Dynamic Control of a Make-to-Order System Under Model Uncertainty

Xu Sun
University of Miami Herbert Business School, Coral Gables, Florida 33146, USA, xxs767@miami.edu

Xiaohan Zhu
School of Management and Engineering, Nanjing University, Nanjing, China, zhuxiaohan@nju.edu.cn

A control policy for a make-to-order manufacturing system is typically derived based on a precise probabilistic model governing demand realization. In practice, however, such a probabilistic model may be a simplification of the actual scenario due to tractability considerations. Consequently, policies derived from such simplifications may perform poorly if the assumed model does not accurately capture reality. To address this issue, this paper proposes a modeling paradigm that can generate control policies based on a simplified model while accounting for possible model errors that may result. The focus is on controlling a multi-product make-to-order manufacturing system with an outsourcing mechanism, and the paper specifically addresses deliberate simplification for the demand model. The robust control formulation takes the form of a two-player zero-sum game. Because the original formulation is not tractable enough, the paper further presents an approximating problem under the heavy-traffic assumption that effectively results in a stochastic differential game. The solution to this game then translates into an implementable control rule that can be used for the original make-to-order system. Additionally, the paper presents a simulation-based method for selecting an appropriate uncertainty set. Numerical experiments are conducted to demonstrate the value of building "robustness" into decision-making.

Key words: make-to-order manufacturing; robust control; heavy-traffic approximation; stochastic differential games

1. Introduction

Embracing the make-to-order (MTO) strategy in manufacturing entails producing the final product only upon customer order placement. Compared to the make-to-stock (MTS) strategy, MTO allows for customizable products tailored to the needs of individual customers, minimizes the need for maintaining inventories of finished goods, and reduces the environmental impact. In recent years, an increasing number of manufacturers have transitioned from traditional MTS practices to more flexible MTO models, a shift due to, in part, the proliferation of online marketplaces such as eBay, Walmart, and Shein, which promise highly customized offerings. Another contributing factor to this shift is technological innovation, including 3D printing, a technology that utilizes 3D printers to produce a variety of printable products on demand and holds the promise for mass customization (Chen et al.

2021). On the flip side, a challenge in implementing the MTO strategy is aligning supply with demand because it cannot stockpile inventory to decouple demand variability from production, potentially resulting in extended waiting times for customers.

The aforementioned trade-off has sparked significant interest in optimizing MTO operations. Researchers in this field often model MTO systems as queueing systems, assuming that the demand model, or more precisely, the probabilistic law governing demand, is known and accurately reflects reality. However, in practice, the demand model may be oversimplified due to tractability considerations. For instance, a doubly stochastic Poisson process with an auto-regressive intensity process might be deemed appropriate for capturing observed auto-correlation among demand arrivals. However, the precise intensity values over time are often unobservable. Even if observable, incorporating such detailed information into the model can increase computational complexity. Thus, while it's often possible to develop a sophisticated "high-fidelity" model to describe demand realization, simplifying assumptions such as "Markovian" and "stationarity" may be adopted to mitigate computational challenges, resulting in a "low-fidelity" model.

In this paper, we explore how decision-makers can effectively utilize simplified models in MTO manufacturing systems while addressing potential model errors and developing strategies to mitigate their negative impacts. Specifically, we focus on a scenario where a manufacturing system produces multiple product types with a shared capacity. While aiming to produce most products in-house, the system can outsource production for certain items at a fixed and proportional cost as needed. For example, for "printable" products, a professional 3D printing bureau may serve as an external subcontractor to support in-house manufacturing efforts as needed. We operate under the stipulation that the decision-maker knows the correct demand model but deems it too complex; meanwhile, he is able to find a simplified model (the nominal model) believed to capture the essential features of reality. This hints at a robust control formulation for the MTO system in question, which we develop in this paper. To infuse "robustness" into decision-making, we leverage a notion stemming from a stream of research pioneered by Hansen and Sargent (2001). The main idea is to extract a nominal model and add nature as a malevolent second player that perturbs the nominal model within some prescribed limits. The malevolence of nature serves as a tool for the decision-maker to explore the fragility of candidate decision rules.

Broadly speaking, the way we incorporate the decision-maker's wariness of model errors due to model simplification on demand conforms to the general philosophy underpinning robust optimization (RO), a popular modeling paradigm well-suited for addressing such an issue. RO seeks to resolve the trade-off between using a high-fidelity model that better reflects reality and adopting a low-fidelity model for tractability. It takes into account the discrepancy between the models by introducing an uncertainty set believed to contain the true model. Recent research contributions in this area, such as those by Bandi et al. (2019), Sun and Van Mieghem (2019), have demonstrated the effectiveness of RO in various applications. Notably, classical RO approaches are often static and may produce overly conservative solutions, especially for problems involving immediate and wait-and-see decisions. In contrast, our formulation allows nature's perturbations to be adaptive to the system's evolution, making the uncertainty set only loosely specified, a feature that distinguishes our work from the classical RO setting.

In greater detail, we assume the demand for each product arrives according to a non-homogeneous Poisson process. However, the decision-maker does not know the true value of the arrival rate beyond knowing that it may fluctuate around some long-term average. Hence, the decision-maker treats the intensity as if it were chosen by nature. The decision-maker then calculates the expected long-run average cost, assuming that nature acts in a way that maximizes costs, creating the worst-case scenario. The decision-maker's objective is to find a joint sequencing and outsourcing control strategy to minimize this cost. The resulting formulation is a stochastic two-player zero-sum game where nature's actions are constrained as she incurs penalties for deviating from the nominal model. It is crucial to make nature's actions costly, and we demonstrate that penalizing nature is equivalent to imposing constraints directly on the magnitude of her perturbations. The severity of the penalties reflects the degree of distrust the decision-maker has towards the nominal model.

Building on a stream of studies employing similar analytical techniques (Çelik and Maglaras 2008, Rubino and Ata 2009, Ata and Barjesteh 2023), we adopt an infinite time horizon to simplify our mathematical analysis. This choice eliminates the need to account for time or time-to-go in the optimality equation, thereby reducing the complexity of the problem. While real-world production facilities may operate indefinitely, product line lifetimes are typically finite. However, the solution to a finite-horizon dynamic programming problem tends to converge towards that of the corresponding long-run average problem as the length of the time horizon increases; see, e.g., (Bertsekas 1995, section 4.3). Therefore, a solution derived from a long-run average problem likely remains relevant for cases where product line lifetimes are relatively long. Additionally, we favor an average cost criterion over a discounted cost criterion because time discounting is less common in manufacturing settings (Ormeci et al. 2008) and an average cost problem is generally easier to handle as it avoids the complexity associated with second-order differential equations.

Our robust control formulation requires specifying a statistical distance to quantify how far away a perturbed model (created by nature) is from its nominal counterpart. For this purpose, we employ Rényi divergence. For one thing, Rényi divergence provides greater flexibility in modeling compared to Kullback-Leibler (KL) divergence, as demonstrated in our numerical experiments, where it leads to better control rules compared to using KL divergence only. Moreover, the use of Rényi divergence leads to a tractable representation of the distance between demand models, simplifying analysis and computation. While other generalizations of KL divergence, such as f-divergence, could potentially be considered, they do not seem to be able to translate into a tractable representation of the model distance in the present context.

Because the original robust control problem is intractable, both analytically and computationally, we adopt the “heavy-traffic approximation” to make further headway. It implicitly assumes that demand and service capacity are both high, resulting in server utilization nearing full capacity. This suggests that the production server must operate almost continuously to meet demand. Admittedly, as we move away from this critical loading condition, solutions obtained under heavy-traffic assumptions may become less effective. However, practical considerations, such as cost constraints for manufacturers, often dictate setting capacity levels that align with demand volumes. As the paper Bradley and Glynn (2002) argues, the high cost of capacity can incentivize production firms to operate at high utilization levels, naturally creating a heavy-traffic regime. Under-provisioning capacity can lead to increased costs from rush orders or the need for outsourced production to meet demand, as well as delays in order processing that can damage a manufacturer’s reputation, whereas over-provisioning capacity may reduce profitability due to high fixed costs and resource waste. From a technical standpoint, our use of heavy-traffic approximation aligns with the well-established solution framework pioneered by Harrison (1988), where a Brownian control problem (BCP) naturally emerges as the heavy-traffic limit of a sequence of queuing control problems. In our setting, the approximation leads to a stochastic differential game (SDG), which we further transform into an equivalent one-dimensional problem, known as the workload problem. This problem uses the system’s workload as the state descriptor. The solution to the workload problem involves a control-band policy for the decision-maker and a drift-rate control for nature. Whenever the workload surpasses an upper threshold, the decision-maker promptly reduces it by outsourcing the manufacturing of a specific product, bringing it back to a lower threshold level. Between consecutive outsourcing operations, nature employs a state-dependent drift-rate control to address the decision-maker’s ambiguity aversion.

Our presentation and analysis of the heavy-traffic approximation are somewhat informal. We choose not to rigorously demonstrate that the SDG arises as the heavy-traffic limit of a sequence of robust control problems. In fact, we do not even employ a sequence of problems in developing the SDG. Moreover, we do not attempt to prove that the control rule derived from the SDG is near-optimal. However, given that much of our approximate analysis relies on results known for simpler systems, we anticipate that the solution derived from the SDG will closely approximate the “exact” solution to the original robust control problem. In our assessment, establishing a proper notion of near optimality would be an arduous task and would deviate from our two primary objectives of using an approximate analysis: (i) exploring the structure of the solution to the robust control problem, treating the SDG as an effective surrogate for the original problem in heavy traffic, and (ii) developing practical and easy-to-implement control strategies.

We summarize the paper's main contributions as follows: First, it investigates joint sequencing and outsourcing control for an MTO system based on a simplified demand model. In particular, outsourcing operations incur both fixed and proportional costs. Despite its clear focus on the control of MTO systems, to the best of our knowledge, this paper is the first to address a long-run average robust control problem with an uncertainty set loosely and adaptively chosen by nature. Second, assuming the underlying system to be critically loaded, we derive and solve an SDG that approximates the original robust control problem. Noting that a pathwise optimal solution does not exist for the decision-maker due to nature controlling the drift of the state process, we identify the solution to the game by solving a nonlinear differential equation with a set of free boundaries. In addition to the aforementioned modeling and methodological contributions, we introduce a simulation-based method for uncertainty set selection. This serves as a proof-of-concept demonstration of the practical applicability of our proposed modeling framework. This method integrates our analytical model with computer simulations to create and assess a range of closely related control rules, ultimately yielding a control rule that is deemed the most suitable for real-world implementation.

2. Literature Review

This work draws on the literature on performance optimization in queues. When arrivals follow Poisson processes and delay costs are linear, the $c\mu$ rule, which assigns static priority levels to jobs based on their $c_i\mu_i$ values, minimizes delay costs (Cox and Smith 1991). Some extensions of this rule take into account more sophisticated cost structures (Van Mieghem 1995) and/or job abandonment from the queue, including Rubino and Ata (2009), Kim and Ward (2013). Several papers have explored the combination of economic levers (e.g., pricing) with operational decisions in managing MTO systems, including Çelik and Maglaras (2008), Ata and Olsen (2013). Our paper differs from these studies in one crucial aspect: these papers assume an accurate probabilistic model for optimization, whereas we consider potential model errors that arise from model simplification, leading to a min-max optimal control problem.

This paper contributes to the literature on sequential decision-making under ambiguous beliefs, a problem category that involves a nominal model considered to be a parsimonious representation of the real-world scenario, and a malevolent agent who can create alternative models by perturbing the nominal model. The idea can be traced back to Hansen and Sargent (2001) and has been applied in various domains, such as revenue management (Lim and Shanthikumar 2007). Aside from tackling different problems, these papers consider either finite-time or infinite-time discounted criteria. In contrast, we consider a long-run average formulation, which is fundamentally different from the aforementioned problem types. Our research bears similarities to the work of Cohen (2019), which examines a differential game derived from the Brownian approximation of a multiclass M/M/1 queue under model uncertainty. However, there are key distinctions between our approaches. Cohen (2019) assume finite queue buffers whose sizes cannot be optimized. Thus, in his workload problem, a barrier-type control is implemented at a predetermined level, so that the sole control lever is job sequencing. The optimal control therein, the $c\mu$ rule, is pathwise optimal and independent of the solution to the optimality equation, rendering the choice of uncertainty set irrelevant to the control strategy. In contrast, our model incorporates outsourcing, and our derived strategy is influenced by the choice of uncertainty set since it is not pathwise optimal. Furthermore, Cohen (2019) considers an infinite-horizon discounted cost criterion, solving a second-order nonlinear differential equation with fixed boundary conditions, while we focus on a long-run average cost criterion, solving a first-order differential equation with free boundary conditions. Additionally, our use of Rényi-type divergence sets our work apart from Cohen (2019). In summary, even if we also employ KL divergence as in Cohen (2019) and eliminate fixed outsourcing costs, his approach is not applicable to our problem.

Methodologically, our paper is related to the impulse control of Brownian systems. Harrison et al. (1983) and Dai and Yao (2013b) consider discounted cost formulations, and Ormeci et al. (2008) and Dai and Yao (2013a) study average cost problems. Via explicit solutions, these studies reveal that a control band policy is optimal. Our problem cannot be solved using their methods since the drift rate is not constant and is subject to control by nature. Thus, obtaining explicit solutions is impossible. Instead, we adopt an approach that does not require explicit solutions to the value function. A few recent papers (Yao 2017, Cao and Yao 2018) have considered joint drift-rate control and impulse control for Brownian models in joint pricing and inventory control problems. Our work differs from theirs in two aspects: (i) their problems belong to the class of cost minimization problems, while ours adopts a min-max criterion; (ii) their problems deal with a single-class model, while ours inherently deals with multiclass models.

3. Nominal Model

All random quantities of interest in this section are on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ endowed with a filtration $\mathbb{F} := (\mathcal{F}_t)$ contained in $\mathcal{F}$. We focus on a single-server MTO manufacturing system that offers I different products indexed by $i = 1, \ldots, I$. Requests for product i, henceforth referred to as class i orders, arrive according to a Poisson process with rate $\bar{\lambda}_i$. We use $A_i(t)$ to denote the number of class i orders placed up to t and $\bar{\lambda} := (\bar{\lambda}_i)$ the arrival rate vector that collects all the arrival rates. The time needed to process a class i order follows an exponential distribution with rate μ_i. The mean processing time for class i orders is therefore $m_i = 1/\mu_i$. Let $S_i(t)$ denote the number of completed class i orders up to time t, assuming that the server is continuously working on class i orders. The decision-maker has discretion over the sequencing of orders but adheres to the head-of-line, or first-in-first-out, sequencing principle within each queue. Aiming to retain tractability while following an existing line of research utilizing a similar approximation procedure (Çelik and Maglaras 2008, Rubino and Ata 2009, Ata and Barjesteh 2023), we assume that switching between the production of different product types is instantaneous, so that there are no setup times between producing different products. Hence, we can conveniently describe sequencing decisions by an I-dimensional process $T := (T_i)$, whose ith component tracks the cumulative amount of time spent producing product i; accordingly, the cumulative idle time of the server up to t is $(t - \sum_i T_i(t))$. We expect that a solution obtained by ignoring them may still prove effective if setup times exist but are short relative to service times. This seems to be supported by numerical studies in Section EC.5. Therein, we also observe that longer setup times can significantly compromise solution quality, a clear indicator that disregarding setup times may overly simplify reality.

Remark 1. The use of a single-server system presents yet another possible oversimplification of an MTO production system in real life, as a real-world system likely involves multiple servers. However, since our analysis essentially follows the conventional heavy-traffic framework, we expect that an N-server system will behave roughly the same as a system with a server that works N times faster than each of the original server. This has been formally justified by Chen and Shanthikumar (1994) under simpler settings (without control) and supported by our simulation study presented in §EC.7. See also (Whitt 2002, chapter 10) for a comprehensive account of the stochastic process limits for systems with multiple but not too many servers. Lastly, we note that similar simplifications are adopted in related studies, such as Huang et al. (2015), where the authors approximate an N-physician system in heavy traffic with a system having one "super" physician, and Liu and Sun (2022), where the authors use one "super" drone to approximate an N-drone system.

The decision-maker has the option to outsource manufacturing needs at a fixed and proportional cost. Specifically, outsourcing a batch size of x class i orders incurs a cost of $\phi_i(x) := (L_i + \ell_i x) \cdot 1_{\{x>0\}}$, where L_i and ℓ_i are fixed and proportional costs, respectively. By doing this, the decision-maker is able to instantly reduce the backlog of class i orders by x units, and we will refer to this type of outsourcing activity as a type i outsourcing operation.

The ability of an outsourcing operation to instantly remove a number of pending orders from a queue without incurring additional holding costs for those orders implies that the subcontractor has a service rate that is infinitely high. In reality, there may be a lead time after the outsourcing decision due to possible delays with order placement and manufacturing times at the subcontractor. Incorporating such a lead time adds an additional layer of complexity, making the already challenging decision problem more complex. As a result, we make the simple assumption that the subcontractor has a zero lead time but will discuss a possible way to relax this assumption in §EC.8.4, drawing upon Wu and Chao (2014), Sun et al. (2024) and references therein.

We define $\Psi_i := (\tau_i(0), \tau_i(1), \ldots; \xi_i(0), \xi_i(1), \ldots)$, where $0 = \tau_i(0) < \tau_i(1) < \cdots$ is a sequence of time epochs at which a type i outsourcing operation is performed, and $\{\xi_i(k); k \geq 0\}$ is the sequence of batch sizes of the consecutive type i outsourcing operations. We can then capture outsourcing decisions by $\Psi := (\Psi_i)$. Let $N_i(t) := \sup\{k \geq 0 : \tau_i(k) \leq t\}$, so that $N_i(t)$ tracks the number of type i outsourcing operations performed up to time t. Denote by $Q_i(t)$ the number of outstanding orders of class i in the system at time t, and let $Q(t) := (Q_i(t))$. Assuming that the system initially has $Q_i(0)$ class i orders, we can describe the dynamics of $Q_i(t)$ using the equation

$$Q_i(t) = Q_i(0) + A_i(t) - S_i(T_i(t)) - \sum_{k=0}^{N_i(t)} \xi_i(k) \quad \text{for} \quad i = 1, \ldots, I. \tag{1}$$

The cost of holding a backlog of class i orders is incurred at a rate of $c_i(Q_i(t))$. Hence, the total backlog penalty is accumulated at the rate of $\sum_i c_i(Q_i(t))$. We also introduce a function that will prove useful in our analysis later on. Specifically, we define

$$h(w) := \min\left\{\sum_{i=1}^{I} c_i(x_i) : m^\top x = w, x \in \mathbb{R}_+^I\right\}. \tag{2}$$

Throughout this paper, we make the following assumption on the function h, which will hold when each $c_i(\cdot)$ is a linear or quadratic function. Such an assumption can be satisfied by most applications of practical interest (Yao et al. 2015). It is important to note that this assumption is weaker than requiring each $c_i(\cdot)$ to possess the stated properties.

Assumption 1. *The function $h(\cdot)$ is continuous, monotonically increasing towards infinity, and bounded from below by an affine function with a positive slope.*

A control policy can be represented by the pair (T, Ψ). Using the long-run average cost criterion, the decision-maker aims to find adaptive control (T, Ψ) that minimizes

$$\limsup_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\sum_{i=1}^{I}\int_0^t c_i(Q_i(u))\mathrm{d}u + \sum_{i=1}^{I}\sum_{k=0}^{N_i(t)} \phi_i(\xi_i(k))\right]. \tag{3}$$

The primary goal of this paper is to formulate and solve robust versions of (3).

4. Robust Control Problem

In reality, the demand rate for a product i may fluctuate over time, possibly in a random manner, hinting to a time-dependent arrival rate function $\lambda_i := \{\lambda_i(t); t \geq 0\}$ instead of a constant rate $\bar{\lambda}_i$. To account for this, the decision-maker is interested in devising control policies that are robust against such variations. To model this scenario, we assume that the decision-maker acts as if there is a second player (i.e., nature) who strategically chooses the vector of demand rate processes $\lambda := (\lambda_i)$. We further define a perturbation process $\theta_i := \{\theta_i(t); t \geq 0\}$ for each i, where $\theta_i(t) := (\lambda_i(t) - \bar{\lambda}_i)/\bar{\lambda}_i$ represents the relative deviation of λ_i from its nominal value $\bar{\lambda}_i$ at time t.

We require that each θ_i is locally integrable and satisfies the condition that $\theta_i(t) \in \Theta_i := [a_i, b_i]$ for all $t \geq 0$, where $-1 < a_i < 0 < b_i < \infty$.[1] Here, we require each θ_i to be strictly bounded away from -1 through the role of a_i because we need to ensure $\ln(1+\theta_i(t))$ is uniformly bounded, a condition needed, for example, in the proof of Proposition 1. We observe, however, that such a requirement should not pose significant harm since each a_i can be made arbitrarily close to -1. The imposition of the upper bound b_i is mostly for technical convenience, and its precise value is inconsequential, provided it is large enough. In fact, aside from its use in making sure that the differential equation obeyed by the relative value function in §6.3 possesses the desired global Lipschitz continuity, nowhere else do we need that upper bound.[2] Hereinafter, we refer to $\theta := (\theta_i)$ as the perturbation process. Because λ and θ determine each other, we will treat θ as nature's decision process instead of λ.

To capture the decision-maker's fear of model errors, we follow the convention established by Hansen and Sargent (2001), which allows nature to select a perturbation strategy within some prescribed limit to maximize the decision-maker's cost, effectively leading to a worst-case analysis. One major benefit of a worst-case analysis is that it provides a conservative estimate of the solution's performance. By assuming that the model input takes on its worst possible values within the specified uncertainty set, the decision-maker can design a solution that performs well even under the most unfavorable scenario. Indeed, if the decision-maker is willing to accept the solution's performance even in the most unfavorable scenario, he can be confident in accepting any other outcomes that may arise due to model simplification. Moreover, as pointed out in the influential survey by Bertsimas et al. (2011), a worst-case analysis frees the decision-maker from the need to fully specify a probability distribution governing uncertain model input while likely providing analytical and computational tractability under suitable types of uncertainty sets.

Existing literature suggests two ways to loosely specify an uncertainty set: imposing constraints on the magnitude of nature's perturbations or penalizing nature's perturbation actions based on their magnitude. While the former aligns better with the core concept of RO, the latter aids in analyzing and computing robust solutions more effectively, as one can easily write down the optimality equation. Importantly, the two approaches are often found to yield equivalent outcomes (Hansen et al.

[1] Requiring θ_i to be no less than -1 is obvious, because otherwise we would allow that $\theta_i(t) < -1$ at time t, in which case $\lambda_i(t)$ will become negative at t, not making practical sense.

[2] It may also be worth pointing out that imposing a sufficiently large upper bound on a drift rate control process to allay technical complexity is not uncommon. For instance, both Ata et al. (2005) and Çelik and Maglaras (2008) have made similar assumptions by introducing a large constant to bound the drift-rate control process; however, the value of the constant has no bearing on the solutions.

2006, section 4); see also Lim and Shanthikumar (2007). Hence, we adopt the latter approach for its computational advantages and its better compatibility with a heavy-traffic analysis, which we will carry out later on. We will, however, present in the next section an alternative formulation drawing on the former approach, which resembles classical RO settings.

To proceed, we need a measure that can be used to quantify the extent to which perturbations are made. For this purpose, we define, for each $\mathbb{F}$-predictable θ_i, the Doléans-Dade exponential:

$$\psi_i(t) := \exp\left\{-\int_0^t \bar{\lambda}_i \theta_i(u)\mathrm{d}u\right\} \prod_{0<u\leq t} (1+\theta_i(u))^{\Delta A_i(u)}, \tag{4}$$

where $\Delta A_i(t) := A_i(t) - A_i(t-)$; see Definition 9.4.3.1 in (Jeanblanc et al. 2009, §9.4.3). Note that if $\theta_i \equiv 0$, then $\psi_i \equiv 1$, implying $\lambda_i \equiv \bar{\lambda}_i$. Also, it is straightforward to verify that the process $\psi_i := \{\psi_i(t); t \geq 0\}$ is a martingale. Denote by $\mathbb{P}_i$ the marginal distribution of A_i in the nominal model, we can define a new measure $\mathbb{Q}_i$ via the Radon-Nikodym derivative:

$$\left.\frac{\mathrm{d}\mathbb{Q}_i}{\mathrm{d}\mathbb{P}_i}\right|_{\mathcal{F}_t} = \psi_i(t) \quad \text{for} \quad t \geq 0. \tag{5}$$

By the Girsanov theorem for filtered Poisson processes (Jeanblanc et al. 2009, Proposition 8.4.5.1), A_i is a filtered Poisson process with intensity $\lambda_i(t) = \bar{\lambda}_i(1+\theta_i(t))$ under the induced measure $\mathbb{Q}_i$.

Identity (5) suggests that changing the intensity of A_i from $\bar{\lambda}_i$ to λ_i is equivalent to changing the measure from $\mathbb{P}_i$ to the induced measure $\mathbb{Q}_i$. This relationship brings to the fore the induced measure $\mathbb{Q}_i$, allowing the use of the notion of distance between the measures $\mathbb{P}_i$ and $\mathbb{Q}_i$ to measure the magnitude of θ_i. Therefore, limiting the size of θ_i would correspond to controlling the distance between the two measures. In this paper, we evaluate the distance between each $(\mathbb{P}_i, \mathbb{Q}_i)$ through the *Rényi divergence*. In general, the Rényi divergence of a measure $\tilde{\mathbb{P}}$ with respect to a reference measure $\mathbb{P}$ of order $\alpha \neq 1$ can be defined as

$$\mathcal{R}^\alpha(\tilde{\mathbb{P}}\|\mathbb{P}) := \frac{1}{\alpha-1}\ln\int\left(\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}}\right)^\alpha \mathrm{d}\mathbb{P} = \frac{1}{\alpha-1}\ln\int\left(\frac{\mathrm{d}\tilde{\mathbb{P}}}{\mathrm{d}\mathbb{P}}\right)^{\alpha-1} \mathrm{d}\tilde{\mathbb{P}}.$$

When $\alpha = 1$, the Rényi divergence reduces to the KL divergence; see, e.g., Van Erven and Harremos (2014). So we may assume $\alpha \in (0, \infty)$. Let $\mathcal{R}_i^\alpha(t)$ denote the Rényi divergence of $\mathbb{Q}_i$ of order α with respect to $\mathbb{P}_i$ on $\mathcal{F}_t$. The following result links $\mathcal{R}_i^\alpha(t)$ to θ_i and its proof can be found in §EC.2.

Proposition 1. *For each fixed $t \geq 0$,*

$$\mathcal{R}_i^\alpha(t) = \frac{\bar{\lambda}_i}{\alpha-1}\int_0^t \left\{(1+\theta_i(u))^\alpha - \alpha\theta_i(u) - 1\right\}\mathrm{d}u$$

if $\alpha \neq 1$ and $\mathcal{R}_i^\alpha(t) = \bar{\lambda}_i \int_0^t \{(1+\theta_i(u))\ln(1+\theta_i(u)) - \theta_i(u)\}\,\mathrm{d}u$ if $\alpha = 1$.

