INSEAD The Business School for the World®

Working Paper
2019/33/TOM

Capacity Choice Game in a Multi-Server Queue: Existence of a Nash Equilibrium

Mor Armony

Stern School of Business, New York University, marmony@stern.nyu.edu

Guillaume Roels

INSEAD, guillaume.roels@insead.edu

Hummy Song

The Wharton School, University of Pennsylvania, hummy@wharton.upenn.edu

In a multi-server, single-queue symmetric capacity choice game, Gopalakrishnan et al. (2016) characterize the existence of a Nash equilibrium under a requirement on the servers' capacity cost functions, which excludes some highly relevant cases where servers have ample discretion over their choice of service rates. Without that requirement and when servers are free to choose any service rate, potentially leading to an unstable queueing system, the servers' cost function is ill-behaved and standard tools for establishing the existence of an equilibrium cannot be applied. In this note, we consider a general power capacity cost function in a two-server capacity choice game with no restriction on the servers' choice of capacity. Relying on a lesser-known result, namely Tarski's intersection theorem, we establish the existence of a Nash equilibrium, thus extending the result by Gopalakrishnan et al. (2016) to our setting. Comparing settings where queue stability is enforceable versus not, we show that there always exists a Nash equilibrium in the former case, unlike in the latter, and that some of the capacity choices that are equilibria in the former case are no longer equilibria in the latter. Our analysis highlights the criticality of the enforceability of system stability on equilibrium outcomes.

Key words : Queueing Theory, Game Theory, Strategic Servers, Tarski's Intersection Theorem

1. Introduction

We consider a symmetric capacity choice game in a multi-server, single-queue system similar to Gopalakrishnan et al. (2016). In this game, symmetric servers choose simultaneously and non-cooperatively their service rates (i.e., capacity) so as to minimize their individual costs, consisting of a convex capacity cost and a penalty that is proportional to their utilization. Unlike Gopalakrishnan et al. (2016) who consider a generic capacity cost function with the requirement that its third derivative be nonnegative, we consider an increasing and convex power cost function, but do not impose restrictions on its elasticity, thus allowing, in addition, for any power function lying between a linear and a quadratic shape. As discussed in Armony et al. (2019), it is precisely in this range of capacity cost functions that servers have a high level of discretion over their choice of service rates, potentially resulting in longer equilibrium expected throughput times in a pooled queueing system than in a dedicated one.

In addition, and in contrast to Armony et al. (2019), we consider a setting where servers are free to choose any service rate they like, potentially resulting in a system that is not stable, consistent with the decentralized nature of the capacity choice process: If a centralized system designer were indeed able to observe (and verify) the stability of a queueing system, what may prevent her from also verifying the servers' choices of capacity and forcing them to work at a specific rate?

From a technical standpoint, establishing the existence of a Nash equilibrium in this setting is challenging because standard tools fail to apply. In particular, Kakutani's or Tarski's well-known fixed point theorems may not apply because the servers' cost functions are neither quasi-convex nor sub- or supermodular (Cachon and Netessine 2006). We rely on a less-known fixed-point theorem, namely Tarski's intersection theorem (Tarski 1955, Amir and De Castro 2017), to establish the existence of a symmetric Nash equilibrium in this two-player game. The application of this theorem relies on a property that the best-response curve is quasi-increasing, i.e., that it has no downward jumps at its points of discontinuity. In particular, we show that, even though the servers' cost function may have multiple local minima, its best response satisfies the quasi-increasing property.

Overall, we show that the existence results obtained by Gopalakrishnan et al. (2016) under the requirement that the third derivative of the servers' capacity cost function be nonnegative are robust, in the sense that they carry over to power cost functions with no such requirement. Extending the analysis to situations where queue stability may not be enforceable shows that there always exists a symmetric equilibrium, unlike the case where the system is required to be stable. Therefore, allowing for non-stable queues results in a well-defined outcome. However, the equilibrium is not necessarily unique; hence, an equilibrium selection rule must be specified. The symmetric equilibria (of which there are a maximum of two) are either null (i.e., both servers choose zero service rate) or are such that the system is stable (i.e., both servers choose a positive service rate with total capacity greater than total demand). In the case in which the equilibrium is positive, it coincides with the equilibrium characterized by Gopalakrishnan et al. (2016). Conversely, some of the capacity choices that were characterized as equilibria under the condition of a stable system are no longer equilibria when stability is not enforced. Our analysis thus highlights the criticality of the assumption of the enforceability of the system stability.

The note is organized as follows. We first introduce the model in §2 and present an outline of the argument to establish existence of a Nash equilibrium in §3, relegating the formal proofs of the side results to the appendix. Section 4 compares the equilibria with or without the requirement on queue stability. We summarize our findings and conclude in §5.

2. Model

We consider a single-queue queueing system with two symmetric servers. Customers (or jobs) arrive according to a Poisson process with rate $2\lambda > 0$. Servers simultaneously and non-cooperatively determine their (static) capacities (i.e., service rates) in order to minimize their individual steady-state costs. For any $i = 1, 2$, let μ_i denote Server i's service rate and μ_{-i} denote the other server's service rate (using the notation $-i \doteq 3 - i$). Customer requests are random; specifically, we assume that service times are independent and exponentially distributed, and the mean service rate is determined by the servers. We thus have an $M/M/2$ system with a single infinite-buffer FCFS queue and an arrival rate 2λ; when a customer arrives into an empty system, she is equally likely to be routed to either server.

Servers' costs consist of two terms: a capacity cost and a cost reflecting servers' preference for idleness. Similar to Armony et al. (2019), we consider a power capacity cost function of the type $c(\mu) = \mu^k$ with $k \geq 1$, with the interpretation of the elasticity k as a measure of servers' lack of discretion over their choice of service rate: When k is small, servers have a lot of discretion, as it often happens in knowledge-intensive services; whereas when k is large, servers have limited discretion, as it typically happens in routine services. For comparison purposes, Gopalakrishnan et al. (2016) consider a generic convex increasing capacity cost function $c(\mu)$, with the requirement that $c'''(\mu) \geq 0$, which excludes the cases where $1 < k < 2$. Although our model is more specific whenever $k \geq 2$, it enables us to also generalize their conditions by capturing situations where servers have a high level of discretion over their choice of service rate, which are of key relevance to these capacity choice games.

The servers' preference for idleness is measured as a penalty, denoted by h, proportional to their long-term utilization. With two servers, the expression for *Server i's utilization* derived by Gopalakrishnan et al. (2016, Theorem 1) simplifies to:

$$1 - \left(1 + \frac{\lambda}{\mu_{-i}}\right)\left(1 + \frac{\lambda}{\mu_1 + \mu_2 - 2\lambda}\frac{(\mu_1 + \mu_2)^2}{\mu_1\mu_2}\right)^{-1} = \frac{\lambda\left(2\lambda\mu_i + \mu_{-i}(\mu_1 + \mu_2)\right)}{\left(\mu_1^2(\lambda + \mu_2) + \mu_2^2(\lambda + \mu_1)\right)}$$

when the system is stable (i.e., $\mu_1 + \mu_2 > 2\lambda$); otherwise, it is equal to 1.

Accordingly, Server i's cost equals:

$$C_i(\mu_i;\mu_{-i}) = \begin{cases} \mu_i^k + h\frac{\lambda\left(2\lambda\mu_i + \mu_{-i}(\mu_1+\mu_2)\right)}{(\mu_1^2(\lambda+\mu_2)+\mu_2^2(\lambda+\mu_1))}, & \text{when } \mu_1+\mu_2 > 2\lambda \\ \mu_i^k + h & \text{otherwise.} \end{cases} \tag{1}$$

Servers simultaneously and non-cooperatively determine their capacities in order to minimize their individual steady-state costs. A pure-strategy Nash equilibrium is thus defined as:

$$\mu_i^\star = \arg\max_{\mu_i \geq 0} C_i(\mu_i;\mu_{-i}^\star) \qquad \text{for } i = 1, 2. \tag{2}$$

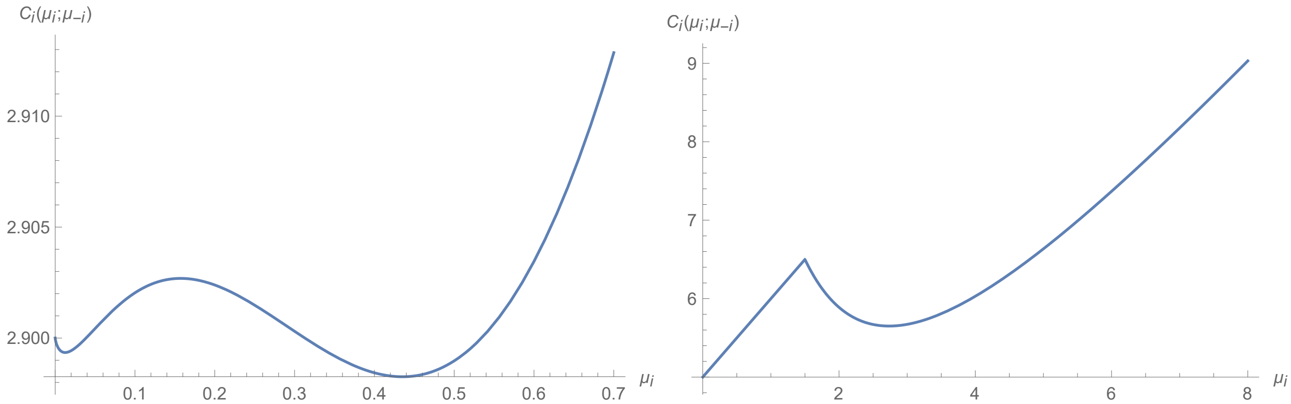
Hence, we assume that servers are free to choose any capacity they like, consistent with the decentralized nature of the capacity choice process. In particular, one possible outcome under our model is that both servers choose a null service rate, resulting in an infinite queue—and this will indeed be an equilibrium. Effectively, workers who go on a strike or who are burned out due to excessive workload work at zero capacity. In practice, overloaded systems may lead to customer abandonments (which we do not model), which may ultimately affect the servers' utility (through intrinsic or extrinsic incentives) and may lead them to choose positive service rates. Hence, reality probably lies between the extreme cases of no system stability enforceability (i.e., $\mu_i \geq 0$) and full system stability enforceability (i.e., $\mu_i > \lambda$). Our model should thus be interpreted as a building block upon which more realistic features can be built. Characterizing the equilibria in this streamlined setting is insightful because it indicates the undesirable outcomes one might reach if proper measures are not put in place to prevent them.

To summarize, our setting is identical to Gopalakrishnan et al. (2016) and Armony et al. (2019) with the exception that (i) the capacity cost function is defined as $c(\mu) = \mu^k$ for any $k \geq 1$ and (ii) the constraint set is defined as $\mu_i \geq 0$ instead of $\mu_i > \lambda$. (When $c'''(\mu) \geq 0$, the analysis of Gopalakrishnan et al. (2016) easily extends to the case where $\mu_i \geq 0$; see Gopalakrishnan et al. (2016, Endnote 2). In contrast, when $c(\mu) = \mu^k$ with $1 < k < 2$, considering $\mu_i \geq 0$ is significantly more challenging than considering $\mu_i > \lambda$; for the proof of the latter, see Armony et al. (2019).) In what follows, we restrict our attention to symmetric equilibria.

3. Existence of a Nash Equilibrium

We first characterize the shape of Server i's cost function and show that it has at most two local minima. If the other server's choice of capacity μ_{-i} is sufficiently large to ensure a stable queueing system (i.e., $\mu_{-i} > 2\lambda$), the two minima are always positive, as is illustrated by the left panel of Figure 1. In contrast, if the other server's choice of capacity is small (i.e., $\mu_{-i} \leq 2\lambda$), one of these local minima is zero, as is illustrated by the right panel of Figure 1.

Figure 1 Server i's cost function.



Note. $k = 1.1$, $\mu_{-i} = 2.4\lambda$, $\lambda = 1$, $h = 2.9$ (left) and $k = 1$, $\mu_{-i} = \lambda/2$, $\lambda = 1$, $h = 5$ (right)

LEMMA 1. *The function $C_i(\mu_i;\mu_{-i})$ defined in (1) has at most two local minima, and those minima are strict. If there are two minima, then (i) when $\mu_{-i} > 2\lambda$, both are positive; and (ii) when $\mu_{-i} \leq 2\lambda$, one is zero and the other greater than $2\lambda - \mu_{-i}$.*

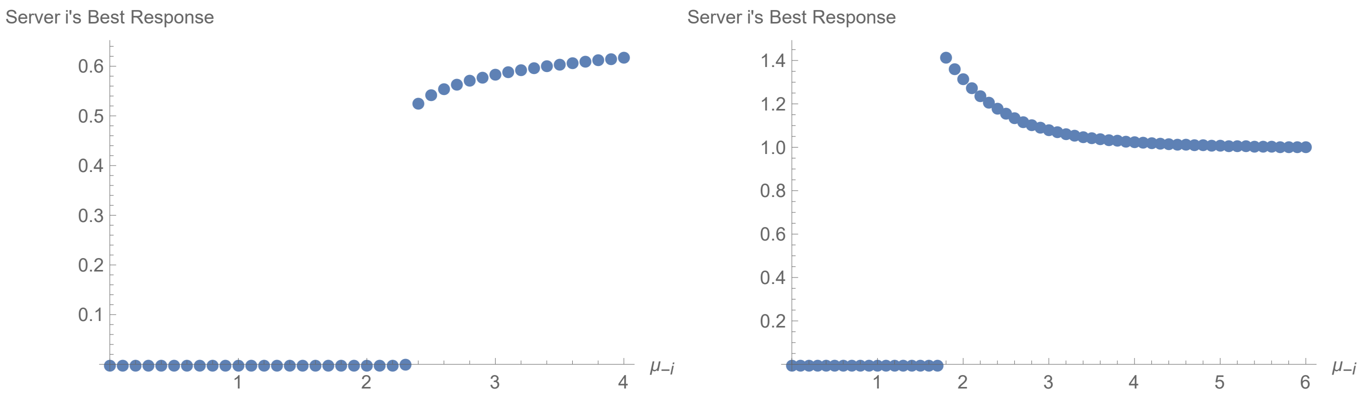
Note that the cost functions are in general not quasiconvex and, thus, one may not directly apply Kakutani's fixed point theorem. For the case where $c'''(\mu) \geq 0$, Gopalakrishnan et al. (2016) find a clever way to circumvent this issue: Because $C_i(\mu_i;\mu_{-i})$ has at most one local minimum when $\mu_i \geq \max\{0, 2\lambda - \mu_{-i}\}$ and is increasing when $\mu_i \leq \max\{0, 2\lambda - \mu_{-i}\}$, $C_i(\mu_i;\mu_{-i})$ is minimized either at that local minimum or at zero; see the right panel of Figure 1. Hence, denoting with μ^e the solution to $C_i'(\mu^e;\mu^e) = 0$, they show that μ^e is an equilibrium if and only if $C_i(\mu^e;\mu^e) \leq C_i(0;\mu^e) = h$. Conversely, $(0,0)$ is an equilibrium if and only if $C_i(0;0) \leq \min_\mu C_i(\mu;0)$, and the right-hand side is easy to compute since it corresponds to the cost function of a one-server, single-queue system. However, their method does not extend to the case where $1 < k < 2$ since $C_i(\mu_i;\mu_{-i})$ may have two interior minima when $\mu_i \geq \max\{0, 2\lambda - \mu_{-i}\}$, as shown on the left panel of Figure 1.

