bigg) \\
&= \frac{b_j\mu_j}{\lambda_j},
\end{aligned} \tag{A.86}$$

where the first equality follows from (A.79), the first inequality follows from (A.77) and the fact that $\theta < 0$ and $\mathrm{Erf}[x] \in (-1, 1)$ for all $x \in \mathbb{R}$, the second equality follows from (A.78), and the third equality follows from (A.84).

Finally, if $w > S$, by (A.77), (A.79), and (A.84),

$$\begin{aligned}
f''(w) &= \frac{2}{\Sigma}\left(\left(\lambda_j(w - S) - \theta\right) f'(w) + \gamma - b_j\mu_j(w - S)\right), \\
f'''(w) &= \left(\frac{1}{w - S - \frac{\theta}{\lambda_j}} + \frac{2}{\Sigma}\left(\lambda_j(w - S) - \theta\right)\right) f''(w) - \frac{2}{\Sigma}\left(\frac{\gamma - \theta\frac{b_j\mu_j}{\lambda_j}}{w - S - \frac{\theta}{\lambda_j}}\right).
\end{aligned} \tag{A.87}$$

By algebra, one can see that $\lim_{w\to\infty} f''(w) = 0$. Because $w > S$ and $\theta < 0$, the coefficient of f'' in (A.87) is strictly positive and the second term in the right-hand-side (RHS) of (A.87) is strictly negative. Therefore, if $f''(w) < 0$ for some $w > S$, then $f'''(w) < 0$ for that w. Therefore, if f'' becomes strictly negative at some point on the interval (S, ∞), it will keep decreasing henceforth. Because $\lim_{w\to\infty} f''(w) = 0$, it must be the case that $f''(w) \geq 0$ for all $w > S$, which in turn implies $f'(w) \geq 0$ for all $w > S$. This result together with (A.85) and (A.86) imply that the pair (f, γ) defined in (A.79) and (A.84) satisfies (12b).

A.8.2. Inventory optimization

For given $S \in \mathbb{R}_+$, the optimal long-run average backorder and emergency repair cost, $\gamma(S)$, is given by (A.80) and (A.84). Therefore, the following optimization problem gives us the optimal inventory level for the EWF (16):

$$\min_{S\in\mathbb{R}_+} \{h_l \mu_l S + \gamma(S)\}. \tag{A.88}$$

Because $\gamma(S)$ is convex in S, by the first order conditions, one can see that the S^* and z^* defined in Theorem 3 are the optimal solution and the optimal objective function value of the optimization problem (A.88), respectively.

A.9. Proof of Theorem 4

For given $\boldsymbol{S} \in \mathbb{R}_+^I$, recall that $S = \sum_{i\in\mathcal{I}} S_i/\mu_i$. Under Assumption 2 and by Lemma 1, (13a) and (13b) are equivalent to

$$\frac{1}{2}\Sigma f''(w) + \left(\theta - \lambda_j (w-S)^+\right) f'(w) + b_j \mu_j (w-S)^+ = \gamma, \qquad \forall w \in [0, B], \tag{A.89a}$$

$$\frac{1}{2}\Sigma f''(w) + \left(\theta - \lambda_j (w-S)^+\right) f'(w) + b_j \mu_j (w-S)^+ \geq \gamma, \qquad \forall w > B. \tag{A.89b}$$

Next, we will derive some properties of the triple (f, B, γ) defined in Lemma 2 Part ii in Section A.9.1. Then, we will optimize the inventory level S in Section A.9.2 and complete the proof of Theorem 4.

A.9.1. Solution to the Bellman equation (13)

For given $B \in \mathbb{R}_{++}$, let

$$f(w) := \begin{cases} \int_0^w g(x)\mathrm{d}x, & \text{if } w \in [0,B], \\ \int_0^B g(x)\mathrm{d}x + c_k\mu_k(w-B), & \text{if } w > B, \end{cases} \tag{A.90}$$

where g is defined in (A.76) and (A.77) depending on θ. We will derive $B \in \mathbb{R}_{++}$ and $\gamma \in \mathbb{R}_+$ such that (f, B, γ) satisfies the condition (A.89). Observe that f satisfies (A.89a) by definition and the fact that g is the solution to the ODE (A.75) with the initial condition $g(0) = 0$.

Observe that $\lim_{w\downarrow B} f'(w) = c_k\mu_k$ and $\lim_{w\downarrow B} f''(w) = 0$. Therefore, we need $\lim_{w\uparrow B} f'(w) = c_k\mu_k$ and $\lim_{w\uparrow B} f''(w) = 0$ for f to be in $\mathbb{C}_2$. There are two cases to consider.

Case 1: $\theta = 0$. By (A.89a) and because we need (13d) and $f''(B) = 0$,

$$\begin{aligned} &\frac{1}{2}\Sigma f''(B) - \lambda_j (B-S)^+ f'(B) + b_j\mu_j(B-S)^+ = \gamma, \\ \Longrightarrow &-\lambda_j (B-S)^+ c_k\mu_k + b_j\mu_j(B-S)^+ = \gamma, \\ \Longrightarrow &\gamma = (b_j\mu_j - \lambda_j c_k\mu_k)(B-S)^+. \end{aligned} \tag{A.91}$$

Henceforth, we let γ defined as in (A.91).