After obtaining the explicit expression for $\mathcal{R}_i^\alpha(t)$, we can put each induced measure $\mathbb{Q}_i$ back into the background and treat the expression as the definition of $\mathcal{R}_i^\alpha(t)$. This allows us to focus solely on the perturbation processes θ_i, rather than the measures they induce. With this, we have all the vocabulary we need to present our robust control problem.[3] In this problem, the decision-maker seeks to find (T,Ψ) to minimize

$$\max_{\theta} \limsup_{t\to\infty} \frac{1}{t}\mathbb{E}^\theta\left[\sum_{i=1}^I \int_0^t c_i(Q_i(u))\mathrm{d}u + \sum_{i=1}^I \sum_{k=0}^{N_i(t)} \phi_i(\xi_i(k)) - \sum_{i=1}^I \gamma_i \mathcal{R}_i^\alpha(t)\right], \tag{6}$$

where we have again followed the literary convention to add the superscript θ to the expectation operator to emphasize that the arrival intensities now follow $\bar\lambda+\theta$ instead of $\bar\lambda$ in the nominal model. It is clear from this formulation that nature's actions are costly in the sense that the further away the perturbed model is from its nominal counterpart, the heavier the penalty will be for nature. Moreover, the severity of the penalty is determined by the parameters $\gamma := (\gamma_i)$. To gain more intuition into the role of γ, it is useful to treat it as a set of design parameters that a user of our modeling framework can choose depending on the decision-maker's level of trust in the nominal model. Higher the value of γ_i, a more cautious nature will be in moving λ_i away from $\bar\lambda_i$, hence the more confidence the decision-maker has in $\mathbb{P}_i$ in its ability to accurately describe the demand realization process for product i. To further demonstrate that Problem (6), in fact, aligns with the core concept of RO, we present an alternative formulation in the next section. This alternative formulation shares high-level similarities with classical RO settings, as it imposes direct constraints to restrict the magnitude of the perturbation caused by nature on the nominal model.

We stipulate that the decision-maker only considers stationary deterministic policies. Then, it suffices for nature to limit herself to only considering stationary deterministic policies because, after the decision-maker chooses a stationary and deterministic strategy, nature faces a Markov decision process (MDP) in which nature aims to optimize a long-run average objective. It is known that there exists a stationary deterministic policy that is optimal among all admissible policies.

5. An Alternative Formulation

Consider the following problem, where the decision-maker seeks to find (T,Ψ) to minimize

$$\begin{aligned} &\max_{\theta} \limsup_{t\to\infty} \frac{1}{t}\mathbb{E}^\theta\left[\sum_{i=1}^I \int_0^t c_i(Q_i(u))\mathrm{d}u + \sum_{i=1}^I \sum_{k=0}^{N_i(t)} \phi_i(\xi_i(k))\right] \\ &\text{subject to } \limsup_{t\to\infty} \frac{1}{t}\mathbb{E}^\theta\left[\mathcal{R}_i^\alpha(t)\right] \le \beta_i \quad \text{for} \quad i=1,\dots,I. \end{aligned} \tag{7}$$

[3] Note that what nature perturbs is the distribution governing demand realization rather than specific realizations of demand. As a result, our notion of "robust" really corresponds to "distributionally robust." For a formal and comprehensive account of the notion of distributional robustness and the corresponding distributionally robust optimization, we refer the reader to the influential survey by Rahimian and Mehrotra (2019).

This formulation involves I model-error constraints, where each β_i measures the degree of mistrust that the decision-maker has in the law governing class i arrivals in the nominal model. A larger value of β_i allows λ_i to deviate more from $\bar{\lambda}_i$, whereas a smaller value of β_i forces λ_i to remain close to its nominal value. Thus, as with γ in (6), we can consider $\beta = (\beta_i)$ as a set of tuning parameters that the decision-maker can use to enhance the robustness of his control strategy.

It is instructive to view Problem (6) as the "Lagrange relaxation" of Problem (7), where parameters $\gamma = (\gamma_i)$ serve as the Lagrange multipliers associated with the I model-error constraints in (7), and we will refer to Problems (6) and (7) as the *penalty problem* and the *constraint problem*, respectively. The next natural question to ask is whether a Lagrange multiplier theorem exists between these two formulations. This turns out to be true, albeit with additional regularity conditions.

As preparation, let $C_c^\star(\beta)$ and $C_p^\star(\gamma)$ denote the optimal values of the constraint and penalty problems, respectively. Proposition 2, stated below, establishes a formal connection between $C_c^\star(\beta)$ and $C_p^\star(\gamma)$. The result suggests that considering and solving the penalty problem is meaningful.

Proposition 2. *Suppose the decision-maker restricts attention to stationary deterministic policies that prevent queues from growing unbounded, and nature is allowed to choose from the class of stationary policies (deterministic and randomized). For a fixed β, if $\gamma^\star \succcurlyeq 0$ minimizes $C_p^\star(\gamma) + \langle \beta, \gamma \rangle$, then $C_c^\star(\beta) = C_p^\star(\gamma^\star) + \langle \beta, \gamma^\star \rangle$. Furthermore, for the penalty problem, $C_p^\star(\gamma^\star)$ can be achieved when both players adopt stationary deterministic policies.*

The proposition allows nature to select from stationary policies essentially because once the decision-maker picks a strategy, nature faces a constrained Markov decision process (CMDP). In this CMDP, nature seeks to optimize a long-run average objective while adhering to specific constraints, which admits an optimal stationary policy among all the admissible policies. While mostly driven by technical considerations, the requirement for queues not to grow unbounded under the decision-maker's control is intuitive. When deciding whether to outsource the production of a product, one must weigh the cost of holding orders in the queue against the cost of outsourcing some of those orders. If there is no limit on queue length, then the holding cost can increase substantially as the queue grows. Consequently, there comes a point where holding orders is no longer cost-effective, making outsourcing a more attractive option. Extensive numerical analyses support this conclusion, as demonstrated in Figures 2 and EC.3. The Lagrange multiplier theorems for the infinite-horizon discounted problem and the finite-time problem were respectively established by Hansen et al. (2006) and Lim and Shanthikumar (2007). However, we cannot directly apply their proofs since we are dealing with a long-run average problem, which differs fundamentally from the settings in their works. Instead, we utilize the "convex analytic method," a technique first introduced in Bhatnagar and Borkar (1995) and further developed in Altman (1999). We provide the detailed proof of Proposition 2 in §EC.1.

6. Heavy-Traffic Analysis

In light of Proposition 1, we define

$$r(\theta) := r^{\alpha}(\theta) := \sum_{i=1}^{I} \frac{\gamma_i \bar{\lambda}_i}{\alpha - 1} \left\{(1+\theta_i)^{\alpha} - \alpha\theta_i - 1\right\}$$

for $\alpha \neq 1$ and $r(\theta) := r^{\alpha}(\theta) := \sum_{i=1}^{I} \gamma_i \bar{\lambda}_i \left\{(1+\theta_i)\ln(1+\theta_i) - \theta_i\right\}$ for $\alpha = 1$. This allows us to express the penalty problem more explicitly, yielding one where the decision-maker aims to minimize

$$\max_{\theta} \limsup_{t\to\infty} \frac{1}{t} \mathbb{E}^{\theta} \left[\sum_{i=1}^{I} \int_0^t c_i(Q_i(u))\mathrm{d}u + \sum_{i=1}^{I} \sum_{k=0}^{N_i(t)} \phi_i(\xi_i(k)) - \int_0^t r(\theta(u))\mathrm{d}u \right]. \tag{8}$$

Our development of the SDG will be based on Problem (8).

6.1. SDG

The operating regime we focus on is the one where both the demand volume and production capacity are large and the capacity balances the supply and demand. To be more specific, we impose the following "critical-loading" assumption:

$$\sum_{i=1}^{I} \rho_i = 1 \quad \text{for} \quad \rho_i := \bar{\lambda}_i m_i, \quad i = 1, \ldots, I. \tag{9}$$

Because the server's long-run proportion of time spent on producing product i is ρ_i, the system can be thought of as critically loaded if the nominal model is correct. Assume that nature employs θ_i to generate the demand rate for product i. Assuming optimistically that $\theta_i\bar{\lambda}_i$ is an order of magnitude smaller than $\bar{\lambda}_i$, we approximate A_i using

$$A_i(t) = \bar{\lambda}_i t + \bar{\lambda}_i \int_0^t \theta_i(u)\mathrm{d}u + \hat{A}_i(t) + \epsilon_i^a(t), \tag{10}$$

where $\hat{A}_i$ is a Brownian motion with zero drift and variance parameter $\bar{\lambda}_i$ and ϵ_i^a is an approximation error term. Define, for each i, the centered time allocation process as $Y_i(t) := \rho_i t - T_i(t)$. Note that $\sum_i Y_i(t)$ tracks the cumulative idleness up to time t. Similarly, we approximate $S_i \circ T_i$ using

$$S_i(T_i(t)) = \mu_i T_i(t) + \hat{S}_i(t) + \epsilon_i^d(t) = \bar{\lambda}_i t - \mu_i Y_i(t) + \hat{S}_i(t) + \epsilon_i^d(t), \tag{11}$$

where $\hat{S}_i$ is a driftless Brownian motion with variance parameter $\bar{\lambda}_i$ and ϵ_i^d is an approximation error term. Plugging (10) and (11) into (1), ignoring the approximation error terms, and replacing Q_i, Y_i, and ξ_i with their respective approximations, denoted as $\hat{Q}_i, \hat{Y}_i$ and $\hat{\xi}_i$, we obtain

$$\hat{Q}_i(t) = \hat{Q}_i(0) + \hat{Z}_i(t) + \int_0^t \bar{\lambda}_i \theta_i(u)\mathrm{d}u + \mu_i \hat{Y}_i(t) - \sum_{k=0}^{N_i(t)} \hat{\xi}_i(k), \quad i = 1, \ldots, I, \tag{12}$$

$$\hat{Q}_i(t) \geq 0 \quad \text{for} \quad t \geq 0, \quad i = 1, \ldots, I, \text{ and} \tag{13}$$

$$\sum_i \hat{Y}_i(t) \text{ is non-decreasing with } \sum_i \hat{Y}_i(0) = 0. \tag{14}$$

where $\hat{Z}_i$ are independent Brownian motions with drift zero and infinitesimal variance $\sigma_i^2 = 2\bar{\lambda}_i$. Denote by $\hat{\Psi}_i$ the approximating type i outsourcing control, i.e.,

$$\hat{\Psi}_i := (\tau_i(0), \tau_i(1), \tau_i(2), \ldots, \tau_i(m), \ldots; \hat{\xi}_i(0), \hat{\xi}_i(1), \hat{\xi}_i(2), \ldots, \hat{\xi}_i(m), \ldots).$$

By writing $\hat{Y} := (\hat{Y}_i)$ and $\hat{\Psi} := (\hat{\Psi}_i)$, we can formally state the decision-maker's problem as one that seeks an adapted control $(\hat{Y}, \hat{\Psi})$ that minimizes

$$\max_{\theta} \limsup_{t\to\infty} \frac{1}{t}\mathbb{E}^{\theta}\left[\int_0^t \left(\sum_{i=1}^{I} c_i(\hat{Q}_i(u)) - r(\theta(u))\right) \mathrm{d}u + \sum_{i=1}^{I}\sum_{k=0}^{N_i(t)} \phi_i(\hat{\xi}_i(k))\right] \tag{15}$$

subject to constraints (12) − (14).

6.2. Dimensional Reduction

Although the SDG is simpler than the original problem it approximates, its solution is not as simple due to the high dimensionality of the state process $\hat{Q} := (\hat{Q}_i)$. For this reason, we seek further simplification, leading to a one-dimensional differential game referred to as the *workload problem.*

To start, we define the one-dimensional workload process W as follows:

$$W(t) := \sum_{i=1}^{I} m_i \hat{Q}_i(t), \quad t \geq 0, \tag{16}$$

which serves as an approximation for the amount of work in the system at time t. As preliminaries to the derivation of the workload problem, we define

$$B(t) := \sum_{i=1}^{I} m_i \hat{Z}_i(t), \quad \zeta(t) := \sum_{i=1}^{I} \rho_i \theta_i(t), \quad U(t) := \sum_{i=1}^{I} \hat{Y}_i(t), \quad \text{and} \quad O(t) := \sum_{i=1}^{I}\sum_{k=0}^{N_i(t)} \tilde{\xi}_i(k), \tag{17}$$

where

$$\tilde{\xi}_i(k) := m_i \hat{\xi}_i(k) \quad \text{for} \quad k \geq 0 \quad \text{and} \quad i = 1 \ldots, I. \tag{18}$$

Intuitively, $\zeta := \{\zeta(t); t \geq 0\}$ corresponds to perturbations nature makes to the workload; $U(t)$ approximates the cumulative idle time up to t, and $O(t)$ approximates the cumulative amount of work outsourced up to t. Multiplying (12) by m_i and summing over i plus utilizing the definitions in (17) and (18), we obtain the system equation for the workload process:

$$W(t) = W(0) + B(t) + \int_0^t \zeta(u)\mathrm{d}u + U(t) - O(t).$$

Recall the function h defined in (2). Note that the function can be naturally interpreted as the effective holding cost rate in the following sense: Given that the total workload w can be instantaneously redistributed across all classes in any way the decision-maker desires, the amount of work will be distributed in such a way that the aggregate holding cost rate is minimized. Similarly, we can define nature's "cost rate function" as

$$r^{\star}(z) := \min\left\{r(y) : \rho^{\top} y = z, y_i \in \Theta_i\right\}, \tag{19}$$

where $z \in [\rho^\top a, \rho^\top b]$. Associated with I different types of outsourcing operations, there are I different outsourcing cost functions, corresponding to I different ways to push down the workload to a desired level. For type i outsourcing operations, we define

$$\tilde{\phi}_i(w) := (L_i + \tilde{\ell}_i w) \cdot 1_{\{w>0\}} \quad \text{for} \quad \tilde{\ell}_i := \ell_i / m_i.$$

We can interpret $\tilde{\ell}_i$ as the proportional cost of outsourcing one unit of work through type i outsourcing operations, and we denote by $\tilde{\Psi}$ the outsourcing rule for the workload process. Slightly overloading the notation (W, B, ζ, U and $\tilde{\xi}_i(k)$ were all previously defined), we can spell out the workload problem, where the decision-maker seeks some adaptive control $(U, \tilde{\Psi})$ to minimize

$$\max_{\zeta} \limsup_{t\to\infty} \frac{1}{t} \mathbb{E}^{\zeta} \left[\int_0^t h(W(u))\mathrm{d}u - \int_0^t r^\star(\zeta(u))\mathrm{d}u + \sum_{i=1}^{I} \sum_{k=0}^{N_i(t)} \tilde{\phi}_i(\tilde{\xi}_i(k)) \right] \tag{20}$$

$$\text{s.t. } W(t) = W(0) + B(t) + \int_0^t \zeta(u)\mathrm{d}u + U(t) - O(t), \tag{21}$$

$$U(t) \text{ is non-decreasing with } U(0) = 0, \text{ and} \tag{22}$$

$$W(t) \geq 0 \quad \text{for} \quad t \geq 0. \tag{23}$$

Here, the superscript ζ in the expectation operator means that there is a state-dependent perturbation ζ to the drift of the underlying process and B is a zero-drift Brownian motion with infinitesimal variance $\sigma^2 = \sum_i m_i^2 \sigma_i^2$. The next result shows the equivalence of the SDG and the workload problem. The corresponding proof can be found in §EC.2.

Proposition 3. *The SDG and the workload problem are equivalent in the following sense: For every admissible control* $(\hat{Y}, \hat{\Psi})$ *for* (12)–(15)*, there exists an admissible control for* (20)–(23) *with at least as good performance; and for every admissible control* $(U, \tilde{\Psi})$ *for* (20)–(23)*, there exists an admissible control for* (12)–(15) *with the same performance.*

6.3. Characterization of the Optimal Solution

Because nature's control only depends on the current workload, we will now write $\zeta(W(t))$ instead of $\zeta(t)$. For the decision-maker, this means that an outsourcing rule would be in the form of a control limit, which we will briefly describe below. It is evident that a deviation from the work-conserving principle can only hurt the decision-maker, so the idleness process U ought to satisfy

$$\int_0^t 1_{\{W(u)>0\}} \mathrm{d}U(u) = 0, \quad t \geq 0.$$

6.3.1. Control Band Policy Following Harrison et al. (1983), we define a relevant class of control rules as follows.

Definition 1. Given some $i \in \{1, \ldots, I\}$ and two parameters q, s with $0 < q < s$, we call (i, q, s) a control band policy of type i with parameters (q, s), if the decision-maker utilizes type i outsourcing operations only, and upon W reaching the upper barrier s, the decision-maker enforces a downward jump to level q, thereby incurring a cost of $\tilde{\phi}_i(s-q)$.

Now, for an arbitrarily given real-valued function $\zeta(\cdot)$, define the differential operator Γ_ζ as $\Gamma_\zeta f(w) = \frac{1}{2}\sigma^2 f''(w) + \zeta(w) f'(w)$. For a fixed $s > 0$, let $\mathcal{C}^2[0, s]$ denote the space of functions that are twice differentiable up to the boundaries. Suppose that there exists some $\eta \in \mathbb{R}$ and $f \in \mathcal{C}^2[0, s]$ collectively satisfying

$$\Gamma_\zeta f(w) + h(w) - r^\star(\zeta(w)) = \eta \quad \text{for} \quad w \in (0, s), \tag{24}$$

subject to the boundary conditions

$$f'(0) = 0 \quad \text{and} \quad f(s) = \tilde{\phi}_i(s-q) + f(q). \tag{25}$$

The proposition presented below motivates the optimality equation to be described in the subsequent subsection. The proof of this identity is a routine application of Itô's formula, which we omit here.

Proposition 4. *Suppose $\eta \in \mathbb{R}$ and $f \in \mathcal{C}^2[0, s]$ jointly satisfy* (24) *and* (25). *Then η is the long-run average cost under the control band policy (i, q, s) and the drift-rate control $\zeta(\cdot)$.*

6.3.2. Optimality Equation The analysis proceeds in three steps. First, using the boundary and smooth pasting conditions while taking nature's strategic behavior into account, we identify a control band policy, denoted as (i, q_i, s_i), that mini-maximizes [4] the long-run average cost within the class of controls, utilizing type i outsourcing operations only. We denote by η_i the resulting long-run average cost of this strategy. Second, we define the candidate solution to the decision-maker's decision problem as the one yielding the lowest long-run average cost under the specified minimax criterion. More formally, we select $i^\star$ so that $\eta_{i^\star} \le \eta_i$ for all $i \ne i^\star$, with the control band policy $(i^\star, q_{i^\star}, s_{i^\star})$ viewed as a potential solution to the decision-maker's problem. Third, by exploiting the structural properties of the value function associated with the control band policy $(i^\star, q_{i^\star}, s_{i^\star})$, we demonstrate that this strategy is indeed average cost optimal for the decision-maker under the *minimax* criterion among all adaptive controls that the decision-maker can take.

Proposition 4 motivates the following optimality equation that facilitates the identification of the control band policy (i, q_i, s_i) as mentioned earlier: find $q_i, s_i, \eta_i \in \mathbb{R}$ and $v \in \mathcal{C}^2[0, s_i]$ such that

$$\max_\zeta \left\{ \frac{1}{2}\sigma^2 v''(w) + \zeta v'(w) + h(w) - r^\star(\zeta) \right\} = \eta_i, \quad w \in (0, s_i), \tag{26}$$

[4] Here, "mini-maximizes" refers to the decision-maker selecting the specific policy to minimize his long-run average cost, while nature employs a drift rate control ζ to maximize the decision-maker's cost.

subject to the boundary conditions

$$v'(0)=0, \quad v(s_i)=\tilde{\phi}_i(s_i-q_i)+v(q_i), \quad \text{and} \quad v(w)=\tilde{\ell}_i(w-s_i)+v(s_i) \quad \text{for} \quad w\geq s_i, \tag{27}$$

plus a set of optimality conditions stemmed from the "principle of smooth fit": $v'(q_i)=v'(s_i)=\tilde{\ell}_i$. Letting $g(x)$ denote the convex conjugate of the function $r^\star(\zeta)$, i.e.,

$$g(x):=\max_{\zeta}\{x\zeta-r^\star(\zeta)\} \quad \text{for} \quad x\in\mathbb{R},$$

we can rewrite (26) as

$$\frac{1}{2}\sigma^2 v''(w)+g(v'(w))+h(w)=\eta_i, \quad w\in(0,s_i). \tag{28}$$

Because (28) does not involve the unknown function v, it is in essence a first-order differential equation. This motivates us to consider the class of functions $\{\pi(\cdot,\eta);\eta\in\mathbb{R}\}$, where $\pi(\cdot,\eta)$ solves

$$\frac{1}{2}\sigma^2\pi_w(w,\eta)+g(\pi(w,\eta))+h(w)-\eta=0, \tag{29}$$

subject to the boundary condition

$$\pi(0,\eta)=0. \tag{30}$$

The parameter pair (q_i,s_i) and the average cost η_i are determined through conditions:

$$\pi(q_i,\eta_i)=\pi(s_i,\eta_i)=\tilde{\ell}_i \quad \text{and} \quad \int_{q_i}^{s_i}\pi(w,\eta_i)\mathrm{d}w=\tilde{\phi}_i(s_i-q_i). \tag{31}$$

The expressions in (30) and (31) effectively make up four constraints. However, two questions remain unanswered. First, does the system of equations given in (29)–(31) yield a solution (i.e., suffice to pin down the four unknowns, q_i,s_i,η_i and $\pi(\cdot,\eta_i)$)? Second, given that the answer to the first question is yes, does the control band policy (i,q_i,s_i) yield the lowest cost possible if the decision-maker were to choose to outsource the manufacturing needs of product i only? Our next result gives positive answers to these questions.

Proposition 5. *Under Assumption 1 holds, the following statements are true. (i) The requirements in* (29)–(31) *uniquely determine q_i, s_i, and η_i. (ii) If we define $v(\cdot)$ such that its first-order derivative is equal to $\pi(\cdot,\eta_i)$ on the interval $[0,s_i)$, and $v(w)=v(s_i)+(w-s_i)\tilde{\ell}_i$ for $w\geq s_i$, then the pair (v,η_i) satisfies the following quasi-variational inequality:*

$$\min\left\{\frac{1}{2}\sigma^2 v''(w)+g(v'(w))+h(w)-\eta_i, \inf_{0\leq z\leq w}\left[v(w-z)+\tilde{\phi}_i(z)\right]-v(w)\right\}=0. \tag{32}$$

(iii) Independent of the initial condition, the control band policy (i,q_i,s_i) mini-maximizes the long-run average cost among the class of adaptive controls only utilizing type i outsourcing operations.

The detailed proof of Proposition 5 is given in §EC.2. In studying a joint impulse and drift rate control problem, Cao and Yao (2018) obtain an optimality equation similar to ours. However, there is a key difference: Cao and Yao (2018) consider impulse control on both sides, introducing an additional degree of freedom related to the value of $v'(0)$. In our case, we focus on "one-sided" impulse control, where the left boundary of $v'(0)$ remains fixed at zero.

Proposition 5 allows us to find a triple (q_i, s_i, η_i) for each fixed i, leading to the optimal control rule when only type i outsourcing operations are available. Notably, there exists a winning type $i^\star \in \{1, \ldots, I\}$ such that $\eta_{i^\star} \leq \eta_i$ for all $i \neq i^\star$. Intuitively, the resulting control band policy $(i^\star, q_{i^\star}, s_{i^\star})$ minimizes the long-run average cost among all adaptive controls that utilize only one type of outsourcing operation. However, a natural question arises: Does this rule also minimize the long-run average cost among *all* possible adaptive controls available to the decision-maker? To investigate, we aim to find a function v that is twice differentiable almost everywhere, has a bounded and continuous first-order derivative, and satisfies the Bellman-Isaacs condition:[5]

$$\min\left\{\frac{1}{2}\sigma^2 v''(w) + g(v'(w)) + h(w) - \eta_{i^\star}, \min_i \inf_{0\leq z\leq w}\left[v(w-z) + \tilde{\phi}_i(z)\right] - v(w)\right\} \geq 0 \tag{33}$$

with $v'(0) = 0$. The subsequent Theorem 1 establishes that if such a function v exists, the control band policy $(i^\star, q_{i^\star}, s_{i^\star})$ is indeed optimal.

THEOREM 1. *If some function v, which is twice differentiable almost everywhere and has a bounded first-order derivative, satisfies* (33) *with $v'(0) = 0$, then the control band policy $(i^\star, q_{i^\star}, s_{i^\star})$ is average cost optimal under the minimax criterion among all adaptive controls.*

To put Theorem 1 to good use, we need to demonstrate the feasibility of constructing a function v that satisfies all the specified properties. In order to achieve this, we let $\tilde{\ell}_* := \min_i \tilde{\ell}_i$ and define $(s_*, \tilde{\ell}_*)$ as the coordinates of the point where the function graph of $\pi(\cdot, \eta_{i^\star})$ intersects the horizontal line $\tilde{\ell}_*$ for the *second* time. By making these preparations, we can now state Proposition 6, which guarantees the existence of a function v with the desired properties.

PROPOSITION 6. *Let v be such that its first-order derivative is equal to $\pi(\cdot, \eta_{i^\star})$ on the interval $[0, s_*)$, and $v(w) = v(s_*) + (w - s_*)\tilde{\ell}_*$ for $w \geq s_*$. Then v satisfies* (33).

The detailed proofs of Theorem 1 and Proposition 6 can be found §EC.2. We will utilize these results to develop practical policy recommendations in §6.4. In closing, we remark that when fixed costs L_i are all zero, outsourcing controls become singular, and one can read off the cheapest product to outsource when the backlog of work is excessive. This simplified problem resembles one in Cohen

[5] For a dynamic programming equation used to characterize the solution to a differential game, the norm is to refer to it as the Hamilton-Jacobi-Bellman-Isaacs equation or the Bellman-Isaacs equation for short (Pham and Zhang 2014); the name "Bellman-Isaacs condition" has been used in Hansen and Sargent (2001) and Hansen et al. (2006).

(2019), where a right boundary plays a role similar to ours. However, our $s^\star$ is a policy parameter that requires optimization, unlike Cohen's predetermined boundary associated with job rejection. Also note that when L_i equals L for all i, which is typically the case when all outsourcing goes to the same contract manufacturer, the class i with the smallest $\tilde{\ell}_i$ becomes the cheapest to outsource.

6.4. Policy Recommendations

To propose an implementable control rule based on the solution to the workload problem, we take both α and γ as given and fixed and calculate $(i^\star, q_{i^\star}, s_{i^\star})$ and $\pi(\cdot, \eta_{i^\star})$. If $q_{i^\star}$ and $s_{i^\star}$ are not integers, we round them to the nearest integers. A procedure regarding how to select γ for a given α in a practical setting efficiently will be the focus of Section 8. We slightly overload the notation $W(t)$ (W was previously defined to denote the approximate workload process) to let it now denote the workload of the actual system at time t. It follows that $W(t) = \sum_i m_i Q_i(t)$, where $Q_i(t)$ is the actual number of class i orders awaiting processing at t. The proposed control has two components, as described below.

Outsourcing. Whenever the workload reaches the upper barrier $s_{i^\star}$, outsource $o_{i^\star} := (s_{i^\star} - q_{i^\star})/m_{i^\star}$ orders of product $i^\star$ immediately, if there are enough $o_{i^\star}$ orders of product $i^\star$ awaiting processing. If the number of outstanding orders of class $i^\star$, say $Q_{i^\star}$, is less than $o_{i^\star}$, then postpone the outsourcing operation until additional $(o_{i^\star} - Q_{i^\star})$ orders of product $i^\star$ arrive. We mention that postponement can be accomplished by creating a virtual queue to hold current orders of class $i^\star$ and routing new orders of this class to this virtual queue until the queue length reaches $o_{i^\star}$, at which point an outsourcing operation is performed to deplete the virtual queue. As a result, from the time those orders are moved to the virtual queue until the outsourcing operation occurs, the $i^\star$th queue is considered empty.

Sequencing. We stipulate that there is a unique solution $(x_i^\star)$ to the optimization problem (2) for each fixed w.[6] We can regard the solution as a function of w serving as the target length of the queues when the workload is at position w. Thus, a desired sequencing rule ought to be one that tries to maintain the actual queue lengths at their respective targets, $(x_i^\star)$. When the waiting cost rates are linear, i.e., $c_i(x) = C_i x$ for some constant $C_i > 0$, we can recover the well-known $c\mu$ priority rule. When all $c_i(\cdot)$ are strictly convex and satisfy $c_i(0) = c_i'(0) = 0$, it leads to the generalized $c\mu$ rule, which states that service priority is given to the job class whose $c_i'(Q_i(t))\mu_i$ index is the largest at time t.

7. Discussion

We comment on some key aspects of the modeling framework proposed in this paper.

