Online Supplement: Proofs

Proof of Proposition 1. Suppose there are N customer types. Without loss of generality, suppose that K price-service classes are chosen by some customers. Write λ_{ik} for the rate of type-i customers who buy class k, where $\lambda_i = \sum_{k=1}^{K} \lambda_{ik}$ is the overall type-i arrival rate.

Part 1. Suppose that for $i \in \{1, 2, ..., N\}$ and $k \in \{1, 2, ..., K\}$, the arrival rates $\{\lambda_{ik}\}$ and delays $\{W_k\}$ are feasible, and that $\{W_k\}$ and the price functions $\{P_k\}$ form a customer equilibrium, i.e., they satisfy the individual rationality (IR) constraints, and if types are indistinguishable by the provider, the incentive-compatibility (IC) constraints. Since the customer equilibrium conditions and the provider's expected revenue rate depend on $\{P_k\}$ only through the expectations $\{E[P_k(W_k)]\}$, setting the flat rates $p_k = E[P_k(W_k)]$ for $k \in \{1, 2, ..., K\}$ establishes Part 1.

Part 2. If customer types have different service requirements, then the lead time in a given class may vary by type and there may not exist flat rates that ensure incentive-compatibility at the desired arrival rates. However, if types are distinguishable, then only the IR constraints need to hold. These constraints are easily satisfied since the provider can charge a different flat rate for each type.■

Proof of Proposition 2. Fix a scheduling policy and a demand vector $\boldsymbol{\lambda}$ such that the steady-state lead times of all classes have finite moments. Let $W_i(\boldsymbol{\lambda})$ denote the lead time of class i, which is targeted to type i. To prove the result we show that: for given $\boldsymbol{\lambda}$ and $W_i(\boldsymbol{\lambda})$ the provider maximizes her expected type-i revenue by charging $P_i(W_i) = \underline{v}_i(\lambda_i) - C_i(W_i)$, and the resulting revenue rate is a function of $\boldsymbol{\lambda}$ that is invariant to customers' risk aversion levels.

Part 1. Because the provider can distinguish among types, he can limit each customer to the price-service class that is targeted to her type. It therefore suffices to prove the claim for a single type. For simplicity we write W_i, suppressing the dependence on $\boldsymbol{\lambda}$. Given λ_i and W_i, the provider maximizes the type-i revenue by solving

$$\max_{P_i} \lambda_i E[P_i(W_i)] \tag{21}$$

$$s.t. \quad E[U(\underline{v}_i(\lambda_i) - C_i(W_i) - P_i(W_i))] = 0. \tag{22}$$

This problem is the multi-type analog of (1)-(2) for given λ_i and W_i. We have

$$U(E[\underline{v}_i(\lambda_i) - C_i(W_i) - P_i(W_i)]) \geq E[U(\underline{v}_i(\lambda_i) - C_i(W_i) - P_i(W_i))] = 0. \tag{23}$$

The inequality follows by Jensen's inequality since U is concave, and the equality follows from (22). Since $U(0) = 0$ and $U' > 0$, we have $U(x) \geq 0 \Leftrightarrow x \geq 0$, so that (23) implies the following upper bound on the expected payment:

$$E[P_i(W_i)] \leq \underline{v}_i(\lambda_i) - E[C_i(W_i)],$$

with equality if and only if (23) holds with equality. If type i is RN, then the inequality in (23) holds with equality, and every tariff P_i is feasible if and only if $E[P_i(W_i)] = \underline{v}_i(\lambda_i) - E[C_i(W_i)]$. If type i is RA, i.e., U is strictly concave, then the inequality in (23) holds with equality if and only if $\underline{v}_i(\lambda_i) - C_i(W_i) - P_i(W_i) \equiv 0$, i.e., the marginal customer's payoff is zero with probability one. It follows that $P_i(W_i) \equiv \underline{v}_i(\lambda_i) - C_i(W_i)$ is the unique optimal tariff.

Since the optimal tariff has this form for arbitrary λ_i and $W_i(\boldsymbol{\lambda})$, it also has this form at the optimal arrival rates and the corresponding lead time distributions.

Part 2. Given a scheduling policy, the tariffs $P_i(W_i) \equiv \underline{v}_i(\lambda_i) - C_i(W_i)$ maximize the revenue rate for fixed $\boldsymbol{\lambda}$ and ensure that (22) holds for all types. The resulting revenue rate function is

$$\sum_{i=1}^{N} \lambda_i(\underline{v}_i(\lambda_i) - E[C_i(W_i(\boldsymbol{\lambda}))]), \tag{24}$$

which is independent of customers' risk aversion and is the same as for RN customers.■

Lemma 1 summarizes useful properties of $\widetilde{W}(\lambda, s)$ and $L(\lambda, s)$ for Propositions 3-6**.**

Lemma 1 *Let $\widetilde{W}(\lambda, s) := E[\exp(sW(\lambda))]$ be the MGF of the r.v. $W(\lambda)$ and $L(\lambda, s) := \ln \widetilde{W}(\lambda, s)$.*