Figure 2 displays two examples of best-response correspondences. From the figure, it appears that the best responses are in general neither increasing nor decreasing, so we also cannot apply Tarski's celebrated fixed-point theorem.

Although the best-response correspondences are discontinuous, all jumps appear to be upward. This property turns out to be sufficient to establish the existence of a Nash equilibrium (Milgrom and Roberts 1994), by Tarski's intersection point theorem (Tarski 1955).

Formally, we will establish that any (minimal or maximal) selection of the best-response correspondences is quasi-increasing (Amir and De Castro 2017). By Lemma 1, Server i's cost function has at most two global minima. We thus need to consider the following four cases, for any fixed value of μ_{-i}:

Figure 2 Best-response correspondences.



Note. $k=1.1$, $\lambda=1$, $h=2.9$ (left) and $k=1$, $\lambda=1$, $h=5$ (right).

1. Server i's cost has only one global strict minimum, and it is equal to zero.
2. Server i's cost has only one global strict minimum, and it is positive.
3. Server i's cost has two global strict minima, and one of these minima is at zero.
4. Server i's cost has two global strict minima, and both minima are positive.

Because Server i's function is continuous, small changes in μ_{-i} will have small effects on the shape of the cost function. Hence, interior local minima will tend to evolve continuously; however, they may suddenly lose their property of being global minima.

In the first case, Server i's best response does not change with a small change in μ_{-i}, i.e., his best response is (locally) constant. In the second case, using a generalized version of the envelope theorem (Milgrom and Segal 2002), we show that Server i's best response changes (locally) continuously, i.e., without a jump.

In the third and fourth cases, we need to show that any selection (e.g., the minimal element or the maximal element) of Server i's best response can only increase whenever it evolves discontinuously. Denote by $\underline{\mu}_i$ and $\overline{\mu}_i$ the global optima, with $\underline{\mu}_i < \overline{\mu}_i$; hence, in the third case, $\underline{\mu}_i = 0$ and in the fourth case $\underline{\mu}_i > 0$. To show that only positive jumps are possible, we show that, when μ_{-i} increases, the local minimum around $\overline{\mu}_i$ achieves a lower cost than the local minimum around $\underline{\mu}_i$. In the third case, the local minimum around $\underline{\mu}_i = 0$ remains equal to zero and thus achieves a constant cost of h, so it is sufficient to show that $C_i(\mu_i;\mu_{-i})$ decreases in μ_{-i} for all $\mu_i > 0$. But in the fourth case, both costs around $\underline{\mu}_i$ and around $\overline{\mu}_i$ may change, so we need the following result:

LEMMA 2. *For any $\mu_{-i} > 2\lambda$, suppose that both $\underline{\mu}_i$ and $\overline{\mu}_i$ globally minimize $C_i(\mu_i;\mu_{-i})$, defined in (1), with $0 < \underline{\mu}_i < \overline{\mu}_i$. Then,*

$$\frac{\partial C_i(\underline{\mu}_i;\mu_{-i})}{\partial \mu_{-i}} > \frac{\partial C_i(\overline{\mu}_i;\mu_{-i})}{\partial \mu_{-i}}.$$

The proof needs to consider two separate cases: (i) when either $k \geq 1.03$ or $\mu_{-i} \geq 2.03\lambda$ and (ii) when $1 \leq k < 1.03$ and $2\lambda < \mu_{-i} < 2.03\lambda$. The second case seems oddly quite specific, but the behavior of the functions to analyze in order to establish the result turns out to be much more convoluted for this range of values; unless we dismiss this case (e.g., by assuming that $k \geq 1.03$), it deserves a special treatment on its own.

Combining these results, we are now ready to establish the result.

THEOREM 1. *There exists a symmetric Nash equilibrium $(\mu^\star, \mu^\star)$ in the capacity choice game (2). The possible equilibria are $(0,0)$ and (μ^e, μ^e), where μ^e is the unique value $\mu > \lambda$ such that*

$$C_i'(\mu;\mu) = k\mu^{k-1} - h\frac{\lambda}{\mu(\lambda+\mu)} = 0. \tag{3}$$

For $(0,0)$ to be an equilibrium, it is necessary and sufficient that $h \leq (2\lambda)^k k^{-k}(1+k)^{1+k}$. For (μ^e, μ^e) to be an equilibrium, it is necessary that $h \geq (\lambda)^k k^{-k}(1+k)^{1+k}$.

Proof. We first show that the servers' action sets can be restricted to compact sets without loss of generality. Because the function $\mu^{k-1} - h/(\lambda+\mu)$ is increasing, it crosses zero at most once, and if it does, the crossing is from below. Hence, the function $\mu \cdot (\mu^{k-1} - h/(\lambda+\mu))$ crosses zero at most once when $\mu > 0$, and the crossing is from below. Let $M = \sup\{\mu | \mu^k - h \cdot \mu/(\lambda+\mu) \leq 0\}$. Because $\mu^k - h \cdot \mu/(\lambda+\mu)$ is continuous and because $\lim_{\mu\to\infty} \mu^k - h\cdot\mu/(\lambda+\mu) = \infty$, we obtain that $M < \infty$. Moreover, $M > 0$.

For any $\mu_i > \max\{0, 2\lambda - \mu_{-i}\}$, we obtain from (1) that

$$\frac{\partial C_i(\mu_i;\mu_{-i})}{\partial \mu_{-i}} = -\frac{h\lambda^2\mu_i\left(4\lambda\mu_{-i} + (\mu_1+\mu_2)^2\right)}{\left(\mu_1^2(\lambda+\mu_2) + \mu_2^2(\lambda+\mu_1)\right)^2} < 0. \tag{4}$$

Hence, for any $\mu_i > \max\{0, 2\lambda - \mu_{-i}\}$, $C_i(\mu_i;\mu_{-i}) > \lim_{\mu_{-i}\to\infty} C_i(\mu_i;\mu_{-i}) = \mu_i^k + h\cdot\lambda/(\lambda+\mu_i)$. Because $\mu_i^k + h\cdot\lambda/(\lambda+\mu_i) > h$ for all $\mu_i > M$, we thus obtain that $C_i(\mu_i;\mu_{-i}) > h = C_i(0;\mu_{-i})$ for any $\mu_i > \max\{M, 2\lambda - \mu_{-i}\}$. Hence, there is no equilibrium such that $\mu_i > \max\{M, 2\lambda\}$ because otherwise setting μ_i to zero would result in a lower cost. Accordingly, one can restrict without loss of generality the strategy set of each server to the compact set $\mathcal{D} := [0, \max\{M, 2\lambda\}]$.

Let $\Phi_i(\mu_{-i})$ be Server i's best-response correspondence, i.e., $\Phi_i(\mu_{-i}) = \arg\min_{\mu_i\in\mathcal{D}} C_i(\mu_i;\mu_{-i})$. Because $C_i(\mu_i;\mu_{-i})$ is continuous and the action sets are compact, $C_i(\mu_i;\mu_{-i})$ attains its minimum on $\mathcal{D}$ by Weierstrass' Theorem, and $\Phi_i(\mu_{-i})$ is nonempty. Let $\varphi_i(\mu_{-i})$ be the minimal or maximal selection of $\Phi_i(\mu_{-i})$. We next show that $\varphi_i(\mu_{-i})$ is quasi-increasing (Amir and De Castro 2017), i.e., that

$$\limsup_{\mu\uparrow\mu_{-i}} \varphi_i(\mu) \leq \varphi_i(\mu_{-i}) \leq \liminf_{\mu\downarrow\mu_{-i}} \varphi_i(\mu). \tag{5}$$

Fix $\hat{\mu}_{-i}$. By Lemma 1, $C_i(\mu_i;\hat{\mu}_{-i})$ has at most two global minima, neither of which is equal to $2\lambda-\hat{\mu}_{-i}$; hence, $\Phi_i(\hat{\mu}_{-i})$ has one or two elements and $2\lambda-\hat{\mu}_{-i}\notin\Phi_i(\hat{\mu}_{-i})$. We next consider four cases: (i) $\Phi_i(\hat{\mu}_{-i})=\{\underline{\mu}_i,\overline{\mu}_i\}$ with $0<\underline{\mu}_i<\overline{\mu}_i$, (ii) $\Phi_i(\hat{\mu}_{-i})=\{\underline{\mu}_i,\overline{\mu}_i\}$ with $0=\underline{\mu}_i<\overline{\mu}_i$, (iii) $\Phi_i(\hat{\mu}_{-i})=\{\hat{\mu}_i\}$ with $\hat{\mu}_i>0$, and (iv) $\Phi_i(\hat{\mu}_{-i})=\{\hat{\mu}_i\}$ with $\hat{\mu}_i=0$.

Case (i): $\Phi_i(\hat{\mu}_{-i})=\{\underline{\mu}_i,\overline{\mu}_i\}$ with $0<\underline{\mu}_i<\overline{\mu}_i$. By Lemma 1, having $\underline{\mu}_i>0$ implies that $\hat{\mu}_{-i}>2\lambda$. By Lemma 1, $\underline{\mu}_i$ and $\overline{\mu}_i$ are strict minima, i.e., there exists an $\epsilon>0$ such that $C_i(\underline{\mu}_i;\hat{\mu}_{-i})<C_i(\mu_i;\hat{\mu}_{-i})$ for all $\mu_i\in B(\underline{\mu}_i,\epsilon)\cap\mathcal{D}$, $\mu_i\neq\underline{\mu}_i$ and such that $C_i(\overline{\mu}_i;\hat{\mu}_{-i})<C_i(\mu_i;\hat{\mu}_{-i})$ for all $\mu_i\in B(\overline{\mu}_i,\epsilon)\cap\mathcal{D}$, $\mu_i\neq\overline{\mu}_i$, in which $B(\mu_i,\epsilon):=\{\mu:|\mu-\mu_i|<\epsilon\}$. From hereon, we fix that ϵ.

Define $N(\mu_i,\epsilon/2):=\{\mu:|\mu-\mu_i|\leq\epsilon/2\}\cap\mathcal{D}$. Let $\underline{V}(\mu_{-i}):=\min_{\mu_i\in N(\underline{\mu}_i,\epsilon/2)}C_i(\mu_i;\mu_{-i})$ and $\overline{V}(\mu_{-i}):=\min_{\mu_i\in N(\overline{\mu}_i,\epsilon/2)}C_i(\mu_i;\mu_{-i})$. Because $C_i(\mu_i;\mu_{-i})$ is continuous in $(\mu_i;\mu_{-i})$ and $N(\underline{\mu}_i,\epsilon/2)$ is compact, $\underline{V}(\mu_{-i})$ is continuous in μ_{-i} by the Maximum Theorem (Sundaram 1996, Theorem 9.14). Moreover, because $\partial C_i(\mu_i;\mu_{-i})/\partial\mu_{-i}$ is continuous in $(\mu_i;\mu_{-i})$ and because $\arg\min_{\mu_i\in N(\underline{\mu}_i,\epsilon/2)}C_i(\mu_i;\hat{\mu}_{-i})=\{\underline{\mu}_i\}$, $\underline{V}(\mu_{-i})$ is differentiable at $\mu_i=\hat{\mu}_i$ and $\underline{V}'(\hat{\mu}_{-i})=\partial C_i(\underline{\mu}_i;\hat{\mu}_{-i})/\partial\mu_{-i}$ (Milgrom and Segal 2002, Corollary 4(iii)). Similarly, we obtain $\overline{V}(\mu_{-i})$ is differentiable at $\mu_i=\hat{\mu}_i$ and $\overline{V}'(\hat{\mu}_{-i})=\partial C_i(\overline{\mu}_i;\hat{\mu}_{-i})/\partial\mu_{-i}$.

Because $\underline{V}(\mu_{-i})$ and $\overline{V}(\mu_{-i})$ are continuously differentiable at $\mu_{-i}=\hat{\mu}_{-i}$, we obtain from Taylor's Theorem (Sundaram 1996, Theorem 1.71),

$$\begin{aligned}
\underline{V}(\mu_{-i}) &= \underline{V}(\hat{\mu}_{-i})+\underline{V}'(\hat{\mu}_{-i})(\mu_i-\hat{\mu}_{-i})+o\left(|\mu_i-\hat{\mu}_{-i}|^2\right)\\
&= C_i(\underline{\mu}_i;\hat{\mu}_i)+\frac{\partial C_i(\underline{\mu}_i;\hat{\mu}_{-i})}{\partial\mu_{-i}}(\mu_i-\hat{\mu}_{-i})+o\left(|\mu_i-\hat{\mu}_{-i}|^2\right)\\
\overline{V}(\mu_{-i}) &= \overline{V}(\hat{\mu}_{-i})+\overline{V}'(\hat{\mu}_{-i})(\mu_i-\hat{\mu}_{-i})+o\left(|\mu_i-\hat{\mu}_{-i}|^2\right)\\
&= C_i(\overline{\mu}_i;\hat{\mu}_i)+\frac{\partial C_i(\overline{\mu}_i;\hat{\mu}_{-i})}{\partial\mu_{-i}}(\mu_i-\hat{\mu}_{-i})+o\left(|\mu_i-\hat{\mu}_{-i}|^2\right).
\end{aligned}$$

Hence, because $C_i(\underline{\mu}_i;\hat{\mu}_i)=C_i(\overline{\mu}_i;\hat{\mu}_i)$ since $\Phi_i(\hat{\mu}_{-i})=\{\underline{\mu}_i,\overline{\mu}_i\}$ and because $\partial C_i(\underline{\mu}_i;\hat{\mu}_{-i})/\partial\mu_{-i}>\partial C_i(\overline{\mu}_i;\hat{\mu}_{-i})/\partial\mu_{-i}$ by Lemma 2, since $\hat{\mu}_{-i}>2\lambda$ and $\overline{\mu}_i>\underline{\mu}_i>0$, we obtain:

$$\begin{aligned}
\frac{\underline{V}(\mu_{-i})-\overline{V}(\mu_{-i})}{\mu_i-\hat{\mu}_{-i}} &= \frac{\partial C_i(\underline{\mu}_i;\hat{\mu}_{-i})}{\partial\mu_{-i}}-\frac{\partial C_i(\overline{\mu}_i;\hat{\mu}_{-i})}{\partial\mu_{-i}}+\frac{o\left(|\mu_i-\hat{\mu}_{-i}|^2\right)}{\mu_i-\hat{\mu}_{-i}}\\
&> \frac{o\left(|\mu_i-\hat{\mu}_{-i}|^2\right)}{\mu_i-\hat{\mu}_{-i}}.
\end{aligned}$$

Hence,

$$\liminf_{\mu_{-i}\to\hat{\mu}_{-i}}\frac{\underline{V}(\mu_{-i})-\overline{V}(\mu_{-i})}{\mu_{-i}-\hat{\mu}_{-i}}\geq\lim_{\mu_{-i}\to\hat{\mu}_{-i}}\frac{o\left(|\mu_i-\hat{\mu}_{-i}|^2\right)}{\mu_{-i}-\hat{\mu}_{-i}}=0. \tag{6}$$

By the Maximum Theorem, $\Phi_i(\mu_{-i})$ is upper semi-continuous (Sundaram 1996, Theorem 9.14). Hence, $\exists\delta>0:|\mu_{-i}-\hat{\mu}_{-i}|<\delta\Rightarrow\Phi_i(\mu_{-i})\subset B(\underline{\mu}_i,\epsilon/3)\cup B(\overline{\mu}_i,\epsilon/3)$.