[6] This stipulation merely seeks to mitigate the potential technical complexity and is satisfied in various settings.

7.1. Connections to Harrison's Framework

We now briefly explain how to scale different model primitives if we were to directly adopt Harrison's framework, which involves a sequence of problems indexed by a scaling parameter, denoted as n. In our context, one can think of the scaling parameter n as the sum of the nominal arrival rates $\bar{\lambda}_i$, growing to infinity to create a sequence of models. To start, each $\bar{\lambda}_i$ should grow proportionally to n; i.e., $\bar{\lambda}_i = \iota_i n$ for some ι_i. To maintain condition (9), we need each service rate μ_i to grow proportionally to n. We also need to scale the cost parameters associated with outsourcing operations and the cost rate functions c_i and r. For the decision-maker, the crux is the trade-off between waiting costs and outsourcing costs. To ensure this trade-off is nontrivial, these two costs need to be of the same order of magnitude. We can achieve this by leaving the fixed outsourcing costs unscaled but scaling down the proportional outsourcing costs by a factor of $\sqrt{n}$. Moreover, we let the waiting cost rate function in the n-th system, denoted as c_i^n, be scaled so that $c_i^n(\cdot) = \hat{c}(\cdot/\sqrt{n})$, where $\hat{c}$ is a baseline function independent of the scaling parameter n, and nature's penalty rate function in the n-th system, denoted as r^n, be scaled in such a way that $r^n(\cdot) = \hat{r}(\sqrt{n}\cdot)$, where $\hat{r}$ is a baseline function independent of the scaling parameter. We can accomplish this by letting $\gamma_i^n = \hat{\gamma}_i/n$ for every i, with each $\hat{\gamma}_i$ being a value independent of the scaling parameter n, so that we can write $\hat{r}$ as

$$\hat{r}(x) = \sum_{i=1}^{I} \frac{\hat{\gamma}_i \iota_i}{\alpha - 1}\{(1 + x_i)^\alpha - \alpha x_i - 1\}$$

for $\alpha \neq 1$ and $\hat{r}(x) = \sum_{i=1}^{I} \hat{\gamma}_i \iota_i \{(1 + x_i)\ln(1 + x_i) - x_i\}$ for $\alpha = 1$.

The scaling condition imposed on the penalty rate function implies that nature's perturbations are moderate, with an order of $1/\sqrt{n}$. This, in turn, implies that the system will remain critically loaded for a sufficiently large value of n, due to (9) and that nature's perturbations are moderate. Then standard heavy-traffic approximation theory suggests that the queue lengths will be of the order of $\sqrt{n}$, commensurate with the scaling condition imposed on the waiting cost rate functions. Moreover, according to the dynamic equation (1), the outsourcing batch size in the n-th system, denoted as ξ_i^n, should be of the same order of magnitude as the queue length. This means that the outsourcing batch size will also be around $\sqrt{n}$, compatible with the scaling condition imposed on the proportional outsourcing costs.

7.2. About the Proposed Control Rule

The derivation of the SDG assumes a critically loaded system with high demand and service rates. Under this assumption, the robust control problem simplifies significantly, reducing the state space dimension from $I + 1$ to one.[7] This enables our solution to decompose over two time scales.

[7] The original system comprises $I + 1$ dimensions: I dimensions representing state processes $\hat{Q}$, alongside an additional dimension dedicated to tracking the current class of product in service. Our sequencing strategy relies on the current class in service, following a non-preemptive service rule.

On the faster time scale, we optimize the distribution of workload across classes based on the current total workload. Whereas instantaneous redistribution is not practical (unlike in the SDG), adjustments to queue lengths can be made relatively quickly, at a rate of approximately n (if we were to borrow the terminology from §7.1). Specifically, if queue lengths fall short of targets, the corresponding classes temporarily lose server access, causing queues to only receive inflow from demand arrivals. As demand arrives at a rate of order n, significantly higher than the queue lengths of order $\sqrt{n}$, actual queue lengths rapidly return to targets within a time length of order $1/\sqrt{n}$. By the same token, classes whose actual queue lengths exceed the targets gain exclusive access to the total service capacity, thereby seeing "an underloaded system" due to the critical-loading condition (9) and the fact that nature's distortions are moderate. In particular, as the net outflow (the actual outflow from these queues minus the inflow) is of order n for these classes, their queue lengths will also quickly return to their targets, again within a time length of order $1/\sqrt{n}$. On the slower time scale, where the total workload evolves, we solve an impulse control problem to determine when and to what extent to reduce the workload, giving rise to the proposed outsourcing rule. Although we allow postponing an outsourcing operation until a sufficient number of class $i^\star$ orders are received, the delay caused by the postponement is expected to be short if the threshold is small. Larger thresholds may result in longer delays, potentially defeating the purpose of reducing an excessive workload, as orders in the virtual queue will continue to accrue holding costs. However, we mention that in heavy traffic, long delays are rare, even with large thresholds. To see this, note that the control parameters $q_{i^\star}$ and $s_{i^\star}$ are supposedly of order $1/\sqrt{n}$, as they are used to regulate the workload, which is also of order $1/\sqrt{n}$, and since $m_{i^\star}$ is of order $1/n$, $o_{i^\star}$ should be of order $\sqrt{n}$. Because the virtual queue faces a demand rate of roughly $\bar{\lambda}_{i^\star}$, which is of order n, the virtual queue will be filled up relatively quickly to the desired target $o_{i^\star}$ at a rate of order $\sqrt{n}$.

One caveat is that accumulating enough outsourceable orders may be slow if the cheapest class to outsource has very low demand. In such cases, this class is deemed to have a "thin arrival," as described in Ata (2006)—the usual assumption that the demand rate for each class is of order n does not hold. Thus, when formulating the SDG and the corresponding workload problem, the role of this class would be limited to contributing to the drift rate in the approximation (Ata 2006), so that this class is not considered a "legitimate class" for outsourcing, potentially resolving the issue.

7.3. Selecting the Uncertainty Set

The main motivation behind adopting a robust control formulation lies in striking a balance between tractability and practicality, a trade-off that hinges on the choice of the underlying uncertainty set. If the set is too restrictive, it may fail to include the true model; conversely, an excessively large set can yield overly conservative solutions. In our context, "model uncertainty" arises because of deliberate model simplification. We thus proceed with the assumption that the true model governing demand realization is known but deemed too complex.

Drawing from our earlier discussions, two potential routes naturally arise for selecting an uncertainty set. The first route can begin with the constraint problem, where one determines an appropriate β by assessing the dissimilarity between the true and nominal models. One can then utilize Proposition 2 to derive a corresponding γ. This route, however, faces two obstacles: (i) Computing the aforementioned dissimilarity may be difficult, especially when the true demand model is very complex. (ii) It requires establishing a functional relationship between β and γ, a task that can be computationally demanding, especially when I is large. The second route is to determine a "good" γ directly. What we will propose in the next section follows this route and draws inspiration from a set of parallels between our penalty problem and regularized regression (RR), a popular model in machine learning. RR has the ability to reduce the sensitivity of linear regression, an effect that can be explained from a robustness perspective (Bertsimas et al. 2011). Indeed, the regularization coefficient thereof, akin to γ in our penalty problem, serves as the dual variable of a corresponding *robust linear regression* problem (Xu et al. 2008). Thus, adjusting the regularization coefficient in RR is tantamount to adjusting robustness levels. The typical approach for determining this coefficient is cross-validation, whose underlying mechanism serves as a direct motivation for our proposed method to be illustrated next.

8. A Simulation-based Method for Uncertainty Set Selection

We describe our method for selecting the parameter γ with α fixed. We note, however, that jointly optimizing α and γ is possible and may provide further performance improvement. In §EC.10, we discuss this issue further, explaining why we choose to downgrade the significance of α and focus on optimizing over γ. We first explain the main idea and then provide an algorithm to operationalize it.

Although the true demand model may be complex, one can often find a simplified one to serve as the nominal model in the formulation of a robust control problem. With each fixed vector γ, we can solve the corresponding SDG to obtain a robust control rule, denoted as $\mathcal{P}(\gamma)$, a step akin to computing all the regression coefficients in an RR model with a given regularization coefficient, resulting in a "predictive model." As mentioned previously, one can use cross-validation to "optimize" the choice of regularization coefficient. Roughly speaking, by splitting the original dataset into training and testing sets, this approach can estimate the performance of each predictive model, and the "optimal" regularization coefficient is the one for which the resulting predictive model exhibits the best estimated performance. The role of the dataset used for solving an RR model parallels that of the true demand model. Assume optimistically that combining the control rule $\mathcal{P}(\gamma)$ with the true demand model results in a long-run average cost, denoted as $\mathcal{C}(\gamma)$.[8] Note that $\mathcal{C}(\gamma)$ serves as a performance indicator for $\mathcal{P}(\gamma)$. Thus, it is natural to pick $\mathcal{P}(\gamma)$ that minimizes $\mathcal{C}(\gamma)$.

If the expression for $\mathcal{C}(\gamma)$ is known, a gradient-descent algorithm is likely to work to locate a minimizer. However, obtaining such an expression may prove difficult. On the other hand, one may be able to evaluate the function value of $\mathcal{C}$ via computer simulations, where we simulate demand realizations based on the true model. The "best" γ is the one that yields the lowest estimated cost from the simulation. This motivates our proposed algorithm, which is close in spirit to a standard gradient-descent algorithm but substitutes a finite difference for gradient (Spall 2005, chapter 6). Algorithm 1, which we refer to as the "quasi-gradient-descent" algorithm, illustrates how to find cost-minimizing γ for a fixed α. In the algorithm, κ represents the learning rate, and $\gamma^{(j)} := \{\gamma_1^{(j)}, \ldots, \gamma_I^{(j)}\}$ represents the γ value at the end of the jth outer iteration. $\hat{\mathcal{C}}(\gamma)$ denotes the simulation estimate of $\mathcal{C}(\gamma)$. By selecting a small $\delta \in \mathbb{R}_+$, we approximate the ith component of the gradient of $\mathcal{C}$ at the point $\gamma^{(j)}$ using

$$\nabla \mathcal{C}_i(\gamma^{(j)}) := \frac{\hat{\mathcal{C}}\left(\gamma^{(j)} + \delta e_i\right) - \hat{\mathcal{C}}\left(\gamma^{(j)} - \delta e_i\right)}{2\delta},$$

where e_i denotes the unit vector whose ith component is one and the remaining components are zero. We present two 2-class examples in §9.2 and a 4-class example in §EC.6 that demonstrate the efficiency of our simulation-based method.

Because the algorithm for finding the minimizing γ relies on simulations, the warm-up period, often defined as the initial simulation period needed for the system to reach a steady state, becomes a crucial parameter that needs to be chosen carefully. The longer the warm-up period, the more likely it is that the simulated system will be in or approaching a steady state. However, a too long warm-up period consumes a significant portion of the total simulation time. In our numerical studies, we determine the appropriate warm-up period through a naive "sensitivity analysis." This involves comparing statistical estimates of the desired output variable (long-run average cost) under two candidate warm-up periods, T_1 and T_2, where $T_1 < T_2$. If there is no noticeable difference between the estimates, we adopt T_2 as the warm-up period. We also explore more sophisticated approaches for the warm-up period selection in §EC.9. The number of replications, which determines how many times the simulation model is executed, is another important parameter requiring judicious selection. More replications improve accuracy but also increase computational time. One approach to determining the appropriate number of replications is by measuring the variability of the output variable. After each

[8] Here, we take an optimistic view because, without a formal specification of the true demand model, there is no guarantee that the demand model, along with the control derived from the SDG with the penalty parameter γ, has a well-defined long-run average cost. This contrasts with settings where the demand model is explicitly defined, such as Poisson or some Markov process. In those cases, it becomes possible to formulate the problem as an MDP for which established theory exists, including regularity conditions that can ensure the well-posedness of the problem.

Algorithm 1 Quasi-Gradient-Descent Algorithm

initialize: κ, $\gamma^{(0)}$, $\gamma^{(1)}$,ϵ
while $\gamma^{(j+1)} \neq \gamma^{(j)}$ **do**
  $\tilde{\gamma}^{(j)} \leftarrow \gamma^{(j)}$
  for $i = 1 : I$ **do**
    while $\left|\nabla \mathcal{C}_i(\tilde{\gamma}^{(j)})\right| > \epsilon$ **do**
      $V^{(j)} \leftarrow \tilde{\gamma}^{(j)}$
      $\tilde{\gamma}_i^{(j)} \leftarrow \max\left\{V^{(j)} - \kappa \nabla \mathcal{C}_i(V^{(j)}), 0.02\right\}$ ▷ Maximum ensures positiveness.
      $\kappa \leftarrow \frac{\left|\left(V_i^{(j)} - \tilde{\gamma}_i^{(j)}\right)\left(\nabla \mathcal{C}_i(V^{(j)}) - \nabla \mathcal{C}_i(\tilde{\gamma}^{(j)})\right)\right|}{\left(\nabla \mathcal{C}_i(V^{(j)}) - \nabla \mathcal{C}_i(\tilde{\gamma}^{(j)})\right)^2}$
    end while
  end for
  $\gamma^{(j+1)} \leftarrow \{\tilde{\gamma}_1^{(j)}, \ldots, \tilde{\gamma}_I^{(j)}\}$
end while

replication, the confidence interval of the desired output variable can be computed, and the iteration can be stopped when the width of the confidence interval falls below a predetermined threshold chosen to be small enough to provide sufficient accuracy.

The proposed procedure thus consists of two main components: an analytical component and a simulation-based evaluation component. The analytical component, developed through Sections 4–6, involves a nominal model, an uncertainty set, and an efficient approach for computing candidate control rules. The simulation-based evaluation component assesses the performance of these candidate control rules in a high-fidelity environment and aims to identify the rule with the lowest long-run average cost based on simulation outputs. Compared to a pure simulation-based optimization method, our proposed method offers two key advantages: (i) The analytical component provides a clear set of control rules as input for the simulation-based evaluation, giving the simulation step a specific objective, whereas a direct simulation-based optimization may lack a clear focus on candidate policies. (ii) The control rules generated by the analytical component possess interpretable structural properties, providing insights into the underlying mechanisms that are hard to obtain from a pure simulation-based approach.

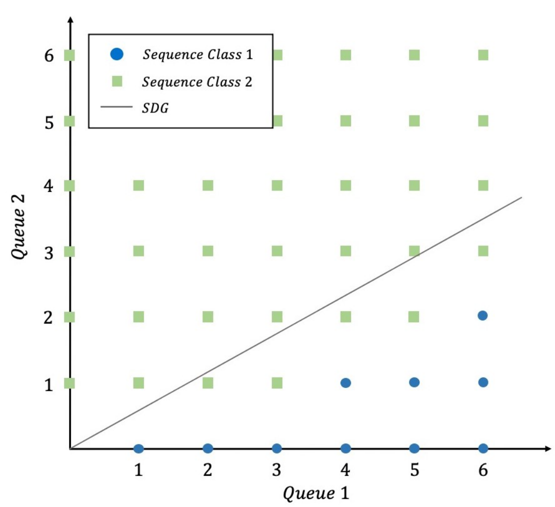
9. Numerical Studies

This section presents various numerical studies. Throughout this section, by mentioning the solution to the SDG, we mean the solution to the workload problem. In §9.1, we compare the solution obtained from the original penalty problem with that from the SDG. In §9.2, we present examples exposing the "value of robustness." In §9.3, we compare the optimal cost attained using the actual demand model (assuming the actual demand model is known to the decision-maker) with the cost attained by the best policy identified.

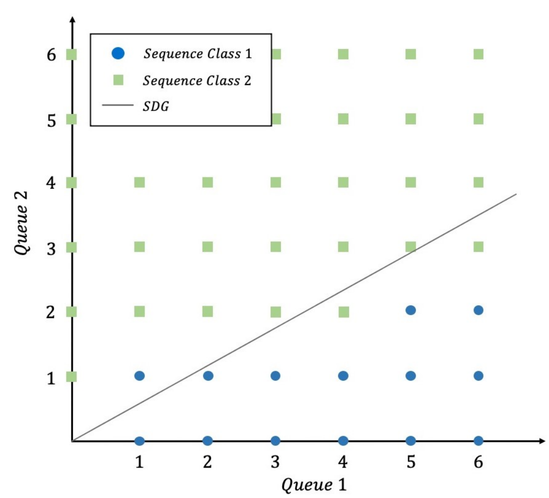
9.1. Solution Comparison Between the Penalty Problem and the SDG

In this subsection, we numerically solve both the original penalty problem and the corresponding SDG and compare their solutions to demonstrate the accuracy and reliability of the heavy-traffic approximation. For the SDG, we compute its solution based on (26)–(27). We provide implementation details in §EC.3. For the original penalty problem, whose state-descriptor can be found in §EC.1, we adopt the "strategy iteration" approach. In each iteration, we fix the strategy of one player and compute the best response of the opponent in the resulting one-player game, resulting in an MDP, which we solve via standard value iteration. We then improve the strategy we originally fixed based on the response of the opponent. As an initialization, we solve the nominal model, which assumes that all arrival rates are exactly equal to their nominal values, hence no participation from nature, allowing us to obtain a strategy for the decision-maker to start with. While we do not intend to establish a formal theory guaranteeing the convergence of the strategy iteration algorithm, our experiments with numerous instances suggest that the algorithm shows quick convergence.

We consider a two-class model with the following parameters: $\bar{\lambda}_1 = 30$, $\bar{\lambda}_2 = 40$, $\mu_1 = 60$, and $\mu_2 = 80$. The cost data includes fixed outsourcing costs $L_1 = 5$ and $L_2 = 15$, proportional outsourcing costs $\ell_1 = \ell_2 = 1.0$, and quadratic holding cost rates $a_1 = 0.4$ and $a_2 = 0.5$. Figure 1 illustrates the optimal sequencing rules under Rényi divergence with $\alpha = 1/2$. For the original problem, the decision-maker would prioritize class 1 (resp. class 2) when the system is in states represented by a blue circle (resp. a green square). For the SDG, the decision-maker would prioritize class 2 (resp. class 1) if the state is above (resp. below) the gray line. The result shows that the sequencing rule derived from the SDG is reasonably close to that obtained from the original penalty problem.

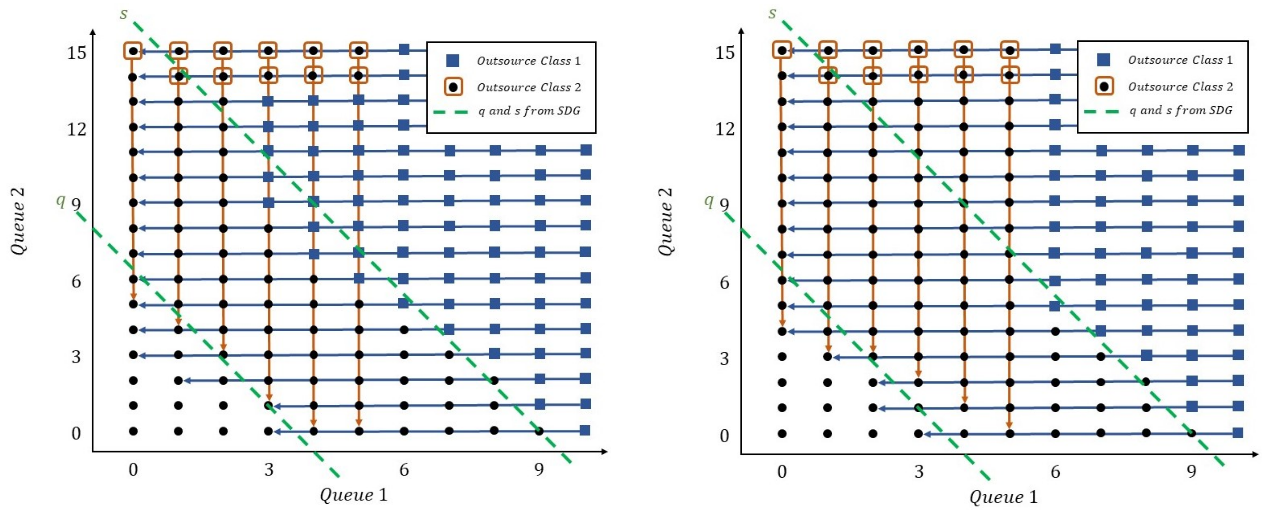


(a) Class 1 is in service



(b) Class 2 is in service

Figure 1 **Sequencing strategies derived from the penalty problem and the SDG with $\gamma_1 = \gamma_2 = 30$ when $\alpha = 1/2$**



(a) Class 1 is in service   (b) Class 2 is in service

Figure 2 Outsourcing strategies derived from the penalty problem and the SDG with $\gamma_1 = \gamma_2 = 30$ when $\alpha = 1/2$

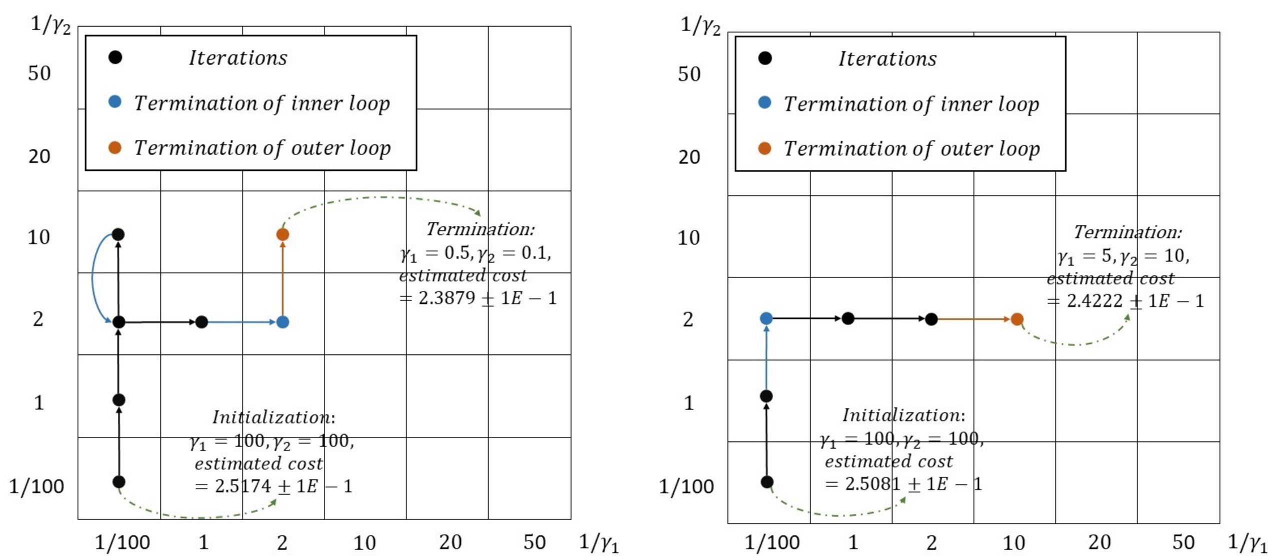
Results from Figure 2 pertain to outsourcing. For the original penalty problem, the blue squares represent the states where product 1 should be outsourced, whereas the orange squares represent the states where product 2 should be outsourced. Each horizontal or vertical arrow in the figures connects the "origin" (the state right before outsourcing) and the "destination" (the state right after outsourcing) associated with an outsourcing operation. The decision-maker is inclined to outsource orders of class 1, unless queue 2 becomes exceedingly large, which is in alignment with the control strategy prescribed by the SDG. The green lines represent the outsourcing thresholds, q_1 and s_1, derived from the solution to the SDG and indicate that the decision-maker should always outsource product 1. Hence, we can conclude that the outsourcing rule derived from the SDG is reasonably close to that obtained from the penalty problem.

In §EC.4.1, we present an additional example involving KL divergence ($\alpha = 1$). Therein, we also compare the long-run average cost generated by the original penalty problem (η_{exact}) and that from the SDG (η_{SDG}) under various parameter settings. Those results show that the long-run average costs derived from the SDG are very close to the values attained from the original penalty problem, further validating the reliability of our approximation scheme.

9.2. Exposing the Value of Robustness

Here, we showcase the importance of incorporating robustness into decision-making. In §9.2.1, we analyze a system where the demand rate of each product follows an auto-regressive integrated moving average (ARIMA) model. In §9.2.2, we consider a system where the demand rate of each product follows a continuous-time Markov chain (CTMC). Simulations are conducted over a time interval of $2,000$ for the stochastic system. In order to estimate the performance of the system, we test the initial warm-up periods with $T_1 = 50$ and $T_2 = 100$. After comparing the statistical estimates of the long-run average simulated cost, we adopt $T_2 = 100$ as the warm-up period. In §EC.9, we also provide evidence demonstrating that setting the warm-up period to 100 is sufficiently long to guarantee reliable performance. In addition, to determine the appropriate number of replications, we set the desired width of the confidence interval to 0.2. After conducting experiments, we observed that each example requires approximately 40 replications to achieve the desired width of the confidence interval, with an average simulation time per replication of less than 1 second.

9.2.1. ARIMA Intensity We demonstrate the simulated average cost $\hat{\mathcal{C}}(\gamma)$, assuming a real-world demand model in which the demand rate of each product follows a non-homogeneous Poisson process with an ARIMA model for intensity. We study a two-class MTO system with the following parameters: $\bar{\lambda}_1 = 80$, $\bar{\lambda}_2 = 60$, $\mu_1 = 100$, and $\mu_2 = 300$. The cost data includes outsourcing costs $L_1 = 0.5, L_2 = 0.8, \ell_1 = \ell_2 = 0.2$, and quadratic holding cost rates $a_1 = 0.01$ and $a_2 = 0.02$. In addition, for both products, we set the two "ARLags" to 0.8 and -0.8 respectively, the MA error coefficient to 0.4 and the variance to 100.



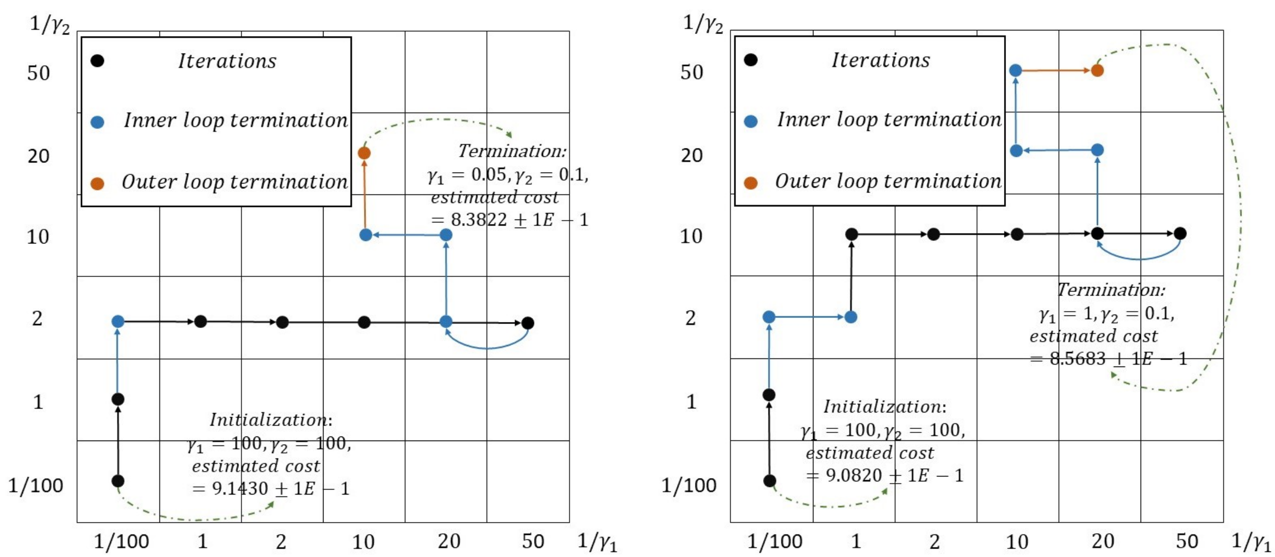
(a) $\alpha = 3$ (b) $\alpha = 1$

Figure 3 Quasi-gradient-descent algorithm for simulations with ARIMA intensity

The quadratic holding costs induce a queue-ratio type sequencing rule: if $a_1\mu_1 Q_1 > a_2\mu_2 Q_2$, priority should be given to producing product 1. Conversely, if the inequality is reversed, priority should be given to producing product 2. For the outsourcing rule, we find that in this example, we should always outsource product 1 because $\eta_1 < \eta_2$ for all pairs of (γ_1, γ_2). Details of optimal control band parameters (q_1, s_1) for different pairs of (γ_1, γ_2) can be found in Table EC.2. Figure 3 depicts the results of our algorithm, with $\hat{\mathcal{C}}(\gamma)$ presented under both Rényi ($\alpha=3$) and KL divergence. The black arrows indicate each step taken during the inner iterations, while the blue arrows correspond to the final step of each inner iteration. The orange arrow denotes the final step of the whole algorithm, while the termination point is marked with an orange circle, indicating the optimal value of γ and the associated simulated cost. It is worth noting that when γ_1 and γ_2 are very large (thus, $1/\gamma_1$ and $1/\gamma_2$ are very small), the corresponding control rule can be viewed as one that completely ignores possible model errors. This implies that the difference between the "limiting value" of each plot and the minimum value on that plot can be seen as the value of robustness. For instance, for $\alpha=3$, the value of robustness is approximately $2.5174-2.3879=0.1295$, which is around 5.42% better than completely ignoring model errors. Similarly, for $\alpha=1$, the value of robustness is approximately 0.0859, which is roughly 3.54% better than ignoring model errors. The comparison between the two values of robustness suggests that a flexible α can lead to cost savings. Table EC.3 in §EC.4.2 displays confidence intervals derived from simulations of the point estimates in this example. In general, we find that our quasi-gradient-descent algorithm is capable of locating the minimum value. This minimum value closely corresponds to the lowest cost presented in Table EC.3.

