Online Appendix

Appendix A - Definitions of Stochastic Orders Used in the Paper

Suppose that X and Y are two non-negative continuous random variables having the same support $\mathcal{S}$ with corresponding cumulative distribution functions $F_X(\cdot)$ and $F_Y(\cdot)$, probability density functions $f_X(\cdot)$ and $f_Y(\cdot)$, and hazard rate functions $r_X(x) = f_X(x)/(1-F_X(x))$ and $r_Y(x) = f_Y(x)/(1-F_Y(x))$, respectively. The following definitions are based on Müller and Stoyan (2002) and Shaked and Shanthikumar (2007).

Definition 1 Suppose that $F_X(x) \leq F_Y(x)$ for all $x \in \mathcal{S}$. Then, we say that F_X is greater than F_Y in the usual stochastic ordering (denoted by $F_X \geq_{st} F_Y$).

Definition 2 If $r_X(x) \leq r_Y(x)$ for all $x \in \mathcal{S}$, then we say that F_X is greater than F_Y in hazard rate ordering (denoted by $F_X \geq_{hr} F_Y$).

Definition 3 Suppose that the following condition holds:

$$P\{X \in A\}P\{Y \in B\} \leq P\{X \in B\}P\{Y \in A\} \tag{17}$$

for all measurable sets A and B in $\mathcal{S}$ such that $A \leq B$, where $A \leq B$ means that $x \in A$ and $y \in B$ implies that $x \leq y$. Then, we say that F_X is greater than F_Y in likelihood ratio ordering (denoted by $F_X \geq_{lr} F_Y$). Condition (17) is equivalent to $\frac{f_X(x)}{f_Y(x)}$ being increasing over $\mathcal{S}$.

Definition 4 Suppose that X and Y both have finite means. If $\mathbb{E}[g(X)] \geq \mathbb{E}[g(Y)]$ for all real convex functions $g(\cdot)$ such that the expectations exist, then, we say that F_X is greater than F_Y in convex ordering (denoted by $F_X \geq_{cx} F_Y$).

Appendix B - Proofs of the Results

Proof of Theorem 1: Consider a policy $\pi \in \Pi_N$ that does not satisfy the conditions of policies described in Theorem 1. Then, there exists an interval $[\ell, u] \subset [0,1]$ and a real number s with $\ell < s < u$ for which π classifies a customer with probability z of being type 1 as class m if $\ell \leq z < s$ and as class n if $s \leq z \leq u$ where $m < n$. We will show that there exists a policy $\tilde{\pi} \in \Pi_N$, which is at least as good as π and which follows π except that customers with a probability z of being type 1 are classified as class m if $\tilde{s} \leq z \leq u$ and as class n if $\ell \leq z < \tilde{s}$ for some $\tilde{s}$ such that $\ell < \tilde{s} < u$. (This will immediately imply the result, since the following can be repeated for any interval such as $[\ell, u]$.)

Suppose that we pick $\tilde{s}$ so that $\Gamma_{n,\tilde{\pi}} = \Gamma_{n,\pi}$, i.e.,

$$\int_{\ell}^{\tilde{s}} (xa_1 + (1-x)a_2)b(x)dx = \int_{s}^{u} (xa_1 + (1-x)a_2)b(x)dx,$$

which can also be written as

$$a_1 \left(\int_{\ell}^{\min(s,\tilde{s})} xb(x)dx - \int_{\max(s,\tilde{s})}^{u} xb(x)dx \right) = -a_2 \left(\int_{\ell}^{\min(s,\tilde{s})} (1-x)b(x)dx - \int_{\max(s,\tilde{s})}^{u} (1-x)b(x)dx \right). \tag{18}$$

(Note that it is always possible to find such an $\tilde{s}$ since $\Gamma_{n,\tilde{\pi}}$ is continuously increasing in $\tilde{s}$ and takes values in $[0, \int_{\ell}^{u}(xa_1 + (1-x)a_2)b(x)dx]$ for $\ell \leq \tilde{s} \leq u$, and $\Gamma_{n,\pi} \in [0, \int_{\ell}^{u}(xa_1 + (1-x)a_2)b(x)dx]$.) The fact that $\Gamma_{n,\pi} = \Gamma_{n,\tilde{\pi}}$ also implies that $\Gamma_{m,\pi} = \Gamma_{m,\tilde{\pi}}$ since $\Gamma_{n,\pi} + \Gamma_{m,\pi} = \Gamma_{n,\tilde{\pi}} + \Gamma_{m,\tilde{\pi}}$. Then, by the construction of policy $\tilde{\pi}$, we have $\Gamma_{j,\tilde{\pi}} = \Gamma_{j,\pi}$ for all $j = 1, 2, \ldots, N$, and hence $W_{j,\tilde{\pi}} = W_{j,\pi}$ for all $j = 1, 2, \ldots, N$, using (1). As we move from policy π to policy $\tilde{\pi}$, the only change would be that some of the customers with probabilities in the interval $[\ell, u]$ will experience different costs of waiting. We will show that $C_\pi \geq C_{\tilde{\pi}}$.