1. *$\widetilde{W}(\lambda, s) \geq 0$, $\widetilde{W}(\lambda, 0) = 1$, and $\widetilde{W}(\lambda, s)$ is strictly increasing in s with $\widetilde{W}_s(\lambda, 0) = E[W(\lambda)]$, and strictly convex in s.*
2. *$L(\lambda, 0) = 0$, $L(\lambda, s)$ is strictly increasing in s with $L_s(\lambda, 0) = E[W(\lambda)]$, and strictly convex in s with $L_{ss}(\lambda, 0) = Var[W(\lambda)]$.*
3. *For any constant $k \neq 0$, $L(\lambda, rk)/r$ is strictly increasing in $r \geq 0$ and $\lim_{r\to 0} L(\lambda, rk)/r = kE[W(\lambda)]$.*
4. *If $W(\lambda)$ is the sojourn time in a M/G/1 queue, then: (a) For $s > 0$ both $\widetilde{W}(\lambda, s)$ and $L(\lambda, s)$ are strictly increasing and strictly convex in λ, and for $s < 0$ both $\widetilde{W}(\lambda, s)$ and $L(\lambda, s)$ are strictly decreasing and strictly concave in λ. (b) The function $L_\lambda(\lambda, s)$ is strictly convex in s.*

Proof. *Parts 1-2.* are standard. E.g., see Gallager (1996), Chapters 1 and 7 (problem 7.7).

*Part 3***.** We first show that

$$\frac{d}{dr}\frac{L(\lambda, rk)}{r} = \frac{rkL_s(\lambda, rk) - L(\lambda, rk)}{r^2} > 0 \text{ for } r \geq 0. \tag{25}$$

For $r > 0$, (25) holds since $L(\lambda, 0) = 0$ and $L(\lambda, s)$ is strictly convex in s by Part 2, which implies

$$rkL_s(\lambda, rk) = \int_0^{rk} L_s(\lambda, rk)\, dx > \int_0^{rk} L_s(\lambda, x)\, dx = L(\lambda, rk).$$

For $r = 0$ we have

$$\lim_{r\to 0} \frac{rkL_s(\lambda, rk) - L(\lambda, rk)}{r^2} = \lim_{r\to 0} \frac{rk^2 L_{ss}(\lambda, rk)}{2r} = \frac{k^2 VAR(W(\lambda))}{2} > 0.$$

The first equality follows since $L(\lambda, 0) = 0$ (Part 2) and from l'Hôpital's rule, and the second since $L_{ss}(\lambda, 0) = Var[W(\lambda)]$ by Part 2. Finally, we have

$$\lim_{r\to 0} \frac{L(\lambda, rk)}{r} = \lim_{r\to 0} \frac{d}{dr} L(\lambda, rk) = k \lim_{r\to 0} L_s(\lambda, rk) = kE[W(\lambda)].$$

The first equality follows since $L(\lambda, 0) = 0$ (Part 2) and from l'Hôpital's rule, and the last since $L_s(\lambda, 0) = E[W(\lambda)]$ by Part 2.

Part 4. Let X be the service time and $\widetilde{X}(s)$ the MGF of its distribution. Let X_e be the equilibrium residual service time and $\widetilde{X}_e(s)$ the MGF of its distribution. Then $\widetilde{X}_e(s) = \mu(\widetilde{X}(s) - 1)/s$. From the Pollaczek-Khinchin formula, the MGF of the sojourn time is

$$\widetilde{W}(\lambda, s) = \widetilde{X}(s) \frac{1-\rho}{1-\rho\widetilde{X}_e(s)} = \widetilde{X}(s) \frac{\mu - \lambda}{\mu - \lambda\widetilde{X}_e(s)},$$

and it is defined only for s such that $1 > \rho\widetilde{X}_e(s)$, or equivalently, $\mu - \lambda\widetilde{X}_e(s) > 0$.

4(a) First consider $\widetilde{W}(\lambda, s)$. We have

$$\begin{aligned}
\widetilde{W}_\lambda(\lambda, s) &= \frac{\mu\widetilde{X}(s)}{\left(\mu - \lambda\widetilde{X}_e(s)\right)^2}\left(\widetilde{X}_e(s) - 1\right) \\
\widetilde{W}_{\lambda\lambda}(\lambda, s) &= \frac{2\mu\widetilde{X}(s)\widetilde{X}_e(s)}{(\mu - \lambda\widetilde{X}_e(s))^3}\left(\widetilde{X}_e(s) - 1\right).
\end{aligned}$$

Both have the same sign as $\left(\widetilde{X}_e(s) - 1\right)$, which has the sign of s because X_e is nonnegative: for $x \geq 0$, $sgn\left(e^{sx} - 1\right) = sgn(s)$. This establishes 4(a) for $\widetilde{W}(\lambda, s)$. For $L(\lambda, s)$ we get

$$\begin{aligned}
L_\lambda(\lambda, s) &= \frac{\widetilde{W}_\lambda(\lambda, s)}{\widetilde{W}(\lambda, s)} = \frac{\mu - \lambda\widetilde{X}_e(s)}{\widetilde{X}(s)(\mu - \lambda)} \frac{\mu\widetilde{X}(s)}{\left(\mu - \lambda\widetilde{X}_e(s)\right)^2}\left(\widetilde{X}_e(s) - 1\right) = \frac{\mu}{\mu - \lambda}\frac{\widetilde{X}_e(s) - 1}{\mu - \lambda\widetilde{X}_e(s)}, \\
L_{\lambda\lambda}(\lambda, s) &= \mu\frac{\mu - \lambda\widetilde{X}_e(s) + \widetilde{X}_e(s)(\mu - \lambda)}{(\mu - \lambda)^2\left(\mu - \lambda\widetilde{X}_e(s)\right)^2}\left(\widetilde{X}_e(s) - 1\right).
\end{aligned}$$

As above both have the sign of s.