9.2.2. CTMC Intensity We now explore an arrival model in which the demand rate for each order adheres to a non-homogeneous Poisson process, with the intensity represented by a CTMC. Specifically, for each product, the arrival intensity is modeled as a CTMC with two distinct states, $\tilde{\lambda}_i$ and $\hat{\lambda}_i$. The demand rate $\tilde{\lambda}_i$ corresponds to an optimistic scenario, whereas $\hat{\lambda}_i$ represents a pessimistic scenario. The nominal demand $\bar{\lambda}_i$ is computed as the weighted sum of the two demand rates. The system parameters are the same as described in §9.2.1. In the CTMC, each state's sojourn time is exponentially distributed at a rate of 10. We further set $\tilde{\lambda}_1=150$ and $\tilde{\lambda}_2=110$, while $\hat{\lambda}_1=\hat{\lambda}_2=10$. Unlike in §9.2.1, we purposely amplify the fluctuations of real-world demand rates beyond what heavy-traffic scaling would entail. This deliberate choice aims to showcase the versatility of our framework in different operating regimes. Although we could have set the parameters to yield moderate fluctuations, thereby adhering closely to heavy-traffic scaling requirements, as will be demonstrated numerically in §9.3, our proposed approach remains effective even when confronted with substantial fluctuations in real-world demand rates.

In this example, the optimal decision rules (sequencing and outsourcing) are also the same as in the previous example. The path taken by the algorithm to locate the minimum value is illustrated in Figure 4. The value of robustness for $\alpha=3$ is approximately 0.7608, which represents a 9.07% improvement over fully ignoring model errors. For $\alpha=1$, the value of robustness is approximately 0.5137, which is approximately 5.99% better than ignoring model errors. Again, these results demonstrate that the flexibility offered by α can result in additional cost savings compared to using KL divergence alone. Table EC.4 provides confidence intervals obtained from all computer simulations.



(a) $\alpha = 3$

(b) $\alpha = 1$

Figure 4 **Quasi-gradient-descent algorithm for simulations with CTMC intensity**

The main takeaway from Figures 3 and 4 is that employing a robust control formulation that accounts for model errors can result in significant cost savings. In the presence of model simplification, following the policy based on the nominal model, thereby completely ignoring model errors, can impair system performance. On the other hand, a too large uncertainty set overemphasizes the worst-case scenario, producing an excessively conservative solution that does not have good performance either. An appropriate choice of the uncertainty set can safeguard against model errors without generating overly conservative solutions.

9.3. Cost Comparison Between Actual and Robust Models

We next compare the cost of the "best" robust control policy, denoted by $\hat{\mathcal{C}}(\gamma^\star)$, with the actual optimal cost $\mathcal{C}^\star$ obtained by conducting optimization based on the true model. For the true model, we use a model setting similar to the one outlined in §9.2.2 (CTMC intensity). To determine the optimal cost $\mathcal{C}^\star$, we must solve an MDP, and we do it through value iteration. However, solving this problem with value iteration is computationally intensive, as the two-class example results in a five-dimensional MDP, including the two queue lengths, the index of the class being served, and the current arrival rates of the two classes. Therefore, we opt to address a single-class problem instead.

We set $\bar{\lambda} = \mu = 100$ and consider the following cost parameters: outsourcing parameters $L = 0.5, \ell = 0.1$, and quadratic holding cost rate $a = 0.01$. In the CTMC, each state's sojourn time is exponentially distributed with a rate of 2.5. Additionally, we set $\tilde{\lambda} = 150$ and $\hat{\lambda} = 50$. In Figure 5, the black straight line at the bottom denotes the actual optimal cost $\mathcal{C}^\star$, obtained by solving the true model through value iteration. In fact, in this example, $\mathcal{C}^\star = \eta_{exact} = 3.717$. The blue curve illustrates the long-run

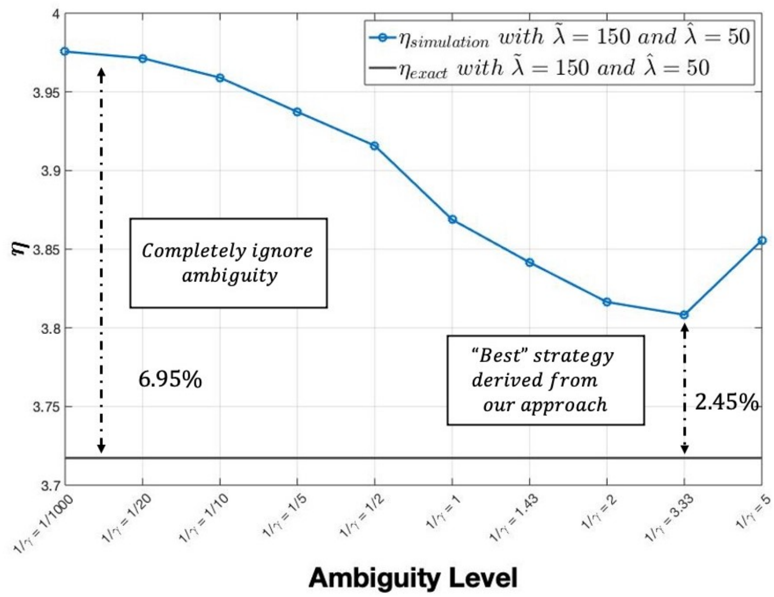


Figure 5 **Actual cost and cost obtained from the "best" robust model with** $\tilde{\lambda} = 150$ **and** $\hat{\lambda} = 50$

average robust cost $\hat{\mathcal{C}}(\gamma)$ for varying γ. Remarkably, the difference between $\hat{\mathcal{C}}(\gamma^\star)$ and $\mathcal{C}^\star$ is only 2.45%. By ignoring model errors, the decision-maker would incur a cost that is 6.95% higher than $\mathcal{C}^\star$. Therefore, our robust control formulation, accompanied by the parameter-determining method, delivers a noticeable value. More detailed information about the outsourcing thresholds (q, s) and the simulated costs in the above example can be found in §EC.4.3.

10. Concluding Remarks

We have considered robust control for a multiclass MTO manufacturing system with an outsourcing mechanism. Our main formulation takes the form of a two-player zero-sum game in which nature, as a malevolent second player, serves to guard against the adverse impact of deliberate model simplification. Considering the system to operate in a suitable operating regime, we have devised an approximating problem that admits a one-dimensional simplification. Combining the decision-maker's strategy from solving the game with simulation programs enables us to develop a practical procedure capable of generating effective joint sequencing and outsourcing rules.

There are several limitations to this study. First, our solution approach is intended for critically loaded systems. Therefore, the analysis, results, and insights primarily apply to systems that fulfill the critical-loading assumption. Analyzing and exploring solutions for underloaded (resp. overloaded) systems, where nominal demand significantly falls short (resp. exceeds) capacity, would be valuable. Second, we do not consider the possibility of learning to update or refine the uncertainty set based on new information as the system evolves over time. Combining online learning with a robust control framework can be valuable, as exemplified by Kim (2016). Third, it is unclear to what extent our solution framework remains effective for more complex systems that are potentially difficult to simulate. For example, simulating a system with many potentially heterogeneous servers would likely pose practical challenges compared to the system considered in this paper. Further investigation into this matter would be worthwhile.

There are multiple avenues for future research. For example, it would be interesting to explore how our approach to achieving "model simplifications" fares in comparison to alternative methods, such as the "state aggregation" method developed by Nadar et al. (2018) to optimize assembly-to-order systems. It would deliver important practical values if we were to actually implement the proposed sequencing and outsourcing policies in a natural test bed to see if they are able to improve upon current practices adopted by MTO manufacturing systems.

References

Altman E (1999) *Constrained Markov Decision Processes: Stochastic Modeling* (Routledge).

Ata B (2006) Dynamic control of a multiclass queue with thin arrival streams. *Operations Research* 54(5):876–892.

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

Ata B, Harrison J, Shepp L (2005) Drift rate control of a brownian processing system. *The Annals of Applied Probability* 15(2):1145–1160.

Ata B, Olsen TL (2013) Congestion-based leadtime quotation and pricing for revenue maximization with heterogeneous customers. *Queueing Systems* 73(1):35–78.

Bandi C, Trichakis N, Vayanos P (2019) Robust multiclass queuing theory for wait time estimation in resource allocation systems. *Management Science* 65(1):152–187.

Bertsekas D (1995) *Dynamic Programming and Optimal Control: Volume 2* (Athena Scientific).

Bertsimas D, Brown DB, Caramanis C (2011) Theory and applications of robust optimization. *SIAM review* 53(3):464–501.

Bhatnagar S, Borkar VS (1995) A convex analytic framework for ergodic control of semi-Markov processes. *Mathematics of Operations Research* 20(4):923–936.

Bradley JR, Glynn PW (2002) Managing capacity and inventory jointly in manufacturing systems. *Management Science* 48(2):273–288.

Cao P, Yao D (2018) Optimal drift rate control and impulse control for a stochastic inventory/production system. *SIAM Journal on Control and Optimization* 56(3):1856–1883.

Çelik S, Maglaras C (2008) Dynamic pricing and lead-time quotation for a multiclass make-to-order queue. *Management Science* 54(6):1132–1146.

Chen H, Shanthikumar JG (1994) Fluid limits and diffusion approximations for networks of multi-server queues in heavy traffic. *Discrete Event Dynamic Systems* 4:269–291.

Chen L, Cui Y, Lee HL (2021) Retailing with 3d printing. *Production and Operations Management* 30(7):1986–2007.

Cohen A (2019) Brownian control problems for a multiclass $M/M/1$ queueing problem with model uncertainty. *Mathematics of Operations Research* 44(2):739–766.

Cox DR, Smith W (1991) *Queues*, volume 2 (CRC Press).

Dai J, Yao D (2013a) Brownian inventory models with convex holding cost, part 1: Average-optimal controls. *Stochastic Systems* 3(2):442–499.

Dai J, Yao D (2013b) Brownian inventory models with convex holding cost, part 2: Discount-optimal controls. *Stochastic Systems* 3(2):500–573.

David G, SCHAEFFER C, John W (2018) *Ordinary Differential Equations: Basics and Beyond* (Springer).

Feinberg EA, Shwartz A (2002) *Handbook of Markov decision processes: methods and applications*, volume 40 (Kluwer Academic Publishers).

Hansen LP, Sargent TJ (2001) Robust control and model uncertainty. *American Economic Review* 91(2):60–66.

Hansen LP, Sargent TJ, Turmuhambetova G, Williams N (2006) Robust control and model misspecification. *Journal of Economic Theory* 128(1):45–90.

Harrison JM (1988) Brownian models of queueing networks with heterogeneous customer populations. *Stochastic Differential Systems, Stochastic Control Theory and Applications*, 147–186 (Springer).

Harrison JM (2013) *Brownian models of performance and control* (Cambridge University Press).

Harrison JM, Sellke TM, Taylor AJ (1983) Impulse control of brownian motion. *Mathematics of Operations Research* 8(3):454–466.

Huang J, Carmeli B, Mandelbaum A (2015) Control of patient flow in emergency departments, or multiclass queues with deadlines and feedback. *Operations Research* 63(4):892–908.

Jeanblanc M, Yor M, Chesney M (2009) *Mathematical Methods for Financial Markets* (Springer Science & Business Media).

Kim J, Ward AR (2013) Dynamic scheduling of a $GI/GI/1+GI$ queue with multiple customer classes. *Queueing Systems* 75(2-4):339–384.

Kim MJ (2016) Robust control of partially observable failing systems. *Operations Research* 64(4):999–1014.

Lim AE, Shanthikumar JG (2007) Relative entropy, exponential utility, and robust dynamic pricing. *Operations Research* 55(2):198–214.

Liu W, Sun X (2022) Energy-aware and delay-sensitive management of a drone delivery system. *Manufacturing & Service Operations Management* 24(3):1294–1310.

Nadar E, Akcay A, Akan M, Scheller-Wolf A (2018) The benefits of state aggregation with extreme-point weighting for assemble-to-order systems. *Operations Research* 66(4):1040–1057.

Ormeci M, Dai J, Vate JV (2008) Impulse control of Brownian motion: the constrained average cost case. *Operations Research* 56(3):618–629.

Pham T, Zhang J (2014) Two person zero-sum game in weak formulation and path dependent bellman–isaacs equation. *SIAM Journal on Control and Optimization* 52(4):2090–2121.

Rahimian H, Mehrotra S (2019) Distributionally robust optimization: A review. *arXiv preprint arXiv:1908.05659* .

Robinson S, Ioannou A (2007) The problem of the initial transient: Techniques for estimating the warm-up period for discrete-event simulation models. *Warwick Business School, University of Warwick, Coventry, UK* 1–30.

Rubino M, Ata B (2009) Dynamic control of a make-to-order, parallel-server system with cancellations. *Operations Research* 57(1):94–108.

Spall JC (2005) *Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control*, volume 65 (John Wiley & Sons).

Sun J, Van Mieghem JA (2019) Robust dual sourcing inventory management: Optimality of capped dual index policies and smoothing. *Manufacturing & Service Operations Management* 21(4):912–931.

Sun X, Chai S, Paul AA, Zhu L (2024) Enhancing make-to-order manufacturing agility: When flexible capacity meets dynamic pricing. *Production and Operations Management (forthcoming)* .

Van Erven T, Harremos P (2014) Rényi divergence and kullback-leibler divergence. *IEEE Transactions on Information Theory* 60(7):3797–3820.

Van Mieghem JA (1995) Dynamic scheduling with convex delay costs: The generalized $c\mu$ rule. *The Annals of Applied Probability* 809–833.

Whitt W (2002) *Stochastic-process limits: an introduction to stochastic-process limits and their application to queues* (Springer).

Wu J, Chao X (2014) Optimal control of a brownian production/inventory system with average cost criterion. *Mathematics of Operations Research* 39(1):163–189.

Xu H, Caramanis C, Mannor S (2008) Robust regression and lasso. *Advances in neural information processing systems* 21.

Yao D (2017) Joint pricing and inventory control for a stochastic inventory system with brownian motion demand. *IISE Transactions* 49(12):1101–1111.

Yao D, Chao X, Wu J (2015) Optimal control policy for a brownian inventory system with concave ordering cost. *Journal of Applied Probability* 52(4):909–925.

E-Companion

The e-companion is organized as follows: In §EC.1, we provide the proof of Proposition 2. In §EC.2, we prove other main results. In §EC.3, we describe our numerical scheme used for solving the optimality equation. We supply additional data on the numerical experiments in §EC.4 and present the numerical results concerning positive setup times in §EC.5, along with results for a 4-class example in §EC.6. In §EC.7, we present numerical results focusing on multi-server systems. In §EC.8, we present a few direct extensions to the model introduced in the main paper, including treatments for slightly overloaded or underloaded systems. The discussion regarding the selection of warm-up periods is provided in §EC.9. We offer some high-level guidance on the choice of the value of α in §EC.10 and provide proofs for some auxiliary results in §EC.11.

EC.1. Proof of Proposition 2

Throughout the proof, we will treat λ rather than θ as nature's decision process. This is done without loss of generality since the two quantities determine each other. Also, note that nature has a bounded action space given as $\Lambda := \{\lambda \in \mathbb{R}^I : \lambda_i \in [\bar{\lambda}_i(1+a_i), \bar{\lambda}_i(1+b_i)]\}$. The proof involves three major steps. Step 1 takes the decision-maker's strategy as given and reformulates nature's decision problem as a discrete-time CMDP. Step 2 establishes a Lagrange multiplier theorem for the CMDP. Step 3 establishes the desired connection between the constraint problem and the penalty problem.

Step 1: Let the decision-maker's strategy be given and fixed. If λ is fixed, the system evolution can be described by the stochastic process $X(t) := (Q(t), J(t))$, where $Q(t)$ is the vector of queue length processes and $J(t)$ is a process taking values from $\{0, 1, \ldots, I\}$, with $J(t) = j$ if a class j job is in service at time t for $j \neq 0$ and $J(t) = 0$ if the server is idle at time t. Note that the decision-maker's strategy not only determines the state space of X, denoted as $\mathcal{S}$, but also allows for the division of $\mathcal{S}$ into two sets: $\tilde{\mathcal{S}}$, which includes all the states that do not trigger outsourcing, and $\bar{\mathcal{S}}$, which includes all the states that will trigger outsourcing. Clearly,

$$\tilde{\mathcal{S}} \cap \bar{\mathcal{S}} = \emptyset \quad \text{and} \quad \tilde{\mathcal{S}} \cup \bar{\mathcal{S}} = \mathcal{S}.$$

Moreover, for each $x \in \bar{\mathcal{S}}$, there is a pair $(\tilde{k}(x), \tilde{\delta}(x))$ stating that whenever X reaches the state x, outsource $\tilde{\delta}(x)$ units of product $\tilde{k}(x)$. If denoting by $e_i \in \mathbb{R}^I$ the unit vector whose ith component is one and remaining components zero, then for each $x := (q, j) \in \tilde{\mathcal{S}}$, we can define two sets:

$$\chi_0(x) := \{i : (q + e_i, j) \notin \bar{\mathcal{S}}\} \quad \text{and} \quad \chi_1(x) := \{i : (q + e_i, j) \in \bar{\mathcal{S}}\}.$$

Intuitively, $\chi_0(x)$ collects the indices of job classes for which a new arrival will not trigger outsourcing, whereas $\chi_1(x)$ gathers the indices of job classes for which a new arrival will trigger outsourcing, given that the current system state is x. Henceforth, we will simply write χ_0 and χ_1 in place of $\chi_0(x)$ and $\chi_1(x)$, respectively, whenever the dependence on x is clear from the context.

Using the standard uniformization technique, we can construct a discrete-time equivalent of the continuous-time process X, denoted as $X(n) = (Q(n), J(n))$. It is worth noting that since outsourcing happens instantaneously, the state space of $\{X(n)\}$ is effectively $\tilde{\mathcal{S}}$. To spell out the transition law of this discrete-time, suppose $X(n) = (q, j)$ for $j \neq 0$. Then (i) with probability $\frac{\lambda_i(n)}{\nu}$, $X(n+1) = (q+e_i, j)$ for $i \in \chi_0$ and $X(n+1) = (q+e_i-\tilde{\delta}(q+e_i,j)e_{\tilde{k}(q+e_i,j)}, j)$ for $i \in \chi_1$; (ii) with probability $\frac{\mu_j}{\nu}$, $X(n+1) = (q-e_j, \tilde{j}(q-e_j))$, where $\tilde{j}$ is determined by the specific sequencing rule chosen by the decision-maker; and (iii) with probability $1 - \frac{\mu_j + \sum_{i=1}^I \lambda_i(n)}{\nu}$, $X(n+1) = (q, j)$. In the above, the constant ν can be chosen arbitrarily as long as it is large enough to make the aforementioned probabilities well-defined; the existence of such a ν is ensured by the boundedness of nature's action space. Also, keep in mind that $\lambda_i(n)$ are nature's decision variables at stage n.

It is straightforward to check that the discrete-time system satisfies the so-called Weak Accessibility condition (Definition 4.2.2 in Bertsekas (1995)). Thus, if ignoring the model-error constraints for now, the Bellman equation characterizing nature's best response admits the following form (Proposition 4.2.3. in Bertsekas (1995)):

$$
\begin{aligned}
\frac{\eta^*}{\nu} + \phi^*(q,j) = \max_{\lambda\in\Lambda} \Bigg[& c(q) + \sum_{i\in\chi_0} \frac{\lambda_i}{\nu}\phi^*(q+e_i,j) + \sum_{i\in\chi_1}\frac{\lambda_i}{\nu}\phi^*\left(q+e_i-\tilde{\delta}(q+e_i,j)e_{\tilde{k}(q+e_i,j)}, j\right) \\
& + \frac{\mu_j}{\nu}\phi^*(q-e_j,\tilde{j}(q-e_j))1_{\{j\neq 0\}} + \left(1 - \frac{\mu_j 1_{\{j\neq 0\}} + \sum_{i=1}^I \lambda_i}{\nu}\right)\phi^*(q,j) \\
& + \sum_{i\in\chi_1}\frac{\lambda_i}{\nu}\left[L_{\tilde{k}(q+e_i,j)} + \ell_{\tilde{k}(q+e_i,j)}\tilde{\delta}(q+e_i,j)\right]\Bigg] \quad \text{for all} \quad (q,j)\in\tilde{\mathcal{S}},
\end{aligned}
$$

where $c(q) := \frac{1}{\nu}\sum_i c_i(q_i)$. The Bellman equation implies that the outsourcing cost can be absorbed into the unit cost, yielding an effective unit cost function:

$$
\tilde{c}(x,\lambda) := c(q) + \sum_{i\in\chi_1}\frac{\lambda_i}{\nu}\left[L_{\tilde{k}(q+e_i,j)} + \ell_{\tilde{k}(q+e_i,j)}\tilde{\delta}(q+e_i,j)\right],
$$

where $x = (q, j)$. Therefore, nature's problem can be cast into a CMDP that seeks λ to maximize

$$
C_{ea}(X,\lambda) := \limsup_{m\to\infty}\frac{1}{m}\mathbb{E}\left[\sum_{n=1}^m \tilde{c}(X(n),\lambda(n))\right]
$$

subject to

$$
D_i(\lambda) := \limsup_{m\to\infty}\frac{1}{m}\mathbb{E}\left[\sum_{n=1}^m d_i^\alpha(\lambda(n))\right] \leq \beta_i, \quad i = 1,\ldots,I, \tag{EC.1}
$$

where

$$
d_i^\alpha(\lambda) := \begin{cases} \frac{\bar{\lambda}_i}{\nu(\alpha-1)}\left\{\left(\frac{\lambda_i}{\bar{\lambda}_i}\right)^\alpha - \alpha\left(\frac{\lambda_i-\bar{\lambda}_i}{\bar{\lambda}_i}\right) - 1\right\} & \text{for} \quad \alpha\neq 1, \\ \frac{1}{\nu}\left\{\lambda_i\ln\left(\frac{\lambda_i}{\bar{\lambda}_i}\right) - \lambda_i + \bar{\lambda}_i\right\} & \text{for} \quad \alpha = 1. \end{cases}
$$

Hereinafter, we will refer to $\mathcal{D}$ as the set containing all *admissible* λ satisfying equation (EC.1). We will also use Φ_S and Φ_D to represent the set of stationary λ and the set of stationary deterministic λ.

Step 2: Recall that β and γ are I-dimensional vectors of real numbers. Now, write $D(\lambda) := (D_i(\lambda))$. Our main task is to establish

$$\max_{\lambda\in\mathcal{D}} C_{ea}(X,\lambda) = \max_{\lambda\in\Phi_S}\min_{\gamma\succcurlyeq 0} J^{\gamma}_{ea}(X,\lambda) = \min_{\gamma\succcurlyeq 0}\max_{\lambda\in\Phi_S} J^{\gamma}_{ea}(X,\lambda) = \min_{\gamma\succcurlyeq 0}\max_{\lambda\in\Phi_D} J^{\gamma}_{ea}(X,\lambda), \tag{EC.2}$$

where $J^{\gamma}_{ea}(X,\lambda) := C_{ea}(X,\lambda) - \langle\gamma, D(\lambda)-\beta\rangle$. Identity (EC.2) will follow from Theorem 12.7 in (Altman 1999, chapter 12), if one can verify two conditions, referred to by Altman (1999) as the moment condition and the boundedness condition. In particular, the moment condition, which corresponds to "the near-monotonic case" in (Feinberg and Shwartz 2002, chapter 11) ensures that Φ_S is a dominating class among all admissible λ, justifying the first equality in (EC.2). The second equality in (EC.2) follows from that Φ_S is a convex set; see, e.g, Lemma 11.2 in (Feinberg and Shwartz 2002, chapter 11). The last equality in (EC.2), which also appears in part (iii) of Theorem 12.7 in (Altman 1999, chapter 12), holds because the relaxed (unconstrained) problem $\max_{\lambda\in\Phi_D} J^{\gamma}_{ea}(X,\lambda)$ can be handled by solving a system of dynamic programming equations, from which one recovers a deterministic policy. We next state and verify the two conditions in turn. (Note that the statements given below are slightly different from but essentially the same as those in Altman (1999), because Altman (1999) considers a minimization problem whereas nature faces a maximization problem.)

Condition 1 *(Moment Condition; Condition 11.21 in Altman (1999)): For all $\tilde{z}\in\mathbb{R}$,*

$$\text{the set } \left\{x\in\tilde{\mathcal{S}} : \max_{\lambda\in\Lambda}\tilde{c}(x,\lambda) > \tilde{z}\right\} \text{ is finite.} \tag{EC.3}$$

Since $\tilde{\mathcal{S}}$ is a finite set, (EC.3) is trivially satisfied.

Condition 2 *(Boundedness Condition; Condition 11.1 in Altman (1999)): $\tilde{c}$ is bounded from above and for each $i=1,2,\cdots I$, d^{α}_i is bounded from below.*

Since $\tilde{\mathcal{S}}$ is a finite set and the demand rate of each product is restricted to a bounded region, we know that $\tilde{c}$ is bounded from above. In addition, we see that for all α,

$$\min_{\lambda\in\Lambda} d^{\alpha}_i(\lambda) = d^{\alpha}_i(\bar{\lambda}) = 0, \tag{EC.4}$$

from which we can conclude each d^{α}_i is bounded from below.

Having verified the two conditions, by Theorem 12.7 in Altman (1999) we conclude that (EC.2) holds (with the decision-maker's strategy being fixed).

Step 3: It is easy to see that the discrete-time equivalent of the constraint problem can be described as one where the decision-maker seeks (T,Ψ) to minimize $\max_{\lambda\in\mathcal{D}} C_{ea}(X,\lambda)$, where, to avoid introducing new notation, we continue to use (T,Ψ) to represent the decision-maker's strategy. Similarly, the discrete-time equivalent of the penalty problem can be described as one where the decision-maker seeks (T,Ψ) to minimize

$$\max_{\lambda\in\Phi_D}\ C_{ea}(X,\lambda)-\langle\gamma,D(\lambda)\rangle.$$

With a slight abuse of notation, let the optimal values of the discrete-time constraint problem and penalty problem be denoted by $C_c^\star(\beta)$ and $C_p^\star(\gamma)$, respectively. We have

$$\begin{aligned}C_c^\star(\beta):=\min_{(T,\Psi)}\max_{\lambda\in\mathcal{D}}C_{ea}(X,\lambda)&\overset{(a)}{=}\min_{(T,\Psi)}\min_{\gamma\succcurlyeq 0}\max_{\lambda\in\Phi_D}J^{\gamma}_{ea}(X,\lambda)\\ =\min_{\gamma\succcurlyeq 0}\min_{(T,\Psi)}\max_{\lambda\in\Phi_D}J^{\gamma}_{ea}(X,\lambda)&\overset{(b)}{=:}\min_{\gamma\succcurlyeq 0}\left[C_p^\star(\gamma)+\langle\beta,\gamma\rangle\right],\end{aligned}$$

where step (a) is due to (EC.2) and step (b) follows by the definition of $C_p^\star(\gamma)$.