From (2), we have

$$\begin{aligned} C_\pi - C_{\tilde{\pi}} &= \lambda \Bigg(W_m \int_{\ell}^{s} (xh_1 + (1-x)h_2)b(x)dx + W_n \int_{s}^{u} (xh_1 + (1-x)h_2)b(x)dx \\ &\quad - W_m \int_{\tilde{s}}^{u} (xh_1 + (1-x)h_2)b(x)dx - W_n \int_{\ell}^{\tilde{s}} (xh_1 + (1-x)h_2)b(x)dx \Bigg), \end{aligned}$$

where $W_j = W_{j,\pi} = W_{j,\tilde{\pi}}$ for $j = m, n$. After a few algebraic manipulations, we find

$$\begin{aligned} C_\pi - C_{\tilde{\pi}} &= \lambda (W_m - W_n) \Bigg(h_1 \left(\int_{\ell}^{\min(s,\tilde{s})} xb(x)dx - \int_{\max(s,\tilde{s})}^{u} xb(x)dx \right) \\ &\quad + h_2 \left(\int_{\ell}^{\min(s,\tilde{s})} (1-x)b(x)dx - \int_{\max(s,\tilde{s})}^{u} (1-x)b(x)dx \right) \Bigg). \end{aligned}$$

Then, using (18), this difference further simplifies to

$$C_\pi - C_{\tilde{\pi}} = \lambda(W_m - W_n)a_1\left(\frac{h_1}{a_1} - \frac{h_2}{a_2}\right)\left(\int_\ell^{\min(s,\tilde{s})} xb(x)dx - \int_{\max(s,\tilde{s})}^{u} xb(x)dx\right).$$

From (1) we know that $W_m \leq W_n$ since $m < n$, and by our modeling assumption, we have $h_1/a_1 > h_2/a_2$. Hence, it is sufficient to show that $\int_\ell^{\min(s,\tilde{s})} xb(x)dx > \int_{\max(s,\tilde{s})}^{u} xb(x)dx$. Suppose for contradiction that

$$\int_\ell^{\min(s,\tilde{s})} xb(x)dx - \int_{\max(s,\tilde{s})}^{u} xb(x)dx > 0. \tag{19}$$

This implies that

$$\min(s,\tilde{s})(B(\min(s,\tilde{s})) - B(\ell)) - \max(s,\tilde{s})(B(u) - B(\max(s,\tilde{s}))) > 0$$

from which it follows that

$$B(u) - B(\min(s,\tilde{s})) < B(\max(s,\tilde{s})) - B(\ell). \tag{20}$$

Now, from (18) and (19), we have

$$\int_{\max(s,\tilde{s})}^{u} (1-x)b(x)dx - \int_\ell^{\min(s,\tilde{s})} (1-x)b(x)dx > 0.$$

This implies that

$$(1-\max(s,\tilde{s}))(B(u) - B(\max(s,\tilde{s}))) - (1-\min(s,\tilde{s}))(B(\min(s,\tilde{s})) - B(\ell)) > 0$$

from which it follows that

$$B(u) - B(\min(s,\tilde{s})) > B(\max(s,\tilde{s})) - B(\ell),$$

which is a contradiction to (20).

Hence, we conclude that $C_\pi \geq C_{\tilde{\pi}}$, which means that $\tilde{\pi}$ performs at least as well as π. □

Proof of Theorem 2: Consider an N-class policy $\pi \in \tilde{\Pi}_N$ with corresponding threshold values $t_0, t_1, \ldots, t_{N-1}, t_N$ such that $1 = t_0 > t_1 > t_2 > \cdots > t_{N-1} > t_N = 0$. Now, let $m \in \{1, 2, \ldots, N\}$ be any priority class and suppose that we obtain an $(N+1)$-class policy $\hat{\pi} \in \tilde{\Pi}_{N+1}$ from π by partitioning the signal interval corresponding to the mth priority class of π, i.e., $[t_m, t_{m-1})$, into two subintervals $[t_m, \bar{t})$ and $[\bar{t}, t_{m-1})$ for some arbitrarily picked $\bar{t} \in (t_m, t_{m-1})$ such that customers whose signals fall into interval $[\bar{t}, t_{m-1})$ (class m of $\hat{\pi}$)

have a higher priority than those with signals in $[t_m, \bar{t})$ (class $m+1$ of $\hat{\pi}$) while all other priority relations with other classes remain the same. Note that if $m > 1$, then customers who belong to any class $j < m$ according to policy π also belong to class j under policy $\hat{\pi}$, i.e., $I_{1,\hat{\pi}} = I_{1,\pi} = [t_1, 1]$ and $I_{j,\hat{\pi}} = I_{j,\pi} = [t_j, t_{j-1})$, for $j = 2, \ldots, m$. Customers who belong to class m under policy π either belong to class m or class $m+1$ under policy $\hat{\pi}$. More specifically, $I_{m,\hat{\pi}} = [\bar{t}, t_{m-1})$ and $I_{m+1,\hat{\pi}} = [t_m, \bar{t})$ so that $I_{m,\hat{\pi}} \cup I_{m+1,\hat{\pi}} = I_{m,\pi}$. Finally, if $m < N$, then customers who belong to class $j > m$ under policy π belong to class $j+1$ under policy $\hat{\pi}$, i.e., $I_{j+1,\hat{\pi}} = I_{j,\pi} = [t_j, t_{j-1})$, for $j = m+1, \ldots, N$.

Using (1), we have

$$W_{j,\pi} = \frac{\lambda\left(p_1 e_1 + p_2 e_2\right)}{2\left(1 - \lambda \sum_{k=1}^{j-1} \Gamma_{k,\pi}\right)\left(1 - \lambda \sum_{k=1}^{j} \Gamma_{k,\pi}\right)} \text{ for } j = 1, 2, \ldots, N, \tag{21}$$

where $\Gamma_{k,\pi} = \int_{t_k}^{t_{k-1}} (xa_1 + (1-x)a_2) b(x) dx$. Similarly, we have

$$W_{j,\hat{\pi}} = \frac{\lambda\left(p_1 e_1 + p_2 e_2\right)}{2\left(1 - \lambda \sum_{k=1}^{j-1} \Gamma_{k,\hat{\pi}}\right)\left(1 - \lambda \sum_{k=1}^{j} \Gamma_{k,\hat{\pi}}\right)} \text{ for } j = 1, 2, \ldots, N+1, \tag{22}$$