4(b) We have

$$L_{\lambda s}(\lambda, s) = \frac{\mu}{\mu - \lambda}\frac{\widetilde{X}_e'(s)\left(\mu - \lambda\widetilde{X}_e\right) + \lambda\left(\widetilde{X}_e(s) - 1\right)\widetilde{X}_e'(s)}{\left(\mu - \lambda\widetilde{X}_e(s)\right)^2} = \frac{\mu\widetilde{X}_e'(s)}{\left(\mu - \lambda\widetilde{X}_e(s)\right)^2}.$$

Noting that $\widetilde{X}_e''(s) > 0$ by Part 1, it follows that

$$L_{\lambda ss}(\lambda, s) = \frac{\mu\widetilde{X}_e''(s)}{\left(\mu - \lambda\widetilde{X}_e(s)\right)^2} + \frac{2\lambda\mu\left(\widetilde{X}_e'(s)\right)^2}{\left(\mu - \lambda\widetilde{X}_e(s)\right)^3} > 0.$$

■

Proof of Proposition 3.

Let $\lambda = \lambda_1 + \lambda_2$. Since customers' processing requirements are i.i.d., the distribution of W depends on the rates λ_i only through their sum λ. The demand system satisfies

$$\alpha \begin{cases} = \underline{v}_i(\lambda_i) - \frac{L(\lambda, r_i(c_i - \beta))}{r_i}, & \lambda_i > 0 \\ \geq \underline{v}_i(\lambda_i) - \frac{L(\lambda, r_i(c_i - \beta))}{r_i}, & \lambda_i = 0 \end{cases}, \tag{26}$$

which is the 2-type version of (4). If it is optimal to serve both types, then the optimal linear price function $\alpha^* - \beta^* W$ and the resulting demand vector $\boldsymbol{\lambda}^*$ must be a solution of the problem

$$\begin{aligned}
\max_{\alpha,\beta,\boldsymbol{\lambda}} \Pi\left(\alpha,\beta,\boldsymbol{\lambda}\right) &= \sum_{i=1,2} \lambda_i \cdot \left(\alpha - \beta E\left[W\left(\lambda\right)\right]\right) && (27)\\
& \quad s.t. \\
0 &\leq \lambda_i \leq \Lambda_i,\ i = 1,2, && (28)\\
\lambda_1 + \lambda_2 &< \mu \\
h_i\left(\alpha,\beta,\boldsymbol{\lambda}\right) &\triangleq \underline{v}_i\left(\lambda_i\right) - \alpha - \frac{L\left(\lambda, r_i\left(c_i - \beta\right)\right)}{r_i} = 0,\ i = 1,2. && (29)
\end{aligned}$$

We first show that the gradients of the constraint functions (29) are linearly independent. We have

$$\nabla h_i\left(\alpha,\beta,\boldsymbol{\lambda}\right) = \begin{bmatrix} \frac{\partial h_i(\alpha,\beta,\boldsymbol{\lambda})}{\partial \alpha} \\ \frac{\partial h_i(\alpha,\beta,\boldsymbol{\lambda})}{\partial \beta} \\ \frac{\partial h_i(\alpha,\beta,\boldsymbol{\lambda})}{\partial \lambda_1} \\ \frac{\partial h_i(\alpha,\beta,\boldsymbol{\lambda})}{\partial \lambda_2} \end{bmatrix} = \begin{bmatrix} -1 \\ L_s\left(\lambda, r_i\left(c_i - \beta\right)\right) \\ \underline{v}_i'\left(\lambda_i\right) 1_{\{i=1\}} - \frac{L_\lambda(\lambda, r_i(c_i-\beta))}{r_i} \\ \underline{v}_i'\left(\lambda_i\right) 1_{\{i=2\}} - \frac{L_\lambda(\lambda, r_i(c_i-\beta))}{r_i} \end{bmatrix}.$$

Let $s_i = r_i\left(c_i - \beta\right)$. $\nabla h_1\left(\alpha,\beta,\boldsymbol{\lambda}\right)$ and $\nabla h_2\left(\alpha,\beta,\boldsymbol{\lambda}\right)$ are linearly dependent if and only if

$$1 = \frac{L_s\left(\lambda, s_1\right)}{L_s\left(\lambda, s_2\right)} = \frac{\underline{v}_1'\left(\lambda_1\right) - \frac{L_\lambda(\lambda, s_1)}{r_1}}{-\frac{L_\lambda(\lambda, s_2)}{r_2}} = \frac{-\frac{L_\lambda(\lambda, s_1)}{r_1}}{\underline{v}_2'\left(\lambda_2\right) - \frac{L_\lambda(\lambda, s_2)}{r_2}}.$$

The first equation holds if and only if $s_1 = s_2$, since $L_s\left(\lambda, s\right)$ is strictly increasing in s by Part 2 of Lemma 1. Suppose that $s_1 = s_2$, and let $k = L_\lambda\left(\lambda, s_1\right) = L_\lambda\left(\lambda, s_2\right)$. If $k = 0$ (which holds iff $s_1 = s_2 = 0$), then the last ratio equals zero, which cannot hold. If $k \neq 0$ then the second-to-last and the last ratio equal one if and only if, respectively,

$$k\left(\frac{1}{r_1} - \frac{1}{r_2}\right) = \underline{v}_1'\left(\lambda_1\right) \text{ and } k\left(\frac{1}{r_1} - \frac{1}{r_2}\right) = -\underline{v}_2'\left(\lambda_2\right).$$

Since $\underline{v}_i'\left(\lambda_i\right) < 0$, these equations cannot both hold, and we conclude that $\nabla h_1\left(\alpha,\beta,\boldsymbol{\lambda}\right)$ and $\nabla h_2\left(\alpha,\beta,\boldsymbol{\lambda}\right)$ are linearly independent.