EC.2. Proofs of Other Main Results

Proof of Proposition 1. The result for $\alpha=1$ has been effectively proved by Lim and Shanthikumar (2007), so we restrict attention to cases where $\alpha\neq 1$. Let $\tilde\alpha:=\alpha-1$. A direct calculation gives

$$\psi_i^{\tilde\alpha}(t)=\exp\left\{\tilde\alpha\int_0^t\ln(1+\theta_i(u))\mathrm{d}A_i(u)\right\}\cdot\exp\left\{-\tilde\alpha\int_0^t\bar\lambda_i\theta_i(u)\mathrm{d}u\right\}. \tag{EC.5}$$

Denote by $\mathbf{u}:=\{u_i\}$ a partition of $[0,t]$, such that $0=u_0<u_1<\cdots<u_m=t$. We get

$$\exp\left\{\tilde\alpha\int_0^t\ln(1+\theta_i(u))\mathrm{d}A_i(u)\right\}=\lim\exp\left\{\sum_k\tilde\alpha\ln(1+\theta_i(u_k))(A_i(u_{k+1})-A_i(u_k))\right\}, \tag{EC.6}$$

where the limit is in probability and taken as $\Delta:=\max_k|u_{k+1}-u_k|\to 0$. By our hypothesis, θ_i is bounded, so we can apply the dominated convergence theorem to conclude that the expectation of left-hand side of (EC.6) with respect to $\mathbb{Q}_i$ equals $\lim\mathbb{E}^{\mathbb{Q}_i}[\mathcal{K}(\mathbf{u})]$, where

$$\mathcal{K}(\mathbf{u}):=\exp\left\{\sum_k\tilde\alpha\ln(1+\theta_i(u_k))(A_i(u_{k+1})-A_i(u_k))\right\}.$$

Here, the limit is again taken as $\Delta:=\max_k|u_{k+1}-u_k|\to 0$. Now, fixing a partition $\mathbf{u}$, we have

$$\begin{aligned}\mathbb{E}^{\mathbb{Q}_i}[\mathcal{K}(\mathbf{u})]&\overset{(a)}{=}\prod_k\mathbb{E}^{\mathbb{Q}_i}\left[\exp\left\{\tilde\alpha\ln(1+\theta_i(u_k))(A_i(u_{k+1})-A_i(u_k))\right\}\right]\\ &\overset{(b)}{=}\prod_k\exp\left\{\lambda_i(u_k)(u_{k+1}-u_k)(e^{\tilde\alpha\ln(1+\theta_i(u_k))}-1)\right\}+o(\Delta)\\ &=\exp\left\{\sum_k\bar\lambda_i(1+\theta_i(u_k))((1+\theta_i(u_k))^{\tilde\alpha}-1)(u_{k+1}-u_k)\right\}+o(\Delta),\end{aligned} \tag{EC.7}$$

where step (a) is due to the fact that increments are conditionally independent given θ_i, and step (b) follows from the piece-wise constant approximation of a non-homogeneous Poisson process plus using the moment generating function for a Poisson random variable. Note that the piece-wise constant approximation is valid due to the local integrability of θ_i. In light of (EC.7),

$$\mathbb{E}^{\mathbb{Q}_i}\left[\exp\left\{\tilde{\alpha}\int_0^t \ln(1+\theta_i(u))\mathrm{d}A_i(u)\right\}\right]=\exp\left\{\int_0^t \bar{\lambda}_i\left[(1+\theta_i(u))^\alpha-(1+\theta_i(u))\right]\mathrm{d}u\right\}.$$

Taking expectation of (EC.5) with respect to $\mathbb{Q}_i$ and substituting for the preceding expression yields

$$\begin{aligned}\mathcal{R}_i^\alpha(t) &:= \frac{1}{\tilde{\alpha}}\ln \mathbb{E}^{\mathbb{Q}_i}\left[\psi_i(t)^{\tilde{\alpha}}\right]\\ &= \frac{1}{\tilde{\alpha}}\left\{\int_0^t \bar{\lambda}_i\left[(1+\theta_i(u))^\alpha-(1+\theta_i(u))\right]\mathrm{d}u-\tilde{\alpha}\int_0^t \bar{\lambda}_i\theta_i(u)\mathrm{d}u\right\},\end{aligned}$$

which, after further simplification, leads to the desired result. □

Proof of Proposition 3. If removing nature from consideration, the line of arguments used to establish the desired equivalence is by now standard; see, e.g., Çelik and Maglaras (2008), Rubino and Ata (2009), Ata and Barjesteh (2023). Thus, a unique aspect of the current proof is dealing with the role of nature. At a high level, we need to argue that an admissible control, along with nature's best response to it, in one formulation can translate into an admissible control, as well as nature's best response to it, in the other formulation, and the translation does not lead to a worse performance from the minimizing player's perspective.

Suppose that $(\hat{Y},\hat{\Psi})$ is an admissible control for the SDG (12)–(15) with Brownian motions $\hat{Z}_i$ and state process $\hat{Q}_i$ and that θ is nature's best response to $(\hat{Y},\hat{\Psi})$. Set $W=\sum_i m_i\hat{Q}_i$. Then, by the definition of h and our construction of W, we have

$$\int_0^t h(W(u))\mathrm{d}u \le \int_0^t \sum_{i=1}^I c_i(\hat{Q}_i(u))\mathrm{d}u \quad \text{for all} \quad t\ge 0. \tag{EC.8}$$

Moreover, $(U,\tilde{\Psi})$ given by (17) and (18), along with Brownian motion B and perturbation ζ defined by (17), fulfill (21)–(23). We claim that

$$\liminf_{t\to\infty}\frac{1}{t}\mathbb{E}\left[\int_0^t r^\star(\zeta(u))\mathrm{d}u\right]=\liminf_{t\to\infty}\frac{1}{t}\mathbb{E}\left[\int_0^t r(\theta(u))\mathrm{d}u\right]. \tag{EC.9}$$

If not, then the left-hand side of (EC.9) is strictly less than the right-hand side. Hence, if we let $\theta':=\{\theta'(t);t\ge 0\}$ be such that $\theta'(t)$ is the minimizer of the right-hand side of (19) with z therein replaced with $\zeta(t)$, then

$$\liminf_{t\to\infty}\frac{1}{t}\mathbb{E}\left[\int_0^t r(\theta'(u))\mathrm{d}u\right]=\liminf_{t\to\infty}\frac{1}{t}\mathbb{E}\left[\int_0^t r^\star(\zeta(u))\mathrm{d}u\right]<\liminf_{t\to\infty}\frac{1}{t}\mathbb{E}\left[\int_0^t r(\theta(u))\mathrm{d}u\right].$$

This, however, contradicts the hypothesis that θ is nature's best response to $(\hat{Y},\hat{\Psi})$. Therefore, we have (EC.9). In view of (EC.8) and (EC.9), to establish the first half of the proposition, we also need to argue that $\zeta = \sum_{i=1}^I \rho_i \theta_i(t)$ is nature's best response to $(U, \tilde{\Psi})$. Suppose not. Then there exists some ζ' such that

$$\liminf_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\int_0^t r^\star(\zeta'(u))\mathrm{d}u\right] < \liminf_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\int_0^t r^\star(\zeta(u))\mathrm{d}u\right].$$

But if we let, by slightly abusing the notation, $\theta' := \{\theta'(t); t \geq 0\}$ be such that $\theta'(t)$ is the minimizer of the right-hand side of (19) with z therein replaced with $\zeta'(t)$, then

$$\liminf_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\int_0^t r(\theta'(u))\mathrm{d}u\right] = \liminf_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\int_0^t r^\star(\zeta'(u))\mathrm{d}u\right] < \liminf_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\int_0^t r(\theta(u))\mathrm{d}u\right].$$

This again contradicts the hypothesis that θ is nature's best response to $(\hat{Y}, \hat{\Psi})$.

Suppose that $(U, \tilde{\Psi})$ is an admissible control for the workload problem (20)–(23) with Brownian motion B and state process W and that ζ is nature's best response to $(U, \tilde{\Psi})$. Then, there exist Brownian motions $\hat{Z}_i$ such that $B(t) = \sum_i m_i \hat{Z}_i(t)$ for all t. Let $\hat{Q}_i$ be such that $\hat{Q}_i(t)$ is the minimizer of the right-hand side of (2) with w therein replaced with $W(t)$ for all $t \geq 0$. Likewise, let θ be such that $\theta(t)$ is the minimizer of the right-hand side of (19) with z therein replaced with $\zeta(t)$ for all $t \geq 0$. We can then construct $\hat{\Psi}$ by setting $\hat{\xi}_i(k) = \mu_i \tilde{\xi}(k)$ for all k and i and control $\hat{Y}_i$ via

$$\hat{Y}_i(t) = m_i \left(\hat{Q}_i(t) - \hat{Q}_i(0) - \hat{Z}_i(t) - \int_0^t \bar{\lambda}_i \theta_i(u)\mathrm{d}u + \sum_{k=0}^{N_i(t)} \hat{\xi}_i(k)\right)$$

for all $t \geq 0$ and $i = 1, \ldots, I$. It is straightforward to verify that $(\hat{Y}, \hat{\Psi})$, along with Brownian motions $\hat{Z}_i$ and the perturbation process θ, fulfill (12)–(14). Note that by construction, the value of

$$\limsup_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\int_0^t \left(\sum_{i=1}^I c_i(\hat{Q}_i(u)) - r(\theta(u))\right)\mathrm{d}u + \sum_{i=1}^I \sum_{k=0}^{N_i(t)} \phi_i(\hat{\xi}_i(k))\right]$$

coincides with that of (20). Therefore, to finish the proof for the second half of the proposition, we need to argue that θ is nature's best response to $(\hat{Y}, \hat{\Psi})$. Suppose not. Then there exists some θ' (again by overloading the notation) such that

$$\hat{Q}_i(t) = \hat{Q}_i(0) + \hat{Z}_i(t) + \int_0^t \bar{\lambda}_i \theta_i'(u)\mathrm{d}u + \mu_i \hat{Y}_i(t) - \sum_{k=0}^{N_i(t)} \hat{\xi}_i(k), \quad i = 1, \ldots, I, \tag{EC.10}$$

where $\hat{Q}, \hat{Z}_i, \hat{Y}_i$, and $\hat{\xi}_i$ are the same as constructed above, and

$$\liminf_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\int_0^t r(\theta'(u))\mathrm{d}u\right] < \liminf_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\int_0^t r(\theta(u))\mathrm{d}u\right] = \liminf_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\int_0^t r^\star(\zeta(u))\mathrm{d}u\right]. \tag{EC.11}$$

Now, multiplying (EC.10) by m_i, summing over all i, and rearranging terms, we obtain

$$\begin{aligned}\int_0^t \sum_{i=1}^I \rho_i \theta_i'(u)\mathrm{d}u &= W(t) - W(0) - B(t) - U(t) + \sum_{i=1}^I \sum_{k=0}^{N_i(t)} \tilde{\xi}_i(k) \\ &= \int_0^t \zeta(u)\mathrm{d}u \quad \text{for all} \quad t \geq 0.\end{aligned} \tag{EC.12}$$

Equation (EC.12), in particular, implies that $\sum_{i=1}^I \rho_i \theta_i'(t) = \zeta(t)$ for almost every t. This, however, contradicts the definition of $r^\star$. Thus, θ must be nature's best response to $(\hat{Y}, \hat{\Psi})$. $\square$

Proof of Proposition 5. To begin, we introduce the following supporting lemmas that are critical for proving this proposition. Below, unless stated otherwise, $\pi'(w,\eta)$ is used to denote the first-order partial derivative of π with respect to its first argument. Lemma EC.1 asserts non-negativity and Lipschitz continuity of the function g.

LEMMA EC.1. *$g(x)$ is non-negative and Lipschitz continuous in $x \in \mathbb{R}$, i.e., for any x_1 and x_2, we have*

$$|g(x_1) - g(x_2)| \leq M\,|x_1 - x_2|\,. \tag{EC.13}$$

Note that the properties claimed by Lemma EC.1 provide the standard (sufficient) condition for equations like (29) to have a unique solution (see, for example, chapter 3 in David et al. (2018)).

LEMMA EC.2. *(i) For any $\eta \in \mathbb{R}$, the ordinary differential equation* (29) *has a unique continuously differentiable solution $\pi(w,\eta)$. (ii) $\pi(w,\eta)$ is continuous in $\eta \in \mathbb{R}$, and $\pi'(w,\eta)$ is continuous in $w \in \mathbb{R}_+$, and $\eta \in \mathbb{R}$, respectively.*

LEMMA EC.3. *For fixed $w > 0$, $\pi(w,\eta)$ is strictly increasing in $\eta \in \mathbb{R}$ and*

$$\lim_{\eta \to \pm\infty} \pi(w,\eta) = \pm\infty. \tag{EC.14}$$

LEMMA EC.4. *There exists an upper bound $\bar{\eta}$ with $\bar{\eta} > 0$ such that the following results hold:*
(i) If $\eta \leq 0$, then $\pi(w,\eta)$ is strictly decreasing in $w \in [0,\infty)$ and

$$\lim_{w \to \infty} \pi(w,\eta) = -\infty. \tag{EC.15}$$

(ii) If $\eta \geq \bar{\eta}$, then $\pi(w,\eta)$ is strictly increasing in $w \in [0,\infty)$ and

$$\lim_{w \to \infty} \pi(w,\eta) = \infty. \tag{EC.16}$$

(iii) If $0 < \eta < \bar{\eta}$, then there exists a unique number $w^\star(\eta)$ such that $\pi(w,\eta)$ is strictly increasing in $[0, w^\star(\eta)]$ and strictly decreasing in $[w^\star(\eta), \infty)$, where

$$w^\star(\eta) := \inf\left\{w \geq 0 : \pi'(w,\eta) \leq 0\right\}.$$

Furthermore, $\lim_{w\to\infty} \pi(w,\eta) = -\infty$.

Lemma EC.4 divides the value of η into three segments, separated by two cut-off points: 0 and $\bar{\eta}$. In particular, the presence of the cut-off point 0 is due to the condition $g(0) = 0$. Mathematically, if $g(0) \neq 0$, then the cut-off point 0 should be modified to $g(0)$. We can then use the previous results to find the unique parameters q_i and s_i for each i. In Lemma EC.5 we show that there exist unique q_i, s_i and η_i with $\eta_i \in (0, \bar{\eta})$ and $0 < q_i < s_i$ such that

$$\pi(q_i,\eta_i)=\pi(s_i,\eta_i)=\tilde{\ell}_i, \tag{EC.17}$$

and

$$\int_{q_i}^{s_i}\left[\pi(w,\eta_i)-\tilde{\ell}_i\right]\mathrm{d}w=L_i. \tag{EC.18}$$

Lemma EC.5. *(i) There exists a finite number $\eta^\ddagger\in(0,\bar{\eta})$ such that for any $\eta\in(\eta^\ddagger,\bar{\eta})$, there exist two unique numbers $q(\eta)$ and $s(\eta)$ with $0<q(\eta)<w^\star(\eta)<s(\eta)$ satisfying*

$$\pi(q(\eta),\eta)=\pi(s(\eta),\eta)=\tilde{\ell}_i.$$

(ii) There exists a unique finite number $\eta_i\in(\eta^\ddagger,\bar{\eta})$ such that

$$\tilde{f}(\eta_i)=L_i, \tag{EC.19}$$

where $\tilde{f}(\eta):=\int_0^\infty(\pi(w,\eta)-\tilde{\ell}_i)^+\mathrm{d}w$ is strictly increasing in $\eta\in(\eta^\ddagger,\bar{\eta})$.

Remark EC.1. Letting $q_i=q(\eta_i)$ and $s_i=s(\eta_i)$, Lemma EC.5 directly implies (EC.17) and (EC.18).

The proof of Lemmas EC.1–EC.5 can be found in §EC.11. We are now going to prove part (i) of Proposition 5. Let $\pi_i=\pi(w,\eta_i)$. Recall that π_i is a continuously differentiable solution to (29) with the initial condition (30), so (EC.17) and (EC.18) ensure the boundary condition (31). Also from Lemma EC.4(iii) we know that $0<q_i<s_i<\infty$. Thus we have completed the proof of part (i).

To prove part (ii) of the proposition, it suffices to show

$$\frac{1}{2}\sigma^2v''(w)+g(v'(w))+h(w)-\eta_i\begin{cases}=0 & \text{for}\quad w\in(0,s_i)\\ \geq 0 & \text{for}\quad w\geq s_i\end{cases}\quad\text{and} \tag{EC.20}$$

$$\inf_{0\leq z\leq w}\left\{v(w-z)+\tilde{\ell}_iz+L_i\right\}-v(w)\begin{cases}\geq 0 & \text{for}\quad w\in(0,s_i)\\ =0 & \text{for}\quad w\geq s_i\end{cases} \tag{EC.21}$$

for any function v defined as in the proposition. By the definition of v, we know that

$$\frac{1}{2}\sigma^2v''(w)+g(v'(w))+h(w)-\eta_i=0\quad\text{for}\quad w\in(0,s_i).$$

According to Lemma EC.4(iii) and Lemma EC.5, we can deduce that $v''(s_i-)=\pi'(s_i-)<0$. Hence, for $w\geq s_i$, we have

$$\frac{1}{2}\sigma^2v''(w)+g(v'(w))+h(w)-\eta_i\geq\frac{1}{2}\sigma^2v''(s_i-)+g(\tilde{\ell}_i)+h(s_i-)-\eta_i=0.$$

This proves (EC.20). To show (EC.21), it suffices to establish the following equivalence:

$$\sup_{0\leq y\leq w}\int_y^w\left[v'(z)-\tilde{\ell}_i\right]\mathrm{d}z\begin{cases}\leq L_i & \text{for}\quad w\in(0,s_i)\\ =L_i & \text{for}\quad w\geq s_i\end{cases}$$

which holds true by the definition of v and the fact that

$$\pi(w,\eta_i)\begin{cases} <\tilde{\ell}_i & \text{for} \quad w<q_i \\ \geq \tilde{\ell}_i & \text{for} \quad w\in[q_i,s_i] \\ <\tilde{\ell}_i & \text{for} \quad w>s_i \end{cases}$$

due to Lemma EC.4(iii) and Lemma EC.5. We thus complete the proof for part (ii) of the proposition.

Towards proving part (iii) of Proposition 5, consider an admissible control of the minimizing player such that under this control

$$J_i(w):=\max_{\zeta}\ \limsup_{t\to\infty}\frac{1}{t}\mathbb{E}^{\zeta}\left[\int_0^t h(W(u))\mathrm{d}u-\int_0^t r^{\star}(\zeta(u))\mathrm{d}u+\sum_{k=0}^{N_i(t)}\tilde{\phi}_i(\tilde{\xi}_i(k))\middle| W(0)=w\right]<\infty.$$

Moreover, consider a special drift-rate control $\zeta^{\#}(W)$, defined as

$$\zeta^{\#}(W):=\inf\arg\max_{\zeta}\left\{v'(W)\zeta-r^{\star}(\zeta)\right\}. \tag{EC.22}$$

Clearly $\zeta^{\#}(\cdot)$ is an adaptive control satisfying

$$g(v'(W))+r^{\star}(\zeta^{\#}(W))-\zeta^{\#}(W)v'(W)=0. \tag{EC.23}$$

Let $\delta_i(k):=v(W(\tau_i(k)))-v(W(\tau_i(k)-))$. From (32) it follows that $v(y)-v(x)\leq\tilde{\phi}_i(y-x)$ for $y>x$, and so

$$-\delta_i(k)\leq\tilde{\phi}_i(W(\tau_i(k)-)-W(\tau_i(k)))=\tilde{\phi}_i(\tilde{\xi}_i(k)) \quad \text{for} \quad k=0,1,2,\ldots. \tag{EC.24}$$

Supposing that the maximizing player adopts $\zeta^{\#}(\cdot)$, then applying the generalized Itô's formula with $W(0)=w$, we obtain, for $t\geq 0$,

$$\begin{aligned} v(W(t))=v(w)&+\int_0^t\left(\frac{\sigma^2}{2}v''(W(u))+\zeta^{\#}(W(u))v'(W(u))\right)\mathrm{d}u \\ &+\sigma\int_0^t v'(W(u))\mathrm{d}B(u)+\sum_{k=0}^{N_i(t)}\delta_i(k)+v'(0)U(t), \end{aligned}$$

which, in turn, implies that

$$\begin{aligned} \Xi(t)&:=\int_0^t h(W(u))\mathrm{d}u-\int_0^t r^{\star}(\zeta^{\#}(W(u)))\mathrm{d}u+\sum_{k=0}^{N_i(t)}\tilde{\phi}_i(\tilde{\xi}_i(k)) \\ &=\eta_i t+v(w)-v(W(t))+\sigma\int_0^t v'(W(u))\mathrm{d}B(u) \\ &\quad+\int_0^t\left(\frac{\sigma^2}{2}v''(W(u))+\zeta^{\#}(W(u))v'(W(u))+h(W(u))-r^{\star}(\zeta^{\#}(W(u)))-\eta_i\right)\mathrm{d}u \\ &\quad+\sum_{k=0}^{N_i(t)}\left(\delta_i(k)+\tilde{\phi}_i(\tilde{\xi}_i(k))\right)+v'(0)U(t). \end{aligned}$$

Because (v, η_i) satisfies the optimality equation and noting (EC.23)–(EC.24), we obtain from the preceding identity that

$$\Xi(t) \geq \eta_i t + v(w) - v(W(t)) + \sigma \int_0^t v'(W(u))\mathrm{d}B(u). \tag{EC.25}$$

Drawing upon our construction of v, it is clear that $v'(w)$ is bounded in $[0, \ell_i]$. Consequently, we can deduce from Corollary 4.8 and Proposition 4.17 in Harrison (2013) that $\mathbb{E}^{\zeta^\#}\left[\int_0^t v'(W(u))\mathrm{d}B(u)\right] = 0$. It thus follows from (EC.25) that

$$\begin{aligned}\frac{1}{t}\mathbb{E}^{\zeta^\#}[\Xi(t)] &\geq \eta_i + \frac{1}{t}v(w) - \frac{1}{t}\mathbb{E}^{\zeta^\#}[v(W(t))] \\ &\geq \eta_i + \frac{1}{t}v(w) - \frac{l}{t}\mathbb{E}^{\zeta^\#}[W(t)]\end{aligned} \tag{EC.26}$$

for some positive constant l, whose existence is guaranteed by the fact that v has bounded first-order derivatives. On account of Assumption 1 and the boundedness of $r^\star$, by a slight abuse of notation, we can deduce that

$$\begin{aligned}\infty > J_i(w) &\geq \limsup_{t\to\infty}\frac{1}{t}\mathbb{E}^{\zeta^\#}\left[\int_0^t h(W(u))\mathrm{d}u - \int_0^t r^\star(\zeta(u))\mathrm{d}u \middle| W(0) = w\right] \\ &\geq C_1 + C_2 \limsup_{t\to\infty}\frac{1}{t}\mathbb{E}^{\zeta^\#}\left[\int_0^t W(u)\mathrm{d}u\right]\end{aligned}$$

for some $C_1 \in \mathbb{R}$ and $C_2 > 0$. With this string of inequalities, we can conclude that

$$\liminf_{t\to\infty}\frac{1}{t}\mathbb{E}^{\zeta^\#}[W(t)] = 0. \tag{EC.27}$$

To see this, suppose that $\liminf_{t\to\infty}\frac{1}{t}\mathbb{E}^{\zeta^\#}[W(t)] > a > 0$. Then there exists some $\bar{t}$ such that $\mathbb{E}^{\zeta^\#}[W(t)] > at/2$ for all $t \geq \bar{t}$. It follows that

$$\limsup_{t\to\infty}\frac{1}{t}\mathbb{E}^{\zeta^\#}\left[\int_0^t W(u)\mathrm{d}u\right] \geq \limsup_{t\to\infty}\frac{1}{t}\int_{\bar{t}}^t \frac{au}{2}\mathrm{d}u = \infty,$$

which is a contradiction. Hence, (EC.27) holds. Finally, taking the limsup on both sides of (EC.26) by passing to the limit $t \to \infty$ and noting (EC.27), we find

$$J_i(w) \geq \limsup_{t\to\infty}\frac{1}{t}\mathbb{E}^{\zeta^\#}[\Xi(t)] \geq \eta_i,$$

as desired.

In the presence of the maximizing player, we still need to verify that $\zeta^\#$ is indeed the maximizer's best response given the decision-maker will commit to the control band policy (i, q_i, s_i). For this purpose, we can easily write down the Bellman equation for the maximizer's problem: seek $v_m \in \mathcal{C}^2(0, s_i)$ and $\eta_m \in \mathbb{R}$ such that

$$\max_\zeta\left\{\frac{1}{2}\sigma^2 v_m''(w) + \zeta v_m'(w) + h(w) - r^\star(\zeta)\right\} = \eta_m, \quad w \in (0, s_i),$$

subject to the boundary conditions

$$v_m'(0) = 0 \quad \text{and} \quad v_m(s_i) = \tilde{\phi}_i(s_i - q_i) + v_m(q_i).$$

Comparing these with (26) and (27), we immediately conclude that $v_m = v$ and $\eta_m = \eta_i$. Therefore, the control rule $\zeta^{\#}$ defined by (EC.22) is the maximizer's best response given the decision-maker chooses to adopt the control band policy (i, q_i, s_i).