If $B \le S$, then $\gamma = 0$ by (A.91) and $f(w) = 0$ for all $w \in [0, B]$ by (A.76) and (A.90), which implies that (13d) is not satisfied. Therefore, we must have $B > S$. In order for (13d) to hold, we need $g(B) = c_k\mu_k$ for some $B > S$ by (A.90). Let us consider the function G defined in (19). Observe that $G(S,S) = 0$ and $\lim_{w\to\infty} G(w,S) = \infty$. Because G is continuous in w, there exists a $w \in (S, \infty)$ such that $G(w,S) = c_k\mu_k$. Hence, we let

$$B := \inf\{w > S : G(w,S) = c_k\mu_k\}. \tag{A.92}$$

Therefore, $\lim_{w\uparrow B} f'(w) = g(B) = G(B,S) = c_k\mu_k$ by (19), (A.76), (A.90), and (A.92). Furthermore, because $f'(0) = g(0) = 0$ by (A.76), (13d) is satisfied by the triple (f, B, γ) defined in (A.90), (A.91), and (A.92). By (A.89a), (13d), (A.91), and (A.92),

$$\lim_{w\uparrow B}\left(\frac{1}{2}\Sigma f''(w) - \lambda_j (w-S)^+ f'(w) + b_j\mu_j(w-S)^+\right) = (b_j\mu_j - \lambda_j c_k\mu_k)(B-S)$$

$$\Longrightarrow \frac{1}{2}\Sigma \lim_{w\uparrow B} f''(w) - \lambda_j\,(B-S)\,c_k\mu_k + b_j\mu_j(B-S) = (b_j\mu_j - \lambda_j c_k\mu_k)\,(B-S)$$

$$\Longrightarrow \lim_{w\uparrow B} f''(B) = 0,$$

which implies $f''(B)=0$ as desired.

Next, we will consider (13c). By (A.76) and (A.90),

$$f''(w) = \begin{cases} \frac{2\gamma}{\Sigma}, & \text{if } w\in[0,S],\\ \frac{2\lambda_j}{\Sigma}(w-S)z(w)+\frac{2\gamma}{\Sigma}, & \text{if } w\in(S,B],\\ 0, & \text{if } w>B, \end{cases}$$

where

$$z(w) := \mathrm{e}^{\frac{\lambda_j}{\Sigma}(w-S)^2}\left(\frac{2\gamma}{\Sigma}S - \frac{b_j\mu_j}{\lambda_j} + \gamma\sqrt{\frac{\pi}{\lambda_j\Sigma}}\mathrm{Erf}\left[\sqrt{\frac{\lambda_j}{\Sigma}}(w-S)\right]\right), \qquad \forall w>S.$$

Because z is nondecreasing and $f''(B)=0$, $z(w)<0$ for all $w\in(S,B]$. By algebra,

$$f'''(w) = \begin{cases} 0, & \text{if } w\in[0,S],\\ \frac{2\lambda_j}{\Sigma}\left((w-S)f''(w)+z(w)\right), & \text{if } w\in(S,B],\\ 0, & \text{if } w>B, \end{cases}$$

Because $z(w)<0$ for all $w\in(S,B]$, if $f''(w)<0$ for some $w\in(S,B)$, then $f'''(w)<0$ for that $w\in(S,B)$. Therefore, if f'' becomes strictly negative at some point on the interval $w\in(S,B)$, it never becomes nonnegative again. Because $f''(B)=0$, it must be the case that $f''(w)\geq 0$ for all $w\in(S,B)$, which in turn implies $f''(w)\geq 0$ for all $w\in\mathbb{R}_+$. Therefore, $f'(w)\geq 0$ for all $w\in\mathbb{R}_+$ by (A.76) and (A.90). By (A.90), $f'(w)\leq f'(B)=c_k\mu_k$ for all $w\in[0,B]$ and $f'(w)=c_k\mu_k$ for all $w>B$. Consequently, the triple (f,B,γ) defined in (A.90), (A.91), and (A.92) satisfies (13c).

Finally, we will consider (A.89b). Recall that $f''(w)=0$ and $f'(w)=c_k\mu_k$ for all $w>B$. Therefore, for all $w>B$,

$$\frac{1}{2}\Sigma f''(w) - \lambda_j\,(w-S)\,f'(w) + b_j\mu_j(w-S) = (b_j\mu_j-\lambda_j c_k\mu_k)\,(w-S) \geq (b_j\mu_j-\lambda_j c_k\mu_k)\,(B-S) = \gamma,$$

which proves that the triple (f,B,γ) defined in (A.90), (A.91), and (A.92) satisfies (A.89b).

Case 2: $\theta<0$. By (A.89a) and because we need $f''(B)=0$ and (13d),

$$\begin{aligned}
&\frac{1}{2}\Sigma f''(B)+\left(\theta-\lambda_j\left(B-S\right)^+\right)f'(B)+b_j\mu_j(B-S)^+=\gamma,\\
\implies&\left(\theta-\lambda_j\left(B-S\right)^+\right)c_k\mu_k+b_j\mu_j(B-S)^+=\gamma,\\
\implies&\gamma=\theta c_k\mu_k+\left(b_j\mu_j-\lambda_j c_k\mu_k\right)(B-S)^+.
\end{aligned}\tag{A.93}$$

Henceforth, we let γ defined as in (A.93).

If $B\le S$, then $\gamma=\theta c_k\mu_k$ by (A.93) and by (A.77) and (A.90) and because we need (13d), we have

$$f'(B)=c_k\mu_k\left(1-\mathrm{e}^{-\frac{2\theta}{\Sigma}B}\right)<c_k\mu_k,\qquad\forall B\in\mathbb{R}_+.$$

Therefore, (13d) is not satisfied and thus we must have $B>S$. In order for (13d) to hold, we need $g(B)=c_k\mu_k$ for some $B>S$ by (A.90). Let us consider the function G defined in (20). Observe that $G(S,S)=c_k\mu_k\left(1-\mathrm{e}^{-\frac{2\theta}{\Sigma}S}\right)<c_k\mu_k$ for all $S\in\mathbb{R}_+$ and $\lim_{w\to\infty}G(w,S)=\infty$. Because G is continuous in w, there exists a $w\in(S,\infty)$ such that $G(w,S)=c_k\mu_k$. Hence, we let

$$B:=\inf\left\{w>S:G(w,S)=c_k\mu_k\right\}.\tag{A.94}$$

Therefore, $\lim_{w\uparrow B}f'(w)=g(B)=G(B,S)=c_k\mu_k$ by (20), (A.77), (A.90), and (A.94) and thus (13d) is satisfied by the triple (f,B,γ) defined in (A.90), (A.93), and (A.94). By (A.89a), (13d), (A.93), and (A.94),

$$\begin{aligned}
&\lim_{w\uparrow B}\left(\frac{1}{2}\Sigma f''(w)+\left(\theta-\lambda_j\left(w-S\right)^+\right)f'(w)+b_j\mu_j(w-S)^+\right)=\theta c_k\mu_k+\left(b_j\mu_j-\lambda_j c_k\mu_k\right)(B-S)\\
\implies&\frac{1}{2}\Sigma\lim_{w\uparrow B}f''(w)+\left(\theta-\lambda_j\left(B-S\right)\right)c_k\mu_k+b_j\mu_j(B-S)=\theta c_k\mu_k+\left(b_j\mu_j-\lambda_j c_k\mu_k\right)(B-S)\\
\implies&\lim_{w\uparrow B}f''(w)=0,
\end{aligned}$$

which implies $f''(B)=0$ as desired.