where $\Gamma_{k,\hat{\pi}} = \Gamma_{k,\pi}$, for $k = 1, 2, \ldots, m-1$, $\Gamma_{m,\hat{\pi}} = \int_{\bar{t}}^{t_{m-1}} (xa_1 + (1-x)a_2) b(x) dx$, $\Gamma_{m+1,\hat{\pi}} = \int_{t_m}^{\bar{t}} (xa_1 + (1-x)a_2) b(x) dx$, $\Gamma_{k,\hat{\pi}} = \Gamma_{k-1,\pi}$, for $k = m+2, \ldots, N+1$. We also have $\Gamma_{m,\pi} = \Gamma_{m,\hat{\pi}} + \Gamma_{m+1,\hat{\pi}}$. Then, one can verify that $W_{j,\hat{\pi}} = W_{j,\pi}$, for $j = 1, 2, \ldots, m-1$, and $W_{j,\hat{\pi}} = W_{j-1,\pi}$, for $j = m+2, \ldots, N+1$. In other words, for $j \in \{1, 2, \ldots, m-1\}$, class j customers under policy π are the same customers as class j customers under policy $\hat{\pi}$ and they experience the same expected waiting times while for $j \in \{m+1, m+2, \ldots, N\}$, class j customers under policy π are the same as class $j+1$ customers under policy $\hat{\pi}$ and they experience the same expected waiting times. Therefore the only difference between the two policies is in costs incurred by class m customers under policy π and class m and class $m+1$ customers under policy $\hat{\pi}$. This gives us

$$\begin{aligned} C_\pi - C_{\hat{\pi}} \;=\; \lambda \Bigg[& W_{m,\pi} \int_{t_m}^{t_{m-1}} (h_1 x + h_2(1-x)) b(x) dx \\ & - W_{m,\hat{\pi}} \int_{\bar{t}}^{t_{m-1}} (h_1 x + h_2(1-x)) b(x) dx - W_{m+1,\hat{\pi}} \int_{t_m}^{\bar{t}} (h_1 x + h_2(1-x)) b(x) dx \Bigg], \end{aligned}$$

which we can also write as

$$\begin{aligned} C_\pi - C_{\hat{\pi}} \;=\; \lambda \Bigg[& \int_{\bar{t}}^{t_{m-1}} (W_{m,\pi} - W_{m,\hat{\pi}})(h_1 x + h_2(1-x)) b(x) dx \\ & + \int_{t_m}^{\bar{t}} (W_{m,\pi} - W_{m+1,\hat{\pi}})(h_1 y + h_2(1-y)) b(y) dy \Bigg]. \end{aligned} \tag{23}$$

Now, from (21) and (22), we can show that

$$W_{m,\pi} - W_{m,\hat{\pi}} = \frac{\lambda \Gamma_{m+1,\hat{\pi}}}{1 - \lambda \sum_{k=1}^{m} \Gamma_{k,\pi}} W_{m,\hat{\pi}} \tag{24}$$

and

$$W_{m,\pi} - W_{m+1,\hat{\pi}} = -\frac{\lambda \Gamma_{m,\hat{\pi}}}{1 - \lambda \sum_{k=1}^{m} \Gamma_{k,\pi}} W_{m,\hat{\pi}}. \tag{25}$$

Then, plugging (24) and (25) into (23), we find

$$\begin{aligned}
C_\pi - C_{\hat{\pi}} &= \frac{\lambda^2 W_{m,\hat{\pi}}}{1 - \lambda \sum_{k=1}^{m} \Gamma_{k,\pi}} \left[\int_{\bar{t}}^{t_{m-1}} \Gamma_{m+1,\hat{\pi}} (h_1 x + h_2(1-x)) b(x) dx \right. \\
&\quad \left. - \int_{t_m}^{\bar{t}} \Gamma_{m,\hat{\pi}} (h_1 y + h_2(1-y)) b(y) dy \right] \\
&= \frac{\lambda^2 W_{m,\hat{\pi}}}{1 - \lambda \sum_{k=1}^{m} \Gamma_{k,\pi}} \left[\int_{\bar{t}}^{t_{m-1}} \int_{t_m}^{\bar{t}} (y a_1 + (1-y) a_2)(h_1 x + h_2(1-x)) b(y) b(x) dy dx \right. \\
&\quad \left. - \int_{t_m}^{\bar{t}} \int_{\bar{t}}^{t_{m-1}} (x a_1 + (1-x) a_2)(h_1 y + h_2(1-y)) b(y) b(x) dx dy \right] \\
&= \frac{\lambda^2 W_{m,\hat{\pi}}}{1 - \lambda \sum_{k=1}^{m} \Gamma_{k,\pi}} \left[\int_{\bar{t}}^{t_{m-1}} \int_{t_m}^{\bar{t}} (y a_1 + (1-y) a_2)(h_1 x + h_2(1-x)) b(y) b(x) dy dx \right. \\
&\quad \left. - \int_{\bar{t}}^{t_{m-1}} \int_{t_m}^{\bar{t}} (x a_1 + (1-x) a_2)(h_1 y + h_2(1-y)) b(y) b(x) dy dx \right],
\end{aligned}$$

which then further simplifies to

$$C_\pi - C_{\hat{\pi}} = \frac{\lambda^2 W_{m,\hat{\pi}}}{1 - \lambda \sum_{k=1}^{m} \Gamma_{k,\pi}} (h_1 a_2 - h_2 a_1) \int_{\bar{t}}^{t_{m-1}} \int_{t_m}^{\bar{t}} (x - y) b(y) b(x) dy dx.$$

We can now conclude that $C_\pi - C_{\hat{\pi}} \geq 0$ since $h_1 a_2 - h_2 a_1 > 0$ and the integrand is non-negative for all x and y values within the integral limits. Hence, the result follows. □

Proof of Corollary 1: Let π be any N-class policy. Then, we know from Theorem 1 that there exists a policy $\tilde{\pi} \in \tilde{\Pi}_N$ under which the long-run average cost is at most the same. From Theorem 2, we know that there exists an $(N+1)$-class policy $\hat{\pi}$ under which the long-run average cost is at most the same as that under $\tilde{\pi}$. Hence, the result follows. □