For any $\mu_{-i} > \hat{\mu}_{-i}$ such that $|\mu_{-i} - \hat{\mu}_{-i}| < \delta$, suppose that $\Phi_i(\mu_{-i}) \cap B(\underline{\mu}_i, \epsilon/3) \neq \emptyset$. By (6), when $\mu_{-i} > \hat{\mu}_{-i}$, if $\Phi_i(\mu_{-i}) \cap B(\underline{\mu}_i, \epsilon/3) \neq \emptyset$, then $\Phi_i(\mu_{-i}) \cap B(\overline{\mu}_i, \epsilon/3) \neq \emptyset$. On the other hand, suppose that $\Phi_i(\mu_{-i}) \cap B(\underline{\mu}_i, \epsilon/3) = \emptyset$. Because $\Phi_i(\mu_{-i}) \subset B(\underline{\mu}_i, \epsilon/3) \cup B(\overline{\mu}_i, \epsilon/3)$, we must thus have that $\Phi_i(\mu_{-i}) \subset B(\overline{\mu}_i, \epsilon/3)$. In either case, $\Phi_i(\mu_{-i}) \cap B(\overline{\mu}_i, \epsilon/3) \neq \emptyset$. Therefore, if $\varphi_i(\mu_{-i})$ is defined as the maximal selection of $\Phi_i(\mu_{-i})$, then $\varphi(\hat{\mu}_i) = \overline{\mu}_i = \liminf_{\mu_{-i} \downarrow \hat{\mu}_i} \varphi_i(\mu_{-i})$. If $\varphi_i(\mu_{-i})$ is defined as the minimal selection of $\Phi_i(\mu_{-i})$, then $\varphi(\hat{\mu}_i) = \underline{\mu}_i \leq \liminf_{\mu_{-i} \downarrow \hat{\mu}_i} \varphi_i(\mu_{-i})$. Summarizing both cases, we obtain that $\varphi(\hat{\mu}_i) \leq \liminf_{\mu_{-i} \downarrow \hat{\mu}_i} \varphi_i(\mu_{-i})$. Similarly, one can show that $\varphi(\hat{\mu}_i) \geq \limsup_{\mu_{-i} \uparrow \hat{\mu}_i} \varphi_i(\mu_{-i})$. Hence, (5) holds.

Case (ii): $\Phi_i(\hat{\mu}_{-i}) = \{\underline{\mu}_i, \overline{\mu}_i\}$ with $0 = \underline{\mu}_i < \overline{\mu}_i$. Because $C_i(0; \mu_{-i}) = h$ for all μ_{-i}, $\partial C_i(\underline{\mu}_i; \hat{\mu}_{-i})/\partial \mu_{-i} = 0$. On the other hand, by (4), $\partial C_i(\overline{\mu}_i; \hat{\mu}_{-i})/\partial \mu_{-i} < 0$. Hence, $\partial C_i(\underline{\mu}_i; \hat{\mu}_{-i})/\partial \mu_{-i} > \partial C_i(\overline{\mu}_i; \hat{\mu}_{-i})/\partial \mu_{-i}$. Applying the same argument as the one above shows that $\limsup_{\mu_{-i} \uparrow \hat{\mu}_{-i}} \varphi_i(\mu_{-i}) = 0 \leq \varphi_i(\hat{\mu}_{-i}) \leq \overline{\mu}_i = \liminf_{\mu_{-i} \uparrow \hat{\mu}_{-i}} \varphi_i(\mu_{-i})$, i.e., (5) holds.

Case (iii): $\Phi_i(\hat{\mu}_{-i}) = \{\hat{\mu}_i\}$ with $\hat{\mu}_i > 0$. Applying Corollary 4 (iii) by (Milgrom and Segal 2002) shows that $C_i(\varphi_i(\hat{\mu}_{-i}); \hat{\mu}_{-i})$ is differentiable at $\hat{\mu}_{-i}$ and its derivative equals $\partial C_i(\hat{\mu}_i; \hat{\mu}_{-i})/\partial \mu_{-i}$. Applying the same argument as the one above shows that $\limsup_{\mu_{-i} \uparrow \hat{\mu}_{-i}} \varphi_i(\mu_{-i}) = \hat{\mu}_i = \varphi_i(\hat{\mu}_{-i}) = \hat{\mu}_i = \liminf_{\mu_{-i} \uparrow \hat{\mu}_{-i}} \varphi_i(\mu_{-i})$, i.e., (5) holds.

Case (iv): $\Phi_i(\hat{\mu}_{-i}) = \{\hat{\mu}_i\}$ with $\hat{\mu}_i = 0$. Because $C_i(0; \mu_{-i}) = h$ for all μ_{-i}, $\partial C_i(\underline{\mu}_i; \hat{\mu}_{-i})/\partial \mu_{-i} = 0$. Hence, $\limsup_{\mu_{-i} \uparrow \hat{\mu}_{-i}} \varphi_i(\mu_{-i}) = 0 = \varphi_i(\hat{\mu}_{-i}) = \liminf_{\mu_{-i} \uparrow \hat{\mu}_{-i}} \varphi_i(\mu_{-i})$, i.e., (5) holds.

Because $\varphi_i(\mu_{-i})$ is quasi-increasing and the strategy sets are compact, there exists a symmetric equilibrium; see Milgrom and Roberts (1994, Corollary 1), Amir and De Castro (2017, Corollary 10), and Vives (1999, p. 41).

By Lemma 1, the symmetric equilibrium capacity investments are either equal to zero or such that $\mu_1 + \mu_2 > 2\lambda$. In the latter case, because $\mu_1 = \mu_2 = \mu^e$, μ^e must solve (3). Because the left-hand side of (3) is increasing in μ, there exists at most one μ^e satisfying (3). Finally, $\mu^e > \lambda$ if and only if $C_i'(\lambda, \lambda) < 0$, i.e., if and only if $h \geq 2k\lambda^k$.

We have that $(0,0)$ is an equilibrium if and only if $C_i(0;0) \leq C_i(\mu;0)$ for all μ, i.e., if and only if $h \leq \min_{\mu > \lambda} \mu^k + h(2\lambda)/\mu$ and if and only if $h \leq (2\lambda)^k k^{-k}(1+k)^{1+k}$. Similarly, (μ^e, μ^e) is an equilibrium with $\mu^e > \lambda$ if and only if $C_i(\mu^e; \mu^e) \leq C_i(\mu; \mu^e)$ for all $\mu \geq 0$. In particular, it is necessary that $C_i(\mu^e; \mu^e) \leq C_i(0; \mu^e)$, i.e., $(\mu^e)^k + h\lambda/\mu^e \leq h$. Suppose that $h < \lambda^k k^{-k}(1+k)^{1+k}$. Then, $h < \min_{\mu > \lambda} \mu^k + h\lambda/\mu \leq (\mu^e)^k + h\lambda/\mu^e$, showing that (μ^e, μ^e) cannot be an equilibrium in this case. Since $h \geq \lambda^k k^{-k}(1+k)^{1+k}$ implies that $h \geq 2k\lambda^k$, $\mu^e > \lambda$ when $h \geq \lambda^k k^{-k}(1+k)^{1+k}$. □

For a power cost function $c(\mu) = \mu^k$, Theorem 1 thus extends the existence result of Gopalakrishnan et al. (2016) to the case where $1 < k < 2$, therefore relaxing the requirement that $c'''(\mu) \geq 0$, and without the requirement that the system be stable. The game has always a Nash equilibrium, and is thus well behaved, even when servers have wide discretion over their choice of capacity, which is common in knowledge-intensive settings (Armony et al. 2019).

4. Comparing Equilibria with or without Queue Stability Requirement

We next compare the equilibria with or without queue stability requirement. First, Theorem 1 shows that, when the queue is not required to be stable, there *always* exists a symmetric equilibrium. In contrast, when the queue is required to be stable, there exists a symmetric equilibrium if and only if $h > 2k\lambda^k$ and $C_i(\mu^e;\mu^e) \leq C_i(\lambda;\mu^e)$ (Gopalakrishnan et al. 2016). Therefore, allowing for non-stable queues brings greater clarity about the outcome of the game when h is small, i.e., when servers exhibit little preference for idleness.

Moreover, Theorem 1 shows that there are only two types of symmetric equilibria: either with capacities equal to zero or equal to μ^e. For comparison, when the queue is required to be stable, the only possible symmetric equilibrium is equal to (μ^e, μ^e) (Gopalakrishnan et al. 2016). Hence, allowing for non-stable queues does not affect the set of possible interior equilibria, even though it opens up the possibility of having a system with infinitely long service times.

The symmetric equilibrium capacity $\mu^\star$ is unique and equal to zero when $h < (\lambda)^k k^{-k}(1+k)^{1+k}$ and unique and equal to μ^e when $h > (2\lambda)^k k^{-k}(1+k)^{1+k}$. Otherwise, there may be more than one symmetric equilibrium. In that case, one needs to specify an equilibrium selection rule. For instance, one possible equilibrium rule could be to choose the productive equilibrium μ^e whenever it exists. Another possible equilibrium selection rule could be to choose the utility-maximizing equilibrium (i.e., $\mu^\star = \mu^e$ if and only if $C_i(\mu^e;\mu^e) \leq C_i(0;0)$) whenever there exist two equilibria.

Four different scenarios emerge from the comparison of the symmetric equilibria between the case where the queue is required to be stable and the case where there is no such requirement:

Scenario I. When $h \in [0, 2k\lambda^k]$ (i.e., when μ^e is ill-defined) or when $h \in (2k\lambda^k, (2\lambda)^k k^{-k}(1+k)^{1+k}]$ and $C_i(\lambda;\mu^e) < C_i(\mu^e;\mu^e)$, $(0,0)$ is the only symmetric equilibrium when servers are free to choose their capacity and there exists no equilibrium when the queue is required to be stable.

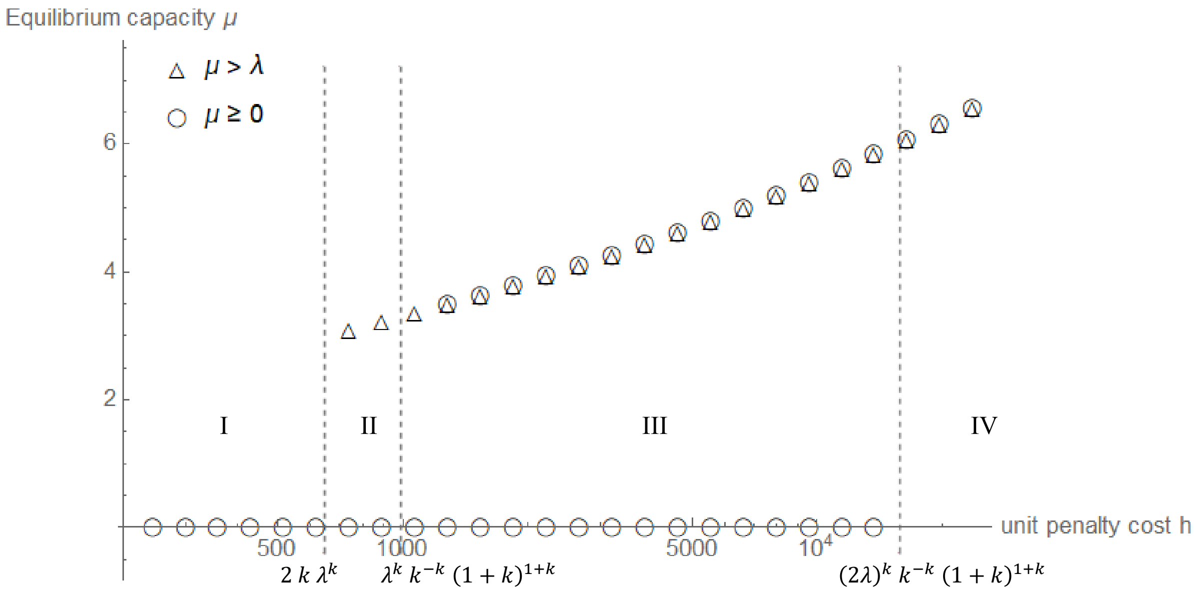
Scenario II. When $h \in (2k\lambda^k, (2\lambda)^k k^{-k}(1+k)^{1+k}]$ and $C_i(\lambda;\mu^e) \geq C_i(\mu^e;\mu^e) > \max_{\mu\in[0,\lambda]} C_i(0;\mu^e)$, $(0,0)$ is the only symmetric equilibrium when servers are not restricted in terms of their choice of capacity and (μ^e,μ^e) is the only symmetric equilibrium when the queue is required to be stable.

Scenario III. When $h \in (2k\lambda^k, (2\lambda)^k k^{-k}(1+k)^{1+k}]$ and $C_i(\mu^e;\mu^e) \leq \max_{\mu\in[0,\lambda]} C_i(0;\mu^e)$ (which happens in particular when $h \geq \lambda^k k^{-k}(1+k)^{1+k}$), both $(0,0)$ and (μ^e,μ^e) are equilibria when servers are not restricted in terms of their choice of capacity and (μ^e,μ^e) is the only symmetric equilibrium when the queue is required to be stable.

Scenario IV. When $h \in ((2\lambda)^k k^{-k}(1+k)^{1+k}, \infty)$, (μ^e,μ^e) is the unique symmetric equilibrium both when servers are free to choose their capacity and when the queue is required to be stable.

Figure 3 shows the set of symmetric equilibria under each scenario when $k = 4$ and $\lambda = 3$, together with the (necessary, but not all sufficient) closed-form thresholds on h determining changes in regimes.

Figure 3 **Set of symmetric equilibria both under the constraint that $\mu_i > \lambda$ for $i = 1, 2$ (triangles) and under the constraint that $\mu_i \geq 0$ for $i = 1, 2$ (circles).**



Note. In Region I, $(0, 0)$ is the only equilibrium when servers are free to choose their capacity and there exists no equilibrium when the queue is required to be stable. In Region II, $(0, 0)$ is the only equilibrium when servers are free to choose their capacity and (μ^e, μ^e) is the only equilibrium when the queue is required to be stable. In Region III, both $(0, 0)$ and (μ^e, μ^e) are equilibria when servers are free to choose their capacity and (μ^e, μ^e) is the only equilibrium when the queue is required to be stable. In Region IV, (μ^e, μ^e) is the unique equilibrium both when servers are free to choose their capacity and when the queue is required to be stable. Parameters are $k = 4$, $\lambda = 3$.

5. Conclusions

The objective of this note was to establish the existence of a Nash equilibrium in a relaxed setting of the single-queue, multi-server symmetric capacity choice game considered by Gopalakrishnan et al. (2016). This allows us to account for situations where servers have a high level of discretion over their choice of service rate.

In our analysis, we also allowed for the possibility of unstable queues. Comparing the equilibria between the case where the queue is required to be stable and the case where there is no such requirement reveals the following points:

- When servers exhibit a small preference for idleness and given the requirement of system stability, the game has an undefined outcome. In contrast, in the absence of such a requirement, there always exists a unique symmetric equilibrium, namely, the null equilibrium.
- When servers exhibit a small-to-moderate preference for idleness and given the requirement of system stability, the game has a unique symmetric interior equilibrium. In contrast, in the absence of such a requirement, the only equilibrium is the null equilibrium, resulting in infinitely long

service times. This shows that the enforceability of system stability is critical in determining the equilibrium outcomes.

