Online Appendix to "Pooling Agents for Customer-Intensive Services"

Appendix A: Technical Details

A.1. Proofs of Results from the Base Model

Proof of Lemma 1 By Figure 1, we have the following balance equations where $\{\pi_0, \pi_{1_1}, \pi_{1_2}, \pi_i := \pi_{1i_1} + \pi_{i_2}, \pi_2, \pi_3, \ldots\}$ are the steady-state probabilities:

$$\pi_0(2\lambda) = \pi_{1_1}\mu_1 + \pi_{1_2}\mu_2, \tag{A.1}$$

$$\pi_{1_1}(2\lambda + \mu_1) = \pi_0\lambda + \pi_2\mu_2, \tag{A.2}$$

$$\pi_{1_2}(2\lambda + \mu_2) = \pi_0\lambda + \pi_2\mu_1, \tag{A.3}$$

$$\pi_i(2\lambda + \mu_1 + \mu_2) = \pi_{i-1}(2\lambda) + \pi_{i+1}(\mu_1 + \mu_2) \text{ for } i \geq 2, \tag{A.4}$$

$$\sum_{i=1}^{\infty} \pi_i = 1. \tag{A.5}$$

It follows from (A.4) that

$$\pi_i = \pi_1 \cdot \overline{\rho}^{\,i-1} \text{ for } i \geq 1 \tag{A.6}$$

where $\overline{\rho} = 2\lambda/(\mu_1 + \mu_2)$. Because $\pi_0 + \sum_{i=1}^{\infty} \pi_i = 1$, we have

$$\pi_1 = (1 - \overline{\rho})(1 - \pi_0). \tag{A.7}$$

Substituting (A.6) and (A.7) into (A.2) and (A.3) gives

$$\pi_{1_1} = \frac{\mu_2\overline{\rho}(1-\overline{\rho})(1-\pi_0) + \lambda\pi_0}{2\lambda + \mu_1}, \tag{A.8}$$

$$\pi_{1_2} = \frac{\mu_1\overline{\rho}(1-\overline{\rho})(1-\pi_0) + \lambda\pi_0}{2\lambda + \mu_2}, \tag{A.9}$$

Combining (A.1), (A.8), and (A.9) gives

$$\pi_0 = \frac{\mu_1\mu_2(1-\overline{\rho})\overline{\rho}}{2\lambda^2 + \mu_1\mu_2(1-\overline{\rho})\overline{\rho}}.$$

Hence, all the other steady-state probabilities can be derived by plugging π_0 into (A.7)-(A.9). The expected queue length is $L^Q(\mu_1, \mu_2) = \sum_{i=3}^{\infty} \pi_i(i-2)$. Thus, by Little's Law, the expected wait time in the queue is given by $W(\mu_1, \mu_2) = L^Q(\mu_1, \mu_2)/(2\lambda) = \frac{\lambda}{2\lambda^2 + \mu_1\mu_2(1-\overline{\rho})\overline{\rho}} \cdot \frac{\overline{\rho}^2}{1-\overline{\rho}}$, which proves part (i). For part (ii), the throughput of Agent 1 is

$$\lambda_1 = \mu_1 \cdot \left[\pi_{1_1} + \sum_{i=2}^{\infty} \pi_i\right].$$

Plugging the steady-state probabilities gives $\lambda_1 = \frac{\lambda\bar{\rho}\mu_1[2\lambda+\mu_2(1-\bar{\rho})]}{2\lambda^2+\mu_1\mu_2(1-\bar{\rho})\bar{\rho}}$. λ_2 can be derived similarly. □

Proof of Proposition 1

First, we search for an equilibrium when there are no customers balking, that is, all customers join in equilibrium or $\lambda_1 = \lambda_2 = \lambda$ in which case the agents will set service rate $> \lambda$. We can derive the first-order and second-order conditions of (2) with respect to μ as follows.

$$\frac{\partial[\Pi_1^D(\mu)]}{\partial\mu} = \lambda\left(\frac{c\lambda(2\mu-\lambda)}{(\mu-\lambda)^2\mu^2} + \frac{e}{\mu^2} - \alpha\right)$$
$$\frac{\partial^2[\Pi_1^D(\mu)]}{\partial\mu^2} = -\frac{2\lambda\left[c\lambda\left(\lambda^2-3\lambda\mu+3\mu^2\right)+e(\mu-\lambda)^3\right]}{\mu^3(\mu-\lambda)^3} < 0 \text{ because } \mu > \lambda.$$

It implies that $\frac{\partial[\Pi_1^D(\mu)]}{\partial\mu}$ is decreasing in $\mu > \lambda$. Since $\frac{\partial[\Pi_1^D(\mu)]}{\partial\mu}|_{\mu=\lambda^+} = +\infty$ and $\frac{\partial[\Pi_1^D(\mu)]}{\partial\mu}|_{\mu=\infty} = -\lambda\alpha$, there exists a unique μ^D that maximizes $\Pi_1^D(\mu)$ and μ^D uniquely solves $\frac{\partial[\Pi_1^D(\mu)]}{\partial\mu} = 0$. Define $\theta_2 := \lambda^2\alpha$ and $\phi_D(\rho,\theta_2) := \frac{\theta_2}{\rho^2} - \frac{c\rho(2-\rho)}{(1-\rho)^2} - e$. It can be shown that $\frac{\partial\Pi_1^D(\mu)}{\partial\mu} = 0 \Leftrightarrow \phi_D(\rho,\theta_2) = 0$.

Therefore, the maximum agent payoff for each dedicated queue can be attained at $\mu^D = \lambda/\rho^D$ for the ρ^D defined in Proposition 1. However, (μ^D,μ^D) can be sustained as an equilibrium if and only if agent payoff is positive, that is,

$$\lambda\left[V - \alpha\mu^D - \frac{c\lambda}{\mu^D(\mu^D-\lambda)} - e/\mu^D\right] > 0 \Leftrightarrow V > V^D(\lambda,\alpha) = \frac{\lambda\alpha}{\rho^D} + \frac{c(\rho^D)^2}{\lambda(1-\rho^D)} + \frac{e\rho^D}{\lambda}.$$

If agent payoff were negative, the agents would have incentives to set service rate $\mu_i \geq V/\alpha$ to leave customers with no incentives to join, and as a result $\lambda_1 = \lambda_2 = 0$ and $\Pi_i(\mu_i) = 0$. This leads to the case of customer balking.

If customers balk in equilibrium, then we must have $(V - \alpha\cdot\mu_i) - c\cdot W_i(\cdot) \leq 0$ from customers' point of view, which implies that the agent payoff satisfies

$$\Pi_i(\mu_i) \leq -e\lambda_i/\mu_i < 0, \quad i = 1,2.$$

Both agents then have incentives to set service rate $\mu_i \geq V/\alpha$ to stop customers from joining, and as a result $\lambda_1 = \lambda_2 = 0$ and $\Pi_i(\mu_i) = 0$. These equilibria are sustained when $V \leq V_D(\lambda,\alpha)$. In this case, $\mu_1 = \mu_2 = V/\alpha$ (with all customers balking) would be sustained as an equilibrium. □

Proof of Proposition 2

We first establish the following Lemma.

LEMMA 3. *For any fixed total throughput $2\bar{\lambda}$, $\lambda_i(\mu_1,\mu_2)$ is strictly increasing and concave in μ_i while $W(\mu_1,\mu_2)$ is strictly decreasing and convex in μ_i for any fixed μ_{3-i}, where $i = 1,2$.*

Proof of Lemma 3

(1) For any fixed throughput $\bar{\lambda}$, when $\bar{\rho} < 1$, i.e., the system is stable, we can verify that $\lambda_1(\mu_1, \mu_2)$ is continuous and differentiable in μ_1. Taking the derivative of $\lambda_1(\mu_1, \mu_2)$ with respect to μ_1 gives

$$\frac{\partial \lambda_1(\mu_1,\mu_2)}{\partial \mu_1} = \frac{\bar{\lambda}^2 \mu_2^2 \left[4\bar{\lambda}\mu_1 + (\mu_1+\mu_2)^2\right]}{\left[\bar{\lambda}\left(\mu_1^2+\mu_2^2\right) + \mu_1\mu_2(\mu_1+\mu_2)\right]^2} > 0,$$

$$\frac{\partial^2 \lambda_1(\mu_1,\mu_2)}{\partial \mu_1^2} = -\frac{2\bar{\lambda}^2 \mu_2^2 (\bar{\lambda}+\mu_2)\left[\bar{\lambda}\left(6\mu_1^2 - 2\mu_2^2\right) + (\mu_1+\mu_2)^3\right]}{\left[\bar{\lambda}\left(\mu_1^2+\mu_2^2\right) + \mu_1\mu_2(\mu_1+\mu_2)\right]^3} < 0.$$

That is, $\lambda_1(\mu_1, \mu_2)$ is strictly increasing and concave in $\mu_1 \in (\lambda, \infty)$ for any fixed μ_2.

(2) For any fixed throughput $\bar{\lambda}$, when $\bar{\rho} < 1$, we can verify that $W(\mu_1, \mu_2)$ is continuous and differentiable in μ_1. By re-writing $W(\mu_1, \mu_2)$ as a function of μ_1 and $\bar{\rho}$, i.e., $W(\mu_1, \bar{\rho})$, we have

$$\frac{\partial W(\mu_1,\bar{\rho})}{\partial \mu_1} = -\frac{\bar{\lambda}\mu_2\bar{\rho}^3}{\left(\mu_1\mu_2(\bar{\rho}-1)\bar{\rho} - 2\bar{\lambda}^2\right)^2} < 0,$$

$$\frac{\partial^2 W(\mu_1,\bar{\rho})}{\partial \mu_1^2} = \frac{2\bar{\lambda}\mu_2^2(1-\bar{\rho})\bar{\rho}^4}{\left(\mu_1\mu_2(1-\bar{\rho})\bar{\rho} + 2\bar{\lambda}^2\right)^3} > 0,$$

$$\frac{\partial W(\mu_1,\bar{\rho})}{\partial \bar{\rho}} = \frac{\bar{\lambda}\bar{\rho}\left(\mu_1\mu_2\bar{\rho}\left(1-\bar{\rho}^2\right) + 2\bar{\lambda}^2(2-\bar{\rho})\right)}{(1-\bar{\rho})^2\left(\mu_1\mu_2(1-\bar{\rho})\bar{\rho} + 2\bar{\lambda}^2\right)^2} > 0,$$

$$\frac{\partial^2 W(\mu_1,\bar{\rho})}{\partial \bar{\rho}^2} = \frac{2\bar{\lambda}\left(4\bar{\lambda}^4 + \mu_1^2\mu_2^2\bar{\rho}^3\left(\bar{\rho}^3 - 3\bar{\rho} + 2\right) + 6\bar{\lambda}^2\mu_1\mu_2(2-\bar{\rho})(1-\bar{\rho})\bar{\rho}^2\right)}{(1-\bar{\rho})^3\left(\mu_1\mu_2(1-\bar{\rho})\bar{\rho} + 2\bar{\lambda}^2\right)^3} > 0,$$

$$\frac{\partial \bar{\rho}}{\partial \mu_1} = -\frac{2\bar{\lambda}}{(\mu_1+\mu_2)^2} < 0, \quad \frac{\partial^2 \bar{\rho}}{\partial \mu_1^2} = \frac{4\bar{\lambda}}{(\mu_1+\mu_2)^3} > 0,$$

$$\frac{\partial^2 W(\mu_1,\bar{\rho})}{\partial \bar{\rho}\partial \mu_1} = -\frac{\bar{\lambda}\mu_2\bar{\rho}^2\left[6\bar{\lambda}^2 + \mu_1\mu_2\bar{\rho}(\bar{\rho}+1)\right]}{\left(2\bar{\lambda}^2 + \mu_1\mu_2(1-\bar{\rho})\bar{\rho}\right)^3} < 0.$$

Combining these results derived above gives

$$\frac{dW(\mu_1,\bar{\rho})}{d\mu_1} = \frac{\partial W(\mu_1,\bar{\rho})}{\partial \mu_1} + \frac{\partial W(\mu_1,\bar{\rho})}{\partial \bar{\rho}} \cdot \frac{\partial \bar{\rho}}{\partial \mu_1} < 0,$$

$$\frac{d^2 W(\mu_1,\bar{\rho})}{d\mu_1^2} = \frac{\partial^2 W(\mu_1,\bar{\rho})}{\partial \mu_1^2} + \frac{\partial^2 W(\mu_1,\bar{\rho})}{\partial \bar{\rho}\partial \mu_1} \cdot \frac{\partial \bar{\rho}}{\partial \mu_1} + \frac{\partial^2 W(\mu_1,\bar{\rho})}{\partial \bar{\rho}^2} \cdot \left(\frac{\partial \bar{\rho}}{\partial \mu_1}\right)^2 + \frac{\partial W(\mu_1,\bar{\rho})}{\partial \bar{\rho}} \cdot \frac{\partial^2 \bar{\rho}}{\partial \mu_1^2} > 0.$$