Proof of Corollary 2: The result immediately follows from Theorem 2. □

Proof of Theorem 3: Let $\hat{\pi}$ be an arbitrary policy in Π_N with $N < \infty$. Then, we know from Theorem 1 that there exists a policy $\pi_N \in \tilde{\Pi}_N$ for which $C_{\hat{\pi}} \geq C_{\pi_N}$. Now, consider a sequence of policies π_n, $n = N, N+1, \ldots$, where π_{n+1} is obtained from an n class policy

π_n by dividing the longest interval (corresponding to a class) into two subintervals of equal length and increasing the total number of classes to $n+1$. (When determining the longest interval, ties can be broken arbitrarily.) It then follows from Theorem 2 that the long-run average cost under π_{n+1} is at most as large as the long-run average cost under π_n for $n \geq N$. Thus, as n increases, the performance of π_n can only improve. We will show that as n goes to infinity, π_n converges to the HSF policy. Suppose that as n goes to infinity, π_n converges to a policy that is not HSF, i.e., a policy under which there exists at least two signals s_1 and s_2 with $s_1 > s_2$ and customers with signal s_2 have higher priority than customers with signal s_1. But this is a contradiction to the construction of the policies π_n as n goes to infinity. To see this, note that there exists an integer K such that for $n \geq K$, all the signal intervals of policy π_n are at most of length $s_1 - s_2$. Hence, a customer with signal s_1 must have a (strictly) higher priority than a customer with signal s_2 under π_n where $n \geq K$. Thus, starting from π_N, it is not possible to converge to a policy that is not HSF. □

Proof of Proposition 1: In this proof, for any function $q(\cdot)$, we use $q'(\cdot)$ to denote the first derivative of $q(\cdot)$. From (3), (4), and (5), it can be shown that

$$C(t) = \frac{\lambda^2(p_1e_1 + p_2e_2)}{2(1-\rho)}\eta(t), \tag{26}$$

where

$$\eta(t) = \frac{h_1(p_1 - \rho E(t)) + h_2(p_2 - \rho \bar{E}(t))}{1 - \lambda(a_1E(t) + a_2\bar{E}(t))}. \tag{27}$$

Then, after some algebraic manipulations, using the fact that $E'(t) = -tb(t)$ and $\bar{E}'(t) = -(1-t)b(t)$, we find

$$\eta'(t) = \frac{\lambda(h_1a_2 - h_2a_1)b(t)\phi(t)}{(1 - \lambda(a_1E(t) + a_2\bar{E}(t)))^2},$$

where

$$\phi(t) = t(p_2 - \rho\bar{E}(t)) - (1-t)(p_1 - \rho E(t))$$

Note that $\phi(0) = -p_1(1-\rho) < 0$, $\phi(1) = p_2(1-\rho) > 0$, and

$$\phi'(t) = p_2 - \rho\bar{E}(t) + p_1 - \rho E(t) = 1 - \rho(1 - B(t)) > 0.$$

Then, since $\phi(t)$ is a continuous function of t, we conclude that there exists a unique $\bar{t} \in (0,1)$ such that $\phi(\bar{t}) = 0$ and consequently since $b(t) > 0$ for $t \in [0,1]$, a unique $\bar{t} \in (0,1)$ such that $\eta'(\bar{t}) = 0$. This implies that $\bar{t}$ is the only candidate for a minimizer of $C(t)$. We next show that $\bar{t}$ is a minimizer. The sign of $\eta'(t)$ is the same as the sign of $\phi(t)$ since $h_1/a_1 > h_2/a_2$

and $b(t) > 0$. Hence, for $\bar{t}$ such that $\phi(\bar{t}) = 0$, it is sufficient to show that $\phi'(\bar{t}) \geq 0$. Since $\rho < 1$, $\phi'(\bar{t}) = 1 - \rho(1 - B(\bar{t})) > 0$, and hence $\bar{t}$ is a minimizer of $C(t)$. □

Proof of Proposition 2: From (8), we find that

$$\frac{\partial g(\rho,t)}{\partial \rho} = \int_t^1 (x-t)b(x)dx \geq 0,$$

for all $t \in [0,1]$, which immediately implies that t^* is a decreasing function of ρ since $g(\rho,0) < 0$, $g(\rho,1) > 0$, and there exists a unique t for which $g(\rho,t) = 0$.

The facts that t^* is monotone in ρ and is bounded imply that it has a limit as ρ converges to 1 or 0. In order to prove that t^* converges to p_1 as ρ approaches 0, it is sufficient to show that for every $\varepsilon > 0$ there exists $0 < \rho^o(\varepsilon) < 1$ such that $p_1 + \varepsilon \geq t^*_\rho \geq p_1 - \varepsilon$ for all $\rho \leq \rho^o(\varepsilon)$. (We changed the notation t^* to t^*_ρ in order to reflect its dependence on ρ.) Since the case for $\varepsilon \geq p_1$ is trivial, we fix $0 < \varepsilon < p_1$. Now, $\lim_{\rho\to 0} g(\rho, p_1 - \varepsilon) = -\varepsilon < 0$. Since $g(\rho,t)$ is continuously increasing in ρ for fixed t, we conclude that there exists $\rho^o(\varepsilon) > 0$ such that $g(\rho^o(\varepsilon), p_1 - \varepsilon) < 0$, which implies that $t^*_{\rho^o(\varepsilon)} \geq p_1 - \varepsilon$. Then, since t^*_ρ decreases with ρ, we have $t^*_\rho \geq p_1 - \varepsilon$ for all $\rho \leq \rho^o(\varepsilon)$. We know that $\lim_{\rho\to 0} g(\rho, p_1 + \varepsilon) = \varepsilon > 0$. Since $g(\rho, p_1 + \varepsilon)$ is continuously increasing in ρ, we have $g(\rho, p_1 + \varepsilon) > 0$ for any $0 < \rho < 1$, which implies that $t^*_\rho \leq p_1 + \varepsilon$. Thus, we conclude that t^* converges to p_1 as ρ approaches 0.