- When servers exhibit a moderate-to-high preference for idleness and given the requirement of system stability, the game has a unique symmetric interior equilibrium. In the absence of such a requirement, there exists another possible equilibrium, which is the null equilibrium. In this case, it is important to specify and enforce equilibrium selection rules.
- When servers exhibit a high preference for idleness, the symmetric equilibrium in the game with the requirement of system stability coincides with the equilibrium in the absence of such a requirement. In this case, the enforceability of system stability is irrelevant.

Establishing the existence of a Nash equilibrium when servers have a high level of discretion over their choice of capacity and when the system may not be required to be stable is challenging because standard tools, such as Kakutani's or Tarski's fixed-point theorems, do not apply given that the servers' cost functions are neither quasi-convex nor sub/supermodular. Similar behavior could arise in other games with economies of scale (Cachon and Harker 2002), e.g., production games with setup costs in which players may have an incentive to either produce beyond their breakeven point or to not produce. To establish the existence of a Nash equilibrium in our queueing game, we have relied on a less-known fixed-point result, namely Tarski's intersection theorem. We hope that other operations researchers will find it useful to establish existence of equilibrium in symmetric games with discontinuous best responses.

References

Amir R, De Castro L (2017) Nash equilibrium in games with quasi-monotonic best-responses. *Journal of Economic Theory* 172:220–246.

Armony M, Roels G, Song H (2019) Pooling queues with strategic servers: The effects of customer ownership, working paper.

Cachon GP, Harker PT (2002) Competition and outsourcing with scale economies. *Management Science* 48(10):1314–1333.

Cachon GP, Netessine S (2006) Game theory in supply chain analysis. *INFORMS TutORials in Operations Research*, 200–233.

Gopalakrishnan R, Doroudi S, Ward AR, Wierman A (2016) Routing and staffing when servers are strategic. *Operations Research* 64(4):1033–1050.

Milgrom P, Roberts J (1994) Comparing equilibria. *American Economic Review* 84(3):441–459.

Milgrom P, Segal I (2002) Envelope theorems for arbitrary choice sets. *Econometrica* 70(2):583–601.

Sundaram RK (1996) *A First Course in Optimization Theory* (Cambridge university press).

Tarski A (1955) A lattice-theoretical fixpoint theorem and its applications. *Pacific Journal of Mathematics* 5(2):285–309.

Vives X (1999) *Oligopoly Pricing: Old Ideas and New Tools* (MIT press).

Appendix: Proofs

LEMMA A-1. *Let $f(x)$ be a twice continuously differentiable function, and suppose that there exists a value $\hat{x}$ such that $f''(x) <_{(>)} 0$ for all $x < \hat{x}$ such that $f'(x) = 0$ and $f''(x) >_{(<)} 0$ for all $x > \hat{x}$ such that $f'(x) = 0$. Then $f(x)$ has at most two interior local optima, and both optima are strict.*

Proof. We consider the case where $f''(x) < 0$ for all $x < \hat{x}$ such that $f'(x) = 0$ and $f''(x) > 0$ for all $x > \hat{x}$ such that $f'(x) = 0$; the other case is symmetric and can be treated similarly.

First, we show that

$$\forall x_1 < \hat{x} \text{ such that } f'(x_1) = 0 \Rightarrow f'(x) > 0, \forall x < x_1 \text{ and } f'(x) < 0, \forall x \in (x_1, \hat{x}). \tag{A-1}$$

Suppose that there exists an $x_1 < \hat{x}$ such that $f'(x_1) = 0$. To obtain a contradiction, suppose that there exists another point, $x_2 \in (x_1, \hat{x})$, such that $f'(x_2) = 0$. If there are more than one such point, we consider the smallest such point x_2 that is greater than x_1. Then, because $f''(x_1) < 0$ and $f'(x)$ is continuous, we must have that $f''(x_2) \geq 0$, a contradiction. Similarly, we can obtain that

$$\forall x_2 > \hat{x} \text{ such that } f'(x_2) = 0 \Rightarrow f'(x) < 0, \forall x \in (\hat{x}, x_2) \text{ and } f'(x) > 0, \forall x > x_2. \tag{A-2}$$

We next consider three cases, depending on the behavior of $f(x)$ to the left of $\hat{x}$.

(i) Suppose first that there exists an $x_1 < \hat{x}$ such that $f'(x_1) = 0$. By (A-1), x_1 is a strict local maximum. We consider the three following cases.

- If $f'(x) > 0$ for all $x > \hat{x}$, then because $f'(x) < 0$ for all $x \in (x_1, \hat{x})$ and $f'(x)$ is continuous, we must have that $f'(\hat{x}) = 0$. Hence, x_1 and $\hat{x}$ are the only stationary points. Because $f'(x) < 0$ for all $x \in (x_1, \hat{x})$ and $f'(x) > 0$ for all $x > \hat{x}$, $\hat{x}$ is a strict local minimum.
- If $f'(x) < 0$ for all $x > \hat{x}$, then because $f'(x) < 0$ for all $x \in (x_1, \hat{x})$ and $f'(x)$ is continuous, we must have that $f'(\hat{x}) \leq 0$. Hence, $f(x)$ is nonincreasing for all $x > x_1$ and x_1 is the only local optimum.
- If there exists some $x_2 > \hat{x}$ such that $f'(x_2) = 0$, then $f'(x) < 0$ for all $x \in (\hat{x}, x_2)$ and $f'(x) > 0$ for all $x > x_2$. Because $f'(x)$ is continuous, we obtain that $f'(\hat{x}) \leq 0$. Hence, $f(x)$ is nonincreasing for all $x \in (x_1, x_2)$, and x_1 and x_2 are the only optima. By (A-2), x_2 is a strict local minimum.

(ii) Suppose next that $f'(x) > 0$ for all $x < \hat{x}$.

- If $f'(x) > 0$ for all $x > \hat{x}$, then because $f'(x)$ is continuous, we must have that $f'(\hat{x}) \geq 0$. Hence, $f(x)$ is always nondecreasing, and there is no interior local optimum.
- If $f'(x) < 0$ for all $x > \hat{x}$, then because $f'(x)$ is continuous, we must have that $f'(\hat{x}) = 0$. Hence, $\hat{x}$ is the only stationary point. Because $f'(x) > 0$ for all $x < \hat{x}$ and $f'(x) < 0$ for all $x > \hat{x}$, $\hat{x}$ is a strict local maximum.
- If there exists some $x_2 > \hat{x}$ such that $f'(x_2) = 0$, then $f'(x) < 0$ for all $x \in (\hat{x}, x_2)$ and $f'(x) > 0$ for all $x > x_2$. Because $f'(x)$ is continuous, we obtain that $f'(\hat{x}) = 0$. Hence, $\hat{x}$ and x_2 are the only stationary points. Because $f'(x) > 0$ for all $x < \hat{x}$ and $f'(x) < 0$ for all $x > \hat{x}$, $\hat{x}$ is a strict local maximum. By (A-2), x_2 is a strict local minimum.

(iii) Finally, suppose that $f'(x) < 0$ for all $x < \hat{x}$. This case is symmetric to the case where $f'(x) > 0$ for all $x < \hat{x}$. and can be treated similarly. The argument is omitted for brevity. □

LEMMA A-2. *The function*

$$\hat{C}_i(\mu_i;\mu_{-i}) = \mu_i^k + h\lambda\frac{2\lambda\mu_i + \mu_{-i}^2 + \mu_1\mu_2}{(\mu_1^2(\lambda+\mu_2)+\mu_2^2(\lambda+\mu_1))}, \tag{A-3}$$

defined over all $\mu_i \geq 0$, has (i) at most three interior local optima, namely, a local minimum, a local maximum, and a local minimum as μ_i increases when $\mu_{-i} > 2\lambda$ and $k > 1$, and (ii) at most two interior local optima, namely, a local maximum and a local minimum as μ_i increases when either $\mu_{-i} \leq 2\lambda$ or $k = 1$. Moreover, all these optima are strict.

Proof. The proof uses Lemma A-1. Define $\tilde{\mu}_i \doteq \mu_i/\lambda$ for $i = 1,2$. Differentiating (A-3), we obtain:

$$\hat{C}_i'(\mu_i;\mu_{-i}) = \hat{C}_i'(\lambda\tilde{\mu}_i;\lambda\tilde{\mu}_{-i}) = k\lambda^{k-1}\tilde{\mu}_i^{k-1} - h\lambda^{-1}\frac{(\tilde{\mu}_1+\tilde{\mu}_2)(1+\tilde{\mu}_{-i})\left(2(\tilde{\mu}_i-\tilde{\mu}_{-i})+\tilde{\mu}_{-i}(\tilde{\mu}_1+\tilde{\mu}_2)\right)}{\left(\tilde{\mu}_1^2(1+\tilde{\mu}_2)+\tilde{\mu}_2^2(1+\tilde{\mu}_1)\right)^2}. \tag{A-4}$$

Suppose that $\hat{C}_i'(\mu_i;\mu_{-i}) = 0$. Taking the second derivative, we obtain

$$\begin{aligned}
&\hat{C}_i''(\mu_i;\mu_{-i})\Big|_{\hat{C}_i'(\mu_i;\mu_{-i})=0}\\
&= \Big(k(k-1)\lambda^{k-1}\tilde{\mu}_i^{k-2} + 2h\lambda^{-1}(1+\tilde{\mu}_{-i})\\
&\qquad \times\frac{(\tilde{\mu}_{-i}-3)\tilde{\mu}_{-i}^4 + 3\tilde{\mu}_1^2\tilde{\mu}_2^2(1+\tilde{\mu}_{-i}) + 3\tilde{\mu}_i(\tilde{\mu}_{-i}-2)\tilde{\mu}_{-i}^2(1+\tilde{\mu}_{-i}) + \tilde{\mu}_i^3(1+\tilde{\mu}_{-i})(2+\tilde{\mu}_{-i})}{(\tilde{\mu}_2^2(1+\tilde{\mu}_1)+\tilde{\mu}_1^2(1+\tilde{\mu}_2))^3}\Big)\Big|_{\hat{C}_i'(\lambda\tilde{\mu}_i;\lambda\tilde{\mu}_{-i})=0}\\
&= (k-1)\tilde{\mu}_i^{-1}h\lambda^{-1}\frac{(\tilde{\mu}_1+\tilde{\mu}_2)(1+\tilde{\mu}_{-i})\left(2(\tilde{\mu}_i-\tilde{\mu}_{-i})+\tilde{\mu}_{-i}(\tilde{\mu}_1+\tilde{\mu}_2)\right)}{\left(\tilde{\mu}_1^2(1+\tilde{\mu}_2)+\tilde{\mu}_2^2(1+\tilde{\mu}_1)\right)^2}\\
&\quad + 2h\lambda^{-1}(1+\tilde{\mu}_{-i})\frac{(\tilde{\mu}_{-i}-3)\tilde{\mu}_{-i}^4 + 3\tilde{\mu}_1^2\tilde{\mu}_2^2(1+\tilde{\mu}_{-i}) + 3\tilde{\mu}_i(\tilde{\mu}_{-i}-2)\tilde{\mu}_{-i}^2(1+\tilde{\mu}_{-i}) + \tilde{\mu}_i^3(1+\tilde{\mu}_{-i})(2+\tilde{\mu}_{-i})}{(\tilde{\mu}_2^2(1+\tilde{\mu}_1)+\tilde{\mu}_1^2(1+\tilde{\mu}_2))^3}\\
&= \frac{h(1+\tilde{\mu}_{-i})}{\lambda\tilde{\mu}_1\left(\tilde{\mu}_2^2(1+\tilde{\mu}_1)+\tilde{\mu}_1^2(1+\tilde{\mu}_2)\right)^3}\\
&\qquad \times\Big((k-1)(\tilde{\mu}_{-i}-2)\tilde{\mu}_{-i}^4 + (1+k)\tilde{\mu}_i^4(1+\tilde{\mu}_{-i})(2+\tilde{\mu}_{-i}) + \tilde{\mu}_i\tilde{\mu}_{-i}^4(\tilde{\mu}_{-i}(1+k)-6)\\
&\qquad\quad + \tilde{\mu}_i^3\tilde{\mu}_{-i}^2(2+3\tilde{\mu}_{-i}+k(4+3\tilde{\mu}_{-i})) + 3\tilde{\mu}_1^2\tilde{\mu}_2^2(\tilde{\mu}_{-i}(\tilde{\mu}_{-i}(k+1)-2)-4)\Big).
\end{aligned}$$

Hence, a stationary point $(\tilde{\mu}_1,\tilde{\mu}_2)$ is a strict local minimum if

$$\begin{aligned}
H(\tilde{\mu}_i,\tilde{\mu}_{-i}) \doteq\ & (k-1)(\tilde{\mu}_{-i}-2)\tilde{\mu}_{-i}^4 + (1+k)\tilde{\mu}_i^4(1+\tilde{\mu}_{-i})(2+\tilde{\mu}_{-i}) + \tilde{\mu}_i\tilde{\mu}_{-i}^4(\tilde{\mu}_{-i}(1+k)-6)\\
& + \tilde{\mu}_i^3\tilde{\mu}_{-i}^2(2+3\tilde{\mu}_{-i}+k(4+3\tilde{\mu}_{-i})) + 3\tilde{\mu}_i^2\tilde{\mu}_{-i}^2(\tilde{\mu}_{-i}(\tilde{\mu}_{-i}(k+1)-2)-4)
\end{aligned} \tag{A-5}$$

is positive, and it is a strict local maximum if $H(\tilde{\mu}_i,\tilde{\mu}_{-i}) < 0$. Note that $\lim_{\tilde{\mu}_i\to\infty} H(\tilde{\mu}_i,\tilde{\mu}_{-i}) = \infty$. Because $H(\tilde{\mu}_i,\tilde{\mu}_{-i})$ is continuous, $H(\tilde{\mu}_i,\tilde{\mu}_{-i})$ must be an increasing in $\tilde{\mu}_i$ as $\tilde{\mu}_i$ gets very large.

Define $G(\tilde{\mu}_i,\tilde{\mu}_{-i}) \doteq \partial H(\tilde{\mu}_i,\tilde{\mu}_{-i})/\partial\tilde{\mu}_i$. It can be checked that $G(\tilde{\mu}_i,\tilde{\mu}_{-i})$ is convex, i.e., that $\partial^2 G(\tilde{\mu}_i,\tilde{\mu}_{-i})/\partial\tilde{\mu}_i^2 \geq 0$ for all $\mu_1,\mu_2 \geq 0$, with equality only if $\mu_1 = \mu_2 = 0$.