Finally, noting that η_i is the long-run average cost when the decision-maker implements (i, q_i, s_i) and the maximizer employs the drift-rate control $\zeta^{\#}$ (cf. Proposition 4) completes the proof. $\square$

Proof of Theorem 1. The proof of this theorem follows closely the steps in the proof of part (iii) in Proposition 5. Thus, we only highlight the key differences. To start, let

$$J(w) := \max_{\zeta} \limsup_{t\to\infty} \frac{1}{t}\mathbb{E}^{\zeta}\left[\int_0^t h(W(u))\mathrm{d}u - \int_0^t r^{\star}(\zeta(u))\mathrm{d}u + \sum_{i=1}^{I}\sum_{k=0}^{N_i(t)} \tilde{\phi}_i(\tilde{\xi}_i(k))\middle| W(0) = w\right],$$

and define $\delta_i(k)$ in the same way as we did in the proof of Proposition 5. Now using (33), we conclude that, for all i, $v(y) - v(x) \leq \tilde{\phi}_i(y - x)$; thus for all $k = 1, 2, \dots$ and $i = 1, \dots, I$, we have

$$-\delta_i(k) \leq \tilde{\phi}_i(\tilde{\xi}_i(k)). \tag{EC.28}$$

Next by applying the generalized Itô's formula, we obtain, for $t \geq 0$,

$$\mathbb{E}^{\zeta}[v(W(t))] = v(w) + \mathbb{E}^{\zeta}\left[\int_0^t \left(\frac{\sigma^2}{2}v''(W(u)) + \zeta(u)v'(W(u))\right)\mathrm{d}u\right] + \mathbb{E}^{\zeta}\left[\sum_{i=1}^{I}\sum_{k=0}^{N_i(t)} \delta_i(k)\right].$$

On substituting (EC.28) into above identity and using (33), we deduce

$$\begin{aligned} v(w) \leq{}& \mathbb{E}^{\zeta}\left[\int_0^t \left(h(W(u)) - r^{\star}(\zeta(u)) - \eta_i\right)\mathrm{d}u\right] + \mathbb{E}^{\zeta}\left[\sum_{i=1}^{I}\sum_{k=0}^{N_i(t)} \tilde{\phi}_i(\tilde{\xi}_i(k))\right] \\ &+ \mathbb{E}^{\zeta}\left[v(W(t))\right] + \mathbb{E}^{\zeta}\left[\int_0^t \left(g(v'(W(u))) + r^{\star}(\zeta(u)) - \zeta(u)v'(W(u))\right)\mathrm{d}u\right]. \end{aligned}$$

The rest of the proof proceeds in exactly the same fashion as the proof of Proposition 5. First, by choosing $\zeta = \zeta^{\#}$ with $\zeta^{\#}$ given as in (EC.22), one can formally show that $\eta_{i^\star} \leq J(w)$. Second, one can easily argue that $\zeta^{\#}$ is the maximizer's best response: when the decision-maker chooses $(i^\star, q_{i^\star}, s_{i^\star})$, the maximizer will follow $\zeta^{\#}$ and never deviate from it. The desired optimality of the control band policy $(i^\star, q_{i^\star}, s_{i^\star})$ then follows immediately from these two points plus the fact that the "lower bound" $\eta_{i^\star}$ is attained with the policy $(i^\star, q_{i^\star}, s_{i^\star})$ as demonstrated by Proposition 5. $\square$

Proof of Proposition 6. We need to show that

$$\frac{1}{2}\sigma^2 v''(w) + g(v'(w)) + h(w) - \eta_{i^\star} \geq 0 \quad \text{and} \tag{EC.29}$$

$$\inf_{0\leq z\leq w}\left[v(w-z)+\tilde{\phi}_i(z)\right] - v(w) \geq 0 \quad \text{for all} \quad i \tag{EC.30}$$

for any v defined as in the proposition. The verification for (EC.29) is very similar to that for (EC.20), so we will only focus on proving (EC.30). To this end, it suffices to argue that the following holds:

$$\sup_{0\leq y\leq w}\int_y^w \left[v'(z)-\tilde{\ell}_i\right]\mathrm{d}z \leq L_i \quad \text{for all} \quad i. \tag{EC.31}$$

Suppose (EC.31) does not hold. Then there exists some k such that

$$\sup_{0\leq y\leq w}\int_y^w \left[v'(z)-\tilde{\ell}_k\right]\mathrm{d}z > L_k. \tag{EC.32}$$

By our definition of v, the function graph of $v'(\cdot)$ coincides with that of $\pi(\cdot,\eta_{i^\star})$ for $w\in[0,s_*)$ and overlaps with the horizontal line $\tilde{\ell}_*$ for $w\in[s_*,\infty)$. Moreover, by Lemma EC.4(iii), the function graph of $v'(\cdot)$ crosses the horizontal line $\tilde{\ell}_*$ once before intersecting it at the coordinate $(s_*,\tilde{\ell}_*)$ for the second time. It follows from the same lemma that the function graph of $v'(\cdot)$ either does not intersect, touches, or intersects twice with any horizontal line *strictly above* the horizontal line $\tilde{\ell}_*$. Because of (EC.32) and the fact that $\tilde{\ell}_k \geq \tilde{\ell}_*$, we must have

$$\sup_{0\leq y\leq w}\int_y^w \left[\pi(z,\eta_{i^\star})-\tilde{\ell}_k\right]\mathrm{d}z = \sup_{0\leq y\leq w}\int_y^w \left[v'(z)-\tilde{\ell}_k\right]\mathrm{d}z > L_k. \tag{EC.33}$$

By the definition of $\tilde{\ell}_*$, there cannot exist $\tilde{\ell}_k < \tilde{\ell}_*$. Figure EC.1 provides for a further graphical justification of the equality in (EC.33). On the other hand, we know that

$$\sup_{0\leq y\leq w}\int_y^w \left[\pi(z,\eta_k)-\tilde{\ell}_k\right]\mathrm{d}z = L_k. \tag{EC.34}$$

In light of Lemma EC.3, we can deduce from (EC.33) and (EC.34) that $\eta_k < \eta_{i^\star}$. This, however, reaches a contradiction given the definition of $\eta_{i^\star}$. Therefore, (EC.31) must be satisfied. The proof is thus complete. □

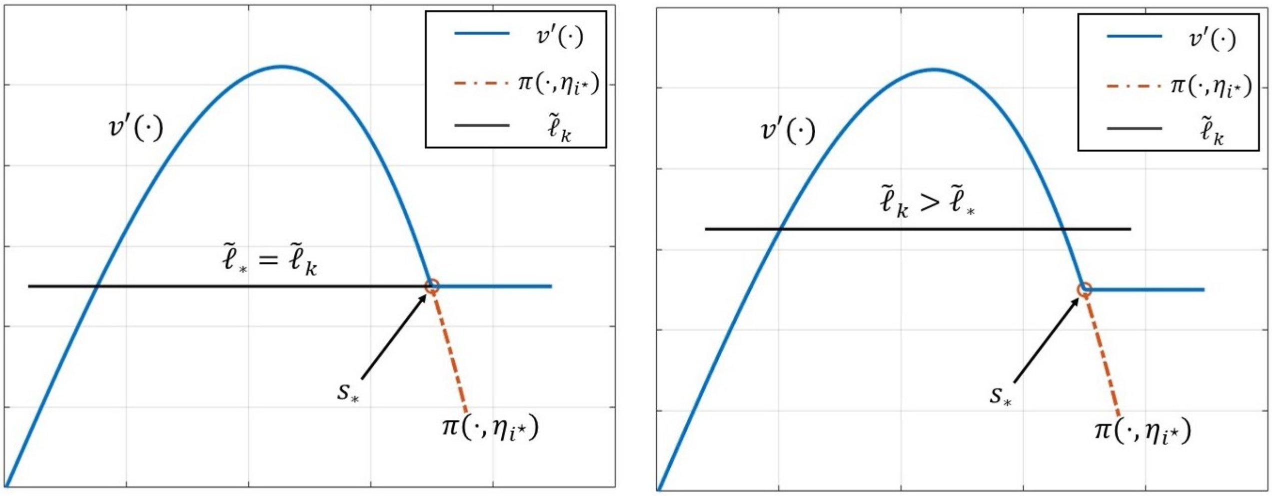
Remark EC.2. Regarding the preceding proof, there is an alternative argument to prove $\eta_k < \eta_{i^\star}$ under the hypothesis (EC.32). The key is to note that (EC.33), in conjunction with EC.4(iii), implies the existence of a new proportional cost parameter $\breve{\ell} > \tilde{\ell}_k$ such that

$$\int_q^s \left[\pi(z,\eta_{i^\star})-\breve{\ell}\right]\mathrm{d}z = \sup_{0\leq y\leq w}\int_y^w \left[\pi(z,\eta_{i^\star})-\breve{\ell}\right]\mathrm{d}z = L_k,$$

where

$$q := \inf\left\{w\geq 0 : \pi(w,\eta_{i^\star}) = \breve{\ell}\right\} \quad \text{and} \quad s := \sup\left\{w\geq 0 : \pi(w,\eta_{i^\star}) = \breve{\ell}\right\}.$$

It follows that the control band policy (k,q,s) with (q,s) specified above yields a long-run average cost of $\eta_{i^\star}$ if the fixed cost remains L_k but the proportional cost is modified to $\breve{\ell}$. However, since $\tilde{\ell}_k < \breve{\ell}$, we must have $\eta_k < \eta_{i^\star}$, again reaching a contradiction.



(a) $\tilde{\ell}_k = \tilde{\ell}_*$ (b) $\tilde{\ell}_k > \tilde{\ell}_*$

Figure EC.1 Visualization of v' and π

EC.3. Numerical Algorithm for Solving the Optimality Equation

To find the solution of the optimality equation (26), we start with an initial guess of v, denoted as v_0, that solves

$$\frac{1}{2}\sigma^2 v_0''(w) + h(w) = \eta_i, \quad w \in (0, s_i) \tag{EC.35}$$

subject to the boundary conditions $v_0'(0) = 0$, $v_0(s_i) = \tilde{\phi}_i(s_i - q_i) + v_0(q_i)$ and necessary optimality conditions $v_0'(q_i) = v_0'(s_i) = \tilde{\ell}_i$. Notice that (EC.35) is a second-order linear ordinary differential equation, so we can solve it analytically. Then for each $w \in (0, s_i)$, we seek $\zeta_0(w)$ that maximizes $\{\zeta_0(w)v_0'(w) - r^\star(\zeta_0(w))\}$. The next step is to find v_1 such that

$$\frac{1}{2}\sigma^2 v_1''(w) + \zeta_0(w)v_1'(w) + h(w) - r^\star(\zeta_0(w)) = \eta_i, \quad w \in (0, s_i) \tag{EC.36}$$

subject to the same boundary conditions and necessary optimality conditions as mentioned previously. We can solve (EC.36) numerically via the finite difference method (FDM).

Using the kth estimate of v, which we denote as v_k, we can find $\zeta_k(w)$ that maximizes $\{\zeta_k(w)v_k'(w) - r^\star(\zeta_k(w))\}$, and further solve the ordinary differential equation

$$\frac{1}{2}\sigma^2 v_{k+1}''(w) + \zeta_k(w)v_{k+1}'(w) + h(w) - r^\star(\zeta_k(w)) = \eta_i, \quad w \in (0, s_i)$$

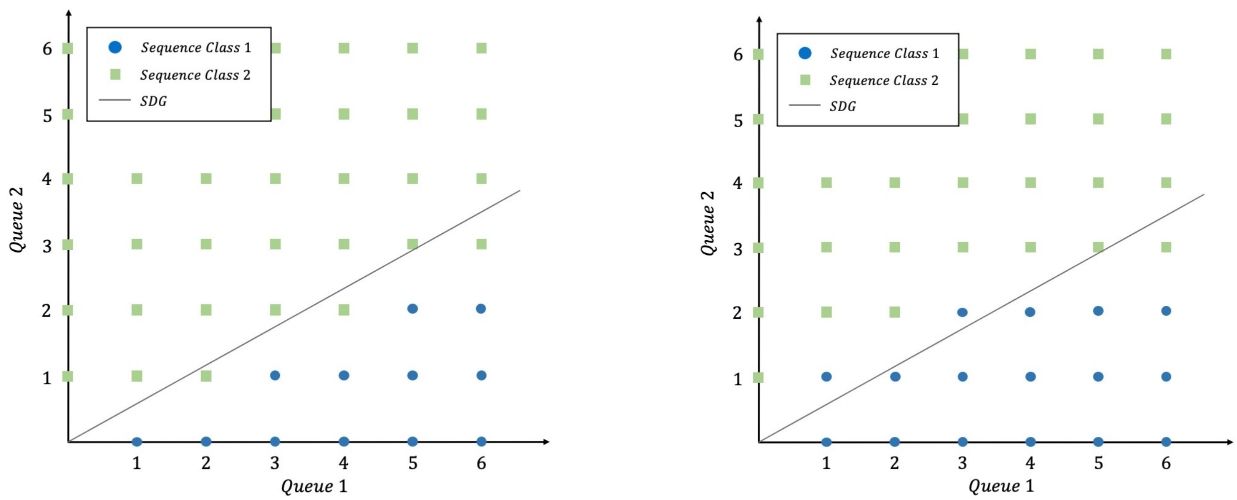
subject to the set of boundary conditions and necessary optimality conditions by using FDM to get $v_{k+1}(w)$, the $(k+1)$th estimate of v. Repeating these steps we obtain an iterative procedure that generates a sequence $\{v_{k+1}(w), \zeta_k(w)\}$ which is expected to converge to the optimal solution when $k \to \infty$. Although we do not attempt to rigorously prove the desired convergence result, our extensive numerical experiments suggest convergence happens after a few iterations. The algorithm terminates when the iteration error $||v_{k+1}(w) - v_k(w)||$ and $||\zeta_{k+1}(w) - \zeta_k(w)||$ become sufficiently small, for $w \in (0, s_i)$.

EC.4. Additional Numerical Results for Two-Product Models

This section presents additional numerical results. In §EC.4.1, we present another example where we compare the solution to the original penalty problem with that to the SDG. In §EC.4.2, we provide the optimal thresholds used in simulations and the corresponding simulated costs. In §EC.4.3, we offer an additional example comparing the optimal cost attained using the actual demand model with the optimal robust cost achieved through our proposed method.

EC.4.1. Solution Comparison Between the Penalty Problem and the SDG

Figures EC.2 and EC.3 show the optimal sequencing and outsourcing rules attained by solving the original penalty problem and the SDG under the KL divergence. We observe that the value of the parameter α also influences the optimal outsourcing rule, such that some states that trigger outsourcing under KL divergence may not trigger outsourcing under the Rényi divergence with $\alpha = 1/2$.



(a) Class 1 is in service

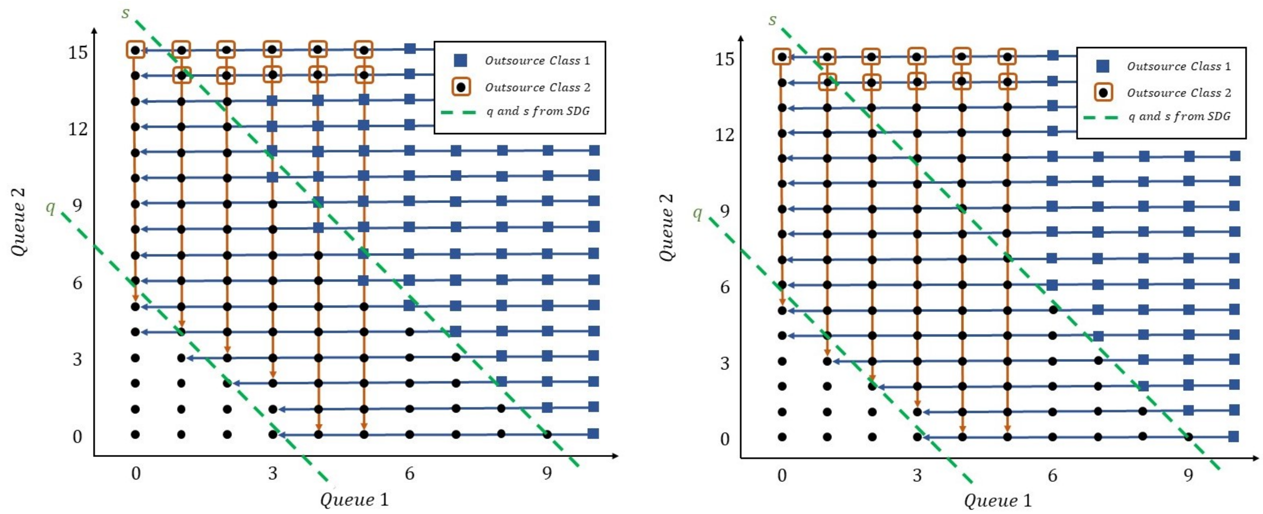
(b) Class 2 is in service

Figure EC.2 Sequencing strategies derived from the penalty problem and the SDG with $\gamma_1 = \gamma_2 = 30$ **when** $\alpha = 1$

Table EC.1 and Figure EC.4 compare the optimal costs obtained from the original penalty problem to those derived from the SDG (workload problem) for various (γ_1, γ_2) pairs. We observe a remarkable similarity between the long-run average costs obtained from the SDG and the values obtained from the original penalty problem.

EC.4.2. Optimal Thresholds and the Simulated Costs in §9.2

We present supplementary numerical results obtained in §9.2 in this subsection. Table EC.2 reports the optimal outsourcing thresholds used in simulations, while Tables EC.3 and EC.4 report the estimated long-run average costs generated from simulations in §9.2.1 and §9.2.2, respectively.



(a) Class 1 is in service

(b) Class 2 is in service

Figure EC.3 **Outsourcing strategies derived from the penalty problem and the SDG with** $\gamma_1 = \gamma_2 = 30$ **when** $\alpha = 1$

Table EC.1 **Long-run average cost** η **obtained from numerical programs when** $\alpha = 1/2$ **and** 1

$\alpha = 1/2$			$\alpha = 1$		
(γ_1, γ_2)	η_{exact}	η_{SDG}	(γ_1, γ_2)	η_{exact}	η_{SDG}
$(7, 7)$	35.1853	35.9829	$(7, 7)$	27.2448	26.7500
$(30, 30)$	22.1417	21.0459	$(30, 30)$	20.3445	19.6836
$(50, 50)$	20.7016	19.7286	$(50, 50)$	19.6659	19.1094
$(500, 500)$	18.6757	17.9884	$(500, 500)$	18.6403	18.2891

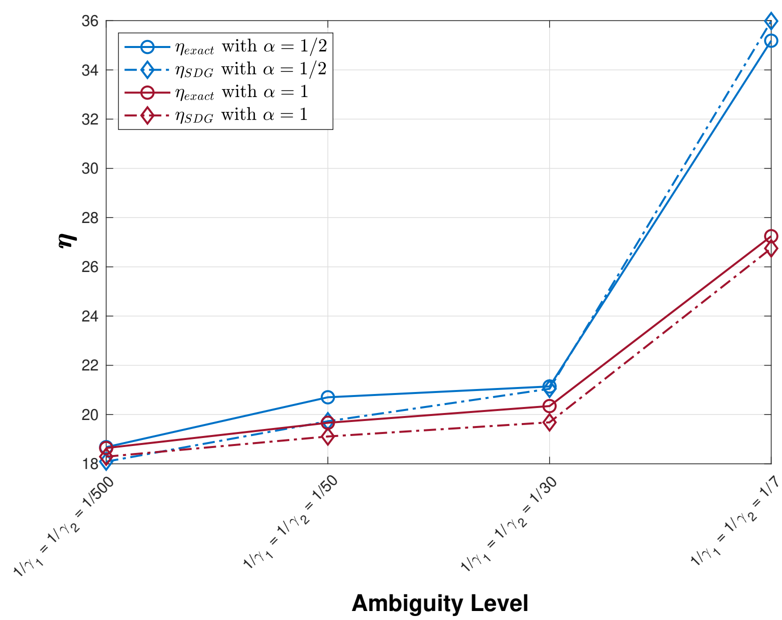


Figure EC.4 **Long-run average costs of the original penalty problem and the SDG with** $\alpha = 1/2$ **and** 1

Table EC.2 Optimal outsourcing parameters for $\alpha = 3$ and 1

$\alpha = 3$			$\alpha = 1$		
(γ_1, γ_2)	q_1	s_1	(γ_1, γ_2)	q_1	s_1
$(0.02, 0.02)$	0.012	0.246	$(0.1, 0.1)$	0.033	0.193
$(0.02, 0.05)$	0.012	0.242	$(0.1, 1)$	0.034	0.192
$(0.02, 0.1)$	0.012	0.239	$(0.1, 5)$	0.035	0.191
$(0.02, 0.5)$	0.013	0.238	$(0.1, 10)$	0.035	0.191
$(0.02, 100)$	0.013	0.240	$(0.1, 100)$	0.035	0.192
$(0.05, 0.02)$	0.019	0.213	$(1, 0.1)$	0.050	0.188
$(0.05, 0.05)$	0.020	0.211	$(1, 1)$	0.052	0.187
$(0.05, 0.1)$	0.021	0.208	$(1, 5)$	0.053	0.189
$(0.05, 0.5)$	0.021	0.207	$(1, 10)$	0.054	0.189
$(0.05, 100)$	0.021	0.207	$(1, 100)$	0.054	0.189
$(0.1, 0.02)$	0.026	0.198	$(5, 0.1)$	0.073	0.200
$(0.1, 0.05)$	0.029	0.194	$(5, 1)$	0.077	0.201
$(0.1, 0.1)$	0.031	0.193	$(5, 5)$	0.079	0.203
$(0.1, 0.5)$	0.032	0.190	$(5, 10)$	0.080	0.204
$(0.1, 100)$	0.035	0.190	$(5, 100)$	0.080	0.204
$(0.5, 0.02)$	0.044	0.187	$(10, 0.1)$	0.077	0.202
$(0.5, 0.05)$	0.054	0.188	$(10, 1)$	0.080	0.205
$(0.5, 0.1)$	0.059	0.191	$(10, 5)$	0.083	0.207
$(0.5, 0.5)$	0.064	0.193	$(10, 10)$	0.084	0.207
$(0.5, 100)$	0.066	0.194	$(10, 100)$	0.084	0.207
$(100, 0.02)$	0.055	0.190	$(100, 0.1)$	0.081	0.203
$(100, 0.05)$	0.071	0.197	$(100, 1)$	0.083	0.206
$(100, 0.1)$	0.078	0.203	$(100, 5)$	0.085	0.208
$(100, 0.5)$	0.085	0.209	$(100, 10)$	0.086	0.209
$(100, 100)$	0.088	0.210	$(100, 100)$	0.088	0.210

Table EC.3 $\hat{\mathcal{C}}(\gamma)$ generated from simulations with ARIMA intensity

γ	$\hat{\mathcal{C}}(\gamma)$	γ	$\hat{\mathcal{C}}(\gamma)$
	$\alpha = 3$		$\alpha = 1$
$(0.02, 0.02)$	$2.750 \pm 1\text{E}{-1}$	$(0.1, 0.1)$	$2.626 \pm 1\text{E}{-1}$
$(0.02, 0.05)$	$2.782 \pm 1\text{E}{-1}$	$(0.1, 1)$	$2.534 \pm 1\text{E}{-1}$
$(0.02, 0.1)$	$2.703 \pm 1\text{E}{-1}$	$(0.1, 5)$	$2.549 \pm 1\text{E}{-1}$
$(0.02, 0.5)$	$2.680 \pm 1\text{E}{-1}$	$(0.1, 10)$	$2.547 \pm 1\text{E}{-1}$
$(0.02, 100)$	$2.760 \pm 1\text{E}{-1}$	$(0.1, 100)$	$2.648 \pm 1\text{E}{-1}$
$(0.05, 0.02)$	$2.689 \pm 1\text{E}{-1}$	$(1, 0.1)$	$2.570 \pm 1\text{E}{-1}$
$(0.05, 0.05)$	$2.646 \pm 1\text{E}{-1}$	$(1, 1)$	$2.551 \pm 1\text{E}{-1}$
$(0.05, 0.1)$	$2.604 \pm 1\text{E}{-1}$	$(1, 5)$	$2.412 \pm 1\text{E}{-1}$
$(0.05, 0.5)$	$2.508 \pm 1\text{E}{-1}$	$(1, 10)$	$2.461 \pm 1\text{E}{-1}$
$(0.05, 100)$	$2.591 \pm 1\text{E}{-1}$	$(1, 100)$	$2.570 \pm 1\text{E}{-1}$
$(0.1, 0.02)$	$2.693 \pm 1\text{E}{-1}$	$(5, 0.1)$	$2.576 \pm 1\text{E}{-1}$
$(0.1, 0.05)$	$2.595 \pm 1\text{E}{-1}$	$(5, 1)$	$2.473 \pm 1\text{E}{-1}$
$(0.1, 0.1)$	$2.619 \pm 1\text{E}{-1}$	$(5, 5)$	$2.534 \pm 1\text{E}{-1}$
$(0.1, 0.5)$	$2.546 \pm 1\text{E}{-1}$	$(5, 10)$	$2.422 \pm 1\text{E}{-1}$
$(0.1, 100)$	$2.494 \pm 1\text{E}{-1}$	$(5, 100)$	$2.498 \pm 1\text{E}{-1}$
$(0.5, 0.02)$	$2.556 \pm 1\text{E}{-1}$	$(10, 0.1)$	$2.539 \pm 1\text{E}{-1}$
$(0.5, 0.05)$	$2.444 \pm 1\text{E}{-1}$	$(10, 1)$	$2.528 \pm 1\text{E}{-1}$
$(0.5, 0.1)$	$2.387 \pm 1\text{E}{-1}$	$(10, 5)$	$2.457 \pm 1\text{E}{-1}$
$(0.5, 0.5)$	$2.404 \pm 1\text{E}{-1}$	$(10, 10)$	$2.472 \pm 1\text{E}{-1}$
$(0.5, 100)$	$2.496 \pm 1\text{E}{-1}$	$(10, 100)$	$2.565 \pm 1\text{E}{-1}$
$(100, 0.02)$	$2.639 \pm 1\text{E}{-1}$	$(100, 0.1)$	$2.562 \pm 1\text{E}{-1}$
$(100, 0.05)$	$2.506 \pm 1\text{E}{-1}$	$(100, 1)$	$2.593 \pm 1\text{E}{-1}$
$(100, 0.1)$	$2.503 \pm 1\text{E}{-1}$	$(100, 5)$	$2.538 \pm 1\text{E}{-1}$
$(100, 0.5)$	$2.487 \pm 1\text{E}{-1}$	$(100, 10)$	$2.496 \pm 1\text{E}{-1}$
$(100, 100)$	$2.517 \pm 1\text{E}{-1}$	$(100, 100)$	$2.508 \pm 1\text{E}{-1}$

Table EC.4 $\hat{\mathcal{C}}(\gamma)$ **generated from simulations with CTMC intensity**

γ	$\hat{\mathcal{C}}(\gamma)$ $\alpha=3$	γ	$\hat{\mathcal{C}}(\gamma)$ $\alpha=1$
$(0.02,0.02)$	$8.908\pm 1\text{E}{-1}$	$(0.1,0.1)$	$8.599\pm 1\text{E}{-1}$
$(0.02,0.05)$	$8.586\pm 1\text{E}{-1}$	$(0.1,1)$	$8.568\pm 1\text{E}{-1}$
$(0.02,0.1)$	$8.632\pm 1\text{E}{-1}$	$(0.1,5)$	$8.588\pm 1\text{E}{-1}$
$(0.02,0.5)$	$8.799\pm 1\text{E}{-1}$	$(0.1,10)$	$8.608\pm 1\text{E}{-1}$
$(0.02,100)$	$8.656\pm 1\text{E}{-1}$	$(0.1,100)$	$8.644\pm 1\text{E}{-1}$
$(0.05,0.02)$	$8.630\pm 1\text{E}{-1}$	$(1,0.1)$	$8.641\pm 1\text{E}{-1}$
$(0.05,0.05)$	$8.584\pm 1\text{E}{-1}$	$(1,1)$	$8.671\pm 1\text{E}{-1}$
$(0.05,0.1)$	$8.382\pm 1\text{E}{-1}$	$(1,5)$	$8.686\pm 1\text{E}{-1}$
$(0.05,0.5)$	$8.545\pm 1\text{E}{-1}$	$(1,10)$	$8.697\pm 1\text{E}{-1}$
$(0.05,100)$	$8.547\pm 1\text{E}{-1}$	$(1,100)$	$8.727\pm 1\text{E}{-1}$
$(0.1,0.02)$	$8.608\pm 1\text{E}{-1}$	$(5,0.1)$	$8.848\pm 1\text{E}{-1}$
$(0.1,0.05)$	$8.478\pm 1\text{E}{-1}$	$(5,1)$	$8.789\pm 1\text{E}{-1}$
$(0.1,0.1)$	$8.487\pm 1\text{E}{-1}$	$(5,5)$	$8.877\pm 1\text{E}{-1}$
$(0.1,0.5)$	$8.543\pm 1\text{E}{-1}$	$(5,10)$	$8.925\pm 1\text{E}{-1}$
$(0.1,100)$	$8.609\pm 1\text{E}{-1}$	$(5,100)$	$8.960\pm 1\text{E}{-1}$
$(0.5,0.02)$	$8.643\pm 1\text{E}{-1}$	$(10,0.1)$	$8.881\pm 1\text{E}{-1}$
$(0.5,0.05)$	$8.460\pm 1\text{E}{-1}$	$(10,1)$	$8.878\pm 1\text{E}{-1}$
$(0.5,0.1)$	$8.654\pm 1\text{E}{-1}$	$(10,5)$	$8.926\pm 1\text{E}{-1}$
$(0.5,0.5)$	$8.618\pm 1\text{E}{-1}$	$(10,10)$	$9.012\pm 1\text{E}{-1}$
$(0.5,100)$	$8.770\pm 1\text{E}{-1}$	$(10,100)$	$9.022\pm 1\text{E}{-1}$
$(100,0.02)$	$8.803\pm 1\text{E}{-1}$	$(100,0.1)$	$8.977\pm 1\text{E}{-1}$
$(100,0.05)$	$8.891\pm 1\text{E}{-1}$	$(100,1)$	$8.910\pm 1\text{E}{-1}$
$(100,0.1)$	$9.046\pm 1\text{E}{-1}$	$(100,5)$	$8.949\pm 1\text{E}{-1}$
$(100,0.5)$	$9.060\pm 1\text{E}{-1}$	$(100,10)$	$9.015\pm 1\text{E}{-1}$
$(100,100)$	$9.143\pm 1\text{E}{-1}$	$(100,100)$	$9.082\pm 1\text{E}{-1}$

EC.4.3. Comparison Between Actual and Robust Models

In this subsection, we present an additional example, aiming to compare the cost of the "best" robust control policy $\hat{\mathcal{C}}(\gamma^\star)$ to the actual optimal cost $\mathcal{C}^\star$. Recall that $\mathcal{C}^\star$ denotes the scenario where we assume the decision-maker has perfect knowledge of the true model.