Now, in order to prove that t^* converges to zero as ρ approaches 1, it is sufficient to show that for every $\varepsilon > 0$, there exists $0 < \rho^o(\varepsilon) < 1$ such that $t^*_\rho \leq \varepsilon$ for all $\rho^o(\varepsilon) \leq \rho < 1$. We first fix $\varepsilon > 0$. Now, note that $g(\rho,\varepsilon)$ is a continuously increasing function of ρ from the proof of the first part of the theorem. Furthermore, $\lim_{\rho\to 1} g(\rho,\varepsilon) = \int_0^\varepsilon (\varepsilon - x)b(x)dx > 0$. This shows that there exists $\rho^o(\varepsilon)$ such that $0 < \rho^o(\varepsilon) < 1$ and $g(\rho^o(\varepsilon),\varepsilon) \geq 0$, and hence $t^*_{\rho^o(\varepsilon)} \leq \varepsilon$ by Proposition 1. Then, since t^*_ρ decreases with ρ, we have $t^*_\rho \leq \varepsilon$ for all $\rho \geq \rho^o(\varepsilon)$. Thus, we conclude that t^* converges to 0 as ρ approaches 1. □

Proof of Proposition 3: From the first-order optimality condition, we have

$$t^*(p_2 - \rho\bar{E}(t^*)) = (1 - t^*)(p_1 - \rho E(t^*)). \tag{28}$$

Then, (27) and (28) yield

$$\begin{aligned} \eta(t^*) &= \frac{(t^*h_1 + (1-t^*)h_2)(p_2 - \rho\bar{E}(t^*))}{(1-t^*)(1-\lambda(a_1E(t^*) + a_2\bar{E}(t^*))} \\ &\geq p_2(t^*/(1-t^*))h_1 + p_2h_2, \end{aligned}$$

where the inequality follows from the fact that $E(t^*) \geq p_1 \bar{E}(t^*)/p_2$, which can be shown by using (28) and the fact that $t^* \leq p_1$ (from Proposition 2). Now, noting that $C(t^*)/C(0) = \eta(t^*)/\eta(0)$ and $\eta(0) = p_1 h_1 + p_2 h_2$ completes the proof. □

Proof of Corollary 3: Proposition 2 implies that t^* converges to p_1 as ρ approaches 0. Using this in Proposition 3, we see that the lower bound on $C(t^*)/C_{FCFS}$ converges to 1. For the second part of the corollary, note that $t^* \leq p_1$ by Proposition 2. This implies that as p_1 approaches 0, t^* approaches to zero as well. Then, using Proposition 3, we conclude that as p_1 nears zero, $C(t^*)/C_{FCFS}$ converges to 1. □

Proof of Proposition 4: In this proof, we follow a similar approach as in the proof of Theorem 1 of Kleinrock (1967). Consider a "tagged" customer with signal x. This customer will possibly wait for the service of the customer who is in service upon her arrival, the service of customers who are already in queue when she arrived and whose signals are above x, and the service of customers who arrive during her sojourn in the queue and whose signals are above x. Let W_0 denote the expected remaining service time for the customer in service when the tagged customer arrives. Then, we know that W_0 is equal to one half of the second moment of the service time multiplied by the total arrival rate (see, e.g., Gross and Harris 1998), which gives us

$$W_0 = \frac{\lambda(p_1 e_1 + p_2 e_2)}{2}. \tag{29}$$

By Little's Law, the expected number of customers in queue with a signal in the interval $(y, y + \Delta y)$ is given by $\lambda b(y) W(y) \Delta y$ as Δy approaches zero. Also, the expected number of arrivals during the queueing time of the tagged customer is $\lambda W(x)$. Hence, we have

$$W(x) = W_0 + \int_x^1 \lambda W(y)(y a_1 + (1-y) a_2) b(y) dy + \int_x^1 \lambda W(x)(y a_1 + (1-y) a_2) b(y) dy$$

It then follows that

$$W(x) = \frac{W_0 + \lambda \int_x^1 W(y)(y a_1 + (1-y) a_2) b(y) dy}{1 - \lambda \int_x^1 (y a_1 + (1-y) a_2) b(y) dy}.$$

The proof then follows verifying the above equality by plugging in (9) above, and using (29) together with the fact that

$$\lambda \int_x^1 \frac{(y a_1 + (1-y) a_2) b(y)}{(1 - \lambda \int_y^1 (z a_1 + (1-z) a_2) b(z) dz)^2} dy = \frac{1}{1 - \lambda \int_x^1 (z a_1 + (1-z) a_2) b(z) dz} - 1, \tag{30}$$

where $x \in [0,1]$. One can prove (30) by a change of variables in the integral on the left-hand side of the equation. □

Proof of Proposition 5: First, we can rewrite (8) as follows.

$$
\begin{aligned}
g(\rho, t) &= t(p_2 - \rho \bar{E}(t)) - (1-t)(p_1 - \rho E(t)) \\
&= t(1 - \rho(1 - B(t))) - p_1 + \rho E(t) \\
&= \rho(E(t) - t(1 - B(t))) + t - p_1 \\
&= \rho \int_t^1 (x - t) b(x) dx + t - p_1 \\
&= \rho \mathbb{E}[\max(0, S - t)] + t - p_1
\end{aligned}
$$

where S is the generic random variable representing the customer signal. Let S_Y and S_Z denote the generic signal random variables for signals Y and Z, respectively. Now, the function $\max(0, s-t)$ is a convex function of s and therefore since $B_Y \geq_{cx} B_Z$ we have $\mathbb{E}[\max(0, S_Y - t)] \geq \mathbb{E}[\max(0, S_Z - t)]$ (from the definition of convex ordering, see Appendix A). Then, from Proposition 1, it follows that $t_Y^* \leq t_Z^*$ since $g(\rho, t)$ is an increasing function of t and $g(\rho, t) = 0$ has a unique solution. □

Proof of Proposition 6: In this proof, for any function $q(\cdot)$, we use $q'(\cdot)$ to denote the first derivative of $q(\cdot)$.