That is, $W(\mu_1, \mu_2)$ is strictly decreasing and convex in μ_1. □

Similar to Proposition 1, we first study the case where all customers join, i.e., $\lambda_1 + \lambda_2 = 2\lambda$. For any fixed μ_2, taking the derivative of $\Pi_1^P(\mu_1, \mu_2)$ with respect to μ_1 gives

$$\frac{\partial \Pi_1^P(\mu_1,\mu_2)}{\partial \mu_1} = \frac{\partial \lambda_1(\mu_1,\mu_2)}{\partial \mu_1}\left[V - \alpha\mu_1 - cW(\mu_1,\mu_2) - e/\mu_1\right] + \lambda_1(\mu_1,\mu_2)\frac{\partial\left[V - \alpha\mu_1 - cW(\mu_1,\mu_2) - e/\mu_1\right]}{\partial \mu_1}$$

$$\frac{\partial^2 \Pi_1^P(\mu_1,\mu_2)}{\partial \mu_1^2} = \frac{\partial^2 \lambda_1(\mu_1,\mu_2)}{\partial \mu_1^2}\left[V - \alpha\mu_1 - cW(\mu_1,\mu_2) - e/\mu_1\right] + 2\frac{\partial \lambda_1(\mu_1,\mu_2)}{\partial \mu_1}\frac{\partial\left[V - \alpha\mu_1 - cW(\mu_1,\mu_2) - e/\mu_1\right]}{\partial \mu_1}$$
$$+ \lambda_1(\mu_1,\mu_2)\frac{\partial^2\left[V - \alpha\mu_1 - cW(\mu_1,\mu_2) - e/\mu_1\right]}{\partial \mu_1^2}.$$

For any fixed μ_2, if $[V-\alpha\mu_1-cW(\mu_1,\mu_2)-e/\mu_1]\leq 0$ for all $\mu_1\in(0,\infty)$, then the best response of Agent 1 WLOG is $\hat{\mu}_1(\mu_2)\geq V/\alpha$ so no customers join and the agent has a zero payoff. If $[V-\alpha\mu_1-cW(\mu_1,\mu_2)-e/\mu_1]>0$ for some $\mu_1\in(0,\infty)$, there must exist $0<\underline{\mu}_1<\bar{\mu}_1<\infty$ such that $[V-\alpha\mu_1-cW(\mu_1,\mu_2)-e/\mu_1]>0$ if and only if $\mu_1\in(\underline{\mu}_1,\bar{\mu}_1)$ since $[V-\alpha\mu_1-cW(\mu_1,\mu_2)-e/\mu_1]$ is strictly concave in μ_1 by Lemma 3. Therefore, there exists a unique cutoff $\tilde{\mu}$ such that $[V-\alpha\mu_1-cW(\mu_1,\mu_2)-e/\mu_1]$ is increasing in $\mu_1\in(\underline{\mu}_1,\tilde{\mu})$ and decreasing in $\mu_1\in(\tilde{\mu},\bar{\mu}_1)$.

Next, we proceed to show that the best response $\hat{\mu}_1(\mu_2)\in(\underline{\mu}_1,\bar{\mu}_1)$ is unique. Since $\frac{\partial\lambda_1(\mu_1,\mu_2)}{\partial\mu_1}>0$ and $[V-\alpha\mu_1-cW(\mu_1,\mu_2)-e/\mu_1]$ is increasing in $\mu_1\in(\underline{\mu}_1,\tilde{\mu})$, it implies that $\Pi_1^P(\mu_1,\mu_2)$ is increasing in $\mu_1\in(\underline{\mu}_1,\tilde{\mu})$. For any $\mu_1\in(\tilde{\mu},\bar{\mu}_1)$, we can verify that $\frac{\partial^2\Pi_1^P(\mu_1,\mu_2)}{\partial\mu_1^2}<0$, i.e., $\Pi_1^P(\mu_1,\mu_2)$ is concave in $\mu_1\in(\tilde{\mu},\bar{\mu}_1)$. Thus, $\Pi_1^P(\mu_1,\mu_2)$ is unimodal in $\mu_1\in(\underline{\mu}_1,\bar{\mu}_1)$. Since $\Pi_1^P(\mu_1,\mu_2)|_{\mu_1=\underline{\mu}_1}=\Pi_1^P(\mu_1,\mu_2)|_{\mu_1=\bar{\mu}_1}=0$, there exists a unique maximizer $\hat{\mu}_1(\mu_2)\in(\underline{\mu}_1,\bar{\mu}_1)$ for any fixed μ_2.

Because we are investigating a symmetric equilibrium between the two servers, in equilibrium, we have $\mu_1^P=\mu_2^P$. Therefore, by letting $\mu_2=\mu_1$ in $\frac{\partial\Pi_1^P(\mu_1,\mu_2)}{\partial\mu_1}$, we can solve for μ_1^P by setting $\frac{\partial\Pi_1^P(\mu_1,\mu_2)}{\partial\mu_1}|_{\mu_2=\mu_1}=0$, i.e.,

$$
\begin{aligned}
&\frac{\lambda\left[c\lambda^2\left(\lambda^2-2\lambda\mu_1+3\mu_1^2\right)-2\mu_1(\lambda-\mu_1)^2(\lambda+\mu_1)(\alpha\mu_1(2\lambda+\mu_1)-e-\lambda V)\right]}{2\mu_1^2(\lambda-\mu_1)^2(\lambda+\mu_1)^2}=0\\
\Leftrightarrow & c\rho_1^4[(1-\rho_1)^2+2]+2\theta_1\rho_1^2(1-\rho_1)^2(1+\rho_1)-2\theta_2(2\rho_1+1)(\rho_1+1)(1-\rho_1)^2=0\\
\Leftrightarrow & \phi_P(\rho_1,\theta_1,\theta_2):=\frac{c\rho_1^4[\rho_1^2-2\rho_1+3]}{(2\rho_1+1)(\rho_1+1)(1-\rho_1)^2}+\frac{2\theta_1\rho_1^2}{2\rho_1+1}-2\theta_2=0,
\end{aligned}
$$

where $\rho_1=\lambda/\mu_1\in[0,1)$, $\theta_1:=\lambda V+e$. Next, we show that $\phi_P(\rho_1,\theta_1,\theta_2)=0$ has a unique solution in ρ_1. First, we can show that $\phi_P(\rho_1,\theta_1,\theta_2)$ is increasing in $\rho_1\in[0,1)$ as follows. It is not difficult to verify that the second term of $\phi_P(\rho_1,\theta_1,\theta_2)$, $\frac{2\theta_1\rho_1^2}{2\rho_1+1}$, is increasing in $\rho_1\in[0,1)$ because

$$\frac{d\left(\frac{\rho_1^2}{2\rho_1+1}\right)}{d\rho_1}=\frac{2\rho_1(2+\rho_1)}{(2\rho_1+1)^2}>0.$$

Then it suffices to show that the first term of $\phi_P(\rho_1,\theta_1,\theta_2)$, $\frac{c\rho_1^4[\rho_1^2-2\rho_1+3]}{(2\rho_1+1)(\rho_1+1)(1-\rho_1)^2}$, is increasing in $\rho_1\in[0,1)$. Note that it can be expressed as $\frac{c\rho_1^4}{(2\rho_1+1)(\rho_1+1)}\cdot\left(1+\frac{2}{(1-\rho_1)^2}\right)$, where the second term of the product is obviously increasing in $\rho_1\in[0,1)$. Taking the derivative of the first term of the product with respect to ρ_1 gives

$$\frac{d\left[\frac{\rho_1^4}{(2\rho_1+1)(\rho_1+1)}\right]}{d\rho_1}=\frac{\rho_1^3\left(4+9\rho_1+4\rho_1^2\right)}{\left(1+3\rho_1+2\rho_1^2\right)^2}>0$$

for all $\rho_1\in[0,1)$. Thus $\phi_P(\rho_1,\theta_1,\theta_2)$ is increasing in $\rho_1\in[0,1)$. Moreover, notice that

$$\phi_P(\rho_1,\theta_1,\theta_2)|_{\rho_1=0}=-2\theta_2<0,$$

$$\phi_P(\rho_1,\theta_1,\theta_2)|_{\rho_1=1^-}=+\infty>0.$$

Thus, there exists a unique ρ^P that solves $\phi_P(\rho,\theta_1,\theta_2)=0$. And the symmetric equilibrium service rate μ^P is given by λ/ρ^P. Taking the derivative of $\phi_P(\rho_1,\theta_1,\theta_2)$ with respect to θ_1 and θ_2 gives

$$\begin{aligned}\frac{\partial\phi_P(\rho_1,\theta_1,\theta_2)}{\partial\theta_1} &= \frac{2\rho_1^2}{2\rho_1+1}\geq 0,\\ \frac{\partial\phi_P(\rho_1,\theta_1,\theta_2)}{\partial\theta_2} &= -2<0.\end{aligned}$$

It gives $\frac{\partial\rho^P}{\partial\theta_1}=-\frac{\partial\phi_P(\rho^P,\theta_1,\theta_2)/\partial\theta_1}{\partial\phi_P(\rho^P,\theta_1,\theta_2)/\partial\rho^P}\leq 0$ and $\frac{d\rho^P}{d\theta_2}=-\frac{\partial\phi_P(\rho^P,\theta_1,\theta_2)/\partial\theta_2}{\partial\phi_P(\rho^P,\theta_1,\theta_2)/\partial\rho^P}>0$. Thus we can establish that ρ^P is decreasing in θ_1 (i.e., $\lambda V+e$) and increasing in θ_2 (i.e., $\lambda^2\alpha$). That is, the symmetric equilibrium service rate (if it exists) $\mu^P=\lambda/\rho^P$ is increasing in V and decreasing in α, respectively. Therefore, if a symmetric equilibrium with all customers joining exists, it must be $\mu^P=\lambda/\rho^P$.

Next, we need to establish the existence of symmetric equilibrium.

- If $\Pi_1^P[\lambda/\rho^P,\lambda/\rho^P]>0$, by the analysis in the beginning of this proof, there exists a unique maximizer when $\mu_2=\lambda/\rho^P$. Notice that $\frac{\partial\Pi_1^P(\mu_1,\mu_P)}{\partial\mu_1}|_{\mu_1=\mu^P}=0$; hence, we can conclude that $\mu_1=\mu^P$ is the unique maximizer when $\mu_2=\mu^P$, which naturally identifies the symmetric equilibrium strategy.
- If $\Pi_1^P[\lambda/\rho^P,\lambda/\rho^P]\leq 0$, $\mu^P=\lambda/\rho^P$ cannot continue to be an equilibrium because agent payoff is negative. Furthermore, by our analysis at the beginning of this proof, customer joining cannot be sustained in equilibrium. Hence, in this case, $\mu_1=\mu_2=V/\alpha$ (with all customers balking) is an equilibrium.

Finally, we will identify the threshold of V, beyond which a symmetric equilibrium with customer joining can be sustained. Notice that

$$\Pi_1^P[\lambda/\rho^P,\lambda/\rho^P]>0\Leftrightarrow V>\frac{\lambda\alpha}{\rho^P}+\frac{c(\rho^P)^3}{\lambda[1-(\rho^P)^2]}+\frac{e\rho^P}{\lambda}\Leftrightarrow V>\frac{c(\rho^P)^3[5-3(\rho^P)^2]}{2\lambda(1-\rho^P)^2(1+\rho^P)^2}+\frac{2e\rho^P}{\lambda}$$

where the last inequality follows from the condition that $\phi_P(\rho_1,\theta_1,\theta_2)=0$. Since $\frac{c(\rho^P)^3[5-3(\rho^P)^2]}{2\lambda(1-\rho^P)^2(1+\rho^P)^2}+\frac{2e\rho^P}{\lambda}$ is increasing in ρ^P and ρ^P is decreasing in V, it follows that $\frac{c(\rho^P)^3[5-3(\rho^P)^2]}{2\lambda(1-\rho^P)^2(1+\rho^P)^2}+\frac{2e\rho^P}{\lambda}$ is decreasing in V. Thus, there exists a unique $V^P(\lambda,\alpha)$ which solves $V=\frac{c(\rho^P)^3[5-3(\rho^P)^2]}{2\lambda(1-\rho^P)^2(1+\rho^P)^2}+\frac{2e\rho^P}{\lambda}$ so that a symmetric equilibrium with customer joining is sustained if and only if $V>V^P(\lambda,\alpha)$. Otherwise (if $V\leq V_P(\lambda,\alpha)$), we have $\mu^P=V/\alpha$ and all customers balk. Denoting $\hat{x}=\rho^P|_{V=V^P(\lambda,\alpha)}$ and eliminating V by combining the following equations

$$V=\frac{c\hat{x}^3[5-3\hat{x}^2]}{2\lambda(1-\hat{x})^2(1+\hat{x})^2}+\frac{2e\hat{x}}{\lambda}\quad\text{and}\quad\Phi_P(\hat{x},\theta_1,\theta_2)=0$$

give $\frac{\lambda\alpha}{\hat{x}^2}-\frac{c\hat{x}^2(3-\hat{x}^2)}{2\lambda(1-\hat{x})^2(1+\hat{x})^2}-\frac{e}{\lambda}=0$. It is not difficult to verify that $\frac{\lambda\alpha}{\hat{x}^2}-\frac{c\hat{x}^2(3-\hat{x}^2)}{2\lambda(1-\hat{x})^2(1+\hat{x})^2}-\frac{e}{\lambda}$ is decreasing in $\hat{x}\in(0,1)$ and $\frac{\lambda\alpha}{\hat{x}^2}-\frac{c\hat{x}^2(3-\hat{x}^2)}{2\lambda(1-\hat{x})^2(1+\hat{x})^2}-\frac{e}{\lambda}|_{\hat{x}=0^+}=+\infty>0>-\infty=\frac{\lambda\alpha}{\hat{x}^2}-\frac{c\hat{x}^2(3-\hat{x}^2)}{2\lambda(1-\hat{x})^2(1+\hat{x})^2}-\frac{e}{\lambda}|_{\hat{x}=1^-}$, which implies that there exists a unique $\hat{x}\in(0,1)$ such that $\frac{\lambda\alpha}{\hat{x}^2}-\frac{c\hat{x}^2(3-\hat{x}^2)}{2\lambda(1-\hat{x})^2(1+\hat{x})^2}-\frac{e}{\lambda}=0$ and we have $V^P(\lambda,\alpha)=\frac{\lambda\alpha}{\hat{x}}+\frac{c\hat{x}^3}{\lambda(1-\hat{x}^2)}+\frac{e\hat{x}}{\lambda}$, which completes this proof. □