The rest of the proof considers two cases separately, namely, when $\tilde{\mu}_{-i} > 2$ or not. Consider first the case where $\tilde{\mu}_{-i} > 2$. In that case,

$$\frac{\partial G(\tilde{\mu}_i,\tilde{\mu}_{-i})}{\partial\tilde{\mu}_i}\Big|_{\tilde{\mu}_i=0} = 6\tilde{\mu}_{-i}^2\left(\tilde{\mu}_{-i}^2(k+1)-2\tilde{\mu}_{-i}-4\right) \geq 6\tilde{\mu}_{-i}^2\left(2\tilde{\mu}_{-i}^2-2\tilde{\mu}_{-i}-4\right) = 12\tilde{\mu}_{-i}^2(1+\tilde{\mu}_{-i})(\tilde{\mu}_{-i}-2) > 0.$$

Because $G(\tilde{\mu}_i,\tilde{\mu}_{-i})$ is convex in $\tilde{\mu}_i$, $\partial G(\tilde{\mu}_i,\tilde{\mu}_{-i})/\partial\tilde{\mu}_i > 0$ for all $\tilde{\mu}_i \geq 0$. Hence, when $\tilde{\mu}_i > 0$, $G(\tilde{\mu}_i,\tilde{\mu}_{-i})$ crosses zero at most once as $\tilde{\mu}_i$ increases, and the crossing is from below. Hence, when $\tilde{\mu}_{-i} > 2$, because $G(\tilde{\mu}_i,\tilde{\mu}_{-i}) \doteq$

$\partial H(\tilde{\mu}_i, \tilde{\mu}_{-i})/\partial \tilde{\mu}_i$, we thus obtain that $H(\tilde{\mu}_i, \tilde{\mu}_{-i})$ is either first decreasing and then increasing in $\tilde{\mu}_i$ or monotone increasing in $\tilde{\mu}_i$, as $\tilde{\mu}_i$ increases, starting from zero.

Suppose first that $k > 1$. Because $H(0, \tilde{\mu}_{-i}) = (k-1)(\tilde{\mu}_{-i} - 2)\tilde{\mu}_{-i}^4 > 0$ and $\lim_{\tilde{\mu}_i \to \infty} H(\tilde{\mu}_i, \tilde{\mu}_{-i}) = \infty$, we obtain that $H(\tilde{\mu}_i, \tilde{\mu}_{-i})$ crosses zero either twice or never, and if it does, it first crosses zero from above and then from below. Therefore, by Lemma A-1, $\hat{C}_i(\lambda\tilde{\mu}_i; \lambda\tilde{\mu}_{-i})$ may have up to three interior local optima, going through a local minimum, a local maximum, and then a local minimum as $\tilde{\mu}_i$ increases, and all optima are strict.

Suppose next that $k = 1$. In that case, because $H(0, \tilde{\mu}_{-i}) = (k-1)(\tilde{\mu}_{-i} - 2)\tilde{\mu}_{-i}^4 = 0$ and $\lim_{\tilde{\mu}_i \to \infty} H(\tilde{\mu}_i, \tilde{\mu}_{-i}) = \infty$, $H(\tilde{\mu}_i, \tilde{\mu}_{-i})$ crosses zero at most once when $\tilde{\mu}_i > 0$, and the crossing, if it happens, is from below. Therefore by Lemma A-1, $\hat{C}_i(\lambda\tilde{\mu}_i; \lambda\tilde{\mu}_{-i})$ may have up to two interior local optima when $\tilde{\mu}_i > 0$, going through a local maximum and then a local minimum as $\tilde{\mu}_i$ increases, and all optima are strict.

Consider next the case where $\tilde{\mu}_{-i} \leq 2$. Because $\left(\tilde{\mu}_{-i}^2 - 4\tilde{\mu}_{-i} - 4\right) < 0$ and $\left(\tilde{\mu}_{-i}^2 - \tilde{\mu}_{-i} - 2\right) \leq 0$ when $\tilde{\mu}_{-i} \leq 2$, we obtain

$$\begin{aligned}
\left.\frac{\partial G(\tilde{\mu}_i, \tilde{\mu}_{-i})}{\partial \tilde{\mu}_i}\right|_{\tilde{\mu}_i = \tilde{\mu}_{-i}\frac{2-\tilde{\mu}_{-i}}{2+\tilde{\mu}_{-i}}} &= -\frac{12\tilde{\mu}_{-i}^2\left(2\tilde{\mu}_{-i} + \tilde{\mu}_{-i}^2 + k\left(\tilde{\mu}_{-i}^2 - 4\tilde{\mu}_{-i} - 4\right)\right)}{2 + \tilde{\mu}_{-i}} \\
&\geq -\frac{12\tilde{\mu}_{-i}^2\left(2\tilde{\mu}_{-i} + \tilde{\mu}_{-i}^2 + \left(\tilde{\mu}_{-i}^2 - 4\tilde{\mu}_{-i} - 4\right)\right)}{2 + \tilde{\mu}_{-i}} \\
&= -\frac{24\tilde{\mu}_{-i}^2\left(\tilde{\mu}_{-i}^2 - \tilde{\mu}_{-i} - 2\right)}{2 + \tilde{\mu}_{-i}} \\
&\geq 0.
\end{aligned}$$

Because $G(\tilde{\mu}_i, \tilde{\mu}_{-i})$ is strictly convex in $\tilde{\mu}_i > 0$, this implies that $G(\tilde{\mu}_i, \tilde{\mu}_{-i})$ is increasing in $\tilde{\mu}_i$ for all $\tilde{\mu}_i > \tilde{\mu}_{-i}\frac{2-\tilde{\mu}_{-i}}{2+\tilde{\mu}_{-i}}$. Hence when $\tilde{\mu}_i \geq \tilde{\mu}_{-i}\frac{2-\tilde{\mu}_{-i}}{2+\tilde{\mu}_{-i}}$, $G(\tilde{\mu}_i, \tilde{\mu}_{-i})$ crosses zero at most once as $\tilde{\mu}_i$ increases, and the crossing is from below. Hence, when $\tilde{\mu}_{-i} \leq 2$, because $G(\tilde{\mu}_i, \tilde{\mu}_{-i}) \doteq \partial H(\tilde{\mu}_i, \tilde{\mu}_{-i})/\partial \tilde{\mu}_i$, we obtain that $H(\tilde{\mu}_i, \tilde{\mu}_{-i})$ is either first decreasing and then increasing in $\tilde{\mu}_i$ or monotone increasing in $\tilde{\mu}_i$, starting from $\tilde{\mu}_{-i}\frac{2-\tilde{\mu}_{-i}}{2+\tilde{\mu}_{-i}}$. Because

$$\left.H(\tilde{\mu}_i, \tilde{\mu}_{-i})\right|_{\tilde{\mu}_i = \tilde{\mu}_{-i}\frac{2-\tilde{\mu}_{-i}}{2+\tilde{\mu}_{-i}}} = -\frac{8\tilde{\mu}_{-i}^4\left(\tilde{\mu}_{-i} - 2\right)\left(\tilde{\mu}_{-i} - 2(1-\sqrt{2})\right)\left(\tilde{\mu}_{-i} - 2(1+\sqrt{2})\right)}{\left(2 + \tilde{\mu}_{-i}\right)^3} \leq 0$$

and $\lim_{\tilde{\mu}_i \to \infty} H(\tilde{\mu}_i, \tilde{\mu}_{-i}) = \infty$, we obtain that $H(\tilde{\mu}_i, \tilde{\mu}_{-i})$ crosses zero exactly once when $\tilde{\mu}_i \geq \tilde{\mu}_{-i}\frac{2-\tilde{\mu}_{-i}}{2+\tilde{\mu}_{-i}}$ and the crossing is from below. Therefore by Lemma A-1, when $\tilde{\mu}_i \geq \tilde{\mu}_{-i}\frac{2-\tilde{\mu}_{-i}}{2+\tilde{\mu}_{-i}}$, $\hat{C}_i(\lambda\tilde{\mu}_i; \lambda\tilde{\mu}_{-i})$ may have up to two interior local optima, going through a local maximum and then a local minimum as $\tilde{\mu}_i$ increases, and all optima are strict. Finally, observe from (A-4) that, for all $\tilde{\mu}_i < \tilde{\mu}_{-i}\frac{2-\tilde{\mu}_{-i}}{2+\tilde{\mu}_{-i}}$, $\hat{C}_i'(\mu_i; \mu_{-i}) > 0$. Hence, when $\tilde{\mu}_i > 0$, $\hat{C}_i(\lambda\tilde{\mu}_i; \lambda\tilde{\mu}_{-i})$ may have up to two interior local optima, going through a local maximum and then a local minimum as $\tilde{\mu}_i$ increases, and all these optima are strict. □

Proof of Lemma 1. The proof uses Lemma A-2. First, note that, for any μ_{-i}, $\lim_{\mu_i \to \infty} C_i(\mu_i; \mu_{-i}) = \infty$, and therefore all minima are either equal to zero or interior.

Suppose that $\mu_{-i} > 2\lambda$. Then $C_i(\mu_i; \mu_{-i}) = \hat{C}_i(\mu_i; \mu_{-i})$ for all $\mu_i \geq 0$, in which $\hat{C}_i(\mu_i; \mu_{-i})$ is defined by (A-3). By Lemma A-2, $C_i(\mu_i; \mu_{-i})$ has at most two interior local minima and at most one when $k = 1$, and these minima are strict. All interior minima must satisfy the following firs-order optimality conditions:

$$C_i'(\mu_i; \mu_{-i}) = k\mu_i^{k-1} - h\lambda\frac{(\mu_1 + \mu_2)(\lambda + \mu_{-i})\left(2\lambda(\mu_i - \mu_{-i}) + \mu_{-i}(\mu_1 + \mu_2)\right)}{\left(\mu_1^2\left(\lambda + \mu_2\right) + \mu_2^2\left(\lambda + \mu_1\right)\right)^2} = 0. \tag{A-6}$$

We thus obtain that, when $k > 1$,

$$k\mu_i^{k-1}\Big|_{\mu_i=0} - h\frac{(\lambda+\mu_{-i})(\mu_{-i}-2\lambda)}{\lambda\mu_{-i}^2} < 0 \Leftrightarrow C_i'(0;\mu_{-i}) < 0,$$

i.e., the function $C_i(\mu_i;\mu_{-i})$ is decreasing at $\mu_i = 0$; hence, zero is a strict local maximum and therefore, all minima are interior.

Suppose next that $\mu_{-i} \le 2\lambda$. By (1), $C_i(\mu_i;\mu_{-i}) = \mu_i^k + h$ for all $\mu_i \in [0, 2\lambda-\mu_{-i}]$, and $C_i(\mu_i;\mu_{-i}) = \hat{C}_i(\mu_i;\mu_{-i})$ for all $\mu_i > 2\lambda - \mu_{-i}$. Note that $C_i(\mu_i;\mu_{-i})$ is continuous at $2\lambda-\mu_{-i}$ and increasing for all $\mu_i \in [0, 2\lambda-\mu_{-i}]$; in particular, zero is a strict local minimum. However, $C_i(\mu_i;\mu_{-i})$ is non-differentiable at $2\lambda - \mu_{-i}$ since

$$\begin{aligned}&\lim_{\mu_i\uparrow 2\lambda-\mu_{-i}} C_i'(\mu_i;\mu_{-i}) = k\left(2\lambda-\mu_{-i}\right)^{k-1}\\ \ge &\lim_{\mu_i\downarrow 2\lambda-\mu_{-i}} C_i'(\mu_i;\mu_{-i}) = k\left(2\lambda-\mu_{-i}\right)^{k-1} - \frac{4h\lambda^3(\lambda+\mu_{-i})(2\lambda-\mu_{-i})}{\left((2\lambda-\mu_{-i})^2(\lambda+\mu_{-i})+\mu_{-i}^2(3\lambda-\mu_{-i})\right)^2}.\end{aligned}$$

Note that $\lim_{\mu_i\uparrow 2\lambda-\mu_{-i}} C_i'(\mu_i;\mu_{-i}) \ge 0$.

We consider two cases: If $\lim_{\mu_i\downarrow 2\lambda-\mu_{-i}} C_i'(\mu_i;\mu_{-i}) \ge 0$, then $C_i(\mu_i;\mu_{-i})$ is nondecreasing at $2\lambda-\mu_{-i}$. Because $\hat{C}_i(\mu_i;\mu_{-i})$ has at most one local maximum and one local minimum when $\mu_{-i} \le 2\lambda$ by Lemma A-2, $C_i(\mu_i;\mu_{-i})$ has at most two local minima, one at zero and another one when $\mu_i > 2\lambda - \mu_{-i}$. Moreover by Lemma A-2, all minima are strict.

If on the other hand $\lim_{\mu_i\downarrow 2\lambda-\mu_{-i}} C_i'(\mu_i;\mu_{-i}) < 0$, then $C_i(\mu_i;\mu_{-i})$ reaches a local maximum at $2\lambda-\mu_{-i}$. However, if $\lim_{\mu_i\downarrow 2\lambda-\mu_{-i}} C_i'(\mu_i;\mu_{-i}) < 0$, then, by Lemma A-2, it must be that $\hat{C}_i(\mu_i;\mu_{-i})$ has only one local optimum greater than $2\lambda-\mu_{-i}$, and that local optimum is a local minimum. In that case, $C_i(\mu_i;\mu_{-i})$ has two local minima, one at zero and another one when $\mu_i > 2\lambda-\mu_{-i}$, and by Lemma A-2, both minima are strict. □

LEMMA A-3. *For any $k < 1.03$ and $2\lambda < \mu_{-i} < 2.03\lambda$, suppose that both $\underline{\mu}_i$ and $\overline{\mu}_i$ globally minimize $C_i(\mu_i;\mu_{-i})$, defined in (1), with $0 < \underline{\mu}_i < \overline{\mu}_i$. Then, $\underline{\mu}_i < \lambda < \overline{\mu}_i$.*

Proof. Because $C_i(0;\mu_{-i}) = h$, if $\underline{\mu}_i$ and $\overline{\mu}_i$ are global minima, then $C_i(\underline{\mu}_i;\mu_{-i}) \le h$ and $C_i(\overline{\mu}_i;\mu_{-i}) \le h$. Because $\underline{\mu}_i$ and $\overline{\mu}_i$ minimize $C_i(\mu_i;\mu_{-i})$ in the interior of its domain by assumption and because $\mu_1+\mu_2 > 2\lambda$, the first-order optimality conditions (A-6) are satisfied at both $\underline{\mu}_i$ and $\overline{\mu}_i$. Therefore, plugging (A-6) into (1), we obtain

$$\begin{aligned}&C_i(\underline{\mu}_i;\mu_{-i})\\ &= \underline{\mu}_i^k + h\lambda\frac{2\lambda\underline{\mu}_i + \mu_{-i}^2 + \underline{\mu}_i\mu_{-i}}{\left(\underline{\mu}_i^2(\lambda+\mu_{-i}) + \mu_{-i}^2\left(\lambda+\underline{\mu}_i\right)\right)}\\ &= \frac{\underline{\mu}_i}{k}\left(h\lambda\frac{(\underline{\mu}_i+\mu_{-i})(\lambda+\mu_{-i})\left(2\lambda(\underline{\mu}_i-\mu_{-i})+\mu_{-i}(\underline{\mu}_i+\mu_{-i})\right)}{\left(\underline{\mu}_i^2(\lambda+\mu_{-i})+\mu_{-i}^2\left(\lambda+\underline{\mu}_i\right)\right)^2}\right) + h\lambda\frac{2\lambda\underline{\mu}_i+\mu_{-i}^2+\underline{\mu}_i\mu_{-i}}{\left(\underline{\mu}_i^2(\lambda+\mu_{-i})+\mu_{-i}^2\left(\lambda+\underline{\mu}_i\right)\right)}\\ &= h\lambda\left(\frac{\underline{\mu}_i}{k}\frac{(\underline{\mu}_i+\mu_{-i})(\lambda+\mu_{-i})\left(2\lambda(\underline{\mu}_i-\mu_{-i})+\mu_{-i}(\underline{\mu}_i+\mu_{-i})\right)}{\left(\underline{\mu}_i^2(\lambda+\mu_{-i})+\mu_{-i}^2\left(\lambda+\underline{\mu}_i\right)\right)^2} + \frac{2\lambda\underline{\mu}_i+\mu_{-i}^2+\underline{\mu}_i\mu_{-i}}{\left(\underline{\mu}_i^2(\lambda+\mu_{-i})+\mu_{-i}^2\left(\lambda+\underline{\mu}_i\right)\right)}\right),\end{aligned}$$

and similarly for $C_i(\overline{\mu}_i;\mu_{-i})$.