Note that $F_1 \geq_{lr} F_2$ if and only if $f_1(x)/f_2(x)$ is non-decreasing in x. Furthermore, $f_1(x)/f_2(x)$ is non-decreasing in x if and only if $f_1'(x)f_2(x) - f_1(x)f_2'(x) \geq 0$ for $x \in [0,1]$. Now, taking the derivative of $p_1(x)$ with respect to x, we find

$$
p_1'(x) = \frac{p_1 p_2 (f_1'(x) f_2(x) - f_1(x) f_2'(x))}{(p_1 f_1(x) + p_2 f_2(x))^2}.
$$

Hence, $p_1(x)$ is non-decreasing if and only if $F_1 \geq_{lr} F_2$. □

Derivation of Equation (16): The proof follows as in the proof of Proposition 4, but is provided for completeness. Following the arguments in the proof of Proposition 4, we obtain

$$
\begin{aligned}
W(x) \quad = \quad & W_0 + \int_x^d \lambda p_1 W(y) a_1 f_1(y) dy + \int_x^d \lambda p_2 W(y) a_2 f_2(y) dy \\
& + \int_x^d \lambda p_1 W(x) a_1 f_1(y) dy + \int_x^d \lambda p_2 W(x) a_2 f_2(y) dy \\
= \quad & W_0 + a_1 \lambda p_1 \int_x^d W(y) f_1(y) dy + a_2 \lambda p_2 \int_x^d W(y) f_2(y) dy \\
& + (a_1 \lambda p_1 (1 - F_1(x)) + a_2 \lambda p_2 (1 - F_2(x)))\, W(x). \qquad (31)
\end{aligned}
$$

Letting $g(y) = (a_1\lambda p_1 f_1(y) + a_2\lambda p_2 f_2(y))/\rho$ and

$$G(x) = \int_c^x g(y)dy = \frac{a_1\lambda p_1 F_1(x) + a_2\lambda p_2 F_2(x)}{\rho}, \tag{32}$$

Equation (31) yields

$$W(x) = \frac{W_0 + \rho\int_x^d W(y)g(y)dy}{1-\rho+\rho G(x)}.$$

The proof then follows verifying the above equality by plugging in (16) above, and using (29) together with the fact that

$$\rho\int_x^d \frac{g(y)dy}{(1-\rho+\rho G(y))^2} = \frac{1}{1-\rho+\rho G(x)} - 1, \tag{33}$$

where $x \in \mathcal{S}$. One can prove (33) by a change of variables in the integral on the left-hand side of the equation. Kleinrock (1967) provides a proof of this equality for the case when $G(x)$ possibly has discontinuities over its support. □

Proof of Proposition 7: Using (14), (15), and (16), we have

$$\begin{aligned} C_{FCFS} - C_{HSF} &= \left(\frac{\lambda p_1 e_1 + \lambda p_2 e_2}{2}\right)\left[\frac{\lambda p_1 h_1 + \lambda p_2 h_2}{1-\rho} - \lambda p_1 h_1 \int_c^d (1-\rho+\rho G(x))^{-2} f_1(x)dx\right. \\ &\quad \left. -\lambda p_2 h_2 \int_c^d (1-\rho+\rho G(x))^{-2} f_2(x)dx\right] \end{aligned} \tag{34}$$

where $G(\cdot)$ is as defined in (32).

Equation (33) yields that

$$\frac{1}{1-\rho} = \int_c^d \frac{1}{(1-\rho+\rho G(x))^2}\left(\frac{\lambda a_1 p_1}{\rho} f_1(x) + \frac{\lambda a_2 p_2}{\rho} f_2(x)\right)dx. \tag{35}$$

Now plugging (35) in (34), we obtain

$$\begin{aligned} &C_{FCFS} - C_{HSF} \\ &= \left(\frac{\lambda p_1 e_1 + \lambda p_2 e_2}{2}\right)\left[(\lambda p_1 h_1 + \lambda p_2 h_2)\int_c^d \frac{1}{(1-\rho+\rho G(x))^2}\left(\frac{\lambda a_1 p_1}{\rho} f_1(x) + \frac{\lambda a_2 p_2}{\rho} f_2(x)\right)dx\right. \\ &\quad \left. - \int_c^d \frac{1}{(1-\rho+\rho G(x))^2}(\lambda p_1 h_1 f_1(x) + \lambda p_2 h_2 f_2(x))dx\right] \\ &= \left(\frac{\lambda p_1 e_1 + \lambda p_2 e_2}{2\rho}\right)\int_c^d \frac{\lambda^2 p_1 p_2}{(1-\rho+\rho G(x))^2}(a_2 h_1 - a_1 h_2)(f_2(x) - f_1(x))dx, \end{aligned}$$

by using $\rho = \lambda(a_1 p_1 + a_2 p_2)$. Then, since $F_1 \geq_{st} F_2$ and $(1-\rho+\rho G(x))^{-2}$ is decreasing in x, and also $h_1/a_1 > h_2/a_2$, we conclude that $C_{FCFS} \geq C_{HSF}$. (Here, we use the fact for any two random variables Y and Z such that Y is larger than Z in the sense of usual stochastic orders, we have $\mathbb{E}[q(Y)] \geq \mathbb{E}[q(Z)]$ for any increasing function $q(\cdot)$ when the expectations exist. See, e.g., Müller and Stoyan 2002.) □

Proof of Proposition 8: Let $\pi_T \in \Theta_N$. Then,

$$C_{\pi_T} = \lambda \sum_{j=1}^{N} W_{j,\pi_T} \sum_{i=1}^{2} p_i h_i (F_i(t_{j-1}) - F_i(t_j))$$

where with a slight abuse of notation, W_{j,π_T} denotes the expected waiting time for the customers in priority class j, as in (1) but under the signal formulation of Section 8. Then, as in (1), it follows from Cobham (1954) that

$$W_{j,\pi_T} = \frac{\lambda(p_1 e_1 + p_2 e_2)}{2\alpha_{j-1}\alpha_j},$$

where $\alpha_j = 1 - \rho + \lambda(a_1 p_1 F_1(t_j) + a_2 p_2 F_2(t_j))$ for $j = 0, 1, \ldots, N$.