Proof of Lemma 2

Recall that $V^P(\lambda,\alpha)=\frac{\lambda\alpha}{\hat{x}}+\frac{c\hat{x}^3}{\lambda(1-\hat{x}^2)}+\frac{e\hat{x}}{\lambda}$, where $\hat{x}$ uniquely solves $\frac{\lambda\alpha}{\hat{x}^2}-\frac{c\hat{x}^2(3-\hat{x}^2)}{2\lambda(1-\hat{x})^2(1+\hat{x})^2}-\frac{e}{\lambda}=0$, and $V^D(\lambda,\alpha)=\frac{\lambda\alpha}{\rho^D}+\frac{c(\rho^D)^2}{\lambda(1-\rho^D)}+\frac{e\rho^D}{\lambda}$, where $\rho^D\in(0,1)$ uniquely solves $\frac{\lambda\alpha}{(\rho^D)^2}-\frac{c\rho^D(2-\rho^D)}{\lambda(1-\rho^D)^2}-\frac{e}{\lambda}=0$. It is not difficult to verify that both $\frac{\lambda\alpha}{x^2}-\frac{cx^2(3-x^2)}{2\lambda(1-x)^2(1+x)^2}-\frac{e}{\lambda}$ and $\frac{\lambda\alpha}{x^2}-\frac{cx(2-x)}{\lambda(1-x)^2}-\frac{e}{\lambda}$ are decreasing in $x\in(0,1)$, and $\frac{\lambda\alpha}{x^2}-\frac{cx^2(3-x^2)}{2\lambda(1-x)^2(1+x)^2}-\frac{e}{\lambda}>\frac{\lambda\alpha}{x^2}-\frac{cx(2-x)}{\lambda(1-x)^2}-\frac{e}{\lambda}$. Then it follows that $\hat{x}>\rho^D$.

Note that

$$\frac{\lambda^2\alpha}{c}=\frac{(\rho^D)^3(2-\rho^D)}{(1-\rho^D)^2}+\frac{e(\rho^D)^2}{c},\quad \frac{\lambda^2\alpha}{c}=\frac{\hat{x}^4(3-\hat{x}^2)}{2(1-\hat{x}^2)^2}+\frac{e\hat{x}^2}{c}. \tag{A.10}$$

It follows that

$$V^D(\lambda,\alpha)=\left[\frac{y^2(3-2y)}{(1-y)^2}+\frac{2ey}{c}\right]\frac{c}{\lambda},\quad V^P(\lambda,\alpha)=\left[\frac{x^3\left(5-3x^2\right)}{2\left(x^2-1\right)^2}+\frac{2ex}{c}\right]\frac{c}{\lambda},$$

where $y=\rho^P\in(0,\hat{x})$ and $x=\hat{x}\in(0,1)$. (A.10) indicates that

$$\frac{e}{c}=\left[\frac{y^3(2-y)}{(1-y)^2}-\frac{x^4(3-x^2)}{(1-x^2)^2}\right]\frac{1}{(x-y)(x+y)}.$$

By plugging e/c into $V^D(\lambda,\alpha)$ and $V^P(\lambda,\alpha)$, it follows that

$$\frac{\lambda}{c}[V^D(\lambda,\alpha)-V^P(\lambda,\alpha)]=\Psi_1(x,y)=\left[\frac{x^4(3-x^2)}{(1-x^2)^2}-\frac{y^3(2-y)}{(1-y)^2}\right]\frac{2}{x+y}+\left[\frac{y^2(3-2y)}{(1-y)^2}-\frac{x^3\left(5-3x^2\right)}{2\left(x^2-1\right)^2}\right].$$

To show $V^D(\lambda,\alpha)-V^P(\lambda,\alpha)>0$, it suffices to prove $\Psi_1(x,y)>0$ for any $0<y<x<1$. Note that for any fixed $x\in(0,1)$, we have

$$\frac{\partial\Psi_1(x,y)}{\partial y}=-\frac{2\left(y-x^2\right)\left(x^4+x^2((4-3y)y-3)+y^2\right)}{\left(x^2-1\right)^2(1-y)^3(x+y)^2}. \tag{A.11}$$

Let $g(x,y)=2\left(y-x^2\right)\left(x^4+x^2((4-3y)y-3)+y^2\right)$, we have

$$\begin{aligned}\frac{\partial g(x,y)}{\partial y}&=3\left(x^4(2y-1)+x^2((2-3y)y-1)+y^2\right),\\ \frac{\partial^2 g(x,y)}{\partial y^2}&=6\left(x^4+x^2(1-3y)+y\right),\\ \frac{\partial^3 g(x,y)}{\partial y^3}&=6(1-3x^2).\end{aligned}$$

We consider two subcases below. (i) If $6(1-3x^2)>0$, then $\min_y\frac{\partial^2 g(x,y)}{\partial y^2}=\frac{\partial^2 g(x,y)}{\partial y^2}|_{y=0}>0$, which implies that $\frac{\partial g(x,y)}{\partial y}$ is increasing in $y\in(0,x)$. (ii) If $6(1-3x^2)\le 0$, then $\min_y\frac{\partial^2 g(x,y)}{\partial y^2}=\frac{\partial^2 g(x,y)}{\partial y^2}|_{y=1}=6x(x^3+x-3x^2+1)>0$ for any $x\in(0,1)$, which also implies that $\frac{\partial g(x,y)}{\partial y}$ is increasing in $y\in$

$(0,x)$. Since $\frac{\partial g(x,y)}{\partial y}|_{y=0} = -3x^2(x^2+1) < 0$ and $\frac{\partial g(x,y)}{\partial y}|_{y=x} = 6(x-1)^2x^3 > 0$, we know $g(x,y)$ first decreases then increases in $y \in [0,x)$. Since $g(x,0) = 2(x-1)^3x^3 > 0 > g(x,x) = 2(x-1)^3x^3$, we know that there exists a unique $y' \in [0,x]$ such that $-g(x,y) < 0$ for $y \in [0,y')$ and $-g(x,y) \geq 0$ for $y \in (y',x]$, which implies that $y' = x^2$ by (A.11). Therefore, it follows that $\Psi_1(x,y) \geq \Psi_1(x,x^2) = \frac{x^3(7+x-3x^2-x^3)}{2(1-x)^2(x+1)^3} > 0$ for any $x \in (0,1)$, which implies that $\Psi_1(x,y) > 0$ for any $0 < y < x < 1$. □

Proof of Theorem 1

(i) We first consider the case where all customers join in equilibrium under both the dedicated and pooled settings, i.e., $V > V^D(\lambda,\alpha)$ (which implies that $V > V^P(\lambda,\alpha)$ by Lemma 2). Recall that ρ^D is determined by $\lambda^2\alpha(1-\rho)^2 - c\rho^3(2-\rho) - e\rho^2(1-\rho)^2 = 0$, which is independent of V, and that ρ^P is determined by $\frac{c\rho^4[\rho^2-2\rho+3]}{(2\rho+1)(\rho+1)(1-\rho)^2} + \frac{2(\lambda V+e)\rho^2}{2\rho+1} - 2\lambda^2\alpha = 0$, which is decreasing in V (see the proof of Proposition 2). In particular, when $V=0$, ρ^P is determined by $\lambda^2\alpha(1-\rho)^2 - \frac{e\rho^2(1-\rho)^2}{2\rho+1} - \frac{c\rho^4[\rho^2-2\rho+3]}{2(2\rho+1)(\rho+1)} = 0$. We can verify that when $V=0$, $\rho^D < \rho^P$ because $\lambda^2\alpha(1-\rho)^2 - c\rho^3(2-\rho) - e\rho^2(1-\rho)^2$ and $\lambda^2\alpha(1-\rho)^2 - \frac{e\rho^2(1-\rho)^2}{2\rho+1} - \frac{c\rho^4[\rho^2-2\rho+3]}{2(2\rho+1)(\rho+1)}$ are both decreasing in $\rho \in [0,1)$, $-e\rho^2(1-\rho)^2 < -\frac{e\rho^2(1-\rho)^2}{2\rho+1}$ and $\lambda^2\alpha(1-\rho)^2 - c\rho^3(2-\rho) < \lambda^2\alpha(1-\rho)^2 - \frac{c\rho^4[\rho^2-2\rho+3]}{2(2\rho+1)(\rho+1)} \Leftrightarrow \rho^4 - 2\rho^3 < -\frac{\rho^4[\rho^2-2\rho+3]}{2(2\rho+1)(\rho+1)} \Leftrightarrow 2(2\rho+1)(\rho+1)(2-\rho) > \rho(\rho^2-2\rho+3) \Leftrightarrow 4+6\rho+\rho(1-\rho)(4\rho+1) > 0$ which is true for $\rho \in [0,1)$.

Note also that ρ^P is decreasing in V and $\lim_{V\to\infty} \rho^P = 0 < \rho^D < \lim_{V\to 0} \rho^P$. Therefore, for any given α and λ, there exists a unique $\tilde{V}(\lambda,\alpha)$ such that $\rho^P < \rho^D$ (i.e., $\mu^P > \mu^D$) if and only if $V > \tilde{V}(\lambda,\alpha)$, where $\tilde{V}(\lambda,\alpha)$ is uniquely determined by $\rho^P|_{V=\tilde{V}(\lambda,\alpha)} = \rho^D$. Recognizing that ρ^D and ρ^P are uniquely determined by $\lambda^2\alpha(1-\rho)^2 - c\rho^3(2-\rho) - e\rho^2(1-\rho)^2 = 0$ and $\frac{c\rho^4[\rho^2-2\rho+3]}{(2\rho+1)(\rho+1)(1-\rho)^2} + \frac{2(\lambda V+e)\rho^2}{2\rho+1} - 2\lambda^2\alpha = 0$, respectively, when $\rho^P = \rho^D$, we have

$$\tilde{V}(\lambda,\alpha) = \frac{c\rho^D\left[4+7\rho^D+4(\rho^D)^2-5(\rho^D)^3\right]}{2\lambda(1-\rho^D)^2(1+\rho^D)} + \frac{2e\rho^D}{\lambda} > 0.$$

Recall that $V^D(\lambda,\alpha) = \left[\frac{y^2(3-2y)}{(1-y)^2} + \frac{2ey}{c}\right]\frac{c}{\lambda}$, we can verify that $\tilde{V}(\lambda,\alpha) > V^D(\lambda,\alpha)$ since $\frac{cx\left[4+7x+4x^2-5x^3\right]}{2\lambda(1-x)^2(1+x)} > \frac{cx^2(3-2x)}{\lambda(1-x)^2} \Leftrightarrow 4 - x^3 + 2x^2 + x > 0$, which holds for any $x \in [0,1]$. Thus we can conclude that $\tilde{V}(\lambda,\alpha) > V^D(\lambda,\alpha) > V^P(\lambda,\alpha)$ because $V^D(\lambda,\alpha) > V^P(\lambda,\alpha)$ (Lemma 2).