Since $h_1\left(\alpha,\beta,\boldsymbol{\lambda}\right) = h_2\left(\alpha,\beta,\boldsymbol{\lambda}\right) = 0$ for any solution that serves both types, and since $\nabla h_1\left(\alpha,\beta,\boldsymbol{\lambda}\right)$ and $\nabla h_2\left(\alpha,\beta,\boldsymbol{\lambda}\right)$ are linearly independent, the solution $\left(\alpha^*,\beta^*,\boldsymbol{\lambda}^*\right)$ must satisfy the Karush-Kuhn-Tucker (KKT) conditions. Suppose that $\left(\alpha^*,\beta^*,\boldsymbol{\lambda}^*\right)$ is a solution that serves some, but not all, customers of each type, so $\lambda_i^* \in \left(0, \Lambda_i\right)$. It cannot be optimal to operate at capacity, so $\lambda_1^* + \lambda_2^* < \mu$. Let γ_i be the Lagrange multiplier of the constraint $h_i\left(\alpha,\beta,\boldsymbol{\lambda}\right) = 0$. Suppressing the superscript $*$ and the arguments of Π and h_i, the KKT conditions are:

$$\begin{aligned}
\lambda = \frac{\partial \Pi}{\partial \alpha} &= \sum_{k=1,2} \gamma_k \frac{\partial h_k}{\partial \alpha} = -\left(\gamma_1 + \gamma_2\right), && (30)\\
-\lambda E\left[W\left(\lambda\right)\right] = \frac{\partial \Pi}{\partial \beta} &= \sum_{k=1,2} \gamma_k \frac{\partial h_k}{\partial \beta} = \gamma_1 L_s\left(\lambda, s_1\right) + \gamma_2 L_s\left(\lambda, s_2\right), && (31)\\
\frac{\partial \Pi}{\partial \lambda_i} &= \sum_{k=1,2} \gamma_k \frac{\partial h_k}{\partial \lambda_i} = \sum_{k=1,2} \gamma_k \left(\underline{v}_i'\left(\lambda_i\right) 1_{\{i=k\}} - \frac{L_\lambda\left(\lambda, s_k\right)}{r_k}\right), \quad i = 1,2, && (32)\\
h_i &= \underline{v}_i\left(\lambda_i\right) - \alpha - \frac{L\left(\lambda, r_i\left(c_i - \beta\right)\right)}{r_i} = 0,\ i = 1,2, && (33)
\end{aligned}$$

where

$$\frac{\partial \Pi}{\partial \lambda_1} = \frac{\partial \Pi}{\partial \lambda_2} = \alpha - \beta EW(\lambda) - \beta \lambda \frac{dE[W(\lambda)]}{d\lambda}.$$

It follows from (32) that

$$\sum_{k=1,2} \gamma_k \frac{\partial h_k}{\partial \lambda_1} = \sum_{k=1,2} \gamma_k \frac{\partial h_k}{\partial \lambda_2}, \text{ which implies } \gamma_1 \underline{v}_1'(\lambda_1) = \gamma_2 \underline{v}_2'(\lambda_2).$$

Together with (30), this implies that $\gamma_i < 0$ for $i = 1, 2$. Specifically,

$$\gamma_2 = -\frac{\lambda}{\underline{v}_2'(\lambda_2)/\underline{v}_1'(\lambda_1) + 1} < 0 \text{ and } \gamma_1 = -\frac{\lambda}{1 + \underline{v}_1'(\lambda_1)/\underline{v}_2'(\lambda_2)} < 0.$$

Next, multiply (30) by $E[W(\lambda)]$ and add it to (31) to obtain

$$\gamma_1 (L_s(\lambda, s_1) - E[W(\lambda)]) + \gamma_2 (L_s(\lambda, s_2) - E[W(\lambda)]) = 0. \tag{34}$$

Noting that $\gamma_i < 0$ for $i = 1, 2$, this holds if and only if one of the following conditions is satisfied:

$$L_s(\lambda, r_1(c_1 - \beta)) > E[W(\lambda)] > L_s(\lambda, r_2(c_2 - \beta))$$

or

$$L_s(\lambda, r_1(c_1 - \beta)) < E[W(\lambda)] < L_s(\lambda, r_2(c_2 - \beta))$$

or

$$L_s(\lambda, r_1(c_1 - \beta)) = E[W(\lambda)] = L_s(\lambda, r_2(c_2 - \beta)).$$

Part 1. We show that for $r_1, r_2 > 0$, (34) holds if and only if $c_2 < \beta < c_1$. If $\beta \geq c_1 > c_2$, we have by Part 2 of Lemma 1:

$$L_s(\lambda, r_1(c_1 - \beta)) \leq E[W(\lambda)] \text{ and } L_s(\lambda, r_2(c_2 - \beta)) < E[W(\lambda)].$$

Since $\gamma_i < 0$, it follows that the LHS of (34) is strictly positive; so that $\beta < c_1$. A similar argument shows that $\beta > c_2$. With $c_2 < \beta < c_1$, we have $L_s(\lambda, r_1(c_1 - \beta)) > 0 > L_s(\lambda, r_2(c_2 - \beta))$, so that $\underline{v}_2 < \alpha < \underline{v}_1$ follows from (29).