Observe that the ODE (A.75) with the initial condition $g(0)=0$ is a special case of the IVP (A.2). Under the γ defined in (A.93), we have $g(B)=G(B,S)=c_k\mu_k>0$, that is, the unique solution to the IVP (A.2) is strictly positive at $B \in \mathbb{R}_+$. Therefore, by Lemma A.3.2 Part ii and Lemma A.3.3, we have $\gamma > 0$.

Next, we will consider (13c). By (A.77) and (A.90),

$$f''(w) = \begin{cases} \frac{2\gamma}{\Sigma} \mathrm{e}^{-\frac{2\theta}{\Sigma} w}, & \text{if } w \in [0, S], \\ \frac{2}{\Sigma} \left((\lambda_j (w - S) - \theta) f'(w) + \gamma - b_j \mu_j (w - S) \right), & \text{if } w \in (S, B], \\ 0, & \text{if } w > B. \end{cases} \tag{A.95}$$

Observe that $f''(w) \geq 0$ for all $w \in [0, S]$ and $f''(B) = 0$. Furthermore, by (A.93), (A.95), and algebra, for all $w \in (S, B)$

$$f'''(w) = \left(\frac{1}{w - S - \frac{\theta}{\lambda_j}} + \frac{2}{\Sigma} (\lambda_j (w - S) - \theta) \right) f''(w) - \frac{2}{\Sigma} (b_j \mu_j - \lambda_j c_k \mu_k) \frac{B - S - \frac{\theta}{\lambda_j}}{w - S - \frac{\theta}{\lambda_j}}. \tag{A.96}$$

Observe that the coefficient of f'' in (A.96) is strictly positive and the second term in the RHS of (A.96) is strictly negative. Therefore, if $f''(w) < 0$ for some $w > S$, then $f'''(w) < 0$ for that w. Therefore, if f'' becomes strictly negative at some point on the interval (S, B), it will keep decreasing henceforth. Because $f''(B) = 0$, it must be the case that $f''(B) \geq 0$ for all $w \in (S, B)$. Therefore, $f''(w) \geq 0$ for all $w \in [0, B]$, which in turn implies $0 \leq f'(w) \leq f'(B) = c_k \mu_k$ for all $w \in [0, B]$. Furthermore, $f'(w) = c_k \mu_k$ for all $w > B$. Consequently, the triple (f, B, γ) defined in (A.90), (A.94), and (A.93) satisfies (13c).

Finally, we will consider (A.89b). Recall that $f''(w) = 0$ and $f'(w) = c_k \mu_k$ for all $w > B$. Therefore, for all $w > B$,

$$\begin{aligned} &\frac{1}{2} \Sigma f''(w) + (\theta - \lambda_j (w - S)) f'(w) + b_j \mu_j (w - S) \\ &\qquad = \theta c_k \mu_k + (b_j \mu_j - \lambda_j c_k \mu_k)(w - S) \geq \theta c_k \mu_k + (b_j \mu_j - \lambda_j c_k \mu_k)(B - S) = \gamma, \end{aligned}$$

which proves that the triple (f, B, γ) defined in (A.90), (A.93), and (A.94) satisfies (A.89b).

A.9.2. Inventory optimization

For given $S \in \mathbb{R}_+$, the optimal long-run average backorder and emergency repair cost, $\gamma(S)$, is given by (A.91) and (A.93) and the associated unique barrier level B is defined in (A.92) and (A.94).

Therefore, the optimization problem (21) finds the optimal inventory level and its associated barrier level to the EWF (16).

Next, we will show that there exists an optimal solution to the optimization problem (21). First, we will prove an auxiliary result about the function $G(w,S)$ defined in (19) and (20). By (19) and algebra, if $\theta = 0$,

$$G'(w,S) := \frac{\partial G(w,S)}{\partial w} = \frac{2\lambda_j}{\Sigma}(w-S)\left(G(w,S) - c_k\mu_k\right) \\ + e^{\frac{\lambda_j}{\Sigma}(w-S)^2}\left(b_j\mu_j - \lambda_j c_k\mu_k\right)\left(\frac{2}{\Sigma}S + \sqrt{\frac{\pi}{\lambda_j\Sigma}}\operatorname{Erf}\left[\sqrt{\frac{\lambda_j}{\Sigma}}(w-S)\right]\right). \tag{A.97}$$

By (20) and algebra, if $\theta < 0$,

$$G'(w,S) = \frac{2(\lambda_j(w-S)-\theta)}{\Sigma}\left(G(w,S) - c_k\mu_k\right) + e^{\frac{1}{\Sigma}\left(\lambda_j(w-S)^2 - 2\theta(w-S)\right)}\left(b_j\mu_j - \lambda_j c_k\mu_k\right) \\ \times\left(\frac{1}{\theta}\left(1 - e^{-\frac{2\theta}{\Sigma}S}\right) + \sqrt{\frac{\pi}{\lambda_j\Sigma}}e^{\frac{\theta^2}{\lambda_j\Sigma}}\left(\operatorname{Erf}\left[\frac{\theta}{\sqrt{\lambda_j\Sigma}}\right] + \operatorname{Erf}\left[\frac{-\theta+\lambda_j(w-S)}{\sqrt{\lambda_j\Sigma}}\right]\right)\right). \tag{A.98}$$

Because $B > S$ and $G(B,S) = c_k\mu_k$ by (A.92) and (A.94), we have $G'(B,S) > 0$ by (A.97) and (A.98). Furthermore, observe that $G'(w,S) > 0$ for all $w > B$ by (A.97) and (A.98). Therefore, after w exceeds the barrier level B, $G(w,S)$ keeps increasing. Consequently, there exists a unique $w > S$ such that $G(w,S) = c_k\mu_k$, that is, the infimum in (21b) is actually redundant.