(ii-1) In equilibrium, the profits for both agents are given by

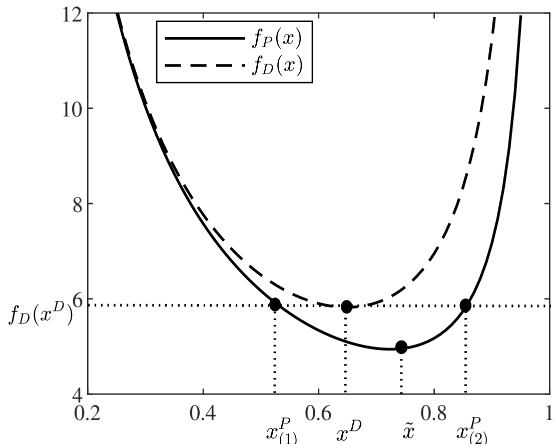
$$\Pi^P = 2\lambda\left[V - \frac{\lambda\alpha}{\rho^P} - \frac{c(\rho^P)^3}{\lambda[1-(\rho^P)^2]} - \frac{e\rho^P}{\lambda}\right], \quad \Pi^D = 2\lambda\left[V - \frac{\lambda\alpha}{\rho^D} - \frac{c(\rho^D)^2}{\lambda[1-\rho^D]} - \frac{e\rho^D}{\lambda}\right]$$

Therefore, we have

$$\Pi^D > \Pi^P \Leftrightarrow \frac{\lambda\alpha}{\rho^P} + \frac{c(\rho^P)^3}{\lambda[1-(\rho^P)^2]} + \frac{e\rho^P}{\lambda} > \frac{\lambda\alpha}{\rho^D} + \frac{c(\rho^D)^2}{\lambda[1-\rho^D]} + \frac{e\rho^D}{\lambda} \Leftrightarrow f_P(\rho^P) > f_D(\rho^D)$$

where $f_P(x) := \frac{\theta_2}{x} + \frac{cx^3}{1-x^2} + ex$ and $f_D(x) := \frac{\theta_2}{x} + \frac{cx^2}{1-x} + ex$. Next, we identify the condition on V when $f_P(\rho^P) > f_D(\rho^D)$ holds.

Figure 4 $\frac{\lambda^2\alpha}{\rho}+\frac{c\rho^3}{1-\rho^2}+e\rho$ **and** $\frac{\lambda^2\alpha}{\rho}+\frac{c\rho^2}{1-\rho}+e\rho$.



First, we show that both $f_P(x)$ and $f_D(x)$ first decrease then increase in $x\in[0,1)$. Taking the derivative of $f_P(x)$ with respect to x gives $\frac{\partial f_P(x)}{\partial x}=e+\frac{-cx^6+x^4(3c-\theta_2)-\theta_2+2cx^2\theta_2}{x^2\left(1-x^2\right)^2}$ where $\frac{\partial[-cx^6+x^4(3c-\theta_2)-\theta+2cx^2\theta_2]}{\partial x}=2x[(3cx^2+2x\theta_2)(1-x^2)+3cx^2]>0$. Because $\frac{-cx^6+x^4(3c-\theta_2)-\theta+2x^2\theta_2}{x^2(1-x^2)^2}+e|_{x=0^+}=-\infty$ and $\frac{-cx^6+x^4(3c-\theta_2)-\theta+2x^2\theta_2}{x^2(1-x^2)^2}+e|_{x=1^-}=+\infty$, there exists a unique $\tilde{x}$ such that $f_P(x)$ decreases in $x\in[0,\tilde{x})$ and increases in $x\in[\tilde{x},1)$. Similarly, taking the derivative of $f_D(x)$ with respect to x gives $\frac{\partial f_D(x)}{\partial x}=\frac{cx^3(2-x)-\theta_2(1-x)^2}{(1-x)^2x^2}+e$ where $\frac{\partial[cx^3(2-x)-\theta_2(1-x)^2]}{\partial x}=2\theta_2(1-x)+4cx^2(1-x)+2cx^2>0$. Because $\frac{cx^3(2-x)-\theta_2(1-x)^2}{x^2(1-x)^2}+e|_{x=0^+}=-\infty<0$ and $\frac{cx^3(2-x)-\theta_2(1-x)^2}{x^2(1-x)^2}+e|_{x=1^-}=\infty$, there exists a unique x^D such that $f_D(x)$ decreases in $x\in[0,x^D)$ and increases in $x\in[x^D,1)$.

On the other hand, it is not difficult to verify that $f_D(x)>f_P(x)$ since $\frac{\theta_2}{x}+\frac{cx^3}{1-x^2}+ex<\frac{\theta_2}{x}+\frac{cx^2}{1-x}+ex$ for any $x\in[0,1)$ and $\theta_2>0$ because $\frac{x^3}{1-x^2}=\frac{x^3}{(1-x)(1+x)}<\frac{x^2}{1-x}$. Therefore, for any fixed θ_2, c and e, there exist two thresholds $x^P_{(1)}<x^D<x^P_{(2)}$ such that $f_P(x)>f_D(x^D)$ if and only if $x<x^P_{(1)}(<x^D=\rho^D)$ or $x>x^P_{(2)}(>x^D=\rho^D)$, where $x^P_{(1)}$ and $x^P_{(2)}$ are the two roots for $f_P(x)=f_D(x^D)$, see Figure 4. Therefore, to find a condition on V such that $f_P(\rho^P)>f_D(\rho^D)$, it suffices to find a condition on V such that $\rho^P<x^P_{(1)}<\rho^D$ or $\rho^P>x^P_{(2)}>\rho^D$.

Finally, we know $\Pi^D>\Pi^P$ if and only if $\rho^P<x^P_{(1)}$ or $\rho^P>x^P_{(2)}$, where $x^P_{(1)}$ and $x^P_{(2)}$ are the smaller and larger root that solve $f_P(x)=f_D(\rho^D)$. Also, ρ^P is decreasing in V and $\lim_{V\to\infty}\rho^P=0<\rho^D<\rho^P|_{\{0<V<\tilde{V}(\lambda,\alpha)\}}$. Therefore, for any given α and λ, there exists a unique $\hat{V}(\lambda,\alpha)>\tilde{V}(\lambda,\alpha)$ (or $\underline{V}(\lambda,\alpha)<\tilde{V}(\lambda,\alpha)$) such that $\rho^P<x^P_{(1)}$ (or $\rho^P>x^P_{(2)}$) if and only if $V>\hat{V}(\lambda,\alpha)$ (or $V<\underline{V}(\lambda,\alpha)$), where $\hat{V}(\lambda,\alpha)$ (or $\underline{V}(\lambda,\alpha)$) can be identified by setting $\rho^P=x^P_{(1)}$ (or $\rho^P=x^P_{(2)}$). Because ρ^P solves $\frac{c\rho^4[\rho^2-2\rho+3]}{(2\rho+1)(\rho+1)(1-\rho)^2}+\frac{2(\lambda V+e)\rho^2}{2\rho+1}-2\lambda^2\alpha=0$ (see Proposition 2), we have

$$\hat{V}(\lambda,\alpha)=\frac{\lambda\alpha(2x^P_{(1)}+1)}{[x^P_{(1)}]^2}-\frac{c[x^P_{(1)}]^2[[x^P_{(1)}]^2-2x^P_{(1)}+3]}{2\lambda(x^P_{(1)}+1)(1-x^P_{(1)})^2}-\frac{e}{\lambda},$$

$$\underline{V}(\lambda,\alpha) = \frac{\lambda\alpha(2x^P_{(2)}+1)}{[x^P_{(2)}]^2} - \frac{c[x^P_{(2)}]^2[[x^P_{(2)}]^2 - 2x^P_{(2)}+3]}{2\lambda(x^P_{(2)}+1)(1-x^P_{(2)})^2} - \frac{e}{\lambda}.$$

Also, we can show $\underline{V}(\lambda,\alpha) \leq V^D(\lambda,\alpha)$ below. Otherwise, if $\underline{V}(\lambda,\alpha) > V^D(\lambda,\alpha)$, then for any $V \in [V^D(\lambda,\alpha), \underline{V}(\lambda,\alpha))$, we have $\Pi^D(V) > \Pi^P(V) > 0$ because $V^D(\lambda,\alpha) > V^P(\lambda,\alpha)$. However, when $V = V^D(\lambda,\alpha)$, we have $\Pi^D(V^D(\lambda,\alpha)) = 0 > \Pi^P(\Pi^P(V)) > 0$, which contradicts our assumption. Then we must have $\underline{V}(\lambda,\alpha) \leq V^D(\lambda,\alpha)$. Since $\underline{V}(\lambda,\alpha) < V^D(\lambda,\alpha)$, the dedicated case cannot support a customer-joining equilibrium when $V < \underline{V}(\lambda,\alpha)$. Then we must have $CW^D \leq CW^P$ and $\Pi^D \leq \Pi^P$ for $V < \underline{V}(\lambda,\alpha)$.

(ii-2) On the other hand, the customer welfare in equilibrium is given by

$$CW^P = 2\lambda\left[V - \frac{\lambda\alpha}{\rho^P} - \frac{c(\rho^P)^3}{\lambda[1-(\rho^P)^2]}\right], \quad CW^D = 2\lambda\left[V - \frac{\lambda\alpha}{\rho^D} - \frac{c(\rho^D)^2}{\lambda[1-\rho^D]}\right].$$

Therefore, we have

$$CW^D > CW^P \Leftrightarrow \frac{\lambda\alpha}{\rho^P} + \frac{c(\rho^P)^3}{\lambda[1-(\rho^P)^2]} > \frac{\lambda\alpha}{\rho^D} + \frac{c(\rho^D)^2}{\lambda[1-\rho^D]} \Leftrightarrow g_P(\rho^P) > g_D(\rho^D)$$

where $g_P(\rho^P) := \frac{\theta_2}{\rho^P} + \frac{c(\rho^P)^3}{1-(\rho^P)^2}$ and $g_D(\rho^D) := \frac{\theta_2}{\rho^D} + \frac{c(\rho^D)^2}{1-\rho^D}$. We will identify the condition on V such that $g_P(\rho^P) > g_D(\rho^D)$.

First, we show that both $g_P(x)$ and $g_D(x)$ first decrease then increase in $x \in [0,1)$. Taking the derivative of $g_P(x)$ with respect to x gives $\frac{\partial g_P(x)}{\partial x} = \frac{-cx^6+x^4(3c-\theta_2)-\theta_2+2cx^2\theta_2}{x^2\left(1-x^2\right)^2}$ where $\frac{\partial[-cx^6+x^4(3c-\theta_2)-\theta+2cx^2\theta_2]}{\partial x} = 2x[(3cx^2+2x\theta_2)(1-x^2)+3cx^2] > 0$. Because $-cx^6+x^4(3c-\theta_2)-\theta+2x^2\theta_2|_{x=0} = -\theta_2 < 0$ and $-cx^6+x^4(3c-\theta_2)-\theta+2x^2\theta_2|_{x=1} = 2c > 0$, there exists a unique $\tilde{x}'$ such that $g_P(x)$ decreases in $x \in [0,\tilde{x}')$ and increases in $x \in [\tilde{x}',1)$. Similarly, taking the derivative of $g_D(x)$ with respect to x gives $\frac{\partial g_D(x)}{\partial x} = \frac{cx^3(2-x)-\theta_2(1-x)^2}{(1-x)^2x^2}$ where $\frac{\partial[cx^3(2-x)-\theta_2(1-x)^2]}{\partial x} = 2\theta_2(1-x)+4cx^2(1-x)+2cx^2 > 0$. Because $cx^3(2-x)-\theta_2(1-x)^2|_{x=0} = -\theta_2 < 0$ and $cx^3(2-x)-\theta_2(1-x)^2|_{x=1} = c > 0$, there exists a unique $\tilde{x}^D$ such that $g_D(x)$ decreases in $x \in [0,\tilde{x}^D)$ and increases in $x \in [\tilde{x}^D,1)$.

On the other hand, it is not difficult to verify that $g_D(x) < g_P(x)$. Therefore, for any fixed θ_2, there exist two thresholds $\tilde{x}^P_{(1)} < \tilde{x}^D < \tilde{x}^P_{(2)}$ such that $g_P(x) > g_D(x^D)$ if and only if $x < \tilde{x}^P_{(1)} (< x^D = \rho^D)$ or $x > \tilde{x}^P_{(2)} (> x^D = \rho^D)$, where $\tilde{x}^P_{(1)}$ and $\tilde{x}^P_{(2)}$ are the two roots for $g_P(x) = g_D(x^D)$, see Figure 4. Therefore, to find a condition on V such that $g_P(\rho^P) > g_D(\rho^D)$, it suffices to find a condition on V such that $\rho^P < \tilde{x}^P_{(1)} < \rho^D$ or $\rho^P > \tilde{x}^P_{(2)} > \rho^D$.