Part 2. If $r_j = 0 < r_i$, then $L_s(\lambda, r_j(c_j - \beta)) = E[W(\lambda)]$, and (34) reduces to

$$\gamma_j (L_s(\lambda, r_i(c_i - \beta)) - E[W(\lambda)]) = 0.$$

Since $\gamma_j < 0$, it holds if and only if $\beta = c_i$. The type-i demand constraint in (29) implies $\alpha = \underline{v}_i$.■

Proof of Proposition 4. The analysis in Section 3.3 implies that the first-best tariff set is IC if and only if $\boldsymbol{\lambda}^*$ satisfies (12) for $\beta_1 = c_1$ and $\beta_2 = c_2$. Noting that $L(\lambda, r_i(c_i - \beta_i))/r_i = 0$ if $\beta_i = c_i$, these conditions are equivalent to

$$-\frac{L(\lambda^*, r_2(c_2 - c_1))}{r_2} \leq \underline{v}_1(\lambda_1^*) - \underline{v}_2(\lambda_2^*) \leq \frac{L(\lambda^*, r_1(c_1 - c_2))}{r_1}, \tag{35}$$

where type-2 (type-1) chooses her targeted tariff if and only if the first (second) inequality holds. Furthermore, since $L(\lambda, s) < 0$ for $s < 0$ by Part 2 of Lemma 1, and since $L(\lambda, r(c_1 - c_2))/r$

is strictly increasing in $r \geq 0$ with $\lim_{r\to 0} L(\lambda, r(c_1 - c_2))/r = (c_1 - c_2) E[W(\lambda)]$ by Part 3 of Lemma 1, it follows that for any r_1 and r_2,

$$0 < -\frac{L(\lambda^*, r_2(c_2 - c_1))}{r_2} \leq (c_1 - c_2) E[W(\lambda^*)] \leq \frac{L(\lambda^*, r_1(c_1 - c_2))}{r_1}, \tag{36}$$

where the second inequality is strict if and only if $r_2 > 0$, and the last if and only if $r_1 > 0$.

We establish the claims by relating them to (35) via (36).

Part 1. Because $\alpha_i^* = \underline{v}_i(\lambda_i^*)$, if $\alpha_1^* \leq \alpha_2^*$, then $\underline{v}_1(\lambda_1^*) - \underline{v}_2(\lambda_2^*) \leq 0$. Then the first inequality in (36) implies that the first inequality in (35) cannot hold, so that IC is violated for any r_i.

Part 2. If $p_1^* = \alpha_1^* - c_1 E[W(\lambda^*)] = p_2^* = \alpha_2^* - c_2 E[W(\lambda^*)]$, then because $c_1 > c_2$, we have $\alpha_1^* > \alpha_2^*$ and $\underline{v}_1(\lambda_1^*) - \underline{v}_2(\lambda_2^*) = (c_1 - c_2) E[W(\lambda^*)]$. From (36) we have that (35) holds for all r_i.

Part 3. If $\alpha_1^* > \alpha_2^*$, then $\underline{v}_1(\lambda_1^*) - \underline{v}_2(\lambda_2^*) > 0$, and we have two cases with $p_1^* \neq p_2^*$.

(i) Suppose that $p_1^* > p_2^*$. Then $\underline{v}_1(\lambda_1^*) - \underline{v}_2(\lambda_2^*) > (c_1 - c_2) E[W(\lambda^*)]$. From (36) we have

$$0 < -\frac{L(\lambda^*, r_2(c_2 - c_1))}{r_2} \leq (c_1 - c_2) E[W(\lambda^*)] < \underline{v}_1(\lambda_1^*) - \underline{v}_2(\lambda_2^*), \tag{37}$$

so the first inequality in (35) holds and type-2 chooses her targeted tariff for all r_2. If type-1 is RN, she also prefers the type-2 tariff: we have $\lim_{r_1\to 0} L(\lambda^*, r_1(c_1 - c_2))/r_1 = (c_1 - c_2) E[W(\lambda^*)] < \underline{v}_1(\lambda_1^*) - \underline{v}_2(\lambda_2^*)$, where the equality follows from (36) and the inequality from (37), so the second inequality in (35) is violated. Because $L(\lambda^*, r_1(c_1 - c_2))/r_1$ strictly increases in r_1 by Part 3 of Lemma 1, and $\lim_{r_1\to\infty} L(\lambda, r_1(c_1 - c_2))/r_1 = \infty$ since $c_1 > c_2$, there is an unique $\underline{r} \in (0, \infty)$ such that

$$\underline{v}_1(\lambda_1^*) - \underline{v}_2(\lambda_2^*) = \left.\frac{L(\lambda^*, r_1(c_1 - c_2))}{r_1}\right|_{r_1 = \underline{r}}$$

and the second inequality in (35) holds if and only if $r_1 \geq \underline{r}$.