For notational convenience, let us define $B : \mathbb{R}_+ \to \mathbb{R}_+$ such that $B(S)$ denotes the unique barrier level satisfying (21b) for all $S \in \mathbb{R}_+$. We will prove that $B(S)$ is continuous in S by the proof by contradiction technique. Suppose that $B(S)$ is not continuous in S. Then, there must exist a sequence $\{S_n, n \in \mathbb{N}\}$ such that $S_n \in \mathbb{R}_+$ for all $n \in \mathbb{N}$, $S_n \to S$ as $n \to \infty$, but $B(S_n) \nrightarrow B(S)$ as $n \to \infty$. There are two cases to consider.

First, suppose that $B(S_n) \to X$ as $n \to \infty$ where $X \neq B(S)$. Because $B(S_n) > S_n$ for all $n \in \mathbb{N}$, we have $X \geq S$. Furthermore, because $G(w,S)$ is continuous in (w,S) and $G(B(S_n), S_n) = c_k\mu_k$ for all $n \in \mathbb{N}$, we have $G(X,S) = c_k\mu_k$. Because $B(S)$ is the unique solution of $G(w,S) = c_k\mu_k$ when $w \geq S$, we must have $X = B(S)$, which is a contradiction.

Second, suppose that $B(S_n)$ does not converge as $n \to \infty$, that is, $\limsup_{n\to\infty} B(S_n) > \liminf_{n\to\infty} B(S_n)$. Because $B(S_n) > S_n$ for all $n \in \mathbb{N}$, we have $\liminf_{n\to\infty} B(S_n) \geq S$. Furthermore, there exists a subsequence $\{n_m, m \in \mathbb{N}\}$ which achieves the lim sup. Specifically, $\lim_{m\to\infty} B(S_{n_m}) = \limsup_{n\to\infty} B(S_n)$. Because $G(w,S)$ is continuous in (w,S) and $G(B(S_{n_m}), S_{n_m}) = c_k\mu_k$ for all $m \in \mathbb{N}$, we have $G(\limsup_{n\to\infty} B(S_n), S) = c_k\mu_k$. Due to the uniqueness of $B(S)$, we must have $B(S) = \limsup_{n\to\infty} B(S_n)$. Similarly, we can prove that $B(S) = \liminf_{n\to\infty} B(S_n)$ implying that $B(S_n) \to B(S)$, which is a contradiction.

Consequently, $B(S)$ is continuous in S implying that the objective function (21a) is also continuous in S. Because $B(S) > S$ for all $S \in \mathbb{R}_+$, the objective function (21a) is bounded below by $\theta c_k \mu_k$ and tends to infinity as $S \to \infty$. Therefore, it attains its minimum and thus there exists an optimal solution to (21).

A.10. Additional Numerical Experiments

We present the MDP model formulation in Section A.10.1 and present the inventory enumeration algorithm in Section A.10.2. In Section A.10.3, we present additional information about the instance from Table 1 in which the performance gap between the LS and the enumeration is the highest under the proposed policy. Next, in Section A.10.4, we repeat the numerical experiments of Table 2 under the proposed policy with the LS. Then, we present numerical experiments in which emergency repair costs are scaled with an order different from n (recall the assumption in (2)) in Section A.10.5. Recall that our proposed policy simplifies significantly under Assumption 2. In Section A.10.6, we repeat the numerical experiments of Table 1 with the simplified version of the proposed policy (specifically, with the PP1) even though Assumption 2 does not hold in those experiments.

A.10.1. MDP formulation

We present the MDP model formulation for a system with two SKUs. Let us fix an arbitrary $(n_1, n_2) \in \mathbb{N}_+^2$ such that $n_1 + n_2 = n$. Let us also fix an arbitrary inventory vector $\boldsymbol{S}^n = (S_1^n, S_2^n) \in \mathbb{N}^2$.

State space. The state of the system at time $t \in \mathbb{R}_+$ is

$$\big((U_1^n(t), Q_1^n(t), OH_1^n(t)),\ (U_2^n(t), Q_2^n(t), OH_2^n(t)),\ Y^n(t)\big), \tag{A.99}$$

where for all $i \in \{1,2\}$ and $t \in \mathbb{R}_+$,

- $U_i^n(t)$ denotes the number of parts of SKU i installed at the capital goods at time t.
- $Q_i^n(t)$ denotes the number of broken parts of SKU i at time t.
- $OH_i^n(t)$ is the number of (on-hand) ready-for-use parts of SKU i at time t. We let

$$(U_i^n(t), Q_i^n(t), OH_i^n(t)) \in \left\{ \begin{array}{l} (0, n_i + S_i^n, 0)\,, \\ (1, n_i + S_i^n - 1, 0)\,, \\ (2, n_i + S_i^n - 2, 0)\,, \\ \quad\vdots \\ (n_i - 1, S_i^n + 1, 0)\,, \\ (n_i, S_i^n, 0)\,, \\ (n_i, S_i^n - 1, 1)\,, \\ (n_i, S_i^n - 2, 2)\,, \\ \quad\vdots \\ (n_i, 1, S_i^n - 1)\,, \\ (n_i, 0, S_i^n)\,. \end{array} \right\} \tag{A.100}$$

- $Y^n(t)$ denotes the server state at time t. We let $Y^n(t) \in \{0,1,2\}$ where $Y^n(t) = 0$ implies that the server is idle, $Y^n(t) = 1$ implies that the server repairs a broken part of SKU 1, and $Y^n(t) = 2$ implies that the server repairs a broken part of SKU 2 at time t.