Now, define a cumulative distribution function $G(\cdot)$ as follows:

$$G(y) = \begin{cases} 0, & y \leq t_N = c, \\ \frac{a_1 p_1 F_1(t_j) + a_2 p_2 F_2(t_j)}{a_1 p_1 + a_2 p_2}, & t_j \leq y < t_{j-1} \text{ for } j = 1, 2, \ldots, N, \\ 1, & t_0 = d < y, \end{cases}$$

Then, using Equation (10) of Kleinrock (1967) (for $x^- = c$), we find

$$\frac{1}{1-\rho} = \sum_{j=1}^{N} \frac{G(t_{j-1}) - G(t_j)}{(1 - \rho + \rho G(t_j))(1 - \rho + \rho G(t_{j-1}))},$$

which we can also write as follows:

$$\frac{1}{1-\rho} = \frac{1}{a_1 p_1 + a_2 p_2} \sum_{j=1}^{N} \frac{a_1 p_1 (F_1(t_{j-1}) - F_1(t_j)) + a_2 p_2 (F_2(t_{j-1}) - F_2(t_j))}{\alpha_{j-1}\alpha_j}. \tag{36}$$

Then, using (36) in (13), and using (14), we find

$$
\begin{aligned}
C_{FCFS} - C_{\pi_T} = & \frac{\lambda^2(p_1e_1+p_2e_2)}{2}\left\{\left(\frac{p_1h_1+p_2h_2}{a_1p_1+a_2p_2}\right)\right. \\
& \times \sum_{j=1}^{N} \frac{a_1p_1(F_1(t_{j-1})-F_1(t_j)) + a_2p_2(F_2(t_{j-1})-F_2(t_j))}{\alpha_{j-1}\alpha_j} \\
& \left. - \sum_{j=1}^{N} \frac{p_1h_1(F_1(t_{j-1})-F_1(t_j)) + p_2h_2(F_2(t_{j-1})-F_2(t_j))}{\alpha_{j-1}\alpha_j}\right\} \\
= & \frac{\lambda^2(p_1e_1+p_2e_2)p_1p_2(h_1a_2-h_2a_1)}{2(a_1p_1+a_2p_2)} \sum_{j=1}^{N} \frac{F_2(t_{j-1})-F_2(t_j)-F_1(t_{j-1})+F_1(t_j)}{\alpha_{j-1}\alpha_j} \\
= & \frac{\lambda^2(p_1e_1+p_2e_2)p_1p_2(h_1a_2-h_2a_1)}{2(a_1p_1+a_2p_2)} \left\{\sum_{j=0}^{N-1} \frac{F_2(t_j)-F_1(t_j)}{\alpha_j\alpha_{j+1}} - \sum_{j=1}^{N} \frac{F_2(t_j)-F_1(t_j)}{\alpha_{j-1}\alpha_j}\right\} \\
= & \frac{\lambda^2(p_1e_1+p_2e_2)p_1p_2(h_1a_2-h_2a_1)}{2(a_1p_1+a_2p_2)} \left\{\frac{F_2(t_0)-F_1(t_0)}{\alpha_0\alpha_1} - \frac{F_2(t_N)-F_1(t_N)}{\alpha_{N-1}\alpha_N}\right. \\
& \left. + \sum_{j=0}^{N-1}\left(\frac{F_2(t_j)-F_1(t_j)}{\alpha_j}\right)\left(\frac{1}{\alpha_{j+1}} - \frac{1}{\alpha_{j-1}}\right)\right\}.
\end{aligned}
$$

Now, the term outside the curly parenthesis is non-negative since $h_1a_2 - h_2a_1 \geq 0$. The first two terms inside the curly parentheses is zero since $F_1(t_0) = F_2(t_0) = 1$ and $F_1(t_N) = F_2(t_N) = 0$. Finally, the summation term inside the curly parentheses is non-negative since $F_1 \geq_{st} F_2$ and $\alpha_{j-1} \geq \alpha_{j+1}$. Hence, $C_{FCFS} \geq C_{\pi_T}$. □

Appendix C - Complementary Material for Section 7

Derivation of Equation (10)**:** Using (26) and (27), we obtain

$$\begin{aligned} C_Y(t_Y) - C_Z(t_Z) &= \frac{\lambda^2(p_1e_1+p_2e_2)}{2(1-\rho)}\left(\frac{h_1(p_1-\rho E_Y(t_Y))+h_2(p_2-\rho\bar{E}_Y(t_Y))}{1-\lambda(a_1E_Y(t_Y)+a_2\bar{E}_Y(t_Y))}\right. \\ &\quad \left. -\frac{h_1(p_1-\rho E_Z(t_Z))+h_2(p_2-\rho\bar{E}_Z(t_Z))}{1-\lambda(a_1E_Z(t_Z)+a_2\bar{E}_Z(t_Z))}\right). \end{aligned}$$

Then, applying several algebraic manipulations, we obtain (10). □

Higher variance does not imply a more useful signal - an example: Suppose that $b_Y(\cdot)$ and $b_Z(\cdot)$ are given by

$$b_Y(x) = \begin{cases} 4.375, & 0 < x \leq 0.2, \\ 0, & 0.2 < x \leq 0.8, \\ 0.625, & 0.8 < x < 1, \end{cases}$$

and

$$b_Z(x) = \begin{cases} 4.86111, & 0 < x \leq 0.16, \\ 0, & 0.16 < x \leq 0.24, \\ 0.29240, & 0.24 < x < 1. \end{cases}$$

It can be shown that signals Y and Z are not ordered in the sense of convex ordering, but variance of Y (which is approximately 0.07333) is larger than the variance of Z (which is approximately 0.06275) and their expected values both equal to 0.2. If we let $h_1 = 4$, $h_2 = 1$, $\lambda = 0.8$, $a_1 = a_2 = 1$, and $e_1 = e_2 = 2$, we obtain that $t_Y^* \approx 0.10448$ while $t_Z^* \approx 0.10094$ and the long-run average cost with signal Y is 4.20302 while the long-run average cost with signal Z is 4.16904, approximately. Thus, the long-run average cost is smaller with the signal that has smaller variance.