We use the same cost parameters as the one in §9.3, but with $\tilde{\lambda}=160$ and $\hat{\lambda}=40$, representing a larger variation of the realized demand rates. Each state's sojourn time is again exponentially distributed with a rate of 2.5. Results are reported in Figure EC.5, where we can see that $\mathcal{C}^\star=\eta_{exact}=4.514$ and the difference between $\hat{\mathcal{C}}(\gamma^\star)$ and $\mathcal{C}^\star$ is 3.44%, which is arguably an acceptable difference. If completely ignoring model errors, the decision-maker would incur a cost that is 8.77% higher than the actual optimal cost $\mathcal{C}^\star$. Table EC.5 further shows the optimal outsourcing thresholds and the corresponding simulated costs for the single-class examples discussed in §9.3 and this subsection.

EC.5. Robustness Against Positive Setup Times

In this section, we investigate scenarios where transitioning between the production of different products incurs a positive setup time. We next provide a modified sequencing rule, considering the setup times, and conduct simulation studies to assess its effectiveness.

We illustrate our idea through a two-class system with quadratic cost rates a_1 and a_2, but the idea can be easily extended to multiclass settings. Let $\bar{s}$ denote the setup time. Suppose the server just

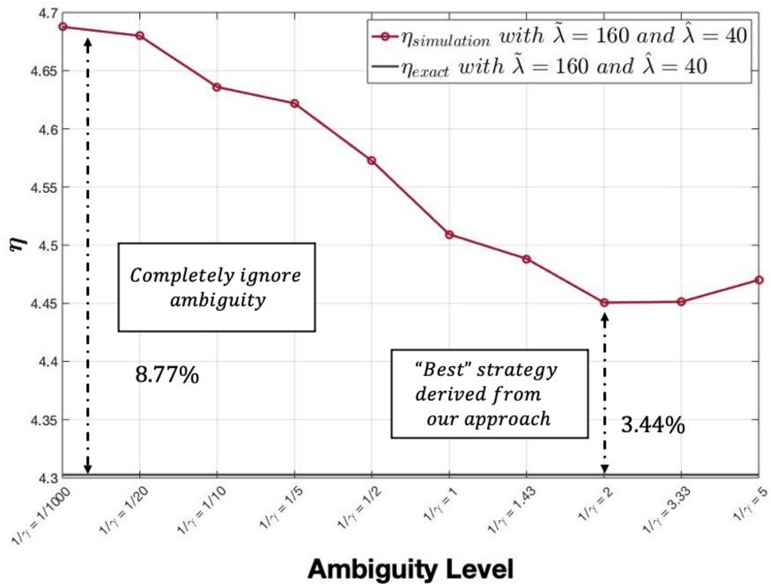


Figure EC.5 **Actual cost and cost obtained from the "best" robust policy with** $\tilde{\lambda} = 160$ **and** $\hat{\lambda} = 40$

Table EC.5 **Outsourcing thresholds and simulated costs for the single-class example**

γ	q	s	$\hat{\mathcal{C}}(\gamma)$	
			$\tilde{\lambda} = 150, \hat{\lambda} = 50$	$\tilde{\lambda} = 160, \hat{\lambda} = 40$
0.2	0.018	0.205	$3.855 \pm 3\text{E}{-2}$	$4.470 \pm 3\text{E}{-2}$
0.3	0.025	0.190	$3.808 \pm 3\text{E}{-2}$	$4.451 \pm 3\text{E}{-2}$
0.5	0.035	0.182	$3.816 \pm 3\text{E}{-2}$	$4.450 \pm 3\text{E}{-2}$
0.7	0.040	0.182	$3.868 \pm 3\text{E}{-2}$	$4.488 \pm 3\text{E}{-2}$
1	0.046	0.184	$3.841 \pm 3\text{E}{-2}$	$4.509 \pm 3\text{E}{-2}$
2	0.053	0.188	$3.937 \pm 3\text{E}{-2}$	$4.572 \pm 3\text{E}{-2}$
5	0.058	0.192	$3.915 \pm 3\text{E}{-2}$	$4.621 \pm 3\text{E}{-2}$
10	0.060	0.193	$3.959 \pm 3\text{E}{-2}$	$4.636 \pm 3\text{E}{-2}$
20	0.061	0.193	$3.971 \pm 3\text{E}{-2}$	$4.680 \pm 3\text{E}{-2}$
1000	0.061	0.194	$3.975 \pm 3\text{E}{-2}$	$4.687 \pm 3\text{E}{-2}$

finishes a class 1 job and sees there are pending jobs in the queues. If following the sequencing rule outlined in the main paper, the production switch occurs if

$$a_1\mu_1 Q_1 < a_2\mu_2 Q_2.$$

In the presence of a setup time, we demand that a switch occur only if

$$a_1\mu_1(Q_1 + \bar{s}\mu_1) < a_2\mu_2 Q_2.$$

Similarly, every moment when the server finishes a class 2 job and sees there are pending jobs in the queues, it switches to produce a class 1 job only if

$$a_1\mu_1 Q_1 > a_2\mu_2(Q_2 + \bar{s}\mu_2).$$

We examine a 2-class MTO system, excluding the need to account for model errors in demand arrivals, to ensure precise comparisons between simulated costs with various setup times and the long-run average cost achieved by solving the Bellman equation. The arrival and service parameters are as follows: $\bar{\lambda}_1 = \bar{\lambda}_2 = 500$ and $\mu_1 = \mu_2 = 1000$. When μ_1 and μ_2 are the same, their mean service times, m_1 and m_2, are equal as well. The cost parameters are specified as $L_1 = 0.5$, $L_2 = 2$, $\ell_1 = 0.2$, and $\ell_2 = 0.5$, with quadratic holding cost rates $a_1 = 0.01$ and $a_2 = 0.05$. By plugging these parameters back into the Bellman equation, we find that $\eta_1 = 10.2512$ and $\eta_2 = 19.0236$. The relation $\eta_1 < \eta_2$ implies that in this example, outsourcing product 1 should always be prioritized over product 2. Moreover, we obtain $q_1 = 0.024$ and $s_1 = 0.045$. With this setup, we can run the simulations. The total duration of each simulation is 500. After comparison, we select the warm-up period to be 50. Computational results are summarized in Figure EC.6, and confidence intervals are reported in Table EC.6.

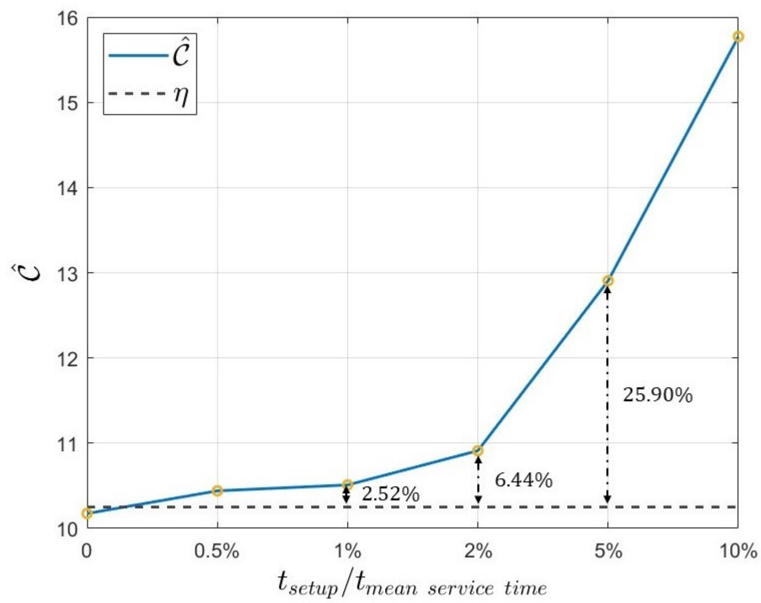


Figure EC.6 **Comparative analysis of simulated costs with positive setup times**

Table EC.6 **Simulated costs for MTO systems with positive setup times**

set up time	*set up time/mean service time*	$\hat{\mathcal{C}}$	*optimality gap*
0	0	$10.175 \pm 1\text{E}{-1}$	0.74%
0.000005	0.5%	$10.441 \pm 1\text{E}{-1}$	1.85%
0.00001	1%	$10.510 \pm 1\text{E}{-1}$	2.52%
0.00002	2%	$10.912 \pm 1\text{E}{-1}$	6.44%
0.00005	5%	$12.906 \pm 1\text{E}{-1}$	25.9%
0.0001	10%	$15.771 \pm 1\text{E}{-1}$	53.8%

Based on Figure EC.6 and Table EC.6, it becomes evident that if the ratio between the setup and the mean service time remains below 2%, the difference between the simulated cost $\hat{\mathcal{C}}$ and the long-run average cost η remains within an acceptable margin, specifically less than 7%. However, as the ratio between the setup and the mean service time increases, the gap between $\hat{\mathcal{C}}$ and η also increases. When the ratio equals 5%, this difference reaches 25.9%. Thus, we can conclude that our solution approach remains effective only when setup times are relatively short.

EC.6. Numerical Results for a Multi-Product System

This section presents an MTO system with four classes, which serves as an example to demonstrate the effectiveness of our quasi-gradient-descent method in enhancing performance. This 4-class example also serves as an extension of the 2-class examples presented in §9.2. It is worth mentioning that, unlike the comparison in §9.3 and §EC.4.3, this section focuses on comparing the value of robustness, rather than analyzing the difference between the cost of the "best" robust policy and actual optimal cost.

The arrival model of each product follows a CTMC, similar to §9.2.2. In the CTMC, the sojourn time of each state is exponentially distributed with a rate of 2. The system parameters include $\bar{\lambda}_1 = 90$, $\bar{\lambda}_2 = 60$, $\bar{\lambda}_3 = 50$, $\bar{\lambda}_4 = 50$, $\mu_1 = 150$, $\mu_2 = 300$, $\mu_3 = 500$, and $\mu_4 = 500$. Furthermore, we set $\tilde{\lambda}_1 = 150$, $\hat{\lambda}_1 = 30$, $\tilde{\lambda}_2 = 100$, $\hat{\lambda}_2 = 20$, $\tilde{\lambda}_3 = 80$, $\hat{\lambda}_3 = 20$, and $\tilde{\lambda}_4 = 80$, $\hat{\lambda}_4 = 20$. The cost data includes fixed outsourcing cost parameters $L_1 = 2$, $L_2 = 3$, $L_3 = 4$, and $L_4 = 5$, proportional outsourcing cost parameters $\ell_1 = 0.2$ and $\ell_2 = \ell_3 = \ell_4 = 0.5$, and quadratic holding cost rates $a_1 = 0.01$, $a_2 = 0.1$, $a_3 = 0.2$, and $a_4 = 0.2$. The total duration of the experiment is 2000. We again set $T_1 = 50$ and $T_2 = 100$ as the warm-up periods. Statistical estimates of the long-run average simulated cost are compared, and no significant difference is observed between the estimates under the two candidate warm-up periods. Therefore, $T_2 = 100$ is selected as the warm-up period. The desired width of the confidence interval in this example is set to 0.2.

In the 4-class example, when $\alpha = 2$, the value of robustness is approximately $9.6714 - 9.1028 = 0.5686$, which is 6.25% better than completely ignoring model errors. Additionally, the performance of our method is highlighted in Figure EC.7, showing that near-minimum value can be achieved within about 20 iterations. On the other hand, if we were to use the exhaustive method, taking only 5 values for each γ_i would require conducting 5^4 simulations, resulting in significantly more intensive computation. Thus, we believe that the quasi-gradient algorithm can deliver significant practical value.

EC.7. Numerical Results for a Multi-Server System

To further substantiate the claim in Remark 1, we present simulation examples to demonstrate that an N-server system will exhibit roughly the same behavior as a system with a single server that operates N times faster than each of the original servers under heavy-traffic conditions.

First, we consider a system featuring a single "super" server. Similar to §EC.5, we omit the consideration of model errors in demand to accurately compare the long-run average cost derived

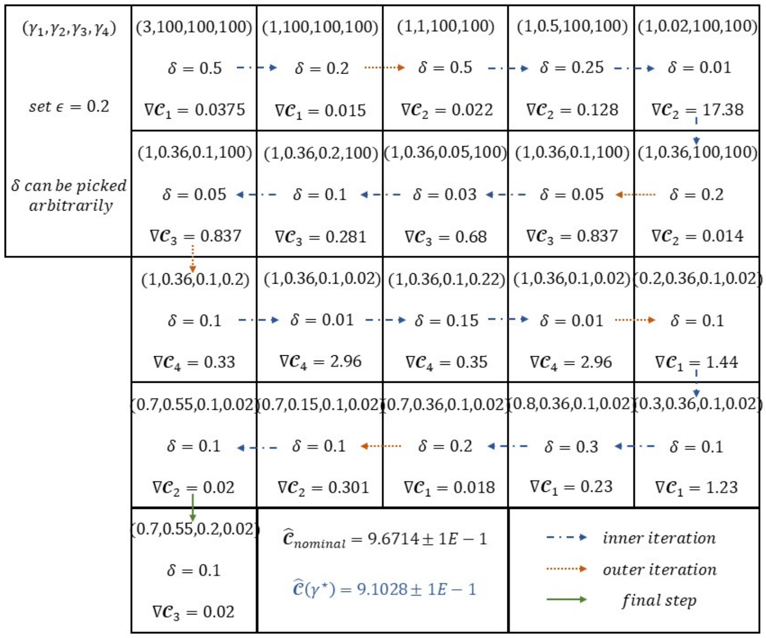


Figure EC.7 Quasi-gradient-descent algorithm for the 4-class example

from our approach and simulations. We set $\bar{\lambda}_1 = 400$, $\bar{\lambda}_2 = 200$, $\mu_1 = 500$, and $\mu_2 = 1000$, and consider the following cost parameters: $L_1 = 1$, $L_2 = 2$, $\ell_1 = \ell_2 = 0.5$, and quadratic holding cost rates $a_1 = 0.02$ and $a_2 = 0.04$. By plugging the parameters back into the Bellman equation with sufficiently large values for γ_1 and γ_2, we observe that $\eta_1 = 14.180$ and $\eta_2 = 22.147$. The relation $\eta_1 < \eta_2$ indicates that we should always outsource product 1 in this example. We also obtain $q_1 = 0.038$ and $s_1 = 0.073$. The sequencing rule remains the generalized $c\mu$ rule.

We next consider a slew of multi-server systems where we maintain all parameters and control rules identical to the single-server system, except that we scale down μ_1 and μ_2 by N, where N is the number of servers. For example, when $N = 2$, each server's service rate for product 1 is 250, and for product 2, it is 500. The total duration of each simulation is 1000. After comparison, we select the warm-up period to be 100. Computational results are summarized in Figure EC.8, and confidence intervals are reported in Table EC.7.

Table EC.7 Simulated costs for multi-server systems

number of servers	$\hat{\mathcal{C}}$	*optimality gap*
1	$14.046 \pm 2E{-}1$	0.95%
2	$14.130 \pm 2E{-}1$	0.35%
3	$14.520 \pm 2E{-}1$	2.40%
4	$14.720 \pm 2E{-}1$	3.81%
5	$15.049 \pm 2E{-}1$	6.13%
6	$15.473 \pm 2E{-}1$	9.12%
7	$15.846 \pm 2E{-}1$	11.75%
8	$16.476 \pm 2E{-}1$	16.20%

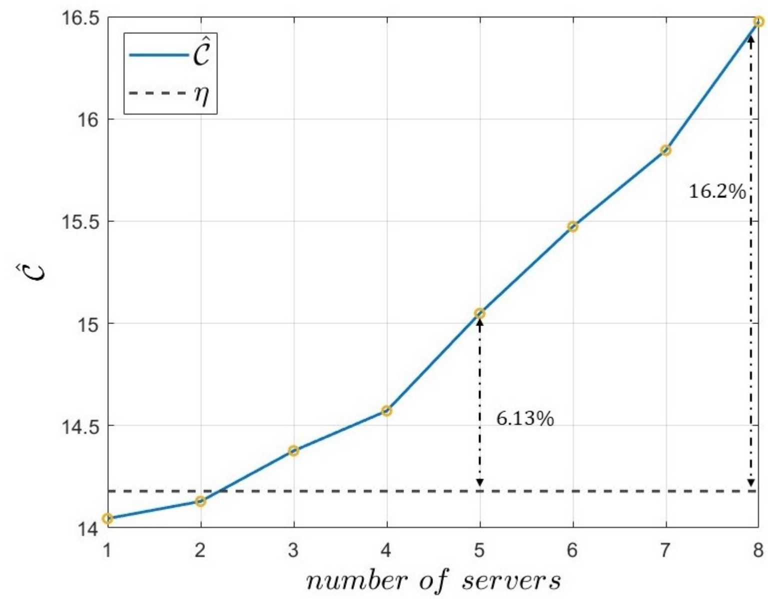


Figure EC.8 Comparative analysis of simulated costs in multi-server systems

We can observe from Figure EC.8 that when the number of servers is relatively small, as is the case when $N<5$ in this example, the single-server-approximation can yield remarkable accuracy, with a gap of typically less than 4%. As the number of servers increases, the gap between the simulated cost $\hat{\mathcal{C}}$ and the long-run average cost η derived from our Bellman equation grows. For example, for $N=7$, the simulated cost is approximately $15.846 \pm 2\text{E}{-1}$, which is 11.7% larger than η. Similarly, for $N=8$, the simulated cost is approximately $16.476 \pm 2\text{E}{-1}$, representing a 16.2% increase over η. We view the results as highly expected because, when N is small, the conventional heavy-traffic regime, as discussed in our main paper, is known to be effective. However, as N increases, the system gradually transitions from the conventional heavy-traffic regime to the many-server heavy-traffic regime. Taken together, we can conclude that when the number of servers is not significantly large, under conventional heavy-traffic conditions, a single-server system can effectively approximate a multi-server system.

EC.8. A Few Direct Extensions

We now present a few immediate extensions to the model introduced in the main paper.

EC.8.1. The Choice of Discrepancy Measure

Rényi divergence provides flexibility in the uncertainty set construction via a single function r that dictates how nature is penalized based on her actions. Importantly, this approach to constructing an uncertainty set does not require a specific penalty rate function. The decision-maker can specify any penalty form, as long as the resulting uncertainty set can allay fear of model errors.

Let $p(\cdot)$ be a penalty function mapping from $\mathbb{R}^I$ to $\mathbb{R}_+$, such that when nature selects $\theta(t)$ at time t, the decision-maker incurs a penalty at the rate of $p(\theta(t))$. This motivates a general robust control formulation, described below. The robust control problem for the penalty function $p(\cdot)$ can be formulated as follows: the decision-maker seeks an adapted strategy (T, Ψ) that minimizes

$$\max_{\theta} \limsup_{t\to\infty} \frac{1}{t}\mathbb{E}^{\theta}\left[\int_0^t \left(\sum_{i=1}^{I} c_i(Q_i(u)) - p(\theta(u))\right) \mathrm{d}u + \sum_{i=1}^{I}\sum_{k=0}^{N_i(t)} \phi_i(\xi_i(k))\right].$$

Replacing p with r recovers problem (8). Following the development in §6.1, we arrive at an approximating SDG, in which the decision-maker chooses $(\hat{Y}, U, \hat{\Psi})$ to minimize

$$\begin{aligned}&\max_{\theta} \limsup_{t\to\infty} \frac{1}{t}\mathbb{E}^{\theta}\left[\int_0^t \left(\sum_{i=1}^{I} c_i(\hat{Q}_i(u)) - p(\theta(u))\right) \mathrm{d}u + \sum_{i=1}^{I}\sum_{k=0}^{N_i(t)} \phi_i(\hat{\xi}_i(k))\right]\\ &\text{subject to constraints } (12)-(13).\end{aligned}$$

By following similar lines of argument to those in §6.2, we can obtain the corresponding workload problem, in which the decision-maker seeks adaptive control $(U, \tilde{\Psi})$ that minimizes

$$\begin{aligned}&\max_{\zeta} \limsup_{t\to\infty} \frac{1}{t}\mathbb{E}^{\zeta}\left[\int_0^t h(W(u))\mathrm{d}u - \int_0^t p^\star(\zeta(u))\mathrm{d}u + \sum_{i=1}^{I}\sum_{k=0}^{N_i(t)} \tilde{\phi}_i(\tilde{\xi}_i(k))\right]\\ &\text{subject to constraints } (21)-(23),\end{aligned}$$

where $p^\star(z) = \min\{p(y) : \rho^\top y = z, y_i \in \Theta_i\}$. At this point, we would like to emphasize that only two properties of $r^\star$ are critical to the proofs of Proposition 5 and Theorem 1. First, $r^\star$ attains its minimum value at $z = 0$ with $r^\star(0) = 0$. Second, its convex conjugate is non-negative and Lipschitz continuous. Consequently, all the analytical results established in the preceding section apply to all $p^\star$ possessing these two properties.

EC.8.2. General Service Times

In the main paper, we have assumed that service times are exponentially distributed. As far as heavy-traffic analysis is concerned, this assumption can be effortlessly relaxed to allow for general service time distributions without affecting the main results established in Section 6. Indeed, with general service times, the process $S_i(t)$ that represents the number of class i products manufactured over time if the server was constantly working on class i orders can be viewed as a renewal process with cycles having mean m_i and coefficient of variation ν_i. As a result, $\hat{S}_i$ in (11) becomes a Brownian motion with zero drift and variance parameter $\bar{\lambda}_i\nu_i^2$, whereas $\hat{Z}_i$ in (12) becomes a Brownian motion with zero drift and infinitesimal variance $\sigma_i^2 = \bar{\lambda}_i(1+\nu_i^2)$.

EC.8.3. (Slightly) "Imbalanced" Systems

In this subsection, we demonstrate how to expand our analysis to a more general scenario by relaxing the assumption of "critical-loading" in Equation (9), allowing for an "imperfectly balanced" system. Specifically, we examine a situation where the capacity does not match the supply and nominal demand level exactly. This can be expressed as:

$$\sum_{i=1}^{I} \rho_i = 1 - \omega \quad \text{for} \quad \rho_i = \bar{\lambda}_i m_i, \quad i = 1, \dots, I,$$

where the constant ω measures the extent to which the capacity exceeds the nominal demand volume. However, we would need ω to be a value of the order of $1/\sqrt{n}$, where n is a parameter that reflects the system scale. (In the main paper, we mention that n can be taken as $\sum_i \bar{\lambda}_i$.) This ensures that the imbalance between capacity and demand, albeit present, is moderate at best.

With this relaxation, the corresponding SDG is modified to minimize (15) subject to (12), (13), and the constraint that

$$\tilde{U} := \sum_i \hat{Y}_i(t) + \omega t \text{ is non-decreasing with } \tilde{U}(0) = 0.$$

As a result, the workload process, again denoted as W, satisfies

$$W(t) = W(0) + B(t) + \int_0^t (\zeta(u) - \omega)\mathrm{d}u + \tilde{U}(t) - O(t),$$

where B, ζ, and O are defined as in (17). Consequently, the Bellman-Isaacs condition associated with the workload problem becomes finding (v, η) that satisfies

$$\min\left\{\frac{1}{2}\sigma^2 v''(w) + g(v'(w)) - \omega v'(w) + h(w) - \eta,\ \min_i \inf_{z\geq 0}\left[v(w-z) + \tilde{\phi}_i(z)\right] - v(w)\right\} \geq 0 \quad \text{(EC.37)}$$

subject to $v'(0) = 0$.

It is noteworthy that we do not require a complete overhaul of the entire analysis for the new setting. To demonstrate this, we can define $\tilde{g}(\cdot) := g(\cdot) - \omega\cdot$. Consequently, Equation (EC.37) will resemble that in (33), and $\tilde{g}$ will possess all the essential properties of g's that are necessary for the mathematical proofs to hold.

On a related note, a control policy based on a heavy-traffic approximation may remain effective even when the actual operating regime deviates considerably from the critical-loading assumption. To illustrate, we present a single-class example without accounting for possible model errors, where we set $\lambda = 300$ and $\mu = 100$, $L = 0.5$, $\ell = 0.1$, and a quadratic holding cost rate function $c(x) = 0.01x^2$. Such a parameter choice implies that the system is overloaded ($\omega = -2$). Absent the role of nature, the SDG simplifies to a BCP, which we can solve for the control band thresholds for the queue length denoted as q and s (by a slight abuse of notation). Since the system is overloaded, existing literature on queuing approximation and control suggests that the system likely evolves like a fluid. In particular, we expect the queue length to increase approximately linearly at a rate of $\lambda - \mu$, which suggests an "EOQ" formula as a result of this deterministic approximation. Indeed, a direct application of the "EOQ" formula (with a slight yet straightforward modification) can give us another upper boundary, at which the decision-maker would like to push the queue length (through outsourcing) to a lower barrier. As outsourcing operations are assumed to occur without delay, the application of the "EOQ" formula should produce a lower barrier of zero. The blue line in Figure EC.9 depicts the sample path of the queue length $Q(t)$, whereas the green and red lines indicate the control band parameters, q and s, respectively, obtained from the BCP. The dash-dotted gray line represents the control band parameter computed from the "EOQ" formula. The plot suggests that the two solutions, one based on Brownian approximation and the other on deterministic approximation, match up, suggesting that a control policy derived from a Brownian approximation may still be practically valuable even if the conditions required to justify its use are not met.

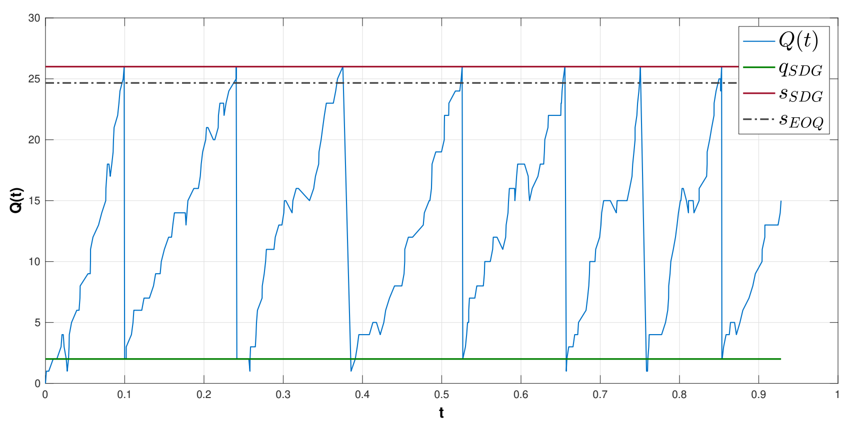


Figure EC.9 **Sample path of the queue length of an overloaded single-class system**

EC.8.4. Incorporating Positive Lead Times for Outsourcing

In the main paper, we operate under the assumption of instantaneous production at the subcontractor. We now explore a possible modeling approach to relaxing this assumption by considering the subcontractor's finite production capacity. It is important to note that our goal is not to present a full extension that incorporates the model in the main paper as a special case, but rather to demonstrate the feasibility of modeling to capture additional realism should one consider finite production capacity at the subcontractor. (In fact, the formulation below will not subsume the model in the main paper because their cost structures for outsourcing are slightly different.)