We define the function

$$G(\tilde{\mu}_i,\tilde{\mu}_{-i},k)=\left(\frac{\tilde{\mu}_i}{k}\frac{(\tilde{\mu}_i+\tilde{\mu}_{-i})(1+\tilde{\mu}_{-i})\left(2(\tilde{\mu}_i-\tilde{\mu}_{-i})+\tilde{\mu}_{-i}(\tilde{\mu}_i+\tilde{\mu}_{-i})\right)}{\left(\tilde{\mu}_i^2\left(1+\tilde{\mu}_{-i}\right)+\tilde{\mu}_{-i}^2\left(1+\tilde{\mu}_i\right)\right)^2}+\frac{2\tilde{\mu}_i+\tilde{\mu}_{-i}^2+\tilde{\mu}_i\tilde{\mu}_{-i}}{\left(\tilde{\mu}_i^2\left(1+\tilde{\mu}_{-i}\right)+\tilde{\mu}_{-i}^2\left(1+\tilde{\mu}_i\right)\right)}\right)$$

such that $C_i(\underline{\mu}_i;\mu_{-i})=hG(\underline{\tilde{\mu}}_i\lambda,\tilde{\mu}_{-i}\lambda,k)$ and $C_i(\overline{\mu}_i;\mu_{-i})=hG(\overline{\tilde{\mu}}_i\lambda,\tilde{\mu}_{-i}\lambda,k)$. Hence, $C_i(\underline{\mu}_i;\mu_{-i})\leq h$ if and only if $G(\underline{\tilde{\mu}}_i\lambda,\tilde{\mu}_{-i}\lambda,k)\leq 1$, and similarly for $\overline{\mu}_i$.

Because

$$\frac{\partial G(\tilde{\mu}_i,\tilde{\mu}_{-i},k)}{\partial k}=-\frac{\tilde{\mu}_i}{k^2}\frac{(\tilde{\mu}_i+\tilde{\mu}_{-i})(1+\tilde{\mu}_{-i})\left(2(\tilde{\mu}_i-\tilde{\mu}_{-i})+\tilde{\mu}_{-i}(\tilde{\mu}_i+\tilde{\mu}_{-i})\right)}{\left(\tilde{\mu}_i^2\left(1+\tilde{\mu}_{-i}\right)+\tilde{\mu}_{-i}^2\left(1+\tilde{\mu}_i\right)\right)^2}<0$$

and because

$$\begin{aligned}&\frac{\partial G(\tilde{\mu}_i,\tilde{\mu}_{-i},k)}{\partial \mu_{-i}}\\&=-\frac{\tilde{\mu}_i\left((1+k)\tilde{\mu}_i^4(1+\tilde{\mu}_{-i})+\tilde{\mu}_i^3\tilde{\mu}_{-i}(4+3\tilde{\mu}_{-i}+k(2+3\tilde{\mu}_{-i}))\right)}{k\left(\tilde{\tilde{\mu}}_i^2\left(1+\tilde{\tilde{\mu}}_{-i}\right)+\tilde{\tilde{\mu}}_{-i}^2\left(1+\tilde{\tilde{\mu}}_i\right)\right)^3}\\&\quad-\frac{\tilde{\mu}_i\left(\tilde{\mu}_i^2\tilde{\mu}_{-i}(3(2+\tilde{\mu}_{-i})^2+k(4+3\tilde{\mu}_{-i}(2+\tilde{\mu}_{-i}))+\tilde{\mu}_i\tilde{\mu}_{-i}^3(\tilde{\mu}_{-i}+k(6+\tilde{\mu}_{-i}))+(k-1)\tilde{\mu}_{-i}^3(4+\tilde{\mu}_{-i})\right)}{k\left(\tilde{\mu}_i^2\left(1+\tilde{\mu}_{-i}\right)+\tilde{\mu}_{-i}^2\left(1+\tilde{\mu}_i\right)\right)^3}\\&<0,\end{aligned}$$

and when $k<1.03$ and $\mu_{-i}<2.03\lambda$,

$$G(\tilde{\mu}_i,\tilde{\mu}_{-i},k)>G(\tilde{\mu}_i,2.03,1.03)=\frac{1+1.99936\tilde{\mu}_i+3.14094\tilde{\mu}_i^2+1.41717\tilde{\mu}_i^3}{(1+\tilde{\mu}_i+0.735276\tilde{\mu}_i^2)^2}.$$

In particular, $G(\tilde{\mu}_i,2.03,1.03)\geq 1$ if and only if

$$\tilde{\mu}_i\left(-0.000642474+0.670386\tilde{\mu}_i-0.0533813\tilde{\mu}_i^2-0.540631\tilde{\mu}_i^3\right)\geq 0.$$

This quartic polynomial has four roots, namely at -1.16448, 0, 0.000958438, and 1.06478. Moreover, the quartic term is negative. Hence when $\tilde{\mu}_i\geq 0$, $G(\tilde{\mu}_i,2.03,1.03)\geq 1$ if and only if $\tilde{\mu}_i\in[0.000958438,1.06478]$. Hence, if $C_i(\underline{\mu}_i;\mu_{-i})\leq h$, then $\underline{\mu}_i\notin[0.000958438,1.06478]$, and similarly for $\overline{\mu}_i$. By Lemma 1, $C_i(\mu_i;\mu_{-i})$ has at most two local minima. Because these two local minima are precisely $\underline{\mu}_i$ and $\overline{\mu}_i$, there must exist some local maximum in between. Because $C_i(\mu;\mu_{-i})\geq h$ for all $\mu\in[0.000958438\lambda,1.06478\lambda]$, we obtain that $\underline{\mu}_i\leq 0.000958438\lambda$ and $\overline{\mu}_i\geq 1.06478\lambda$. □

LEMMA A-4. *For any $\mu_i\geq 0$ and $\mu_{-i}\geq 2\lambda$, the function*

$$\begin{aligned}G(\mu_i,\mu_{-i})&\doteq 4\lambda\mu_i\mu_{-i}\left(4\lambda\mu_i+(\mu_i+\mu_{-i})(\mu_i+3\mu_{-i})\right)\\&\quad-k(\mu_i+\mu_{-i})\left(2\lambda(\mu_i-\mu_{-i})+\mu_{-i}(\mu_i+\mu_{-i})\right)\left(4\lambda\mu_{-i}+(\mu_i+\mu_{-i})^2\right)\end{aligned}$$

is decreasing in μ_{-i}.

Proof. We first show that $G(\mu_i,\mu_{-i})$ is concave in μ_{-i} when $\mu_i\geq 0$ and $\mu_{-i}\geq 2\lambda$:

$$\begin{aligned}\frac{\partial^2 G(\mu_i,\mu_{-i})}{\partial\mu_{-i}^2}&=-8k\mu_i^3-4(2k\lambda+9k\mu_{-i}-8\lambda)\mu_i^2-24\mu_{-i}(2k\mu_{-i}+k\lambda-3\lambda)\mu_i-4k\mu_{-i}(5\mu_{-i}^2+6\lambda\mu_{-i}-12\lambda^2)\\&\leq -8\mu_i^3-4(2\lambda+18\lambda-8\lambda)\mu_i^2-24\mu_{-i}(4\lambda+\lambda-3\lambda)\mu_i-4k\mu_{-i}(20\lambda^2+12\lambda^2-12\lambda^2)\\&=-8\mu_i^3-48\lambda\mu_i^2-48\lambda\mu_{-i}\mu_i-80k\mu_{-i}\lambda^2\\&\leq 0.\end{aligned}$$

We next show that $G(\mu_i,\mu_{-i})$ is decreasing in μ_{-i} at $\mu_{-i}=2\lambda$ when $\mu_i \geq 0$:

$$\begin{aligned}\frac{\partial G(\mu_i,2\lambda)}{\partial \mu_{-i}} &= 4\lambda\mu_i(2\lambda+\mu_i)(18\lambda+\mu_i) - k(48\lambda^4+176\lambda^3\mu_i+96\lambda^2\mu_i^2+20\lambda\mu_i^3+\mu_i^4)\\ &\leq 4\lambda\mu_i(2\lambda+\mu_i)(18\lambda+\mu_i) - (48\lambda^4+176\lambda^3\mu_i+96\lambda^2\mu_i^2+20\lambda\mu_i^3+\mu_i^4)\\ &= -48\lambda^4-32\lambda^3\mu_i-16\lambda^2\mu_i^2-16\lambda\mu_i^3-\mu_i^4\\ &< 0.\end{aligned}$$

Because $G(\mu_i,\mu_{-i})$ is concave in μ_{-i}, we obtain that $G(\mu_i,\mu_{-i})$ is decreasing in μ_{-i} for all $\mu_{-i}\geq 2\lambda$. □

LEMMA A-5. *For any $\mu_i>0$, $\mu_{-i}>2\lambda$, if either (i) $k\geq 1.03$ or (ii) $\mu_{-i}\geq 2.03\lambda$, the function*

$$F(\mu_i,\mu_{-i}) = \frac{(\mu_i+\mu_{-i})\left(2\lambda(\mu_i-\mu_{-i})+\mu_{-i}(\mu_i+\mu_{-i})\right)}{\mu_i^k(4\lambda\mu_{-i}+(\mu_i+\mu_{-i})^2)} \tag{A-7}$$

is decreasing in μ_i.

Proof. The proof uses Lemma A-4. We have

$$\begin{aligned}&\frac{\partial F(\mu_i,\mu_{-i})}{\partial \mu_i}\times \mu_i^{1+k}\left(4\lambda\mu_{-i}+(\mu_1+\mu_{-i})^2\right)^2\\ &= 4\lambda\mu_i\mu_{-i}\left(4\lambda\mu_i+(\mu_i+\mu_{-i})(\mu_i+3\mu_{-i})\right)\\ &\quad -k(\mu_i+\mu_{-i})\left(2\lambda(\mu_i-\mu_{-i})+\mu_{-i}(\mu_i+\mu_{-i})\right)\left(4\lambda\mu_{-i}+(\mu_i+\mu_{-i})^2\right)\\ &\doteq G(\mu_i,\mu_{-i})\end{aligned} \tag{A-8}$$

and therefore $F(\mu_i,\mu_{-i})$ is nonincreasing in μ_i if and only if $G(\mu_i,\mu_{-i})\leq 0$.
(i) Suppose first that $k\geq 1.03$. Because $G(\mu_i,\mu_{-i})$ is decreasing in μ_{-i} by Lemma A-4, we obtain, for any $\mu_{-i}>2\lambda$,

$$G(\mu_i,\mu_{-i}) < G(\mu_i,2\lambda) = -4\lambda^4\mu_i\left(24(k-1)+4(5k-6)\frac{\mu_i}{\lambda}+2(3k-1)\left(\frac{\mu_i}{\lambda}\right)^2+k\left(\frac{\mu_i}{\lambda}\right)^3\right).$$

The term in parentheses in the right-hand side is a cubic function of μ_i/λ, it is convex for all $\mu_i\geq 0$ and $k\geq 1$. When $k\leq 1+\sqrt{42}/6$, it reaches a local minimum at $\mu_i/\lambda = 2\left(1-3k+\sqrt{1+12k-6k^2}\right)/(3k)\geq 0$; otherwise, it is always increasing for any $\mu_i/\lambda\geq 0$. Suppose first that $k>1+\sqrt{42}/6$. For any $\mu_i>0$, we have

$$G(\mu_i,\mu_{-i}) < G(\mu_i,2\lambda) \leq -4\lambda^4\mu_i\left(24(k-1)\right)\leq 0.$$

Suppose next that $k\leq 1+\sqrt{42}/6$. For any $\mu_i\geq 0$, we have:

$$\begin{aligned}G(\mu_i,\mu_{-i}) &< G(\mu_i,2\lambda)\cdot\frac{\mu_i}{\frac{2\lambda\left(1-3k+\sqrt{1+12k-6k^2}\right)}{3k}}\\ &\leq G\left(2\lambda\frac{1-3k+\sqrt{1+12k-6k^2}}{3k},2\lambda\right)\\ &= -4\lambda^4\mu_i\frac{16\left(-1-\sqrt{1-6(k-2)k}+6k\left(-3-2\sqrt{1-6(k-2)k}+k\left(6+\sqrt{1-6(k-2)k}\right)\right)\right)}{27k^2}.\end{aligned}$$

Define $H(k)\doteq\left(-1-\sqrt{1-6(k-2)k}+6k\left(-3-2\sqrt{1-6(k-2)k}+k\left(6+\sqrt{1-6(k-2)k}\right)\right)\right)$. Since

$$\begin{aligned}H'(k) &= 18\left(-1-\sqrt{1-6(k-2)k}+k\left(4+\sqrt{1-6(k-2)k}\right)\right)\\ &\geq 18\left(-1-\sqrt{1-6(k-2)k}+\left(4+\sqrt{1-6(k-2)k}\right)\right)\\ &= 18\times 3>0,\end{aligned}$$

and since $H(k)=0$ when

$$\begin{aligned}
k &= 3/2-\frac{1}{4\sqrt{\frac{3}{-32+(111448-11376\sqrt{79})^{\frac{1}{3}}+2(13931+1422\sqrt{79})^{\frac{1}{3}}}}} \\
&\quad +1/2\sqrt{-\frac{16}{3}-\frac{(111448-11376\sqrt{79})^{\frac{1}{3}}}{12}-\frac{(13931+1422\sqrt{79})^{\frac{1}{3}}}{6}+54\sqrt{\frac{3}{-32+(111448-11376\sqrt{79})^{\frac{1}{3}}+2(13931+1422\sqrt{79})^{\frac{1}{3}}}}} \\
&\approx 1.0273259295534587,
\end{aligned}$$

we conclude that $H(k)>0$ for all $k\geq 1.03$. Hence, for any $k\in[1.03,1+\sqrt{42}/6]$ and $\mu_i\geq 0$,

$$G(\mu_i,\mu_{-i}) < G\left(2\lambda\frac{1-3k+\sqrt{1+12k-6k^2}}{3k},2\lambda\right)\cdot\frac{\mu_i}{\frac{2\lambda\left(1-3k+\sqrt{1+12k-6k^2}\right)}{3k}}=-4\lambda^4\mu_i\frac{16}{27k^2}H(k)\leq 0.$$

Summarizing both cases, we thus find that $G(\mu_i,\mu_{-i})<0$ for all $\mu_i\geq 0$, $\mu_{-i}>2\lambda$ and $k\geq 1.03$.