Recall that only the work-conserving and non-preemptive repair policies are allowed in the repair facility. Because the repair facility operates in a work-conserving fashion,

$$Y^n(t) \neq i \text{ if } Q_i^n(t) = 0 \text{ for all } i \in \{1,2\}, \tag{A.101a}$$

$$Y^n(t) = 0 \text{ if and only if } Q_1^n(t) = Q_2^n(t) = 0. \tag{A.101b}$$

By (A.99), (A.100), and (A.101), the number of states in the MDP model is

$$2(n_1 + S_1^n)(n_2 + S_2^n) + n_1 + S_1^n + n_2 + S_2^n + 1,$$

where the multiplier 2 in the first term captures the type of the part that is under repair. We let $\mathcal{S}$ denote the state space of the MDP model. Moreover, we let $x_0 := ((n_1, 0, S_1^n), (n_2, 0, S_2^n), 0)$ denote the only state in which there are no broken parts and thus the server is idle.

In the MDP model for a system with three SKUs, the number of states is equal to

$$3\prod_{i=1}^{3}(n_i+S_i^n)+2(n_1+S_1^n)(n_2+S_2^n)+2(n_1+S_1^n)(n_3+S_3^n)+2(n_2+S_2^n)(n_3+S_3^n)+\sum_{i=1}^{3}(n_i+S_i^n)+1.$$

For example, if there are three SKUs and $n_i + S_i^n = 33$ for all $i \in \{1, 2, 3\}$, then the number of states is equal to 114,445 and thus the optimal cost computation is computationally challenging. Because we enumerate the inventory vector $\boldsymbol{S}^n$ to find the optimal inventory levels, we need to solve the MDP multiple times. Therefore, we consider two SKUs in the numerical experiments. For example, for many instances with $n_1 = n_2 = 50$ in Table 1, we enumerate hundreds of inventory vectors per instance.

Action space. Actions are taken at breakdown and repair epochs but no action is taken between consecutive event epochs. The system controller makes two types of decisions. First, when the server becomes available for repair, if there are broken parts of both SKU 1 and SKU 2, then the system controller should choose which part to repair. We let $r \in \{0, 1, 2\}$ denote the action of choosing the type of the broken part for repair in the repair facility. Specifically, $r = 1$ ($r = 2$) denotes the action of repairing a broken part of SKU 1 (2) and $r = 0$ denotes the action of idling the repair facility, which happens only when there is no broken part in the repair facility.

Second, if a part breaks down when there is no on-hand inventory, then the system controller should decide whether to backorder the demand or use an emergency repair to fulfill the demand. For all $i \in \{1, 2\}$, We let $\xi_i \in \{0, 1, 2\}$ denote the associated action for SKU i such that $\xi_i = 0$ denotes the action of fulfilling the demand of SKU i from the on-hand inventory, $\xi_i = 1$ denotes the action of backordering the demand for SKU i, and $\xi_i = 2$ denotes the action of using an emergency repair to satisfy the demand for SKU i. Let $A(x)$ denote the set of feasible actions in state $x=((U_1,Q_1,OH_1),(U_2,Q_2,OH_2),Y)$ for all $x\in\mathcal{S}$. Then, for all $x\in\mathcal{S}\setminus\{x_0\}$, we have

$$A(x)=\left\{ r\in \begin{Bmatrix} \{0\}, & \text{if } Q_1-\mathbb{I}\{Y=1\}=Q_2-\mathbb{I}\{Y=2\}=0 \\ \{1\}, & \text{if } Q_1-\mathbb{I}\{Y=1\}>0 \text{ and } Q_2-\mathbb{I}\{Y=2\}=0 \\ \{2\}, & \text{if } Q_1-\mathbb{I}\{Y=1\}=0 \text{ and } Q_2-\mathbb{I}\{Y=2\}>0 \\ \{1,2\}, & \text{if } Q_1-\mathbb{I}\{Y=1\}>0 \text{ and } Q_2-\mathbb{I}\{Y=2\}>0 \end{Bmatrix}, \right.$$
$$\left. \xi_1\in\Big\{\mathbb{I}\{OH_1=0\},2\mathbb{I}\{OH_1=0\}\Big\},\ \xi_2\in\Big\{\mathbb{I}\{OH_2=0\},2\mathbb{I}\{OH_2=0\}\Big\}\right\},$$

and

$$A(x_0)=\left\{ r\in \begin{Bmatrix} \{1\}, & \text{if the next event is a breakdown of a part of SKU 1} \\ \{2\}, & \text{if the next event is a breakdown of a part of SKU 2} \end{Bmatrix}, \right.$$
$$\left. \xi_1\in\Big\{\mathbb{I}\{OH_1=0\},2\mathbb{I}\{OH_1=0\}\Big\},\ \xi_2\in\Big\{\mathbb{I}\{OH_2=0\},2\mathbb{I}\{OH_2=0\}\Big\}\right\}.$$

Uniformization. We uniformize the continuous time MDP model to obtain an equivalent discrete time MDP model. Let $\beta:=n_1\lambda_1^n+\mu_1+n_2\lambda_2^n+\mu_2$ so that β is an upper bound on the transition rate at all states. In the uniformized MDP model, in state $x=((U_1,Q_1,OH_1),(U_2,Q_2,OH_2),Y)$, the next event is a breakdown of a part of SKU i with probability $U_i\lambda_i^n/\beta$ and the repair of a broken part of SKU i with probability $\mu_i\mathbb{I}\{Y=i\}/\beta$ for all $i\in\{1,2\}$. The next event is a fictitious epoch in which the system state does not change with probability

$$P_f(x):=1-\frac{U_1\lambda_1^n+\mu_1\mathbb{I}\{Y=1\}+U_2\lambda_2^n+\mu_2\mathbb{I}\{Y=2\}}{\beta}.$$