Noting that impulse control in a production environment implicitly assumes infinite production capacity and "real-world production/inventory systems all have finite capacity," Wu and Chao (2014) study a continuous-review production system. In this system, the net inventory process is modeled as a Brownian motion having two operating modes (due to finite production capacity). They frame the control problem within this context as an optimal switching problem. It has long been recognized that, compared to impulse control, an optimal switching problem is generally more difficult to solve because of the need to track the status of the control (Wu and Chao 2014).

To incorporate non-zero "service times" at the subcontractor, it is instructive to conceptualize the subcontractor as a secondary server capable of switching between two modes: on and off. These two modes correspond to the decisions to initiate or halt the outsourcing of orders, respectively. Following the modeling approach in Wu and Chao (2014), we can assume that activating the secondary server entails a fixed cost L and deactivating it is free. It has been known in the literature on optimal switching that the case where deactivation also incurs a fixed cost, say L', is mathematically equivalent to having a fixed cost of $L+L'$ for activation and zero cost for deactivation. Moreover, while operational, this server accrues a usage cost at a rate of ℓ, mirroring but differing from the proportional cost structure in the impulse control setting. Note that we might be able to capture additional realism by allowing different products to have different "usage costs." However, doing so will make the resulting problem more complex, blurring the main message we try to convey here.

To incorporate this idea into a multiclass MTO scenario, we can follow the development in Sun et al. (2024) by assuming that the secondary server has a service rate of υ_i for product i. To draw a parallel, one can think of the proportional outsourcing cost ℓ_i introduced in the main paper as being $\ell\upsilon_i$, given that ℓ represents the per-time-unit outsourcing cost in the present context. Next, by keeping all other model ingredients the same as in the main paper and adapting the derivations in Sun et al. (2024) to our robust control setting, we will arrive at the following SDG: The decision-maker seeks adaptive control $(\hat{Y},(G_i),H)$ to minimize

$$\begin{aligned}
&\max_{\theta}\ \limsup_{t\to\infty}\frac{1}{t}\mathbb{E}^{\theta}\left[\int_0^t\left(\sum_{i=1}^{I}c_i(\hat{Q}_i(u))-r(\theta(u))\right)\mathrm{d}u+\ell\int_0^t H(u)\mathrm{d}u+L\sum_{u\le t}[\Delta H(u)]^+\right]\\
&\text{s.t. } \hat{Q}_i(t)=\hat{Q}_i(0)+\hat{Z}_i(t)+\int_0^t\bar{\lambda}_i\theta_i(u)\mathrm{d}u+\mu_i\hat{Y}_i(t)-\upsilon_i\int_0^t G_i(u)\mathrm{d}u,\quad i=1,\dots,I,\\
&\qquad \hat{Q}_i(t)\ge 0\quad\text{for}\quad t\ge 0,\quad i=1,\dots,I,\\
&\qquad G_1(t),\dots,G_I(t),H(t)\in\{0,1\}\quad\text{for}\quad t\ge 0,\\
&\qquad \sum_i G_i(t)\le H(t)\quad\text{for}\quad t\ge 0,\quad\text{and}\\
&\qquad \sum_i\hat{Y}_i(t)+\int_0^t\left(H(u)-\sum_i G_i(u)\right)\mathrm{d}u\text{ is non-decreasing with }\sum_i\hat{Y}_i(0)=0.
\end{aligned}\tag{EC.38}$$

Note that Problem (EC.38) differs from the SDG (12)–(15) in that it involves additional control processes (G_i) and H. Intuitively, process H tracks the status of the secondary server; $H(t)=1$ if the server is operational (but could potentially be idle) at time t and $H(t)=0$ otherwise. Similarly, $G_i(t)$ is an indicator assuming the value of one if the server is utilized to process class i orders at time t and zero otherwise.

To effectively reduce Problem (EC.38) to a lower-dimensional equivalent workload formulation, we need additional assumptions regarding the service rates (υ_i). In particular, we would require (υ_i) to be proportional to (μ_I), so there exists some constant $\bar{\upsilon}$ such that

$$\frac{\upsilon_i}{\mu_i}=\bar{\upsilon} \quad \text{for} \quad i=1,\dots,I. \tag{EC.39}$$

In studying a joint pricing and capacity adjustment problem, Sun et al. (2024) makes a similar assumption in order to devise an approximate control policy. In addition to offering analytical tractability, assumption (EC.39) also makes practical sense. If a product requires more production time from the in-house server, it should consume more machine time on the secondary server, since the length of time it takes to manufacture a product is typically determined by its inherited characteristics.

Let W,B, and ζ be defined as before and redefine "idleness process" by

$$U(t):=\sum_i \hat{Y}_i(t)+\bar{\upsilon}\int_0^t \left(H(u)-\sum_i G_i(u)\right)\mathrm{d}u.$$

From (EC.38) and utilizing (EC.39), we arrive at the following workload problem: The decision-maker seeks adaptive control (U,H) to minimize

$$\begin{aligned}
&\max_{\zeta}\ \limsup_{t\to\infty}\frac{1}{t}\mathbb{E}^{\zeta}\left[\int_0^t \left(h(W(u))-r^{\star}(\zeta(u))\right)\mathrm{d}u+\ell\int_0^t H(u)\mathrm{d}u+L\sum_{u\le t}[\Delta H(u)]^{+}\right]\\
&\text{s.t. } W(t)=W(0)+B(t)+\int_0^t \zeta(u)\mathrm{d}u-\bar{\upsilon}\int_0^t H(u)\mathrm{d}u+U(t), \qquad\qquad \text{(EC.40)}\\
&\qquad U(t) \text{ is non-decreasing with } U(0)=0, \text{ and}\\
&\qquad W(t)\ge 0 \quad \text{for} \quad t\ge 0.
\end{aligned}$$

Unlike the workload problem in the main paper, Problem (EC.40) is a two-dimensional differential game, where H serves as both state and control processes. Letting $v(y,w)$ denote the relative value function and the function g be defined as before, with reference to the general control theory, we expect v, along with some constant η, to satisfy

$$\begin{aligned}
\min\Bigg\{&\frac{1}{2}\sigma^2 v_{ww}(y,w)-\bar{\upsilon}v_w(y,w)+g(v_w(y,w))+h(w)+\ell y-\eta,\\
&v(1,w)+L-v(0,w),\ v(0,w)-v(1,w)\Bigg\}=0
\end{aligned} \tag{EC.41}$$

Although the techniques we develop for the main model are not applicable to establishing the well-posedness of (EC.41), some of the technical tools developed in Sun et al. (2024) may be useful for this purpose, given that (EC.41) and the optimality equation in Sun et al. (2024) are structurally similar.

EC.9. Further Discussion on the Selection of the Warm-up Period

In the main paper, we present an easy-to-implement yet somewhat naive method for determining the warm-up period. Therein we compare two candidate warm-up periods, $T_1 = 50$ and $T_2 = 100$, to determine if there exists a noticeable difference between the simulation estimates. There are, however, other well-adopted methods in the literature for selecting the warm-up period; refer to Robinson and Ioannou (2007) for a comprehensive summary. In this section, we contrast various methods that are applicable to our problem for assessing the impact of warm-up period selection. All selected approaches exhibit relatively good performance, as illustrated in Table 1 of Robinson and Ioannou (2007). We use a system similar to that in §9.2.2 with $\alpha = 1$ and $\gamma_1 = \gamma_2 = 1$, wherein the cost data includes outsourcing costs $L_1 = 0.5$, $L_2 = 0.8$, $\ell_1 = \ell_2 = 0.2$, and quadratic holding cost rates $a_1 = 0.01$ and $a_2 = 0.02$. For the CTMC, each state's sojourn time is exponentially distributed at a rate of 5, with $\tilde{\lambda}_1 = 150$, $\tilde{\lambda}_2 = 110$, $\hat{\lambda}_1 = \hat{\lambda}_2 = 10$, while $\mu_1 = 100$ and $\mu_2 = 300$. It can be achieved that $q_1 = 0.052$ and $s_1 = 0.187$.

The first approach we want to explore is time-series inspection, which can be seen as more general compared to the one utilized in the main paper. Using this method, the warm-up period is determined by examining the time series of the queue length. The length of the warm-up period is chosen as the point where the time series seems to stabilize around a consistent mean. Specifically, by a slight abuse of notation, we consider I candidate warm-up periods, denoted as $T_1 < T_2 < \cdots < T_I$. Each iteration involves selecting T_i as the warm-up period, followed by recording the queue lengths $Q_1(T_i)$ and $Q_2(T_i)$ corresponding to T_i. In our examples, we set $T_i = i \times 5$. The number of simulation replications is 300. Without specific clarification, in the following context, $Q_1(T_i)$ and $Q_2(T_i)$ represent the sample mean of each queue length over 300 replications. Computational results are reported in Table EC.8. It is evident from Table EC.8 that when the warm-up period is equal to or exceeds $T_4 = 20$, the simulated queue lengths remain relatively stable around their respective means. Specifically, after $i = 4$, the mean of the time series for Q_1 is 4.880, while for Q_2, it is 0.984. This outcome further validates the effectiveness of the chosen warm-up period $T = 100$ in our simulations, as discussed in the main paper.

Table EC.8 Computational results for the selection of the warm-up period

i	T_i	$Q_1(T_i)$	$Q_2(T_i)$	$O_{\mathrm{MCR}}(i)$
1	5	3.773	0.753	0.0436
2	10	4.777	0.913	0.0335
3	15	4.497	0.803	0.0445
4	20	4.887	0.873	0.0559
5	25	4.926	0.920	0.0871
6	30	4.875	1.047	0.1545
7	35	4.900	1.070	0.2970
8	40	4.813	1.011	1.1482

We can extend the simple time-series inspection to other similar approaches, such as ensemble average plots, the cumulative-mean rule, and the deleting-the-cumulative-mean rule. The performance of these approaches closely resembles that of time-series inspection; therefore, we omit detailed discussion of them in this section.

The second approach we want to test involves the forward data-interval rule and the backward data-interval rule. In our examples, applying the forward data-interval rule to a time-series of I observations of Q_1 and Q_2 entails identifying the warm-up period as the first point in the series that is neither the maximum nor the minimum of the remaining observations. Conversely, employing the backward data-interval rule suggests selecting the warm-up period as the last point that is neither the maximum nor the minimum of all the preceding observations of both Q_1 and Q_2. We can observe from Table EC.8 that applying the forward data-interval rule results in a warm-up period of 30, while applying the backward data-interval rule results in a warm-up period of 15. Once again, applying these two approaches can demonstrate the effectiveness of the selected warm-up period discussed in the main paper.

Another approach is the so-called marginal confidence rule (MCR), which requires the following judgment: if the initial observations deviate significantly from the sample mean and exert a notable influence on the calculation of confidence intervals, then those observations should be removed. In our context, for a series of observations, $Q_1(T_i)$ and $Q_2(T_i)$, we can choose the warm-up period at the point d that minimizes the following expression:

$$\frac{1}{(I-d)^2}\sum_{i=d+1}^{I}\left[\left(Q_1(T_i)-Q_1(T_{I-d})\right)^2+\left(Q_2(T_i)-Q_2(T_{I-d})\right)^2\right].$$

Denote the objective of the above equation for each i as $O_{\mathrm{MCR}}(i)$. The details of these results are presented in Table EC.8. It is evident that setting the warm-up period to 10 yields the minimum objective of MCR. This reaffirms the conclusion that the selected warm-up period discussed in the main paper is sufficiently long to ensure reliable performance.

EC.10. Further Discussion on the Choice of Parameter α

In the main paper, we focus on varying γ to create uncertainty sets that reflect diverse ambiguity levels and demonstrate that the choice directly impacts the control rule derived for the decision-maker. Here, we first offer some clue, based on our intuition, as to when a larger or smaller value of α may produce more benefits. We then conduct some numerical experiments, which serve as initial validation for our intuition. We should note, however, that we do not claim that the numerical insights are fully generalizable, as they are based on limited test cases.

The system parameters as follows: $\bar{\lambda}_1 = 200$, $\mu_1 = 250$, $\bar{\lambda}_2 = 100$, and $\mu_2 = 500$. The cost parameters are fixed outsourcing costs of $L_1 = 0.5$ and $L_2 = 0.8$, proportional outsourcing costs of $\ell_1 = \ell_2 = 0.2$, and quadratic holding cost rates of $a_1 = 0.01$ and $a_2 = 0.02$. Similar to §9.2.2 in the main paper, we use an arrival model where the demand rate for each order follows a non-homogeneous Poisson process with a CTMC intensity. Below, we present two scenarios: one where the actual demand rate of class 1 is moderately higher than the average value of $\bar{\lambda}_1 = 200$ most of the time but can occasionally drop to a very low value, and another where the actual demand rate of class 1 is moderately lower than the average value of $\bar{\lambda}_1 = 200$ most of the time but can occasionally rise to a very high value.

When the actual arrival rate is moderately higher than the average value for most of the time but significantly lower for a small portion of the time, we expect that a larger value of α in the Rényi divergence can slightly benefit the decision-maker. This is because a larger value of α penalizes the right tail of the demand rate distribution (i.e., values that are greater than the nominal demand rate) more heavily than the left tail (i.e., values that are smaller than the nominal demand rate). In contrast, when the actual demand rate is moderately lower than its nominal value for most of the time but significantly higher for a small portion of the time, we expect that the decision-maker can benefit more from choosing a smaller value for α. This is because a smaller value of α would penalize the left tail more heavily, reflecting the decision-maker's prior belief that larger demand rates are more likely than smaller ones.

To validate our intuition, we consider two scenarios, both assuming CTMC intensity. In the first scenario, we set $\tilde{\lambda}_1 = 260$ and let its sojourn time be exponentially distributed with a rate of 5. We set $\hat{\lambda}_1 = 20$ and let its sojourn time be exponentially distributed with a rate of 15. Furthermore, we set $\tilde{\lambda}_2 = \hat{\lambda}_2 = 100$, so that the demand rate of class 2 stays at the average value $\bar{\lambda}_2 = 100$ at all times. The optimal outsourcing rule is always to outsource product 1, and the optimal sequencing rule is the generalized $c\mu$ rule. Under this setting, if the decision-maker does not account for model errors and uses the outsourcing thresholds obtained from the nominal model, the long-run average simulated cost is approximately $12.394 \pm 1\text{E}{-1}$. However, if the decision-maker uses KL divergence and varies γ, the lowest simulated cost is around $11.753 \pm 1\text{E}{-1}$, resulting in a 5.17% difference. If the decision-maker uses Rényi divergence with $\alpha = 5$, the lowest simulated cost is approximately $11.581 \pm 1\text{E}{-1}$, yielding a 6.56% difference. So, using a Rényi divergence with a larger value of α in this scenario seems to deliver slightly better performance. In the second scenario, we set $\tilde{\lambda}_1 = 380$ with the sojourn time exponentially distributed with a rate of 15, while $\hat{\lambda}_1 = 140$ with the sojourn time exponentially distributed with a rate of 5. Similarly, we set $\tilde{\lambda}_2 = \hat{\lambda}_2 = 100$, so that the demand rate of class 2 stays at the average value $\bar{\lambda}_2 = 100$ at all times. In this case, if the decision-maker does not consider model errors and uses the outsourcing thresholds obtained from the nominal model, the long-run average simulated cost is approximately $11.306 \pm 1\text{E}{-1}$. If using KL divergence and varying γ, the lowest simulated cost is around $10.486 \pm 1\text{E}{-1}$, giving a 7.25% difference. If, however, the decision-maker uses Rényi divergence with $\alpha = 1/2$, the lowest simulated cost is approximately

$10.269 \pm 1\text{E}{-1}$; the difference is 9.17%. Therefore, in this scenario, using Rényi divergence with a smaller α seems to deliver a slightly better performance.

The two scenarios above illustrate that the parameter α allows for potential cost savings when the real-world model deviates from the nominal model. However, it seems clear that, in comparison to choosing the value of γ, the choice of α appears to have limited potential for improving cost savings.

EC.11. Proof of Auxiliary Results

Proof of Lemma EC.1. The non-negativity of g is immediate. We then need to establish the Lipschitz continuity of g. For any $x_1, x_2 \in \mathbb{R}$ with $x_1 < x_2$, let $\zeta_1, \zeta_2 \in [\rho^\top a, \rho^\top b]$ be the corresponding maximizers for $g(x_1)$ and $g(x_2)$. The definition of g yields two inequalities:

$$g(x_1) = x_1\zeta_1 - r^\star(\zeta_1) \geq x_1\zeta_2 - r^\star(\zeta_2) \quad \text{and} \quad g(x_2) = x_2\zeta_2 - r^\star(\zeta_2) \geq x_2\zeta_1 - r^\star(\zeta_1).$$

Combining the two inequalities gives us

$$(x_1 - x_2)\zeta_2 \leq g(x_1) - g(x_2) \leq (x_1 - x_2)\zeta_1,$$

establishing the desired Lipschitz continuity of g. $\square$

Proof of Lemma EC.2. For part (i), since g is Lipschitz continuous and h is continuous, we can invoke the Picard–Lindelöf theorem to conclude that there exists a unique continuous solution $\pi(w, \eta)$ to (29) on the interval $[0, \infty)$.

For part (ii), to show the continuity of $\pi(w, \eta)$ in $\eta \in \mathbb{R}$ and the continuity of $\pi'(w, \eta)$ in $w \in \mathbb{R}^+$ and $\eta \in \mathbb{R}$, we can refer to Lemma 5 in Cao and Yao (2018), along with part (i) of the present lemma and the continuity of h, g, and π. $\square$

Proof of Lemma EC.3. We first argue that, if $\eta_1 < \eta_2$, then $\pi(w, \eta_1) < \pi(w, \eta_2)$ for any fixed $w > 0$. To that end, suppose for the sake of contradiction that $\pi(w, \eta_1) > \pi(w, \eta_2)$ for some $w > 0$. By a slight abuse of notation, we define:

$$f(w) := \pi(w, \eta_2) - \pi(w, \eta_1) \quad \text{and} \quad w_0 := \inf\{w > 0 : f(w) \leq 0\}.$$

It is clear that $w_0 > 0$ due to the fact that $f'(0) > 0$. It follows from the definition and continuity of π that $f(w_0) = 0 = f(0)$ and $f(w) > 0$ for all $w \in (0, w_0)$. By the continuity of $f(w)$ around w_0, there exist two real numbers $w_1, w_2 \in (0, w_0)$ with $w_1 < w_2$ such that

$$f(w_1) > f(w_2) \quad \text{and} \quad Mf(w) < \eta_2 - \eta_1 \quad \text{for all} \quad w \in [w_1, w_2]. \tag{EC.42}$$

It is clear from (29) that

$$\frac{1}{2}\sigma^2 f'(w) + g(\pi(w, \eta_2)) - g(\pi(w, \eta_1)) = \eta_2 - \eta_1. \tag{EC.43}$$

Integrating (EC.43) from w_1 to w_2 yields that

$$\begin{aligned}
&(\eta_2-\eta_1)(w_2-w_1)\\
&\quad=\frac{1}{2}\sigma^2(f(w_2)-f(w_1))+\int_{w_1}^{w_2}\left[g(\pi(w,\eta_2))-g(\pi(w,\eta_1))\right]\mathrm{d}w\\
&\quad<\int_{w_1}^{w_2}\left[g(\pi(w,\eta_2))-g(\pi(w,\eta_1))\right]\mathrm{d}w\\
&\quad\leq\int_{w_1}^{w_2}Mf(w)\mathrm{d}w\\
&\quad<(\eta_2-\eta_1)(w_2-w_1),
\end{aligned}\tag{EC.44}$$

where the first and last inequalities follow from (EC.42) and the second inequality is due to (EC.13). Equation (EC.44) yields a contradiction. Therefore, $\pi(w,\eta_1)<\pi(w,\eta_2)$ holds for any fixed $w>0$, if $\eta_1<\eta_2$.

Next, we show that $\lim_{\eta\to\infty}\pi(w,\eta)=\infty$ for any given $w>0$. Note that there must exist a number $\hat{\eta}$ (dependent on w) such that for all $\eta\geq\hat{\eta}$,

$$\eta>h(w).\tag{EC.45}$$

We claim that for any fixed $y\in(0,w)$,

$$\pi(y,\eta)\geq 0\quad\text{for all}\quad\eta\geq\hat{\eta}.\tag{EC.46}$$

To prove this claim, suppose for the sake of contradiction that (EC.46) is not true. Then, there must exist some $z\in(0,w)$ such that $\pi(z,\eta)=0$ and $\pi'(z,\eta)<0$. It follows that

$$0>\frac{1}{2}\sigma^2\pi'(z,\eta)=\eta-h(z)>\eta-h(w)>0,$$

which is a contradiction. Therefore, (EC.46) holds, which in particular implies that for all $y\in(0,w)$

$$\frac{1}{2}\sigma^2\pi'(y,\eta)+g(0)+M\pi(y,\eta)\geq\eta-h(y)\quad\text{for all}\quad\eta\geq\hat{\eta}.$$

It follows that

$$\pi(w,\eta)\geq\frac{2}{\sigma^2}\int_0^w\left[\eta-h(y)\right]e^{-\frac{2M}{\sigma^2}y}\mathrm{d}y\quad\text{for all}\quad\eta\geq\hat{\eta}.$$

Letting $\eta\to\infty$ in the inequality above allows us to conclude that

$$\lim_{\eta\to\infty}\pi(w,\eta)=\infty.$$

The proof of $\lim_{\eta\to-\infty}\pi(w,\eta)=-\infty$ is similar and thus is omitted. $\square$

Proof of Lemma EC.4. Note that $h(w_1) < h(w_2)$ for $w_1 < w_2$. This, along with (29), allows us to conclude that there do not exist two numbers $w_1 < w_2$ such that

$$\pi(w_1,\eta) = \pi(w_2,\eta) \quad \text{and} \quad \pi'(w_1,\eta) \le 0 \le \pi'(w_2,\eta). \tag{EC.47}$$

Therefore, (a) $\pi(w,\eta)$ cannot have a local minimizer in $w \in (0,\infty)$, and (b) $\pi(w,\eta)$ cannot be a constant in any interval in $(0,\infty)$. We will then employ properties (a) and (b) to prove parts (i)–(iii).

We first prove part (i). Note that when $\eta \le 0$, we must have $\pi'(0,\eta) < 0$. The continuity of $\pi'(0,\eta)$ and properties (a) and (b) immediately imply that $\pi(w,\eta)$ is strictly decreasing in w for $w > 0$. Next, we show that $\lim_{w\to\infty}\pi(w,\eta) = -\infty$. We can prove this by contradiction. If the statement is not correct, there must exist a finite number $\underline{\pi}$ such that $\lim_{w\to\infty}\pi(w,\eta) = \underline{\pi}$ and $\lim_{w\to\infty}\pi'(w,\eta) = 0$. Letting $w \to \infty$ in (29) yields that $\lim_{w\to\infty} h(w) = \eta - g(\underline{\pi})$, which contradicts Assumption 1 that $\lim_{w\to\infty} h(w) = \infty$.

For (ii), we define

$$\bar{\eta} := \sup\left\{\eta \in \mathbb{R} : \text{there exists a } w > 0 \text{ such that } \pi'(w,\eta) < 0\right\}. \tag{EC.48}$$

Note that $\bar{\eta}$ is well defined since $\bar{\eta} > 0$. If $\eta \ge \bar{\eta}$, by the definition of $\bar{\eta}$ and the continuity of $\pi'(w,\eta)$ in η, we can conclude that $\pi(w,\eta)$ is strictly increasing in w. The proof of $\lim_{w\to\infty}\pi(w,\eta) = \infty$ is similar to that of (EC.15) and thus is omitted.

For (iii), we begin by claiming that for each $\eta \in (0,\bar{\eta})$, there exists a number w such that $\pi'(w,\eta) < 0$. If not, we have $\pi'(w,\eta) > 0$ for all $w > 0$. Similar to the proof of (i), we can obtain that

$$\lim_{w\to\infty}\pi(w,\eta) = \infty. \tag{EC.49}$$

On the other hand, the definition of $\bar{\eta}$ and the continuity of $\pi(w,\eta)$ indicate that there exist a $\eta^\dagger > \eta$ such that $\pi'(w^\dagger,\eta^\dagger) < 0$ for some $w^\dagger$. For $w > w^\dagger$, $\pi(w,\eta^\dagger)$ is decreasing and

$$\lim_{w\to\infty}\pi(w,\eta^\dagger) = -\infty. \tag{EC.50}$$

However, (EC.49) and (EC.50) contradict Lemma EC.3 with $\eta^\dagger > \eta$.

By the continuity of $\pi'(w,\eta)$ in w and the definition of $w^\star(\eta)$, we can conclude that $\pi'(w^\star(\eta),\eta) = 0$. Furthermore, properties (a) and (b) imply that $\pi(w,\eta)$ is strictly increasing in $[0,w^\star(\eta)]$ and strictly decreasing in $[w^\star(\eta),\infty)$. The proof of $\lim_{w\to\infty}\pi(w,\eta) = -\infty$ is very similar to that of (EC.15) and thus is omitted. $\square$

Proof of Lemma EC.5. For (i), define

$$\eta^{\ddagger} := \inf\left\{\eta \in (0,\bar{\eta}) : \pi(w^{\star}(\eta),\eta) \geq \tilde{\ell}_i\right\}.$$

It follows from Lemma EC.4(iii) that

$$\pi(w^{\star}(\eta),\eta) = \max_{w\geq 0} \pi(w,\eta), \tag{EC.51}$$

when $\eta \in (0,\bar{\eta})$. Recall that Lemma EC.3 implies that $\pi(w^{\star}(\eta),\eta)$ is increasing in η. Furthermore, Lemma EC.4 shows that

$$\lim_{\eta\downarrow 0} \pi(w^{\star}(\eta),\eta) = 0 < \tilde{\ell}_i \quad \text{and} \quad \lim_{\eta\uparrow\bar{\eta}} \pi(w^{\star}(\eta),\eta) = \infty.$$

Hence, $\eta^{\ddagger}$ is well-defined, and we also have

$$\pi(w^{\star}(\eta^{\ddagger}),\eta^{\ddagger}) = \tilde{\ell}_i \quad \text{and} \quad \pi(w^{\star}(\eta^{\ddagger}),\eta) > \tilde{\ell}_i, \quad \text{for} \quad \eta \in (\eta^{\ddagger},\bar{\eta}).$$

Furthermore, let

$$q(\eta) := \inf\left\{w \geq 0 : \pi(w,\eta) = \tilde{\ell}_i\right\} \quad \text{and} \quad s(\eta) := \sup\left\{w \geq 0 : \pi(w,\eta) = \tilde{\ell}_i\right\}.$$

It follows from Lemma EC.4(iii) that both $q(\eta)$ and $s(\eta)$ are well-defined, finite, and unique.

For part (b), we can rewrite the function $\tilde{f}(\eta)$ as

$$\tilde{f}(\eta) = \begin{cases} 0 & \text{for} \quad \eta \in (-\infty,\eta^{\ddagger}], \\ \int_{q(\eta)}^{s(\eta)} [\pi(w,\eta) - \tilde{\ell}_i]\mathrm{d}w & \text{for} \quad \eta \in (\eta^{\ddagger},\bar{\eta}), \\ \infty & \text{for} \quad \eta \in [\bar{\eta},\infty). \end{cases}$$

Also, $\tilde{f}(\eta)$ is strictly increasing in $\eta \in (\eta^{\ddagger},\bar{\eta})$. We can thus obtain that

$$\lim_{\eta\downarrow\eta^{\ddagger}} \tilde{f}(\eta) = 0 \quad \text{and} \quad \lim_{\eta\uparrow\bar{\eta}} \tilde{f}(\eta) = \infty.$$

Combining this with the continuity of $\tilde{f}$ in $\eta \in (\eta^{\ddagger},\bar{\eta})$, we conclude that there exists a unique $\eta_i \in (\eta^{\ddagger},\bar{\eta})$ such that (EC.19) holds. $\square$