(ii) Suppose that $\mu_i\geq 2.03\lambda$. For this part, we include k in the arguments of the function $G(\mu_i,\mu_{-i})$ defined in (A-8). Because $G(\mu_i,\mu_{-i},k)$ is decreasing in μ_{-i} by Lemma A-4 and because it is decreasing in k when $\mu_{-i}>2\lambda$, we obtain, for any $\mu_{-i}\geq 2.03\lambda$ and $k\geq 1$,

$$\begin{aligned}
G(\mu_i,\mu_{-i},k) &\leq G(\mu_i,2.03\lambda,1) \\
&\leq \lambda^5\left(-1.51331-1.00385\frac{\mu_i}{\lambda}+15.4982\left(\frac{\mu_i}{\lambda}\right)^2-16.4836\left(\frac{\mu_i}{\lambda}\right)^3-4.03\left(\frac{\mu_i}{\lambda}\right)^4\right).
\end{aligned}$$

The term in parentheses is a quartic polynomial in μ_i/λ with three stationary points: at $\mu_i/\lambda=-3.60574$ (local maximum), at $\mu_i/\lambda=0.0342819$ (local minimum), and at $\mu_i/\lambda=0.503789$ (local maximum) and tends to negative infinity when $\mu_i/\lambda\to\infty$. Hence, for any $\mu_i>0$,

$$G(\mu_i,\mu_{-i},k) < G(0.503789\lambda,2.03\lambda,1)\leq -0.452775\lambda^5<0.$$

Hence, $G(\mu_i,\mu_{-i},k)<0$ for all $\mu_i>0$, $k\geq 1$ and $\mu_{-i}\geq 2.03\lambda$. □

LEMMA A-6. *For any $\mu_{-i}>2\lambda$, the function $F(\mu_i,\mu_{-i})$ defined in (A-7) is decreasing in μ_i for all $\mu_i\geq\lambda$.*

Proof. Let $G(\mu_i,\mu_{-i})$ defined as in (A-8). Therefore $F(\mu_i,\mu_{-i})$ is decreasing in μ_i if and only if $G(\mu_i,\mu_{-i})<0$.

When $k\geq 1$ and $\mu_{-i}>2\lambda$, we have, for all $\mu_i\geq\lambda$,

$$\begin{aligned}
\frac{\partial^2 G(\mu_i,\mu_{-i})}{\partial\mu_i^2} &= -4\left(-2\lambda\mu_{-i}(4\lambda+3\mu_i+4\mu_{-i})+k\left(4\lambda^2\mu_{-i}+3\mu_{-i}(\mu_i+\mu_{-i})^2+2\lambda(3\mu_i^2+3\mu_i\mu_{-i}+\mu_{-i}^2)\right)\right) \\
&\leq -4\left(-2\lambda\mu_{-i}(4\lambda+3\mu_i+4\mu_{-i})+\left(4\lambda^2\mu_{-i}+3\mu_{-i}(\mu_i+\mu_{-i})^2+2\lambda(3\mu_i^2+3\mu_i\mu_{-i}+\mu_{-i}^2)\right)\right) \\
&= -4\mu_i^2(6\lambda+3\mu_{-i})-24\mu_i\mu_{-i}^2+4(4\lambda^2\mu_{-i}+6\lambda\mu_{-i}^2-3\mu_{-i}^3) \\
&\leq -4\lambda^2(6\lambda+3\mu_{-i})-24\lambda\mu_{-i}^2+4(4\lambda^2\mu_{-i}+6\lambda\mu_{-i}^2-3\mu_{-i}^3) \\
&= 4(-6\lambda^3+\lambda^2\mu_{-i}-3\mu_{-i}^3).
\end{aligned}$$

Let $K(\mu_{-i})=(-6\lambda^3+\lambda^2\mu_{-i}-3\mu_{-i}^3)$. Because $K(2\lambda)=-28\lambda^3<0$ and, for all $\mu_{-i}>2\lambda$, $K'(\mu_{-i})=\lambda^2-9\mu_{-i}^2<-35\lambda^2<0$, we obtain that $K(\mu_{-i})<0$ for all $\mu_{-i}>2\lambda$, and therefore $G(\mu_i,\mu_{-i})$ is strictly concave in μ_i for all $\mu_i\geq\lambda$ and $\mu_{-i}>2\lambda$.

Next, we obtain, using the facts that $k \geq 1$ and $\mu_{-i} > 2\lambda$,

$$\begin{aligned}
\frac{\partial G(\lambda,\mu_{-i})}{\partial \mu_i} &= 4\mu_{-i}\lambda(11\lambda^2+8\lambda\mu_{-i}+3\mu_{-i}^2) - 4k(2\lambda^4+8\lambda^3\mu_{-i}+5\lambda^2\mu_{-i}^2+4\lambda\mu_{-i}^3+\mu_{-i}^4)\\
&\leq 4\mu_{-i}\lambda(11\lambda^2+8\lambda\mu_{-i}+3\mu_{-i}^2) - 4(2\lambda^4+8\lambda^3\mu_{-i}+5\lambda^2\mu_{-i}^2+4\lambda\mu_{-i}^3+\mu_{-i}^4)\\
&= -4(2\lambda^4-3\lambda^3\mu_{-i}-3\lambda^2\mu_{-i}^2+\lambda\mu_{-i}^3+\mu_{-i}^4)\\
&\leq -4(2\lambda^4-3\lambda^3\mu_{-i}-3\lambda^2\mu_{-i}^2+3\lambda\mu_{-i}^3)\\
&\leq -4(2\lambda^4-3\lambda^3\mu_{-i}+3\lambda^2\mu_{-i}^2)\\
&\leq -4(2\lambda^4+3\lambda^3\mu_{-i})\\
&< 0.
\end{aligned}$$

Because $G(\mu_i,\mu_{-i})$ is strictly concave in μ_i for all $\mu_i \geq \lambda$ and $\mu_{-i} > 2\lambda$, this implies that $G(\mu_i,\mu_{-i})$ is decreasing in μ_i for all $\mu_i \geq \lambda$ and $\mu_{-i} > 2\lambda$.

Finally, using the facts that $k \geq 1$ and $\mu_{-i} > 2\lambda$, observe that

$$\begin{aligned}
G(\lambda,\mu_{-i}) &= 4\lambda^2\mu_{-i}(5\lambda^2+4\lambda\mu_{-i}+3\mu_{-i}^2) - k(\lambda+\mu_{-i})(2\lambda^2-\lambda\mu_{-i}+\mu_{-i}^2)(\lambda^2+6\lambda\mu_{-i}+\mu_{-i}^2)\\
&\leq 4\lambda^2\mu_{-i}(5\lambda^2+4\lambda\mu_{-i}+3\mu_{-i}^2) - (\lambda+\mu_{-i})(2\lambda^2-\lambda\mu_{-i}+\mu_{-i}^2)(\lambda^2+6\lambda\mu_{-i}+\mu_{-i}^2)\\
&= -2\lambda^5+7\lambda^4\mu_{-i}+8\lambda^3\mu_{-i}^2+10\lambda^2\mu_{-i}^3-6\lambda\mu_{-i}^4-\mu_{-i}^5\\
&\leq -2\lambda^5+7\lambda^4\mu_{-i}+8\lambda^3\mu_{-i}^2+10\lambda^2\mu_{-i}^3-8\lambda\mu_{-i}^4\\
&\leq -2\lambda^5+7\lambda^4\mu_{-i}+8\lambda^3\mu_{-i}^2-6\lambda^2\mu_{-i}^3\\
&\leq -2\lambda^5+7\lambda^4\mu_{-i}-4\lambda^3\mu_{-i}^2\\
&\leq -2\lambda^5-\lambda^4\mu_{-i} < 0.
\end{aligned}$$

Because $G(\mu_i,\mu_{-i})$ is decreasing in μ_i for all $\mu_i \geq \lambda$, this implies that $G(\mu_i,\mu_{-i}) < 0$ for all $\mu_i \geq \lambda$ and $\mu_{-i} > 2\lambda$. □

LEMMA A-7. *For any $k < 1.03$ and $\mu_{-i} > 2\lambda$, suppose that both $\underline{\mu}_i$ and $\overline{\mu}_i$ are local minima of $C_i(\mu_i;\mu_{-i})$, defined in (1), with $0 < \underline{\mu}_i < \lambda < \overline{\mu}_i$. Then, $F(\underline{\mu}_i,\mu_{-i}) > F(\overline{\mu}_i,\mu_{-i})$, where $F(\mu_i,\mu_{-i})$ is defined in (A-7).*

Proof. The proof uses Lemmas A-2 and A-6. Suppose, to obtain a contradiction, that $F(\underline{\mu}_i,\mu_{-i}) \leq F(\overline{\mu}_i,\mu_{-i})$. Define $\tilde{\mu}_i = \mu_i/\lambda$ and $\tilde{\mu}_{-i} = \mu_{-i}/\lambda$. Moreover, define $\tilde{F}(\mu_i,\mu_{-i})$ as identical to $F(\mu_i,\mu_{-i})$ with $\lambda = 1$. Hence, using (A-7),

$$\begin{aligned}
&F(\underline{\mu}_i,\mu_{-i}) \leq F(\overline{\mu}_i,\mu_{-i})\\
\Leftrightarrow\ &\frac{(\underline{\mu}_i+\mu_{-i})\left(2\lambda(\underline{\mu}_i-\mu_{-i})+\mu_{-i}(\underline{\mu}_i+\mu_{-i})\right)}{\underline{\mu}_i^k(4\lambda\mu_{-i}+(\underline{\mu}_i+\mu_{-i})^2)} \leq \frac{(\overline{\mu}_i+\mu_{-i})\left(2\lambda(\overline{\mu}_i-\mu_{-i})+\mu_{-i}(\overline{\mu}_i+\mu_{-i})\right)}{\overline{\mu}_i^k(4\lambda\mu_{-i}+(\overline{\mu}_i+\mu_{-i})^2)}\\
\Leftrightarrow\ &\frac{(\underline{\tilde{\mu}}_i+\tilde{\mu}_{-i})\left(2(\underline{\tilde{\mu}}_i-\tilde{\mu}_{-i})+\tilde{\mu}_{-i}(\underline{\tilde{\mu}}_i+\tilde{\mu}_{-i})\right)}{\underline{\tilde{\mu}}_i^k(4\tilde{\mu}_{-i}+(\underline{\tilde{\mu}}_i+\tilde{\mu}_{-i})^2)} \leq \frac{(\overline{\tilde{\mu}}_i+\tilde{\mu}_{-i})\left(2(\overline{\tilde{\mu}}_i-\tilde{\mu}_{-i})+\tilde{\mu}_{-i}(\overline{\tilde{\mu}}_i+\tilde{\mu}_{-i})\right)}{\overline{\tilde{\mu}}_i^k(4\tilde{\mu}_{-i}+(\overline{\tilde{\mu}}_i+\tilde{\mu}_{-i})^2)}\\
\Leftrightarrow\ &\tilde{F}(\underline{\tilde{\mu}}_i,\tilde{\mu}_{-i}) \leq \tilde{F}(\overline{\tilde{\mu}}_i,\tilde{\mu}_{-i}).
\end{aligned}$$

Because $\overline{\tilde{\mu}}_i > 1$, we obtain from Lemma A-6 that $\tilde{F}(\overline{\tilde{\mu}}_i, \tilde{\mu}_{-i}) < \tilde{F}(1, \tilde{\mu}_{-i})$. Moreover, it can easily be checked that $\tilde{F}(\tilde{\mu}_i, \tilde{\mu}_{-i})$ is increasing in k for all $\tilde{\mu}_i \leq 1$ when $\tilde{\mu}_{-i} > 2$. Hence, because $k \geq 1$,

$$\frac{(\underline{\tilde{\mu}}_i + \tilde{\mu}_{-i})\left(2(\underline{\tilde{\mu}}_i - \tilde{\mu}_{-i}) + \tilde{\mu}_{-i}(\underline{\tilde{\mu}}_i + \tilde{\mu}_{-i})\right)}{\underline{\tilde{\mu}}_i(4\tilde{\mu}_{-i} + (\underline{\tilde{\mu}}_i + \tilde{\mu}_{-i})^2)} \leq \tilde{F}(\underline{\tilde{\mu}}_i, \tilde{\mu}_{-i})$$
$$\leq \tilde{F}(\overline{\tilde{\mu}}_i, \tilde{\mu}_{-i}) < \frac{(1 + \tilde{\mu}_{-i})(2(1 - \tilde{\mu}_{-i}) + \tilde{\mu}_{-i}(1 + \tilde{\mu}_{-i}))}{4\tilde{\mu}_{-i} + (1 + \tilde{\mu}_{-i})^2},$$

and therefore, combining the extreme ends of these inequalities,

$$(\underline{\tilde{\mu}}_i + \tilde{\mu}_{-i})\left(2(\underline{\tilde{\mu}}_i - \tilde{\mu}_{-i}) + \tilde{\mu}_{-i}(\underline{\tilde{\mu}}_i + \tilde{\mu}_{-i})\right)(4\tilde{\mu}_{-i} + (1 + \tilde{\mu}_{-i})^2)$$
$$< (1 + \tilde{\mu}_{-i})(2(1 - \tilde{\mu}_{-i}) + \tilde{\mu}_{-i}(1 + \tilde{\mu}_{-i}))\,\underline{\tilde{\mu}}_i(4\tilde{\mu}_{-i} + (\underline{\tilde{\mu}}_i + \tilde{\mu}_{-i})^2),$$

which is equivalent to requiring that the following function

$$K(\underline{\tilde{\mu}}_i; \tilde{\mu}_{-i}) \doteq -\underline{\tilde{\mu}}_i^3(1 + \tilde{\mu}_{-i})(2 - \tilde{\mu}_{-i} + \tilde{\mu}_{-i}^2) + \underline{\tilde{\mu}}_i^2\left(2 + 9\tilde{\mu}_{-i} + 6\tilde{\mu}_{-i}^2 + \tilde{\mu}_{-i}^3 - 2\tilde{\mu}_{-i}^4\right)$$
$$+\underline{\tilde{\mu}}_i\tilde{\mu}_{-i}\left(-8 - 4\tilde{\mu}_{-i} + 11\tilde{\mu}_{-i}^2 - 2\tilde{\mu}_{-i}^3 - \tilde{\mu}_{-i}^4\right) + \tilde{\mu}_{-i}^2(\tilde{\mu}_{-i} - 2)(1 + 6\tilde{\mu}_{-i} + \tilde{\mu}_{-i}^2) \quad \text{(A-9)}$$

be negative. Hence, we will obtain a contradiction if we can show that $K(\underline{\tilde{\mu}}_i; \tilde{\mu}_{-i}) \geq 0$ for all $0 < \underline{\mu}_i < \lambda$ and $\mu_{-i} > 2\lambda$.