Transition probabilities. Let $e_j\in\mathbb{R}^7$ denote the vector whose jth component is equal to 1 and all the other components are equal to 0 for all $j\in\{1,2,\ldots,7\}$. In the uniformized MDP, the transition probabilities from state $x = ((U_1, Q_1, OH_1), (U_2, Q_2, OH_2), Y) \in \mathcal{S}$ to $x' \in \mathcal{S}$ at the next event under action $a \in A(x)$ is

$$P_{xx'}(a) = \begin{cases} U_1\lambda_1^n/\beta, & \text{if } (OH_1 > 0 \text{ and } x' = x + e_2 - e_3) \\ & \quad \text{or } (OH_1 = 0,\ \xi_1 = 1, \text{ and } x' = x - e_1 + e_2), \\ U_2\lambda_2^n/\beta, & \text{if } (OH_2 > 0 \text{ and } x' = x + e_5 - e_6) \\ & \quad \text{or } (OH_2 = 0,\ \xi_2 = 1, \text{ and } x' = x - e_4 + e_5), \\ \mu_1 \mathbb{I}\{Y = 1\}/\beta, & \text{if } (U_1 = n_1 \text{ and } x' = x - e_2 + e_3 + (r-1)e_7) \\ & \quad \text{or } (U_1 < n_1 \text{ and } x' = x + e_1 - e_2 + (r-1)e_7), \\ \mu_2 \mathbb{I}\{Y = 2\}/\beta, & \text{if } (U_2 = n_2 \text{ and } x' = x - e_5 + e_6 + (r-2)e_7) \\ & \quad \text{or } (U_2 < n_2 \text{ and } x' = x + e_4 - e_5 + (r-2)e_7), \\ P_f(x) + \sum_{i=1}^2 \frac{U_i\lambda_i^n}{\beta} \mathbb{I}\{OH_i = 0,\ \xi_i = 2\}, & \text{if } x' = x. \end{cases}$$

Costs. Under state $x = ((U_1, Q_1, OH_1), (U_2, Q_2, OH_2), Y) \in \mathcal{S}$ and action $a \in A(x)$, the expected one-stage cost until the next event is given by

$$C(x, a) = \sum_{i=1}^{2} \big(b_i(n_i - U_i) + c_i^n U_i \lambda_i^n \mathbb{I}\{OH_i = 0,\ \xi_i = 2\}\big)/\beta. \tag{A.102}$$

Recall that the inventory vector (S_1^n, S_2^n) is a fixed parameter. Because the inventory holding cost per unit time, which is equal to $h_1 S_1^n + h_2 S_2^n$, is a fixed cost, we do not include it in (A.102).

Bellman equation. The optimality equations are as follows.

$$\gamma + J(x) = \min_{a \in A(x)} \left\{ C(x, a) + \sum_{x' \in \mathcal{S}} P_{xx'}(a) J(x') \right\}, \qquad \forall x \in \mathcal{S}, \tag{A.103a}$$

$$J(x_0) := 0, \tag{A.103b}$$

where γ denotes the optimal long-run average backorder and emergency repair cost and $J : \mathcal{S} \to \mathbb{R}$ denotes the relative value function. We make the definition in (A.103b) without loss of generality.

The uniformized MDP model has the following properties.

- Both the number of states and the number of actions at each state are finite.
- The transition probabilities and the expected one-stage costs are stationary.
- The expected one-stage costs are bounded.

For all $x \in \mathcal{S}$, let $a^{(1)}, a^{(2)} \in A(x)$. Consider the randomized action $a(\varphi) := \varphi a^{(1)} + (1-\varphi)a^{(2)}$ for all $\varphi \in [0,1]$, that is, the action $a(\varphi)$ is equal to $a^{(1)}$ with probability φ and equal to $a^{(2)}$ with probability $1-\varphi$. One can see that both $P_{xx'}(a(\varphi))$ and $C(x, a(\varphi))$ are continuous in φ for all $x, x' \in \mathcal{S}$ and $a^{(1)}, a^{(2)} \in A(x)$. Furthermore, observe that, each deterministic Markov control (DMC) policy results in a single recurrent class of states and a possibly empty set of transient states. Therefore, the uniformized MDP is unichain (see page 348 of Puterman (2005)). Consequently, by theorem 8.4.3 of Puterman (2005), the Bellman equation (A.103) has a unique solution which is associated with a DMC policy that is optimal among all DMC policies. Furthermore, by theorem 8.4.7 of Puterman (2005), an optimal DMC policy is also optimal among the randomized and history-dependent control policies. Finally, by theorem 8.6.6 of Puterman (2005), the policy iteration algorithm (see section 8.6.1 of Puterman (2005)) finds an optimal DMC policy in finite number of iterations.

A.10.2. Inventory Enumeration

We present the pseudocode that we use for the inventory enumeration. Let $\tilde{C}(S_1, S_2)$ denote the long-run average cost under the inventory vector (S_1, S_2), under a given policy, and under a parameter instance, which is computed by solving the uniformized MDP model.

```
C~_2 <- +inf and S_2 <- 0
While 1
        C~_1 <- +inf and S_1 <- 0
        While 1
                If C~_1 > C~(S_1, S_2)
                        C~_1 <- C~(S_1, S_2) and S_1^** <- S_1
                Else
                        Break loop
                End if
                S_1 <- S_1 + 1.
        End while
        If C~_2 > C~_1
```

$\tilde{C}_2 \leftarrow \tilde{C}_1$, $S_2^* \leftarrow S_2$, and $S_1^* \leftarrow S_1^{**}$
Else
Break loop
End if
$S_2 \leftarrow S_2 + 1$.
End while
Return $\tilde{C}_2$, S_1^*, and S_2^*

A.10.3. Illustration of the LS under an instance from Table 1

We present additional information about the instance from Table 1 in which the performance gap between the LS and the enumeration is the highest (1.56%) under the proposed policy. Specifically, Table A.10.2 presents the costs under the proposed policy for various given inventory vectors under a specific instance from Table 1. The yellow cells show the path followed by the LS and the red cell shows the optimal cost. We make the following observations from Table A.10.2.