Finally, we know $CW^D > CW^P$ if and only if $\rho^P < \tilde{x}^P_{(1)}$ or $\rho^P > \tilde{x}^P_{(2)}$, where $\tilde{x}^P_{(1)}$ and $\tilde{x}^P_{(2)}$ are the smaller and larger root that solve $g_P(x) = g_D(\rho^D)$. Also, ρ^P is decreasing in V and $\lim_{V\to\infty} \rho^P = 0 < \rho^D < \rho^P|_{\{0<V<\tilde{V}(\lambda,\alpha)\}}$. Therefore, for any given α and λ, there exists a unique $\hat{V}'(\lambda,\alpha) > \tilde{V}(\lambda,\alpha)$ (or $\underline{V}'(\lambda,\alpha) < \tilde{V}(\lambda,\alpha)$) such that $\rho^P < \tilde{x}^P_{(1)}$ (or $\rho^P > \tilde{x}^P_{(2)}$) if and only if $V > \hat{V}'(\lambda,\alpha)$ (or $V < \underline{V}'(\lambda,\alpha)$), where $\hat{V}'(\lambda,\alpha)$ (or $\underline{V}'(\lambda,\alpha)$) can be identified by setting $\rho^P=\tilde{x}^P_{(1)}$ (or $\rho^P=\tilde{x}^P_{(2)}$). Because ρ^P solves $\frac{c\rho^4[\rho^2-2\rho+3]}{(2\rho+1)(\rho+1)(1-\rho)^2}+\frac{2(\lambda V+e)\rho^2}{2\rho+1}-2\lambda^2\alpha=0$ (see Proposition 2), we have

$$\hat{V}'(\lambda,\alpha)=\frac{\lambda\alpha(2\tilde{x}^P_{(1)}+1)}{[\tilde{x}^P_{(1)}]^2}-\frac{c[\tilde{x}^P_{(1)}]^2[[\tilde{x}^P_{(1)}]^2-2\tilde{x}^P_{(1)}+3]}{2\lambda(\tilde{x}^P_{(1)}+1)(1-\tilde{x}^P_{(1)})^2}-\frac{e}{\lambda},$$
$$\underline{V}'(\lambda,\alpha)=\frac{\lambda\alpha(2\tilde{x}^P_{(2)}+1)}{[\tilde{x}^P_{(2)}]^2}-\frac{c[\tilde{x}^P_{(2)}]^2[[\tilde{x}^P_{(2)}]^2-2\tilde{x}^P_{(2)}+3]}{2\lambda(\tilde{x}^P_{(2)}+1)(1-\tilde{x}^P_{(2)})^2}-\frac{e}{\lambda}.$$

Notice that $x^P_{(1)}$ and $\tilde{x}^P_{(1)}$ are the unique solutions of $f_P(x)-f_D(\rho^D)=0$ and $g_P(x)-g_D(\rho^D)=0$ in $x\in(0,x^D)$. Since $[f_P(x)-f_D(\rho^D)]-[g_P(x)-g_D(\rho^D)]=e(x-\rho^D)<0$, it implies that $x^P_{(1)}<\tilde{x}^P_{(1)}$. Because $\frac{\lambda\alpha(2x+1)}{x^2}-\frac{cx^2[x^2-2x+3]}{2\lambda(x+1)(1-x)^2}-\frac{e}{\lambda}$ is decreasing in x, it follows that $\hat{V}(\lambda,\alpha)>\hat{V}'(\lambda,\alpha)$.

Similarly, since $x^P_{(2)}$ and $\tilde{x}^P_{(2)}$ are the unique solutions of $f_P(x)-f_D(\rho^D)=0$ and $g_P(x)-g_D(\rho^D)=0$ in $x\in(x^D,1)$ respectively and $[f_P(x)-f_D(\rho^D)]-[g_P(x)-g_D(\rho^D)]=e(x-\rho^D)>0$, it implies that $x^P_{(2)}<\tilde{x}^P_{(2)}$. Because $\frac{\lambda\alpha(2x+1)}{x^2}-\frac{cx^2[x^2-2x+3]}{2\lambda(x+1)(1-x)^2}-\frac{e}{\lambda}$ is decreasing in x, it follows that $\underline{V}(\lambda,\alpha)>\underline{V}'(\lambda,\alpha)$.

Since $\hat{V}(\lambda,\alpha)>\hat{V}'(\lambda,\alpha)>\tilde{V}(\lambda,\alpha)>V^D(\lambda,\alpha)>\underline{V}(\lambda,\alpha)>\underline{V}'(\lambda,\alpha)$, we can have the following three cases by summarizing the results above: (1) When $V\le\hat{V}'(\lambda,\alpha)$, we have $CW^D\le CW^P$ and $\Pi^D\le\Pi^P$; (2) When $\hat{V}'(\lambda,\alpha)<V\le\hat{V}(\lambda,\alpha)$, it follows that $CW^D>CW^P$ and $\Pi^D\le\Pi^P$; (3) When $V>\hat{V}(\lambda,\alpha)$, it follows that $CW^D>CW^P$ and $\Pi^D>\Pi^P$, which competes this proof. □

Proof of Proposition 3

When all customers join, under the bonus pooling policy with parameter t, the equilibrium service rates of the two agents (μ_1^*,μ_2^*) satisfy the following conditions:

$$\mu_1^*\in\arg\max_{\mu_1}\lambda_1\left[V'-\alpha\mu_1-cW(\mu_1,\mu_2)-e/\mu_1\right]+\lambda t,$$
$$\mu_2^*\in\arg\max_{\mu_2}\lambda_2\left[V'-\alpha\mu_2-cW(\mu_1,\mu_2)-e/\mu_2\right]+\lambda t,$$

where $V'=V-t$. Thus, when all customers join, the equilibrium results follow from those of Proposition 2. Let $\mu^B=\lambda/\rho^B$ denote the symmetric equilibrium where ρ^B uniquely solves

$$\frac{c(\rho^B)^4[(\rho^B)^2-2(\rho^B)+3]}{(2\rho^B+1)(\rho^B+1)(1-\rho^B)^2}+\frac{2(\lambda(V-t)+e)(\rho^B)^2}{2\rho^B+1}-2\lambda^2\alpha=0.$$

Similar to the proof of Proposition 2, we can verify that $\frac{c(\rho^B)^4[(\rho^B)^2-2(\rho^B)+3]}{(2\rho^B+1)(\rho^B+1)(1-\rho^B)^2}+\frac{2(\lambda(V-t)+e)(\rho^B)^2}{2\rho^B+1}-2\lambda^2\alpha$ is increasing and decreasing in ρ^B and t, respectively. By the implicit function theorem, ρ^B is increasing in t and $\lim_{t\to\infty}\rho^B=1$. Since $\Pi^B=2\left[\lambda V-\frac{\alpha\lambda^2}{\rho^B}-\frac{c(\rho^P)^3}{1-(\rho^B)^2}-e\rho^B\right]$ is concave in ρ^B, there exist two thresholds $0<\rho_1^B<\rho_2^B<1$ such that $\Pi^B(\lambda/\rho_1^B,\lambda/\rho_1^B)=\Pi^B(\lambda/\rho_2^B,\lambda/\rho_2^B)=0$. Note that $\Pi^B(\lambda/\rho^B,\lambda/\rho)|_{t=0}>0$ because $V>V^P(\lambda,\alpha)$, then there exists unique t^B such that $\Pi^B(\lambda/\rho^B,\lambda/\rho^B)>0$ for $t\in[0,t^B)$, where t^B uniquely solves

$$\frac{c(\rho_2^B)^4[(\rho_2^B)^2-2(\rho_2^B)+3]}{(2\rho_2^B+1)(\rho_2^B+1)(1-\rho_2^B)^2}+\frac{2(\lambda(V-t)+e)(\rho_2^B)^2}{2\rho_2^B+1}-2\lambda^2\alpha=0.$$

Therefore, if $t < t^B$, μ^B is an equilibrium under which all customers join. If $t \geq t^B$, similar to Proposition 2, $\mu_B^P = V/\alpha$ with all customers balking is an equilibrium; in this case, $CW_B^P = \Pi_B^P = 0$. $\square$

Proof of Theorem 2

When $V > \hat{V}'(\lambda,\alpha) > V^P(\lambda,\alpha)$, by Proposition 3, $\mu^B = \lambda/\rho^B$ is the unique equilibrium if $t < t^B$, and ρ^B uniquely solves

$$\frac{c(\rho^B)^4[(\rho^B)^2 - 2(\rho^B) + 3]}{(2\rho^B+1)(\rho^B+1)(1-\rho^B)^2} + \frac{2(\lambda(V-t)+e)(\rho^B)^2}{2\rho^B+1} - 2\lambda^2\alpha = 0.$$

For any $t \in (V - \hat{V}'(\lambda,\alpha), V - V^P(\lambda,\alpha))$, we have $V' \in (V^P(\lambda,\alpha), \hat{V}'(\lambda,\alpha))$. Thus, by Theorem 1 and Proposition 2, the performance metrics under the new induced symmetric equilibrium (μ^B, μ^B) satisfy $CW^B > CW^D$ and $\Pi^B > \Pi^D$.

If $t \leq V - \hat{V}'(\lambda,\alpha)$, we have $V' \geq \hat{V}'(\lambda,\alpha)$, which implies that $CW^P \leq CW^D$.

If $t \geq V - V^P(\lambda,\alpha)$, we have $\rho^B > \rho^D$ and by the proof of Theorem 1, $\Pi^B > \Pi^D$ if and only if $\rho^B < x_2^P$ and $CW^B > CW^D$ if and only if $\rho^B < \tilde{x}_2^P$, where $x_{(2)}^P$ and $\tilde{x}_{(2)}^P$ are the unique solutions of $f_P(x) - f_D(\rho^D) = 0$ and $g_P(x) - g_D(\rho^D) = 0$ in $x \in (x^D, 1)$ respectively and $x_{(2)}^P < \tilde{x}_{(2)}^P$. Thus, $\Pi^B > \Pi^D$ and $CW^B > CW^D$ if and only if $\rho^B < x_2^P$. Let

$$\bar{t} = V + \frac{c(x_{(2)}^P)^2[(x_{(2)}^P)^2 - 2x_{(2)}^P + 3]}{2\lambda(1+x_{(2)}^P)(1-x_{(2)}^P)^2} + \frac{e}{\lambda} - \frac{\lambda\alpha(2(x_{(2)}^P)+1)}{(x_{(2)}^P)^2},$$

we can conclude that $\Pi^B > \Pi^D$ and $CW^B > CW^D$ if and only if $t \in (V - V^P(\lambda,\alpha), \bar{t})$ by noticing that ρ^B is increasing in t. Combining the cases above gives that $\Pi^B > \Pi^D$ and $CW^B > CW^D$ if and only if $t \in (V - \hat{V}'(\lambda,\alpha), \bar{t})$, where $\bar{t} > V - V^P(\lambda,\alpha)$. $\square$

A.2. More Details about the Extensions

A.2.1. Waiting Cost Incurred During Service (§4.1)

In what follows, we re-solve for the equilibrium under the dedicated, (agent) pooling, and bonus pooling settings. We continue to use superscripts D, P and B to denote these settings, and also use subscript S to denote the case when waiting cost is incurred during service. In the dedicated case, WLOG, we consider Agent 1's optimal service rate choice μ_S^D which maximizes $\Pi_{S,1}^D(\mu_1)$ given by

$$\Pi_{S,1}^D[\mu_1] = \lambda\left[V - \alpha\mu_1 - cW^D(\mu_1) - s/\mu_1 - e/\mu_1\right] = \lambda\left[V - \alpha\mu_1 - \frac{c\lambda}{\mu_1(\mu_1-\lambda)} - (e+s)/\mu_1\right]$$

where s/μ_1 is customers' expected waiting cost during service.

PROPOSITION 5. *In the dedicated setting, there exists a unique* $\mu_S^D \in (\lambda,\infty)$ *that maximizes* $\Pi_{S,1}^D[\mu_1]$ *when* $V > V_S^D(\lambda,\alpha) := \frac{\lambda\alpha}{\rho_S^D} + \frac{c(\rho_S^D)^2}{\lambda(1-\rho_S^D)} + \frac{(e+s)\rho_S^D}{\lambda}$ *where* $\rho_S^D = \lambda/\mu_S^D \in (0,1)$ *uniquely solves* $\frac{\lambda\alpha}{(\rho_S^D)^2} - \frac{c\rho_S^D(2-\rho_S^D)}{\lambda(1-\rho_S^D)^2} - \frac{(e+s)}{\lambda} = 0$. *Accordingly,* $CW_S^D = 2\left[\lambda V - \frac{\lambda^2\alpha}{\rho_S^D} - \frac{c(\rho_S^D)^2}{1-\rho_S^D} - s\rho_S^D\right]$, $\Pi_S^D = 2\left[\lambda V - \frac{\alpha\lambda^2}{\rho_S^D} - \frac{c(\rho_S^D)^2}{1-\rho_S^D} - (e+s)\rho_S^D\right]$. *Otherwise (i.e., when* $V \le V_S^D(\lambda,\alpha)$*), all customers balk and* $CW_S^D = \Pi_S^D = 0$.