This cubic polynomial $K(\underline{\tilde{\mu}}_i; \tilde{\mu}_{-i})$ has one root at $\tilde{\mu}_i = 1$ when $\tilde{\mu}_{-i} > 2.01635$. When $\tilde{\mu}_{-i} \leq 2.01635$, it has three roots, at $\tilde{\mu}_i = 1$ and at

$$\tilde{\mu}_i^L(\tilde{\mu}_{-i}) = \frac{4\tilde{\mu}_{-i} + 3\tilde{\mu}_{-i}^2 - \tilde{\mu}_{-i}^4 - 2\tilde{\mu}_{-i}\sqrt{5 + 12\tilde{\mu}_{-i} + 3\tilde{\mu}_{-i}^2 - 3\tilde{\mu}_{-i}^3 + \tilde{\mu}_{-i}^4 - \tilde{\mu}_{-i}^5}}{2 + \tilde{\mu}_{-i} + \tilde{\mu}_{-i}^3}$$
$$\tilde{\mu}_i^R(\tilde{\mu}_{-i}) = \frac{4\tilde{\mu}_{-i} + 3\tilde{\mu}_{-i}^2 - \tilde{\mu}_{-i}^4 + 2\tilde{\mu}_{-i}\sqrt{5 + 12\tilde{\mu}_{-i} + 3\tilde{\mu}_{-i}^2 - 3\tilde{\mu}_{-i}^3 + \tilde{\mu}_{-i}^4 - \tilde{\mu}_{-i}^5}}{2 + \tilde{\mu}_{-i} + \tilde{\mu}_{-i}^3}.$$

We next show that when $\tilde{\mu}_{-i} > 2$, $\tilde{\mu}_i^R(\tilde{\mu}_{-i}) < 1$. Indeed,

$$\tilde{\mu}_i^R(\tilde{\mu}_{-i}) < 1 \Leftrightarrow 4\tilde{\mu}_{-i} + 3\tilde{\mu}_{-i}^2 - \tilde{\mu}_{-i}^4 + 2\tilde{\mu}_{-i}\sqrt{5 + 12\tilde{\mu}_{-i} + 3\tilde{\mu}_{-i}^2 - 3\tilde{\mu}_{-i}^3 + \tilde{\mu}_{-i}^4 - \tilde{\mu}_{-i}^5} < 2 + \tilde{\mu}_{-i} + \tilde{\mu}_{-i}^3$$
$$\Leftrightarrow 2\tilde{\mu}_{-i}\sqrt{5 + 12\tilde{\mu}_{-i} + 3\tilde{\mu}_{-i}^2 - 3\tilde{\mu}_{-i}^3 + \tilde{\mu}_{-i}^4 - \tilde{\mu}_{-i}^5} < 2 - 3\tilde{\mu}_{-i} - 3\tilde{\mu}_{-i}^2 + \tilde{\mu}_{-i}^3 + \tilde{\mu}_{-i}^4$$
$$\Leftrightarrow 4\tilde{\mu}_{-i}^2\left(5 + 12\tilde{\mu}_{-i} + 3\tilde{\mu}_{-i}^2 - 3\tilde{\mu}_{-i}^3 + \tilde{\mu}_{-i}^4 - \tilde{\mu}_{-i}^5\right) < \left(2 - 3\tilde{\mu}_{-i} - 3\tilde{\mu}_{-i}^2 + \tilde{\mu}_{-i}^3 + \tilde{\mu}_{-i}^4\right)^2$$
$$\Leftrightarrow -4 + 12\tilde{\mu}_{-i} + 23\tilde{\mu}_{-i}^2 + 26\tilde{\mu}_{-i}^3 + 5\tilde{\mu}_{-i}^4 + 9\tilde{\mu}_{-i}^6 - 6\tilde{\mu}_{-i}^7 - \tilde{\mu}_{-i}^8 < 0.$$

Because $\tilde{\mu}_{-i} > 2$, we obtain

$$-4 + 12\tilde{\mu}_{-i} + 23\tilde{\mu}_{-i}^2 + 26\tilde{\mu}_{-i}^3 + 5\tilde{\mu}_{-i}^4 + 9\tilde{\mu}_{-i}^6 - 6\tilde{\mu}_{-i}^7 - \tilde{\mu}_{-i}^8$$
$$\leq -4 + 12\tilde{\mu}_{-i} + 23\tilde{\mu}_{-i}^2 + 26\tilde{\mu}_{-i}^3 + 5\tilde{\mu}_{-i}^4 + 9\tilde{\mu}_{-i}^6 - 8\tilde{\mu}_{-i}^7$$
$$\leq -4 + 12\tilde{\mu}_{-i} + 23\tilde{\mu}_{-i}^2 + 26\tilde{\mu}_{-i}^3 + 5\tilde{\mu}_{-i}^4 - 7\tilde{\mu}_{-i}^6$$
$$\leq -4 + 12\tilde{\mu}_{-i} + 23\tilde{\mu}_{-i}^2 + 26\tilde{\mu}_{-i}^3 - 23\tilde{\mu}_{-i}^4$$
$$\leq -4 + 12\tilde{\mu}_{-i} + 23\tilde{\mu}_{-i}^2 - 20\tilde{\mu}_{-i}^3$$
$$\leq -4 + 12\tilde{\mu}_{-i} - 17\tilde{\mu}_{-i}^2$$
$$\leq -4 - 22\tilde{\mu}_{-i}$$
$$< 0.$$

Hence, $\tilde{\mu}_i^L(\tilde{\mu}_{-i}) < \tilde{\mu}_i^R(\tilde{\mu}_{-i}) < 1$. Because the cubic term in $\underline{\tilde{\mu}}_i$ in the expression $K(\underline{\tilde{\mu}}_i;\tilde{\mu}_{-i})$ has a negative coefficient, $K(\underline{\tilde{\mu}}_i;\tilde{\mu}_{-i}) < 0$ for all $\underline{\tilde{\mu}}_i > 1$ when $\tilde{\mu}_{-i} > 2.01635$, and for all $\underline{\tilde{\mu}}_i \in (\tilde{\mu}_i^L(\tilde{\mu}_{-i}), \tilde{\mu}_i^R(\tilde{\mu}_{-i})) \cup (1,\infty)$ when $2 < \tilde{\mu}_{-i} \leq 2.01635$.

By assumption, $\underline{\tilde{\mu}}_i \leq 1$, so we can have $K(\underline{\tilde{\mu}}_i;\tilde{\mu}_{-i}) < 0$ only if $2 < \tilde{\mu}_{-i} \leq 2.01635$, which we assume in the sequel. Under that condition, $K(\underline{\tilde{\mu}}_i;\tilde{\mu}_{-i}) < 0$ if $\underline{\tilde{\mu}}_i \in (\tilde{\mu}_i^L(\tilde{\mu}_{-i}), \tilde{\mu}_i^R(\tilde{\mu}_{-i})) \cup (1,\infty)$.

We next consider the function $H(\tilde{\mu}_i, \tilde{\mu}_{-i}, k)$, which was defined in (A-5) in the proof of Lemma A-2, but extend its set of arguments to include k. It can easily be checked that, when $\tilde{\mu}_{-i} > 2$, $H(\tilde{\mu}_i, \tilde{\mu}_{-i}, k)$ is increasing in k. Similar to the proof of Lemma A-2, every stationary point of $\hat{C}_i(\tilde{\mu}_i\lambda;\tilde{\mu}_{-i}\lambda)$, defined in (A-3), such that $H(\tilde{\mu}_i, \tilde{\mu}_{-i}, k) > 0$ is a strict local minimum and such that $H(\tilde{\mu}_i, \tilde{\mu}_{-i}, k) < 0$ is a strict local maximum. When $\tilde{\mu}_{-i} > 2$, $C_i(\tilde{\mu}_i\lambda;\tilde{\mu}_{-i}\lambda) = \hat{C}_i(\tilde{\mu}_i\lambda;\tilde{\mu}_{-i}\lambda)$ for all $\tilde{\mu}_i$, so the same holds for the stationary points of $C_i(\tilde{\mu}_i\lambda;\tilde{\mu}_{-i}\lambda)$. The proof of Lemma A-2 shows that, when $\tilde{\mu}_{-i} > 2$, $H(\tilde{\mu}_i, \tilde{\mu}_{-i})$ is either decreasing and then increasing, or always increasing. Because $\underline{\tilde{\mu}}_i$ and $\overline{\tilde{\mu}}_i$ are local minima of $C_i(\tilde{\mu}_i\lambda;\tilde{\mu}_{-i}\lambda)$, there must be some local maximum lying between these two minima; hence, $H(\tilde{\mu}_i, \tilde{\mu}_{-i})$ must cross zero twice, first from above and then from below, and the first crossing happens at or after $\underline{\tilde{\mu}}_i$ and the second crossing happens at or before $\overline{\tilde{\mu}}_i$. In particular, $H(\tilde{\mu}_i, \tilde{\mu}_{-i})$ is decreasing in $\tilde{\mu}_i$ for all $\tilde{\mu}_i \leq \underline{\tilde{\mu}}_i$.

Because $\tilde{\mu}_i^L(\tilde{\mu}_{-i}) < \underline{\tilde{\mu}}_i$ and because $k \leq 1.03$, we must thus have

$$0 \leq H(\underline{\tilde{\mu}}_i, \tilde{\mu}_{-i}, k) \leq H(\tilde{\mu}_i^L(\tilde{\mu}_{-i}), \tilde{\mu}_{-i}, k) < H(\tilde{\mu}_i^L(\tilde{\mu}_{-i}), \tilde{\mu}_{-i}, 1.03).$$

It can be numerically checked that the univariate function $H(\tilde{\mu}_i^L(\tilde{\mu}_{-i}), \tilde{\mu}_{-i}, 1.03)$ is nonpositive and decreasing in $\tilde{\mu}_{-i}$ when $\tilde{\mu}_{-i} \in [2, 2.01635]$. Hence, $0 < H(\tilde{\mu}_i^L(\tilde{\mu}_{-i}), \tilde{\mu}_{-i}, 1.03) < H(\tilde{\mu}_i^L(2), 2, 1.03) = 0$, a contradiction. □

Proof of Lemma 2. The proof uses Lemmas A-3, A-5, and A-7. Because $\underline{\mu}_i$ and $\overline{\mu}_i$ minimize $C_i(\mu_i;\mu_{-i})$ in the interior of its domain and because $\mu_1 + \mu_2 > 2\lambda$, the first-order optimality conditions (A-6) are satisfied at both $\underline{\mu}_i$ and $\overline{\mu}_i$. Therefore,

$$\frac{k}{h\lambda} = \frac{(\underline{\mu}_i + \mu_{-i})\left(2\lambda(\underline{\mu}_i - \mu_{-i}) + \mu_{-i}(\underline{\mu}_i + \mu_{-i})\right)}{\underline{\mu}_i^{k-1}\left(\underline{\mu}_i^2(\lambda + \mu_{-i}) + \mu_{-i}^2\left(\lambda + \underline{\mu}_i\right)\right)^2} = \frac{(\overline{\mu}_i + \mu_{-i})(2\lambda(\overline{\mu}_i - \mu_{-i}) + \mu_{-i}(\overline{\mu}_i + \mu_{-i}))}{\overline{\mu}_i^{k-1}(\overline{\mu}_i^2(\lambda + \mu_{-i}) + \mu_{-i}^2(\lambda + \overline{\mu}_i))^2}$$

$$\Leftrightarrow \frac{\underline{\mu}_i^{k-1}(\overline{\mu}_i + \mu_{-i})(2\lambda(\overline{\mu}_i - \mu_{-i}) + \mu_{-i}(\overline{\mu}_i + \mu_{-i}))}{\overline{\mu}_i^{k-1}(\underline{\mu}_i + \mu_{-i})\left(2\lambda(\underline{\mu}_i - \mu_{-i}) + \mu_{-i}(\underline{\mu}_i + \mu_{-i})\right)} = \frac{\left(\overline{\mu}_i^2(\lambda + \mu_{-i}) + \mu_{-i}^2(\lambda + \overline{\mu}_i)\right)^2}{\left(\underline{\mu}_i^2(\lambda + \mu_{-i}) + \mu_{-i}^2\left(\lambda + \underline{\mu}_i\right)\right)^2}. \quad \text{(A-10)}$$

Suppose, for contradiction, that

$$\frac{\partial C_i(\underline{\mu}_i;\mu_{-i})}{\partial \mu_{-i}} \leq \frac{\partial C_i(\overline{\mu}_i;\mu_{-i})}{\partial \mu_{-i}},$$

i.e., after taking the derivative of (1) with respect to μ_{-i}, that

$$-h\lambda^2 \frac{\underline{\mu}_i(4\lambda\mu_{-i} + (\underline{\mu}_i + \mu_{-i})^2)}{\left(\underline{\mu}_i^2(\lambda + \mu_{-i}) + \mu_{-i}^2\left(\lambda + \underline{\mu}_i\right)\right)^2} \leq -h\lambda^2 \frac{\overline{\mu}_i(4\lambda\mu_{-i} + (\overline{\mu}_i + \mu_{-i})^2)}{(\overline{\mu}_i^2(\lambda + \mu_{-i}) + \mu_{-i}^2(\lambda + \overline{\mu}_i))^2}$$

$$\Leftrightarrow \frac{\left(\overline{\mu}_i^2(\lambda + \mu_{-i}) + \mu_{-i}^2(\lambda + \overline{\mu}_i)\right)^2}{\left(\underline{\mu}_i^2(\lambda + \mu_{-i}) + \mu_{-i}^2\left(\lambda + \underline{\mu}_i\right)\right)^2} \geq \frac{\overline{\mu}_i(4\lambda\mu_{-i} + (\overline{\mu}_i + \mu_{-i})^2)}{\underline{\mu}_i(4\lambda\mu_{-i} + (\underline{\mu}_i + \mu_{-i})^2)}.$$

In that case, using (A-10), we obtain

$$\frac{\underline{\mu}_i^{k-1}(\overline{\mu}_i+\mu_{-i})\left(2\lambda(\overline{\mu}_i-\mu_{-i})+\mu_{-i}(\overline{\mu}_i+\mu_{-i})\right)}{\overline{\mu}_i^{k-1}(\underline{\mu}_i+\mu_{-i})\Big(2\lambda(\underline{\mu}_i-\mu_{-i})+\mu_{-i}(\underline{\mu}_i+\mu_{-i})\Big)} \geq \frac{\overline{\mu}_i(4\lambda\mu_{-i}+(\overline{\mu}_i+\mu_{-i})^2)}{\underline{\mu}_i(4\lambda\mu_{-i}+(\underline{\mu}_i+\mu_{-i})^2)}$$

$$\Leftrightarrow \frac{(\overline{\mu}_i+\mu_{-i})\left(2\lambda(\overline{\mu}_i-\mu_{-i})+\mu_{-i}(\overline{\mu}_i+\mu_{-i})\right)}{\overline{\mu}_i^k(4\lambda\mu_{-i}+(\overline{\mu}_i+\mu_{-i})^2)} \geq \frac{(\underline{\mu}_i+\mu_{-i})\Big(2\lambda(\underline{\mu}_i-\mu_{-i})+\mu_{-i}(\underline{\mu}_i+\mu_{-i})\Big)}{\underline{\mu}_i^k(4\lambda\mu_{-i}+(\underline{\mu}_i+\mu_{-i})^2)}$$

$$\Leftrightarrow F(\overline{\mu}_i,\mu_{-i}) \geq F(\underline{\mu}_i,\mu_{-i}),$$

in which $F(\mu_i,\mu_{-i})$ is defined in (A-7).

We next consider two cases, depending on whether (i) $k<1.03$ and $\mu_{-i}<2.03\lambda$ or (ii) either $k\geq 1.03$ or $\mu_{-i}\geq 2.03\lambda$. Recall that, in both cases, we assume that $\mu_{-i}>2\lambda$. Suppose first that $k<1.03$ and $\mu_{-i}<2.03\lambda$. By Lemma A-3, $\underline{\mu}_i<\lambda<\overline{\mu}_i$. Hence, by Lemma A-7, $F(\overline{\mu}_i,\mu_{-i})<F(\underline{\mu}_i,\mu_{-i})$, a contradiction. Suppose next that either $k\geq 1.03$ or $\mu_{-i}\geq 2.03\lambda$. Because $F(\mu_i,\mu_{-i})$ is decreasing in μ_i when either $k\geq 1.03$ or $\mu_{-i}\geq 2.03\lambda$ by Lemma A-5, and $0<\underline{\mu}_i<\overline{\mu}_i$, we must have $F(\overline{\mu}_i,\mu_{-i})<F(\underline{\mu}_i,\mu_{-i})$, a contradiction. $\square$