(ii) Suppose that $p_1^* < p_2^*$. Then $\underline{v}_1(\lambda_1^*) - \underline{v}_2(\lambda_2^*) < (c_1 - c_2) E[W(\lambda^*)]$. From (36) we have

$$0 < \underline{v}_1(\lambda_1^*) - \underline{v}_2(\lambda_2^*) < (c_1 - c_2) E[W(\lambda^*)] \leq \frac{L(\lambda^*, r_1(c_1 - c_2))}{r_1}, \tag{38}$$

where $\alpha_1^* > \alpha_2^*$ implies the first inequality. It follows that the second inequality in (35) holds and type-1 chooses her targeted tariff for all r_1. If type-2 is RN, she also prefers the type-1 tariff: we have $\lim_{r_2\to 0} -L(\lambda^*, r_2(c_2 - c_1))/r_2 = (c_1 - c_2) E[W(\lambda^*)] > \underline{v}_1(\lambda_1^*) - \underline{v}_2(\lambda_2^*)$, where the equality follows from (36) and the inequality from (38), so that the first inequality in (35) is violated. Since $-L(\lambda^*, r_2(c_2 - c_1))/r_2$ strictly decreases in r_2 by Part 3 of Lemma 1, and $\lim_{r_2\to\infty} L(\lambda, r_2(c_2 - c_1))/r_2 = 0$ because $c_2 < c_1$, there is an unique $\underline{r} \in (0, \infty)$ such that

$$\underline{v}_1(\lambda_1^*) - \underline{v}_2(\lambda_2^*) = -\left.\frac{L(\lambda^*, r_2(c_2 - c_1))}{r_2}\right|_{r_2 = \underline{r}}$$

and the first inequality in (35) holds if and only if $r_2 \geq \underline{r}$.■

Sufficient conditions for Assumptions A2 and A3.

A2. Two mild conditions are sufficient for $\Pi(\lambda)$ and $\Pi^f(\lambda; r)$ to be strictly concave in λ.

(i) The gross revenue function $\lambda\underline{v}(\lambda)$ is strictly concave. Many common distributions satisfy this assumption, e.g., the uniform, normal, logistic, Laplace and power function distributions, and the gamma and Weibull distributions with shape parameter ≥ 1.

(ii) The function $\lambda L(\lambda, s)$ is convex in λ for $s \geq 0$, i.e., $2L_\lambda(\lambda, s) + \lambda L_{\lambda\lambda}(\lambda, s) \geq 0$. This assumption also implies that $\lambda E[W(\lambda)]$ is convex in λ since $L(\lambda, r)/r$ converges uniformly to $E[W(\lambda)]$ as $r \to 0$. This assumption seems reasonable for many queueing models. A sufficient condition is that $W(\lambda)$ is stochastically increasing in λ and that its c.d.f. is concave in λ; for $G/G/1$ queues both properties follow from a similar argument as in Weber (1983).

A3. $L_\lambda(\lambda, s)/s$ increases in s, i.e., $sL_{\lambda s}(\lambda, s) - L_\lambda(\lambda, s) \geq 0$. By Lemma 1, this holds for the M/G/1 queue: $L_\lambda(\lambda, 0) = 0$ by Part 2 of Lemma 1 and $L_\lambda(\lambda, s)$ is strictly convex in s by Part 4(b).

Proof of Proposition 5. *Part 1.* That $\Pi^f(r) < \Pi^*$ for $r > 0$ follows from the uniqueness of the optimal price function by Part 1 of Proposition 1. We next show that $\lambda^f(r) < \lambda^*$ for $r > 0$. We first establish that $\Pi'(\lambda) > \Pi_\lambda^f(\lambda; r)$. From (14) and (15)

$$\Pi'(\lambda) = \underline{v}(\lambda) + \lambda \underline{v}'(\lambda) - c\left(E[W(\lambda)] + \lambda \frac{dE[W(\lambda)]}{d\lambda}\right), \tag{39}$$

$$\Pi_\lambda^f(\lambda; r) = \underline{v}(\lambda) + \lambda \underline{v}'(\lambda) - \frac{L(\lambda, rc) + \lambda L_\lambda(\lambda, rc)}{r}. \tag{40}$$

The difference satisfies

$$\Pi'(\lambda) - \Pi_\lambda^f(\lambda; r) = \left(\frac{L(\lambda, rc)}{r} - cE[W(\lambda)]\right) + \lambda\left(\frac{L_\lambda(\lambda, rc)}{r} - c\frac{dE[W(\lambda)]}{d\lambda}\right) > 0, \tag{41}$$

because the first bracket is positive by Part 3 of Lemma 1, and the second is non-negative as we show next. *A3* implies that $L_\lambda(\lambda, rc)/r$ increases in r. Furthermore,

$$\lim_{r \to 0} \frac{L_\lambda(\lambda, rc)}{r} = \frac{d}{d\lambda} \lim_{r \to 0} \frac{L(\lambda, rc)}{r} = c\frac{dE[W(\lambda)]}{d\lambda}, \tag{42}$$

where the first equality holds since $L_\lambda(\lambda, 0) = 0$ (Part 2 of Lemma 1) and from l'Hôpital's rule, and the second by Part 3 of Lemma 1. Using (41) we have $\Pi_\lambda^f(\lambda^*; r) < \Pi'(\lambda^*) = 0$ where the equality follows by *A1*. Finally, because $\Pi(\lambda)$ is strictly concave by *A2*, the optimal $\Pi^f(\lambda; r)$ is obtained for $\lambda < \lambda^*$.