- The optimal solution is a diagonal neighbor of the solution provided by the LS and thus the LS solution is very close to the optimal one.
- When S_1^n (S_2^n) is fixed, the cost is "almost" convex in S_2^n (S_1^n). This is why the LS performs very close to the enumeration in the numerical experiments. However, in some cases, the aforementioned convexity does not hold and the LS can stuck in a local optimum solution. For example, in Table A.10.2, $\tilde{C}(15,10) = 3.9487$, $\tilde{C}(16,10) = 3.9582$, and $\tilde{C}(17,10) = 3.9069$, where $\tilde{C}(S_1^n, S_2^n)$ denotes the cost under the proposed policy under the inventory vector (S_1^n, S_2^n).

A.10.4. Experiments in Section 6.3 under the proposed policy with the LS

We repeat the experiments in Section 6.3 under the PP2. On the one hand, the PP1 never outperforms the PP2. On the other hand, the performance gap is small. In Table A.10.1, we present the average and maximum absolute % deviations of the costs under the PP2 from the PP1, that is, at each instance (among the 204 instances), we compute

$$100 \times \frac{|\text{cost under the PP2} - \text{cost under the PP1}|}{\text{cost under the PP1}}.$$

Table A.10.1 Average and maximum absolute % deviations of the costs under the PP2 from the PP1 under the experiments in Table 2.

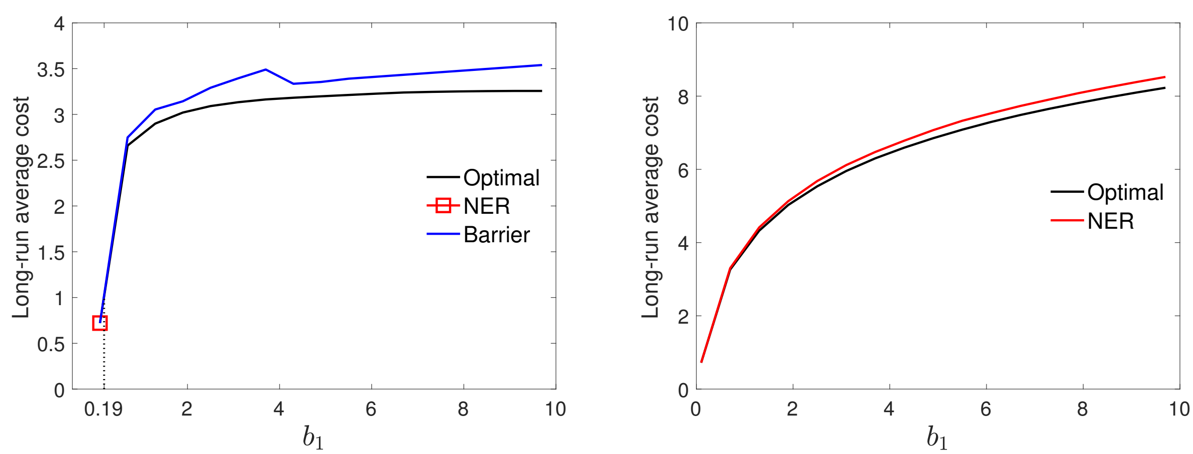
	$n_1 = n_2 = 10$		$n_1 = n_2 = 25$		$n_1 = n_2 = 50$	
ρ^n	Avg.	Max.	Avg.	Max.	Avg.	Max.
0.8	0.34	3.43	0.16	1.96	0.18	2.27
0.9	0.22	2.32	0.17	1.9	0.12	1.23
0.95	0.49	1.9	0.16	1.3	0.08	0.57
1	0.87	2.95	0.1	0.84	0.05	0.43

Table A.10.2 (Color online) Costs under the proposed policy for various (S_1^n, S_2^n) values when $n_1 = n_2 = 10$, $\rho^n = 1$, $\mu_2 = 4$, and $b_1 = 9.1$.

$S_2^n \backslash S_1^n$	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
0	18.7637	14.5136	12.3872	11.3320	10.7170	10.3295	10.1110	9.9759	9.8849	9.8491	9.8430	9.8590	9.8920	9.9380	9.9943	10.0588	10.1297	10.2059
1	17.0992	12.9808	10.7531	9.5495	8.8094	8.3578	8.1067	7.9401	7.7936	7.7327	7.7031	7.6971	7.7096	7.7366	7.7753	7.8234	7.8793	7.9417
2	16.0079	11.8753	9.6324	8.3624	7.7118	7.1394	6.8497	6.6484	6.4493	6.3616	6.3074	6.2789	6.2705	6.2782	6.2989	6.3303	6.3706	6.4183
3	14.8502	10.9384	9.0414	7.7193	7.0335	6.4046	6.0844	5.8555	5.6135	5.5037	5.4294	5.3825	5.3572	5.3492	5.3553	5.3731	5.4006	5.4363
4	13.8599	10.4118	8.6000	7.2813	6.4031	5.9345	5.5998	5.2579	5.0872	4.9642	4.8775	4.8190	4.7829	4.7648	4.7614	4.7703	4.7893	4.8170
5	13.3536	9.9788	8.2511	6.9670	6.0954	5.6283	5.2906	4.9351	4.7581	4.6285	4.5352	4.4703	4.4279	4.4038	4.3948	4.3982	4.4121	4.3811
6	12.6358	9.3077	7.6937	6.7311	5.8811	5.4245	5.0913	4.7323	4.5543	4.4226	4.3267	4.2588	4.2134	4.1862	4.1740	4.1744	4.1854	4.2053
7	12.0588	9.0029	7.4562	6.5475	5.7268	5.2862	4.9620	4.6057	4.4301	4.2994	4.2035	4.1350	4.0886	4.0602	4.0468	4.0457	4.0552	4.0091
8	11.7807	8.7431	7.2554	6.1983	5.6127	5.1907	4.7564	4.5284	4.3574	4.2295	4.1352	4.0677	4.0217	3.9934	3.9797	3.9782	3.9872	3.9379
9	11.5484	8.5201	7.0840	6.0810	5.5266	5.1242	4.7030	4.4835	4.3182	4.1943	4.1028	4.0371	3.9924	3. 9649	3.9517	3.9505	3.8880	3.9087
10	11.1272	8.0642	6.9366	5.9848	5.4605	5.0777	4.6709	4.4603	4.3014	4.1821	4.0939	4.0307	3.9877	3.9615	3.9492	3.9487	3.9582	3.9069
11	10.9681	7.9025	6.8092	5.9052	5.4095	5.0458	4.6534	4.4521	4.2997	4.1853	4.1007	4.0402	3.9993	3.9747	3.9636	3.9640	3.9018	3.9233
12	10.8339	7.7629	6.5031	5.8391	5.3701	4.9008	4.6466	4.4542	4.3085	4.1990	4.1182	4.0605	4.0219	3.9989	3.9892	3.9907	3.9296	3.9516
13	10.7205	7.6421	6.4109	5.7843	5.3401	4.8899	4.6477	4.4640	4.3247	4.2201	4.1430	4.0882	4.0518	4.0307	4.0224	3.9505	3.9653	3.9879
14	10.4483	7.5374	6.3315	5.7390	5.3177	4.8856	4.6548	4.4794	4.3463	4.2464	4.1729	4.1211	4.0869	4.0676	4.0607	4.0645	4.0063	4.0297