Definition of signal distributions used in the numerical analysis of Section 7: We considered four different families of distributions. In the following, let $b^s(\cdot)$ denote the probability density function and $B^s(\cdot)$ denote the cumulative distribution function for the signal distribution parameterized by s.

Family 1: The distribution $B^s(\cdot)$ is Uniform over the interval $[p_1 - s, p_1 + s]$ where $s \in (0, \min(p_1, 1-p_1))$. In Figure 3, observations from this family of distributions are indicated by markers of square shape.

Family 2: For $p_1 \leq 0.5$, we have

$$b^s(x) = \begin{cases} \frac{1}{2(p_1-s)}, & 0 < x \leq p_1 - s, \\ 0, & p_1 - s < x \leq p_1 + s, \\ \frac{1}{2(p_1-s)}, & p_1 + s < x \leq 2p_1, \\ 0, & 2p_1 < x < 1, \end{cases}$$

while for $p_1 \geq 0.5$, we have

$$b^s(x) = \begin{cases} 0, & 0 < x \leq 2p_1 - 1, \\ \frac{1}{2(1-p_1-s)}, & 2p_1 - 1 < x \leq p_1 - s, \\ 0, & p_1 - s < x \leq p_1 + s, \\ \frac{1}{2(1-p_1-s)}, & p_1 + s < x < 1, \end{cases}$$

where $s \in (0, \min(p_1, 1-p_1))$. In Figure 3, observations from this family of distributions are indicated by markers of triangle shape.

Family 3: We have

$$b^s(x) = \begin{cases} \frac{1-(p_1-s)}{(1+2s)(p_1-s)}, & 0 < x \leq p_1 - s, \\ 0, & p_1 - s < x \leq p_1 + s, \\ \frac{p_1+s}{(1+2s)(1-p_1-s)}, & p_1 + s < x \leq 1, \end{cases}$$

where $s \in (0, \min(p_1, 1-p_1))$. In Figure 3, observations from this family of distributions are indicated by markers of diamond shape.

Family 4: We have

$$b^s(x) = \begin{cases} \frac{2p_1-(2-s)}{2s(s-1)}, & 0 < x \leq s, \\ 0, & s < x \leq 1 - s, \\ \frac{s-2p_1}{2s(s-1)}, & 1 - s < x \leq 1, \end{cases}$$

where $s \in (0, \min(p_1, 1-p_1))$. In Figure 3, observations from this family of distributions are indicated by markers of "×" shape.

Appendix D - Sketches of Proofs for the Case with Multiple Servers

Assume that there are $K \geq 2$ servers in the system. If we assume that service times are i.i.d. exponential random variables with mean $1/\mu$ for all customers regardless of their types, then we can show that all of our analytical results in this paper still hold after some minor modifications to our proofs. In the following, we provide an outline of these modifications.

Firstly, from Cobham (1954), our Equation (1) becomes

$$W_{j,\pi} = \frac{\bar{W}_0}{(1 - \lambda \sum_{k=1}^{j-1} \Gamma_{k,\pi})(1 - \lambda \sum_{k=1}^{j} \Gamma_{k,\pi})}, \tag{37}$$

where $\Gamma_{k,\pi} = (\int_{I_{k,\pi}} b(x)dx)/(K\mu)$ for $k = 1, \ldots, N$,

$$\bar{W}_0 = \frac{(\frac{\lambda}{\mu})^K}{K!K\mu(1-\rho)(\sum_{j=0}^{K-1} \frac{(\lambda/\mu)^j}{j!} + \frac{(\lambda/\mu)^K}{K!(1-\rho)})},$$

and $\rho = \lambda/(K\mu) < 1$. Using this expression along with the condition that $h_1 > h_2$, the proofs of Theorems 1 and 2 follow as before, which in turn yields, Corollary 1, Corollary 2, and Theorem 3. Similarly, by using (37), we find that Proposition 1 also holds with the same first-order condition. (Note that the only change in the first-order condition would be that now $\rho = \lambda/(K\mu)$.) Since the expression for the optimal threshold does not change with multiple servers, Propositions 2 and 3 as well as Corollary 3 continue to hold as well. Similarly, it is possible to show that Theorem 4 is also valid.

We were also able to obtain an expression for $W(x)$ as in Proposition 4 under the case with multiple servers:

$$W(x) = \frac{\bar{W}_0}{(1 - \rho + \rho B(x))^2}.$$

The proof of this result is very similar to the proof of Proposition 4 except that in several places we replace a_1 and a_2 with $1/(K\mu)$ and also use the fact that service times are exponentially distributed. Similarly, (16) can be also revised as

$$W(x) = \frac{\bar{W}_0}{(1 - \rho + \rho(p_1 F_1(x) + p_2 F_2(x)))^2},$$

for the multiple-server setting. Furthermore, the expression for the long-run average waiting time under the FCFS policy becomes $W_{FCFS} = \bar{W}_0/(1-\rho)$, see, e.g., page 71 in Gross and Harris (1998). Then, we use these two expressions to show that Proposition 7 also holds when there are multiple servers. Finally, our proof of Proposition 8 can be extended to the multiple-server case by noting that the expression for W_{j,π_T}, $j = 1, 2, \ldots, N$ now becomes

$$W_{j,\pi_T} = \frac{\bar{W}_0}{(1-\rho+\rho(p_1F_1(t_{j-1})+p_2F_2(t_{j-1})))(1-\rho+\rho(p_1F_1(t_j)+p_2F_2(t_j)))}$$

using (37). □