Since Part 3 of Lemma 1 and (42) imply that $\lim_{r\to 0} \Pi_\lambda^f(\lambda; r) = \Pi'(\lambda)$ for all λ, $\lim_{r\to 0} \lambda^f(r) = \lambda^*$ and $\lim_{r\to 0} \Pi^f(r) = \Pi^*$. That $\lim_{r\to 0} \alpha^f(r) < \alpha^*$ follows from (16) because

$$\lim_{r \to 0} \alpha^f(r) = \lim_{r \to 0}\left(\underline{v}\left(\lambda^f(r)\right) - \frac{L\left(\lambda^f(r), rc\right)}{r}\right) = \underline{v}(\lambda^*) - cE[W(\lambda^*)] < \alpha^* = \underline{v}(\lambda^*).$$

Part 2. First note that since $\lambda^f(r) < \lambda^*$, if $\lambda^f(r) > 0$ it must satisfy $\Pi_\lambda^f\left(\lambda^f(r); r\right) = 0$.

Next, since $\Pi^f(\lambda; r)$ is strictly concave in λ for fixed r (by *A2*) and since $\Pi_{\lambda r}^f(\lambda; r) < 0$ as shown in (43), it follows that if $\lambda^f(r) > 0$ then $\lambda^f(r)$ and $\Pi^f(r)$ are strictly decreasing at r, and that if $\lambda^f(r^\circ) = 0$ for some r° then $\lambda^f(r) = \Pi^f(r) = 0$ for all $r \geq r^\circ$. From (40) we have

$$\begin{aligned}\Pi_{\lambda r}^f(\lambda; r) &= -\frac{cr\left(L_s(\lambda, rc) + \lambda L_{\lambda s}(\lambda, rc)\right) - \left(L(\lambda, rc) + \lambda L_\lambda(\lambda, rc)\right)}{r^2} \\ &= -\frac{\left(crL_s(\lambda, rc) - L(\lambda, rc)\right) + \lambda\left(crL_{\lambda s}(\lambda, rc) - L_\lambda(\lambda, rc)\right)}{r^2} < 0, \end{aligned} \tag{43}$$

where the first bracket in the numerator is positive since $L(\lambda, rc)/r$ is strictly increasing in $r \geq 0$ by Part 3 of Lemma 1, and the second is non-negative by *A3*.

It remains to establish the threshold $\overline{r}$ in (17). Since $\Pi^f(\lambda;r)$ is strictly concave in λ for fixed r, the arrival rate $\lambda^f(r)>0$ if and only if $\Pi^f_\lambda(0;r)>0$. Note that $\lim_{r\to 0}\Pi^f_\lambda(0;r)=\Pi'(0)>0$ (the equality follows from Part 1, the inequality from *A1*), and that from (40)

$$\lim_{r\to\infty}\Pi^f_\lambda(0;r)=\underline{v}(0)-\lim_{r\to\infty}\frac{L(0,rc)}{r}=\underline{v}(0)-c\lim_{r\to\infty}L_s(0,rc)=-\infty.$$

Since $\Pi^f_{\lambda r}(0;r)<0$, there is an unique risk aversion parameter $\overline{r}\in(0,\infty)$ such that $\Pi^f_\lambda(0,\overline{r})=0$, or equivalently, $L(0,\overline{r}c)/\overline{r}=\underline{v}(0)$, and $\Pi^f_\lambda(0;r)>0$ if and only if $r<\overline{r}$.

Part 3. See Figure 2 for Example 1 which shows that $\alpha^f(r)>0$ and increases in $r\in[0,2]$. That $\alpha^f(r)$ is nonmonotone in r follows from (16) because

$$\lim_{r\to\overline{r}}\alpha^f(r)=\lim_{r\to\overline{r}}\left(\underline{v}\left(\lambda^f(r)\right)-\frac{L\left(\lambda^f(r),rc\right)}{r}\right)=\underline{v}(0)-\frac{L(0,\overline{r}c)}{\overline{r}}=0.$$

The second equality holds since $\liminf_{r\to\overline{r}}\lambda^f(r)=0$ by Part 2, and the third by (17).■

Proof of Proposition 6. Let $\Pi^f(\lambda,\mu;r)$ be the revenue function under flat rate pricing. With RN customers, flat rate and optimal lead time-dependent pricing yield the same revenue, so $\Pi^f(\lambda,\mu;0)$ is the revenue function under optimal lead time-dependent pricing. In the proof we first derive properties of $\Pi^f(\lambda,\mu;r)$ for $r>0$ and then take limits as $r\to 0$.

Let the r.v. $W(\lambda,\mu)$ denote the steady-state lead time, $\widetilde{W}(\lambda,\mu,s):=E[\exp(sW(\lambda,\mu))]$ its MGF, and $L(\lambda,\mu,s):=\ln\widetilde{W}(\lambda,\mu,s)$ its semi-invariant MGF. We have from (14) and (15) that

$$\begin{aligned}\Pi^f(\lambda,\mu;0) &= \lambda\left(\underline{v}(\lambda)-cE[W(\lambda,\mu)]\right), &(44)\\ \Pi^f(\lambda,\mu;r) &= \lambda\left(\underline{v}(\lambda)-\frac{1}{r}L(\lambda,\mu,rc)\right), &(45)\\ \Pi^f_\lambda(\lambda,\mu;r) &= \underline{v}(\lambda)+\lambda\underline{v}'(\lambda)-\frac{L(\lambda,\mu,rc)+\lambda L_\lambda(\lambda,\mu,rc)}{r}, &(46)\\ \Pi^f_\mu(\lambda,\mu;r) &= -\lambda\frac{L_\mu(\lambda,\mu,rc)}{r}. &(47)\end{aligned}$$