A.10.5. Experiments with different emergency repair cost scalings

We assume that the emergency repair costs are scaled with the order of n in the assumption in (2). Because both n and c_i^n are fixed in a pre-limit system for all $i \in \mathcal{I}$, any c_i^n value can be scaled by n (e.g., $c_i^n = nc_i$ where $c_i := c_i^n/n$). Therefore, this assumption is not restrictive in a pre-limit system. To illustrate this, we present numerical experiments in which the emergency repair costs are not scaled with the order of n. Specifically, we repeat the numerical experiments depicted in Figure 4b by updating the emergency repair cost with $c_2^n = 2c_1^n = 4\sqrt{n}$ or $c_2^n = 2c_1^n = 4n^{1.5}$.



(a) $c_2^n = 2c_1^n = 4\sqrt{n}$ and the switch point is 0.19 (b) $c_2^n = 2c_1^n = 4n^{1.5}$ and the switch point is 19

Figure A.10.1 (Color online) Performances of the optimal policy and the PP1 when $n_1 = n_2 = 50$, $\mu_1 = \mu_2 = 1$, $\lambda_1^n = \lambda_2^n = 0.95/n$, $h_2 = 2h_1 = 0.2$, $b_2 = 2b_1$, and $b_1 \in \{0.1, 0.7, 0.13, \ldots, 9.7\}$.

Figure A.10.1a presents the numerical experiments in which $c_2^n = 2c_1^n = 4\sqrt{n}$ and the switch point is $\lambda_1^n c_1^n = 1.9/\sqrt{n} = 0.19$. Observe that as n increases, the switch point decreases in the order of $\sqrt{n}$ and thus the proposed policy becomes the barrier policy in more cases. For example, the proposed policy is the NER policy only when $b_1 = 0.1$ in the experiments depicted in Figure A.10.1a. This is not surprising because as n increases, the emergency repairs become relatively less expensive and thus the barrier policy becomes more preferable.

Figure A.10.1b presents the numerical experiments in which $c_2^n = 2c_1^n = 4n^{1.5}$ and the switch point is $\lambda_1^n c_1^n = 1.9\sqrt{n}$. Observe that as n increases, the switch point increases in the order of $\sqrt{n}$

and thus the proposed policy becomes the NER policy in more cases. For example, the switch point is 19 and thus the proposed policy is always the NER policy in the experiments depicted in Figure A.10.1b. As n increases, the emergency repairs become more expensive and thus the NER policy becomes more preferable. Nevertheless, the proposed policy performs reasonably well in the experiments depicted in Figure A.10.1.

A.10.6. Experiments in Table 1 with a simplified version of the proposed policy

Recall that the optimal EWF (10) solution and thus the proposed policy simplify significantly under Assumption 2. For example, under Assumption 2, the optimal EWF (10) solution backorders only the demands of a particular SKU and keeps non-zero inventory only for a particular SKU, and the proposed policy simplifies accordingly. Therefore, the proposed policy under Assumption 2 is more static than it is in the general case.

We repeat the numerical experiments presented in Table 1 with the simplified version of the proposed policy (specifically, with the PP1) even though Assumption 2 does not hold in those instances. Specifically, we implement the PP1 by assuming that $\lambda_1^n > \lambda_2^n$ even though $\lambda_1^n < \lambda_2^n$ in those experiments. The results are presented in Table A.10.3. According to the results, the PP1 performs poorly. Therefore, when Assumption 2 does not hold, it is crucial to implement the dynamic version of the proposed policy that is designed for the general case.

Table A.10.3 Average and maximum % deviations of the costs under the PP1 from the optimal costs under the numerical experiments in Table 1.

		$n_1 = n_2 = 10$		$n_1 = n_2 = 25$		$n_1 = n_2 = 50$	
	ρ^n	Avg.	Max.	Avg.	Max.	Avg.	Max.
$\mu_2 = 3$	0.8	6.78	12.59	4.76	6.45	4.67	6.37
	0.9	9.29	19.82	4.72	5.86	4.30	5.59
	0.95	9.02	15.98	4.60	7.94	4.10	6.26
	1	7.81	14.23	4.50	13.95	4.72	15.12
$\mu_2 = 4$	0.8	10.03	20.45	4.78	6.32	4.51	5.84
	0.9	13.82	27.40	6.50	9.15	5.20	6.31
	0.95	13.76	22.79	7.92	13.87	6.72	11.84
	1	12.96	22.26	10.06	26.37	10.74	29.10
$\mu_2 = 5$	0.8	13.83	28.20	5.84	8.70	4.65	5.85
	0.9	18.12	34.28	9.53	11.73	6.86	9.02
	0.95	18.47	28.41	11.98	18.97	9.91	16.55
	1	18.66	32.85	17.22	36.65	18.47	40.06
$\mu_2 = 6$	0.8	17.52	35.13	7.92	12.75	5.17	7.48
	0.9	22.15	39.68	13.12	15.58	9.36	11.19
	0.95	23.73	34.07	17.05	24.67	14.26	21.10
	1	22.69	37.29	22.57	42.77	24.54	46.96