For the pooling setting, Agent 1's goal is to maximize

$$\Pi_{S,1}^P[\mu_1,\mu_2] := \lambda_1(\mu_1,\mu_2)\left[V - \alpha\mu_1 - cW(\mu_1,\mu_2) - (e+s)/\mu_1\right]$$

while Agent 2's goal is to maximize

$$\Pi_{S,2}^P[\mu_1,\mu_2] := \lambda_2(\mu_1,\mu_2)\left[V - \alpha\mu_2 - cW(\mu_1,\mu_2) - (e+s)/\mu_2\right]$$

where $W(\mu_1,\mu_2)$ is given by Lemma 1/(ii) and $\lambda_1(\mu_1,\mu_2)$ and $\lambda_2(\mu_1,\mu_2)$ are given by Lemma 1/(iii). In equilibrium, the service rates of the two agents $(\mu_{S,1}^P, \mu_{S_2}^P)$ satisfy the following conditions:

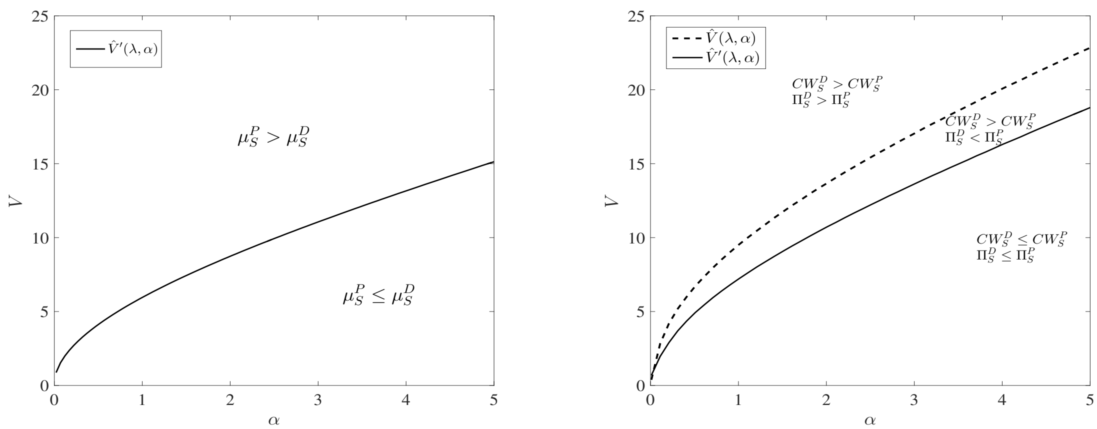
$$\mu_{S,1}^P \in \arg\max_{\mu_1} \Pi_{S,1}^P[\mu_1,\mu_{S,2}^P], \quad \mu_{S,2}^P \in \arg\max_{\mu_2} \Pi_{S,2}^P[\mu_{S,1}^P,\mu_2].$$

PROPOSITION 6. *In the pooling setting, there exists a unique symmetric equilibrium* $\mu_S^P \in (\lambda,\infty)$ *when* $V > V_S^P(\lambda,\alpha) := \frac{\lambda\alpha}{\hat{x}_s} + \frac{c\hat{x}_s^3}{\lambda(1-\hat{x}_s^2)} + \frac{(e+s)\hat{x}_s}{\lambda}$ *where* $\hat{x}_s$ *uniquely solves* $\frac{\lambda\alpha}{\hat{x}_s^2} - \frac{c\hat{x}_s^2(3-\hat{x}_s^2)}{2\lambda(1-\hat{x}_s)^2(1+\hat{x}_s)^2} - \frac{e+s}{\lambda} = 0$ *and* $\rho_S^P = \lambda/\mu^P \in (0,1)$ *uniquely solves the following equation:*

$$\frac{c(\rho_S^P)^4[(\rho_S^P)^2 - 2(\rho_S^P) + 3]}{(2\rho_S^P+1)(\rho_S^P+1)(1-\rho_S^P)^2} + \frac{2(\lambda V + e + s)(\rho_S^P)^2}{2\rho_S^P+1} - 2\lambda^2\alpha = 0.$$

Accordingly, $CW_S^P = 2\left[\lambda V - \frac{\lambda^2\alpha}{\rho_S^P} - \frac{c(\rho_S^P)^3}{1-(\rho_S^P)^2} - s\rho_S^P\right]$, $\Pi_S^P = 2\left[\lambda V - \frac{\alpha\lambda^2}{\rho_S^P} - \frac{c(\rho_S^P)^3}{1-(\rho_S^P)^2} - (e+s)\rho_S^P\right]$. *Otherwise (i.e., when* $V \le V_S^P(\lambda,\alpha)$*), all customers balk and* $CW_S^P = \Pi_S^P = 0$.

Figure 5 **Comparison of the dedicated and pooling settings under different (α, V) for the extension in §4.1**



(a) Service rate

(b) Performance metrics

Note. $\lambda = 0.5$, $c = s = 1$, $e = 5$.

PROPOSITION 7. *When $V > V_S^P(\lambda, \alpha)$, in the bonus pooling setting, there exists a unique t_S^B such that $\mu_S^B \in (\lambda, \infty)$ is the unique equilibrium if $t \in [0, t_S^B)$, where $\mu_S^B = \lambda/\rho_S^B$ and $\rho_S^B \in (0,1)$ uniquely solves:*

$$\frac{c(\rho_S^B)^4[(\rho_S^B)^2 - 2(\rho_S^B) + 3]}{(2\rho_S^B + 1)(\rho_S^B + 1)(1 - \rho_S^B)^2} + \frac{2(\lambda(V - t) + e + s)(\rho_S^B)^2}{2\rho_S^B + 1} - 2\lambda^2\alpha = 0.$$

Accordingly, $CW_S^B = 2\left[\lambda V - \frac{\lambda^2\alpha}{\rho_S^B} - \frac{c(\rho_S^B)^3}{1-(\rho_S^B)^2} - s\rho_S^B\right]$, $\Pi_S^B = 2\left[\lambda V - \frac{\alpha\lambda^2}{\rho_S^B} - \frac{c(\rho_S^B)^3}{1-(\rho_S^B)^2} - e\rho_S^B\right]$. *Otherwise (i.e., if $t \geq t_S^B$), all customers balk and $CW_S^B = \Pi_S^B = 0$.*

Based on the equilibria characterized in Propositions 5 and 6, we numerically compare the dedicated setting and the pooling setting in terms of agents' service rate choice and system performance. We present the results in Figure 5. We observe that Figure 5 is very similar to its counterpart of the main model (Figure 2). In particular, agents can speed up in the pooling setting (i.e., $\mu_S^P > \mu_S^D$) and consequently, customers and/or agents can be worse off (i.e., $CW_S^D > CW_S^P$ and/or $\Pi_S^D > \Pi_S^P$).

Finally, we numerically demonstrate that bonus pooling (which requires each agent to contribute an amount t per served customer to a common pool) can continue to result in a win-win situation for both customers and agents ($CW_S^B > CW_S^D$ and $\Pi_S^B > \Pi_S^D$). Table 5 shows the ranges of t that can lead to a win-win situation for a specific service value potential ($V = 30$) and different values of sensitivity to speed (α).

Table 5 **Range of t for the bonus pooling policy to ensure $CW_S^B > CW_S^P$ and $\Pi_S^B > \Pi_S^P$**

Sensitivity to speed α	1	2	3	4	5	6	7	8
Efficient tip policy t	(22.8,25.1)	(19.3,23.0)	(16.4,21.4)	(13.7,20.1)	(11.2,18.8)	(8.8,17.7)	(6.5,16.7)	(4.2,15.7)

Note. $\lambda = 0.5$, $c = e = s = 1$, $V = 30$.

Proof of Propositions 5-7

By replacing e with $e+s$ in Propositions 1-3, we can obtain μ_S^D, μ_S^P and μ_S^B, accordingly. Also, the waiting cost during service s/μ_S^X affects customer welfare; CW_S^X can be obtained by subtracting s/μ_S^X from CW^X, where $X = D, S, B$. □

A.2.2. Dedicated Setting with Competing Agents (§4.2)

Proof of Proposition 4

Similar to the base model, we first identify the equilibrium strategy when all customers join, i.e., $\lambda_1 + \lambda_2 = 2\lambda$. Together with the second constraint (6), the optimization problem for Agent 1 for any fixed μ_2 can be rewritten as

$$\begin{aligned}
&\max_{\mu_1>0} \quad F(\lambda_1,\mu_1) = \lambda_1\left[V - \alpha\mu_2 - \frac{c(2\lambda-\lambda_1)}{\mu_2(\mu_2-2\lambda+\lambda_1)} - \frac{e}{\mu_1}\right],\\
s.t. \quad & V - \alpha\mu_1 - \frac{c\lambda_1}{\mu_1(\mu_1-\lambda_1)} = V - \alpha\mu_2 - \frac{c\lambda_2}{\mu_2(\mu_2-\lambda_2)} \geq 0,\\
& \lambda_1+\lambda_2 = 2\lambda, \quad 0\leq\lambda_1<\mu_1, \quad 0\leq\lambda_2<\mu_2,\\
& V - \alpha\mu_1 - \frac{c\lambda_1}{\mu_1(\mu_1-\lambda_1)} - \frac{e}{\mu_1} \geq 0.
\end{aligned} \tag{A.12}$$

Assume there exist some μ_1, μ_2 that satisfy (A.12), i.e., the feasible region is non-empty, and denote the feasible region of μ_1 for any fixed μ_2 by $\Omega_1(\mu_2)$. Then, for any fixed μ_2, λ_1 is uniquely determined by

$$f(\lambda_1,\mu_1) = \alpha(\mu_2-\mu_1) + \frac{c(2\lambda-\lambda_1)}{\mu_2(\mu_2-2\lambda+\lambda_1)} - \frac{c\lambda_1}{\mu_1(\mu_1-\lambda_1)} = 0.$$

It is not difficult to verify that $f(\lambda_1,\mu_1)$ is continuous and differentiable in (λ_1,μ_1) for any $\mu_1>\lambda_1$ and $\mu_2 > 2\lambda-\lambda_1$. Since $\frac{\partial f(\lambda_1,\mu_1)}{\partial\lambda_1} < 0$, by implicit function theorem, we have

$$\frac{\partial\lambda_1}{\partial\mu_1} > 0 \Leftrightarrow -\frac{\partial f(\lambda_1,\mu_1)/\partial\mu_1}{\partial f(\lambda_1,\mu_1)/\partial\lambda_1} > 0 \Leftrightarrow \frac{\partial f(\lambda_1,\mu_1)}{\partial\mu_1} > 0 \Leftrightarrow \frac{c\lambda_1(2\mu_1-\lambda_1)}{\mu_1^2(\lambda_1-\mu_1)^2} > \alpha \Leftrightarrow \mu_1 \in (\lambda_1, \bar{\mu}_1)$$

by noticing that $\frac{c\lambda_1(2\mu_1-\lambda_1)}{\mu_1^2(\lambda_1-\mu_1)^2}$ is decreasing in $\mu_1 > \lambda_1$. Therefore, if $V - \alpha\mu_2 - \frac{c(2\lambda-\lambda_1)}{\mu_2(\mu_2-2\lambda+\lambda_1)} - \frac{e}{\mu_1} \leq 0$ for all $\mu_1 \in (\lambda_1,\bar{\mu}_1)$, the best response of Agent 1 satisfies $\hat{\mu}_1(\mu_2) \geq \bar{\mu}_1$. Otherwise, if $V - \alpha\mu_2 - \frac{c(2\lambda-\lambda_1)}{\mu_2(\mu_2-2\lambda+\lambda_1)} - \frac{e}{\mu_1} > 0$ for some $\mu_1 \in (\lambda_1,\bar{\mu}_1)$, there exists a certain threshold $\underline{\mu}_1 < \bar{\mu}_1$ such that $V - \alpha\mu_2 - \frac{c(2\lambda-\lambda_1)}{\mu_2(\mu_2-2\lambda+\lambda_1)} - \frac{e}{\mu_1} > 0$ if and only if $\mu_1 \in (\underline{\mu}_1, \bar{\mu}_1)$, under which we have

$$\frac{dF(\lambda_1,\mu_1)}{d\mu_1} = \frac{\partial F(\lambda_1,\mu_1)}{\partial\lambda_1}\frac{\partial\lambda_1}{\partial\mu_1} + \frac{\partial F(\lambda_1,\mu_1)}{\partial\mu_1} > 0,$$

which implies that $F(\lambda_1,\mu_1)$ is increasing in $\mu_1 \in (\underline{\mu}_1,\bar{\mu}_1)$, which also implies that $\hat{\mu}_1(\mu_2) \geq \bar{\mu}_1$. Then we just need to consider the case that $\mu_1 \in [\bar{\mu}_1, \infty)$, under which we have $\frac{\partial\lambda_1}{\partial\mu_1} \leq 0$. Next, we will show that $F(\lambda_1,\mu_1)$ is concave in $\mu_1 \in [\bar{\mu}_1,\infty)\bigcap\Omega_1(\mu_2)$. We can verify that

$$\frac{\partial^2 f(\lambda_1,\mu_1)}{\partial\mu_1^2} = 2c\left(\frac{1}{\mu_1^3} - \frac{1}{(\mu_1-\lambda_1)^3}\right) < 0, \quad \frac{\partial^2 f(\lambda_1,\mu_1)}{\partial\lambda_1\partial\mu_1} = \frac{2c}{(\mu_1-\lambda_1)^3} > 0.$$