Let $\lambda^{f*}(\mu;r):=\arg\max_\lambda\Pi^f(\lambda,\mu;r)$, so $\Pi^{f*}(\mu)$=$\Pi^f\left(\lambda^{f*}(\mu;r),\mu;r\right)$ and $\Pi^*(\mu)$=$\Pi^f\left(\lambda^{f*}(\mu;0),\mu;0\right)$.
Parts 1(a)/2(a). We have $\lim\limits_{\mu\to\infty}\Pi^{f*}(\mu)=\lim\limits_{\mu\to\infty}\Pi^*(\mu)=\max_{\lambda\in[0,\Lambda]}\lambda\underline{v}(\lambda)$ since $\lim\limits_{\mu\to\infty}L(\lambda,\mu,rc)=0$ and so $\lim\limits_{\mu\to\infty}\Pi^f(\lambda,\mu;r)=\lambda\underline{v}(\lambda)$ for all λ,r. We have that

$$L_\lambda(\lambda,\mu,rc)\geq 0\geq L_\mu(\lambda,\mu,rc)\ \text{ for }\lambda<\mu,\ rc\geq 0, \tag{48}$$

because $W(\lambda,\mu)$ is stochastically increasing in λ (decreasing in μ), and therefore the expectation of every increasing function of $W(\lambda,\mu)$, such as the MGF with a positive argument ($rc\geq 0$), is increasing in λ (decreasing in μ).

It follows that the bracketed term in (45) decreases in λ and increases in μ, so $\Pi^{f*}(\mu)=0$ for

$$\mu\leq\underline{\mu}^f:=\arg\left\{\mu\geq 0:\underline{v}(0)-\frac{1}{r}L(0,\mu,rc)=0\right\}. \tag{49}$$

(Note that $L(0,\mu,rc)$ is the semi-invariant MGF of the service time, evaluated at rc.) Similarly, we have from (44) that $\Pi^*(\mu)=0$ for

$$\mu\leq\underline{\mu}:=\arg\left\{\mu\geq 0:\underline{v}(0)-cE[W(0,\mu)]=0\right\}=c/\underline{v}(0), \tag{50}$$

where the last equality holds because $E\left[W\left(0,\mu\right)\right]=1/\mu$. Note that $\underline{\mu}<\underline{\mu}^{f}$ because $cE\left[W\left(0,\mu\right)\right]<\frac{1}{r}L\left(0,\mu,rc\right)$ for $r>0$ by Part 3 of Lemma 1.

For $\mu>\underline{\mu}^{f}$ we have $\lambda^{f*}\left(\mu;r\right)>0$, and furthermore

$$\frac{d\Pi^{f}\left(\lambda^{f*}\left(\mu;r\right),\mu;r\right)}{d\mu}=\Pi_{\mu}\left(\lambda^{f*}\left(\mu;r\right),\mu;r\right)=-\lambda^{f*}\left(\mu;r\right)\frac{L_{\mu}\left(\lambda^{f*}\left(\mu;r\right),\mu,rc\right)}{r}\geq 0. \tag{51}$$

The first equality in (51) holds since $\Pi_{\lambda}^{f}\left(\lambda^{*}\left(\mu;r\right),\mu;r\right)=0$ by *A1,* and $\left|d\lambda^{f*}\left(\mu;r\right)/d\mu\right|<\infty$ since $\Pi_{\lambda\lambda}^{f}\left(\lambda,\mu;r\right)<0$ by *A2,* the second equality holds by (47), and the inequality follows from (48).

Furthermore, we have

$$\lim_{\mu\to\underline{\mu}^{f}}\Pi^{f*\prime}\left(\mu\right)=0=\lim_{\mu\to\infty}\Pi^{f*\prime}\left(\mu\right). \tag{52}$$

The first equality holds by (51) since $\lambda^{f*}(\underline{\mu}^{f};r)=0$, and $L(0,\underline{\mu}^{f},rc)=\underline{v}\left(0\right)<\infty$ implies that $\left|L_{\mu}(0,\underline{\mu}^{f},rc)\right|<\infty$. The second equality holds since $\Pi^{f*}\left(\mu\right)$ is increasing with $\lim_{\mu\to\infty}\Pi^{f*}\left(\mu\right)=\max_{\lambda\in[0,\Lambda]}\lambda\underline{v}\left(\lambda\right)$.

A similar argument (as $r\to 0$) shows $\Pi^{*\prime}\left(\mu\right)\geq 0$ for $\mu>\underline{\mu}$ and $\lim_{\mu\to\underline{\mu}}\Pi^{*\prime}\left(\mu\right)=0=\lim_{\mu\to\infty}\Pi^{*\prime}\left(\mu\right)$.

Parts 1(b)/2(b) We omit the straightforward proof that $0<\overline{b}^{f}<\overline{b}<\infty$. That $\mu^{*}\left(b\right)>\underline{\mu}$ for $b<\overline{b}$ ($\mu^{f*}\left(b\right)>\underline{\mu}^{f}$ for $b<\overline{b}^{f}$) is immediate from 1(a) (from 2(a)). That $\mu^{*}\left(b\right)$ and $\mu^{f*}\left(b\right)$ decrease in b follows since the profit functions $\left(\Pi^{*}\left(\mu\right)-b\mu\right)$ and $\left(\Pi^{f*}\left(\mu\right)-b\mu\right)$ are submodular in $\left(b,\mu\right)$.■

References

Gallager, R. G. 1996. *Discrete Stochastic Processes.* Kluwer, Boston, MA.

Weber, R.R. 1983. A note on waiting times in single-server queues. *Oper. Res.* **31** 950-951.