Thus it follows that

$$\frac{\partial^2 F(\lambda_1,\mu_1)}{\partial \mu_1^2} = -\frac{2\lambda_1 e}{\mu_1^3} < 0, \quad \frac{\partial^2 F(\lambda_1,\mu_1)}{\partial \lambda_1 \partial \mu_1} = \frac{e}{\mu_1^2} > 0,$$

$$\frac{\partial^2 \lambda_1}{\partial \mu_1^2} = -\frac{\frac{\partial^2 f(\lambda_1,\mu_1)}{\partial \mu_1^2} + \frac{\partial^2 f(\lambda_1,\mu_1)}{\partial \lambda_1 \partial \mu_1} \cdot \frac{\partial \lambda_1}{\partial \mu_1}}{\frac{\partial f(\lambda_1,\mu_1)}{\partial \lambda_1}} < 0, \quad \frac{\partial F(\lambda_1,\mu_1)}{\partial \lambda_1} = \left[V - \alpha\mu_2 - \frac{c(2\lambda-\lambda_1)}{\mu_2(\mu_2 - 2\lambda + \lambda_1)} - \frac{e}{\mu_1}\right] + \frac{c}{(\lambda_1 + \mu_2 - 2\lambda)^2} > 0$$

by noticing that $V - \alpha\mu_2 - \frac{c(2\lambda-\lambda_1)}{\mu_2(\mu_2-2\lambda+\lambda_1)} - \frac{e}{\mu_1} > 0$ when $\mu_1 \in \Omega_1(\mu_2)$. Then it gives

$$\frac{d^2 F(\lambda_1,\mu_1)}{d\mu_1^2} = \frac{\partial^2 F(\lambda_1,\mu_1)}{\partial \mu_1^2} + \frac{\partial^2 F(\lambda_1,\mu_1)}{\partial \lambda_1 \partial \mu_1} \cdot \frac{\partial \lambda_1}{\partial \mu_1} + \frac{\partial^2 \lambda_1}{\partial \mu_1^2} \cdot \frac{\partial F(\lambda_1,\mu_1)}{\partial \lambda_1} < 0,$$

i.e., $F(\lambda_1,\mu_1)$ is concave in $\mu_1 \in [\bar{\mu}_1, \infty) \bigcap \Omega_1(\mu_2)$. Next, we will show that there exists exactly only one interval $I(\mu_2) = (a,b)$ such that $(a,b) = [\bar{\mu}_1,\infty) \bigcap \Omega_1(\mu_2)$. It is sufficient to show that $F(\lambda_1,\mu_1)$ is unimodal in $\mu_1 \in [\bar{\mu}_1,\infty) \bigcap \Omega_1(\mu_2)$. That is, for any $\mu_1 \in [\bar{\mu}_1,\infty) \bigcap \Omega_1(\mu_2)$,

$$\frac{d[V - \alpha\mu_2 - \frac{c(2\lambda-\lambda_1)}{\mu_2(\mu_2-2\lambda+\lambda_1)} - \frac{e}{\mu_1}]}{d\mu_1}|_{\mu_1 = x} < 0$$

implies that

$$\frac{d[V - \alpha\mu_2 - \frac{c(2\lambda-\lambda_1)}{\mu_2(\mu_2-2\lambda+\lambda_1)} - \frac{e}{\mu_1}]}{d\mu_1}|_{\mu_1 = x+\delta} < 0$$

for any $\delta > 0$. Notice that

$$\frac{d[V - \alpha\mu_2 - \frac{c(2\lambda-\lambda_1)}{\mu_2(\mu_2-2\lambda+\lambda_1)} - \frac{e}{\mu_1}]}{d\mu_1} < 0 \Leftrightarrow \frac{e}{\mu_1^2} < \frac{-c\partial\lambda_1/\partial\mu_1}{(\mu_2 - 2\lambda + \lambda_1)^2}.$$

Since $\partial\lambda_1/\partial\mu_1 < 0$ and $\partial^2\lambda_1/\partial\mu_1^2 < 0$, it follows that

$$\frac{\partial \frac{-c\partial\lambda_1/\partial\mu_1}{(\mu_2-2\lambda+\lambda_1)^2}}{\partial\mu_1} = \frac{(-\partial\lambda_1^2/\partial\mu_1^2)(\mu_2 - 2\lambda + \lambda_1)^2 + 2(\mu_2 - 2\lambda + \lambda_1)(\partial\lambda_1/\partial\mu_1)^2}{(\mu_2 - 2\lambda + \lambda_1)^4} > 0.$$

Thus, $\frac{-\partial\lambda_1/\partial\mu_1}{(\mu_2-2\lambda+\lambda_1)^2}|_{\mu_1=x} < \frac{-\partial\lambda_1/\partial\mu_1}{(\mu_2-2\lambda+\lambda_1)^2}|_{\mu_1=x+\delta}$. On the other hand, since $\frac{e}{(\mu_1+\delta)^2} < \frac{e}{\mu_1^2}$, we have

$$\frac{e}{(\mu_1+\delta)^2} < \frac{e}{\mu_1^2} < \frac{-c\partial\lambda_1/\partial\mu_1}{\mu_2(\mu_2 - 2\lambda + \lambda_1)^2}|_{\mu_1=x} < \frac{-c\partial\lambda_1/\partial\mu_1}{\mu_2(\mu_2 - 2\lambda + \lambda_1)^2}|_{\mu_1=x+\delta}.$$

That is, $d[V - \alpha\mu_2 - \frac{c(2\lambda-\lambda_1)}{\mu_2(\mu_2-2\lambda+\lambda_1)} - e/\mu_1]/d\mu_1|_{\mu_1=x+\delta} < 0$ for any $\delta > 0$. Therefore, $F(\lambda_1,\mu_1)$ is concave and $F(\lambda_1,\mu_1) > 0$ in $\mu_1 \in (a,b)$. It follows that there exists a unique maximizer $\hat{\mu}_1(\mu_2) \in (a,b)$ for any fixed μ_2, which satisfies $\frac{dF(\lambda_1,\mu_1)}{d\mu_1} = 0$. Then the symmetric equilibrium can be attained at $\mu_C^D = \hat{\mu}_1(\mu_C^D)$. Note that

$$\frac{\partial\lambda_1}{\partial\mu_1}|_{\mu_2=\mu_1} = -\frac{\alpha\mu_1^2(\lambda-\mu_1)^2 + c\lambda(\lambda - 2\mu_1)}{2c\mu_1^2}.$$

Then the equilibrium satisfies

$$\frac{dF(\lambda_1,\mu_1)}{d\mu_1}|_{\mu_2=\mu_1} = \frac{\partial\lambda_1}{\partial\mu_1}\left[V-\alpha\mu_1-\frac{c\lambda}{\mu_1(\mu_1-\lambda)}-\frac{e}{\mu_1}\right]|_{\mu_2=\mu_1}+\lambda\left[\frac{\partial\lambda_1}{\partial\mu_1}\cdot\frac{c}{(\mu_1-\lambda)^2}+\frac{e}{\mu_1^2}\right]|_{\mu_2=\mu_1}=0$$

$$\Leftrightarrow \frac{\alpha\mu_1^2(\lambda-\mu_1)^2+c\lambda(\lambda-2\mu_1)}{2c\mu_1^2}+\frac{e\lambda(\lambda-\mu_1)^2}{\mu_1\left[(\lambda-\mu_1)^2(e-\mu_1 V+\mu_1^2\alpha)-c\lambda^2\right]}=0$$

$$\Leftrightarrow \Phi(\rho)=\frac{\lambda\alpha}{\rho^2}-\frac{c\rho(2-\rho)}{\lambda(1-\rho)^2}-\frac{2ce\rho^3}{\lambda[c\rho^4+(1-\rho)^2(\lambda\rho V-\lambda^2\alpha-e\rho^2)]}=0.$$

We can verify that $\Phi(\rho)|_{\rho=1^-}=-\infty<0$ and $\Phi(\rho)|_{\rho\to 0^+}=+\infty>0$, then there exists a certain solution ρ^C such that $\Phi(\rho_C^D)=0$. Similar to the proof of Proposition 2, the symmetric equilibrium exists if and only if $F(\lambda,\mu_C^D)>0$, i.e., $V>V_C^D(\lambda,\alpha)=\alpha\lambda/\rho_C^D+\frac{c(\rho_C^D)^2}{\lambda^2(1-\rho_C^D)}+\frac{e\rho_C^D}{\lambda}$. Otherwise, if $V\le V_C^D(\lambda,\alpha)$, $\mu_C^D=V/\alpha$ with all customers balking is an equilibrium under which we have $CW_C^D=\Pi_C^D=0$. □

A.2.3. Alternative Effort Cost Structure (§4.3)

In what follows, we re-solve for the equilibria in the dedicated, (agent) pooling, and bonus pooling settings, respectively, based on (9). We continue to use superscripts D, P and B to denote these settings, and also use subscript A to denote the alternative effort cost structure.

Proposition 8. *In the dedicated setting with the alternative effort cost structure, there exists a unique $\mu_A^D\in(\lambda,\infty)$ that maximizes $\Pi_{A,1}^D[\mu_1]$ when $V>V_A^D(\lambda,\alpha)$, where $\mu_A^D=\lambda/\rho_A^D$ and $\rho_A^D\in(0,1)$ uniquely solves $\frac{\lambda\alpha+e}{(\rho_A^D)^2}-\frac{c\rho_A^D(2-\rho_A^D)}{\lambda(1-\rho_A^D)^2}=0$; $V_A^D(\lambda,\alpha)=\frac{\lambda\alpha+e}{\rho_A^D}+\frac{c(\rho_A^D)^2}{\lambda(1-\rho_A^D)}$. Accordingly, $CW_A^D=2\lambda\left[V-\frac{\lambda\alpha}{\rho_A^D}-\frac{c(\rho_A^D)^2}{\lambda[1-\rho_A^D]}\right]$, $\Pi_A^D=2\left[\lambda V-\frac{\alpha\lambda^2}{\rho_A^D}-\frac{c(\rho_A^D)^2}{1-\rho_A^D}-e\rho_A^D\right]$. Otherwise (i.e., when $V\le V_A^D(\lambda,\alpha)$), all customers balk and $CW_A^D=\Pi_A^D=0$.*

Proposition 9. *In the pooling setting with the alternative effort cost structure, there exists a unique symmetric equilibrium $\mu_A^P\in(\lambda,\infty)$ when $V>V_A^P(\lambda,\alpha):=\frac{\lambda\alpha+e}{\rho_A^P}+\frac{c(\rho_A^P)^3}{\lambda[1-(\rho_A^P)^2]}$ where $\rho_A^P=\lambda/\mu_A^P\in(0,1)$ uniquely solves the following equation:*

$$\frac{c(\rho^P)^4[(\rho^P)^2-2(\rho^P)+3]}{(2\rho^P+1)(\rho^P+1)(1-\rho^P)^2}+\frac{2\lambda[(\rho^P)^2V-e(1+\rho^P)]}{2\rho^P+1}-2\lambda^2\alpha=0.$$

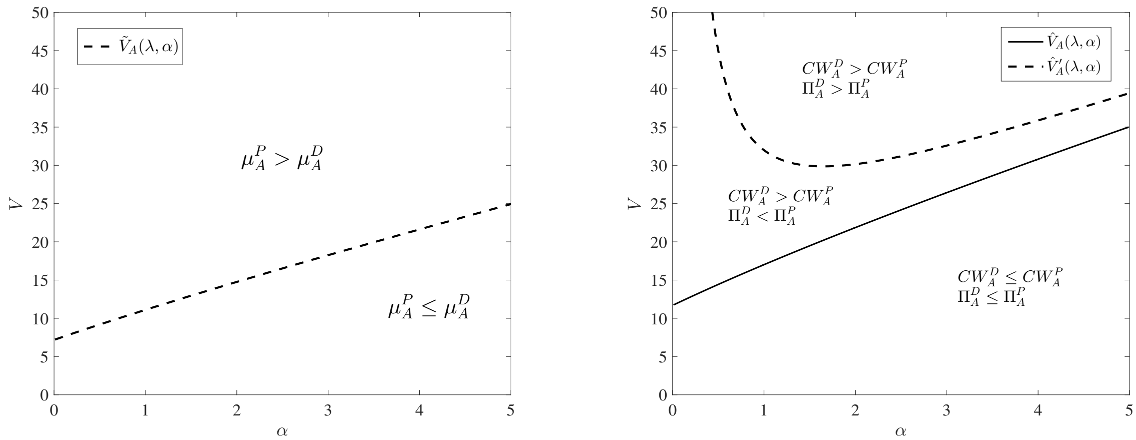
Accordingly, $CW_A^P=2\lambda\left[V-\frac{\lambda\alpha}{\rho_A^P}-\frac{c(\rho_A^P)^3}{\lambda[1-(\rho_A^P)^2]}\right]$, $\Pi_A^P=2\left[\lambda V-\frac{\alpha\lambda^2}{\rho_A^P}-\frac{c(\rho_A^P)^3}{1-(\rho_A^P)^2}-e\rho_A^P\right]$. Otherwise (i.e., when $V\le V_A^P(\lambda,\alpha)$), all customers balk and $CW_A^P=\Pi_A^P=0$.

Proposition 10. *When $V>V_A^P(\lambda,\alpha)$, in the bonus pooling setting, there exists a unique t_A^B such that $\mu_A^B\in(\lambda,\infty)$ is the unique equilibrium if $t\in[0,t_A^B)$, where $\mu_A^B=\lambda/\rho_A^B$ and $\rho_A^B\in(0,1)$ uniquely solves:*

$$\frac{c(\rho_A^B)^4[(\rho_A^B)^2-2(\rho_A^B)+3]}{(2\rho_A^B+1)(\rho_A^B+1)(1-\rho_A^B)^2}+\frac{2\lambda[(\rho_A^B)^2(V-t)-e(1+\rho_A^B)]}{2\rho_A^B+1}-2\lambda^2\alpha=0.$$

Accordingly, $CW_A^B=2\lambda\left[V-t-\frac{\lambda\alpha}{\rho_A^B}-\frac{c(\rho_A^B)^3}{\lambda[1-(\rho_A^B)^2]}\right]$, $\Pi_A^B=2\left[\lambda(V-t)-\frac{\alpha\lambda^2}{\rho_A^B}-\frac{c(\rho_A^B)^3}{1-(\rho_A^B)^2}-e\rho_A^B\right]$. Otherwise (i.e., if $t\ge t_A^B$), all customers balk and $CW_A^B=\Pi_A^B=0$.

Figure 6 Comparison of the dedicated and pooling settings under different (α, V) for §4.3.



(a) Service rate (b) Performance metrics

Note. $\lambda = c = e = 1$

Based on the equilibria characterized in Propositions 8 and 9, we numerically compare the dedicated setting and the pooling setting in terms of agents' service rate choice and system performance. We present the results in Figure 6. We observe from Figure 6 that, with the alternative effort cost structure, customers and/or agents can still be worse off under the pooling setting compared to the dedicated setting (i.e., $CW_A^D > CW_A^P$ and/or $\Pi_A^D > \Pi_A^P$), and this continues to be driven by agents' speedup behavior under pooling ($\mu_A^P > \mu_A^D$) in an attempt to seize more of the available customers (when the service value potential V is sufficiently large). Such an effect is not limited to any specific effort cost structure.

Table 6 Range of t for the bonus pooling policy to ensure $CW_A^P > CW_A^D$ and $\Pi_A^P > \Pi_A^D$

$\alpha =$	1	2	3	4	5	6	7	8
$t \in$	(16.3,25.8)	(11.1,24.2)	(6.4,22.7)	(1.9,21.2)	(0,19.8)	(0,18.4)	(0,17.1)	(0,15.8)

Note. $\lambda = c = e = 1$, $V = 30$.

Finally, we numerically demonstrate that bonus pooling (which requires each agent to contribute an amount t per served customer to a common pool) can continue to result in a win-win situation for both customers and agents ($CW^P > CW_C^D$ and $\Pi^P > \Pi_C^D$). Table 6 shows the ranges of t that can lead to a win-win situation for a specific service value potential ($V = 30$) and different values of sensitivity to speed (α).

Proof of Proposition 8

When all customers join, we consider Agent 1's optimal service rate choice μ_A^D which maximizes $\Pi_{A,1}^D(\mu)$ given by

$$\Pi_{A,1}^D[\mu] = \lambda\left[V - \alpha\mu - cW^D(\mu)\right] - e\mu = \lambda\left[V - \alpha\mu - \frac{c\lambda}{\mu(\mu-\lambda)}\right] - e\mu, \tag{A.13}$$

where $W^D(\mu) = \frac{\lambda}{\mu(\mu-\lambda)}$ for an $M/M/1$ queue. We can derive the first-order and second-order conditions of (A.13) with respect to μ as follows.

$$\begin{aligned}\frac{\partial[\Pi_{A,1}^D(\mu)]}{\partial\mu} &= \lambda\left(\frac{c\lambda(2\mu-\lambda)}{(\mu-\lambda)^2\mu^2} - \alpha\right) - e,\\ \frac{\partial^2[\Pi_{A,1}^D(\mu)]}{\partial\mu^2} &= -\frac{2c\lambda^2\left(\lambda^2 - 3\lambda\mu + 3\mu^2\right)}{\mu^3(\mu-\lambda)^3} < 0 \text{ because } \mu > \lambda.\end{aligned}$$

Then $\frac{\partial[\Pi_{A,1}^D(\mu)]}{\partial\mu}$ is decreasing in $\mu > \lambda$. Since $\frac{\partial[\Pi_{A,1}^D(\mu)]}{\partial\mu}|_{\mu=\lambda^+} = +\infty$ and $\frac{\partial[\Pi_{A,1}^D(\mu)]}{\partial\mu}|_{\mu=\infty} = -\lambda\alpha - e$, there exists a unique μ_A^D that maximizes $\Pi_{A,1}^D(\mu)$ and uniquely solves $\frac{\partial[\Pi_{A,1}^D(\mu)]}{\partial\mu} = 0 \Leftrightarrow \frac{\lambda\alpha}{(\rho_A^D)^2} - \frac{c\rho_A^D(2-\rho_A^D)}{\lambda(1-\rho_A^D)^2} + \frac{e}{(\rho_A^D)^2} = 0$, where $\rho_A^D = \lambda/\mu_A^D$. Similar to the proof of Proposition 2, the symmetric equilibrium with all customers joining exists if and only if

$$\lambda\left[V - \alpha\mu_A^D - \frac{c\lambda}{\mu_A^D(\mu_A^D-\lambda)}\right] - e\mu_A^D > 0 \Leftrightarrow V > V_A^D(\lambda,\alpha) = \frac{\lambda\alpha}{\rho_A^D} + \frac{c(\rho_A^D)^2}{\lambda(1-\rho_A^D)} + \frac{e}{\rho_A^D}.$$

Otherwise, if $V \le V_A^D(\lambda,\alpha)$, $\mu_A^D = V/\alpha$ with all customers balking is an equilibrium under which we have $CW_A^D = \Pi_A^D = 0$. □

Proof of Proposition 9

We first search for the equilibrium when all customers join, i.e., $\lambda_1 + \lambda_2 = 2\lambda$. For any fixed μ_2, taking the derivative of $\Pi_{A,1}^P(\mu_1,\mu_2)$ with respect to μ_1 gives

$$\begin{aligned}\frac{\partial\Pi_{A,1}^P(\mu_1,\mu_2)}{\partial\mu_1} &= \frac{\partial\lambda_1(\mu_1,\mu_2)}{\partial\mu_1}\left[V - \alpha\mu_1 - cW(\mu_1,\mu_2)\right] - e - \lambda_1\left[\alpha + c\frac{\partial W(\mu_1,\mu_2)}{\partial\mu_1}\right],\\ \frac{\partial^2\Pi_{A,1}^P(\mu_1,\mu_2)}{\partial\mu_1^2} &= \frac{\partial^2\lambda_1(\mu_1,\mu_2)}{\partial\mu_1^2}\left[V - \alpha\mu_1 - cW(\mu_1,\mu_2)\right] - \lambda_1 c\frac{\partial^2 W(\mu_1,\mu_2)}{\partial\mu_1^2}.\end{aligned}$$

If $\lambda_1\left[V - \alpha\mu_1 - cW(\mu_1,\mu_2)\right] - e/\mu_1 \le 0$ for all $\mu_1 \in (0,\infty)$, then the best response of Agent 1 is $\mu_1 \ge V/\alpha$. If $\lambda_1\left[V - \alpha\mu_1 - cW(\mu_1,\mu_2)\right] - e/\mu_1 > 0$ for some $\mu_1 \in (0,\infty)$, there must exist $0 < \underline{\mu}_1 < \bar{\mu}_1 < \infty$ such that $\lambda_1\left[V - \alpha\mu_1 - cW(\mu_1,\mu_2)\right] - e/\mu_1 > 0$ if and only if $\mu_1 \in (\underline{\mu}_1, \bar{\mu}_1)$ since $\left[V - \alpha\mu_1 - cW(\mu_1,\mu_2)\right]$ is strictly concave in μ_1 by Lemma 3. Therefore, the best response $\hat{\mu}_1(\mu_2)$ can only be attained at $\hat{\mu}_1(\mu_2) \in (\underline{\mu}_1, \bar{\mu}_1)$. Since $\frac{\partial^2\lambda_1(\mu_1,\mu_2)}{\partial\mu_1^2} < 0$ and $c\frac{\partial^2 W(\mu_1,\mu_2)}{\partial\mu_1^2} + \frac{2e}{\mu_1^3} > 0$, it follows that $\frac{\partial^2\Pi_1^P(\mu_1,\mu_2)}{\partial\mu_1^2} < 0$ for $\mu_1 \in (\underline{\mu}_1, \bar{\mu}_1)$, i.e., $\Pi_{A,1}^P(\mu_1,\mu_2)$ is strictly concave in $\mu_1 \in (\underline{\mu}_1, \bar{\mu}_1)$. Since

$\Pi^P_{A,1}(\mu_1,\mu_2)|_{\mu_1=\underline{\mu}_1} = \Pi^P_{A,1}(\mu_1,\mu_2)|_{\mu_1=\bar{\mu}_1} = 0$, there exists a unique maximizer $\hat{\mu}_1(\mu_2) \in (\underline{\mu}_1, \bar{\mu}_1)$ for any fixed μ_2.

In a symmetric equilibrium, $\mu^P_{A,1} = \mu^P_{A,2}$. Therefore, by letting $\mu_2 = \mu_1$ in $\frac{\partial \Pi^P_{A,1}(\mu_1,\mu_2)}{\partial \mu_1}$, we can solve for $\mu^P_{A,1}$ by setting $\frac{\partial \Pi^P_{A,1}(\mu_1,\mu_2)}{\partial \mu_1}|_{\mu_2=\mu_1} = 0$, i.e.,

$$\frac{c\lambda^3\left(\lambda^2 - 2\lambda\mu_1 + 3\mu_1^2\right) - 2\mu_1(\lambda-\mu_1)^2(\lambda+\mu_1)\left(\mu_1[\alpha\lambda(2\lambda+\mu_1) + e(\lambda+\mu_1)] - \lambda^2 V\right)}{2\left(\mu_1^3 - \lambda^2\mu_1\right)^2} = 0$$

$$\Leftrightarrow \bar{\phi}_{A,P}(\rho_1) := \frac{c\rho_1^4[\rho_1^2 - 2\rho_1 + 3]}{(2\rho_1+1)(\rho_1+1)(1-\rho_1)^2} + \frac{2\lambda[\rho_1^2 V - e(1+\rho_1)]}{2\rho_1+1} - 2\lambda^2\alpha = 0,$$

where $\rho_1 = \lambda/\mu_1 \in [0,1)$. Similar to the proof of Proposition 2, we can show that $\bar{\phi}_{A,P}(\rho_1)$ is increasing in ρ_1 and $\bar{\phi}_{A,P}(\rho_1) = 0$ has a unique solution in ρ_1, which determines the unique symmetric equilibrium. The existence of the equilibrium can be similarly guaranteed. Taking derivatives of $\bar{\phi}_{A,P}(\rho_1)$ with respect to V gives

$$\frac{\partial \bar{\phi}_{A,P}(\rho_1)}{\partial V} = \frac{2\lambda\rho_1^2}{2\rho_1+1} > 0.$$

Thus we can establish that ρ^P_A is decreasing in V. That is, the symmetric equilibrium service rate $\mu^P_A = \lambda/\rho^P_A$ is increasing in V. Finally, similar to the proof of Proposition 2, the symmetric equilibrium with all customers joining exists if and only if

$$\Pi^P_{A,1}[\mu^P_A, \mu^P_A] > 0 \Leftrightarrow V > \frac{\lambda\alpha}{\rho^P_A} + \frac{c(\rho^P_A)^3}{\lambda[1-(\rho^P_A)^2]} + \frac{e}{\rho^P_A}.$$

Otherwise, if $V \le V^P_A(\lambda,\alpha)$, $\mu^P_A = V/\alpha$ with all customers balking is an equilibrium under which we have $CW^P_A = \Pi^P_A = 0$. □

Proof of Proposition 10

Same as the proof for Proposition 3. □