A. Proofs

Proof of Proposition 1: We first prove part (i) and then prove part (ii).

Proof of part (i): It is helpful to first obtain a lower bound on the cost of any staffing policy, regardless of the routing policy $\pi \in \Pi$ that is followed. For this, let $S_\pi(l)$ be the expected number of busy servers when the arrival rate realization is $l$, and let $S_\pi(\Lambda)$ be the associated random variable. Since $N \geq S_\pi(l)$,

$$C(N,\pi) \geq cE[S_\pi(\Lambda)] + pE[\Lambda P_\pi(\text{out}; N, \Lambda)] + aE[\Lambda P_\pi(\text{ab}; N, \Lambda)],$$

so that the assumption $c \geq \min(a,p)$ then yields

$$C(N,\pi) \geq \min(a,p)E[S_\pi(\Lambda) + \Lambda (P_\pi(\text{out}; N, \Lambda) + P_\pi(\text{ab}; N, \Lambda))]. \tag{A.1}$$

For any realization $l$ of $\Lambda$, the arrival rate into the system must equal the departure rate from the system (due to both abandonments and service completion), so that

$$l(1 - P_\pi(\text{out}; N, l)) = S_\pi(N,l) + lP_\pi(\text{ab}; N, l),$$

or, equivalently,

$$l = S_\pi(N,l) + l (P_\pi(\text{out}; N, l) + P_\pi(\text{ab}; N, l)).$$

Noting that the equation above is valid for any realization $l$ of $\Lambda$ and taking expectations with respect to $\Lambda$ shows that (recall that $E[\Lambda] = \lambda$)

$$\lambda = E[S_\pi(\Lambda) + \Lambda (P_\pi(\text{out}; N, \Lambda) + P_\pi(\text{ab}; N, \Lambda))]. \tag{A.2}$$

Substituting the equality (A.2) into (A.1) shows

$$C(N,\Lambda) \geq \min(a,p)\lambda.$$

Suppose $a \leq p$ ($a > p$). Then, the minimum in (A.1) is attained by setting a zero staffing level and letting everyone abandon (outsourcing everyone), because the cost associated with that policy is $C(0,\Lambda) = a\lambda$ ($C(0,\Lambda) = p\lambda$). Hence $N^{\text{opt}} = 0$ and

$$\pi^{\text{opt}}(0,\Lambda) = (0,0,\ldots,0) \text{ if } a > p,$$

$$\pi^{\text{opt}}(0,\Lambda) = (1,1,\ldots,1) \text{ if } a \leq p.$$

Proof of part (ii): We must show that

$$E[z_\pi(N,\Lambda)] = pE[\Lambda P_\pi(\text{out}; N, \Lambda)] + aE[\Lambda P_\pi(\text{ab}; N, \Lambda)].$$
is lower bounded by the cost associated with the policy $\tau(\infty)$ that does not outsource any calls (has $P_{\tau(\infty)}(\text{out}; \Lambda) = 0$); i.e., we must show that

$$pE[\Lambda P_{\pi}(\text{out}; N, \Lambda)] + aE[\Lambda P_{\pi}(\text{ab}; N, \Lambda)] \geq aE[\Lambda P_{\tau}(\infty)(\text{ab}; N, \Lambda)].$$

Since $p \geq a$ by assumption, it is sufficient to show

$$E[\Lambda (P_{\pi}(\text{out}; N, \Lambda) + P_{\pi}(\text{ab}; N, \Lambda))] \geq E[\Lambda P_{\tau}(\infty)(\text{ab}; N, \Lambda)].$$

(A.3)

It can be seen through a coupling argument that the number of busy servers is stochastically larger in the system that does not outsource any calls, and so

$$E[S_{\pi}(N, \Lambda)] \leq E[S_{\tau}(\infty)(N, \Lambda)].$$

(A.4)

The inequality (A.4) combined with the equality (A.2) in the proof of part (i) implies (A.3), and so the proof is complete. ■

Proof of Proposition 2: We first prove part (i) and then prove part (ii).

Proof of part (i): Since $E[z_{\lambda}(N, \Lambda^{\lambda}(X))] \geq 0$ for all $\lambda$,

$$\liminf_{\lambda \to \infty} \frac{cN^{\lambda} + E[z_{\pi}(N, \Lambda^{\lambda}(X))]}{\lambda} \geq c \liminf_{\lambda \to \infty} \frac{N^{\lambda}}{\lambda}.$$

Hence if $\liminf_{\lambda \to \infty} N^{\lambda}/\lambda \geq 1$, the proof is complete. Assume otherwise, that $\liminf_{\lambda \to \infty} N^{\lambda}/\lambda = \delta < 1$. As in the proof of Proposition 1, define $S^{\lambda}_{\pi}(N, l^{\lambda})$ as the expected number of busy servers in the system with mean arrival rate $\lambda$ and arrival rate realization $l^{\lambda} = \lambda + x\sqrt{\lambda}$, and let $S^{\lambda}_{\pi}(N, \Lambda^{\lambda})$ be the associated random variable. Recall the equality (A.2) in the proof of Proposition 1 and note that it holds for each $\lambda$, and so,

$$E[\Lambda^{\lambda}(P_{\pi}(\text{out}; N, \Lambda^{\lambda}) + P_{\pi}(\text{ab}; N, \Lambda^{\lambda}))] = \lambda - E[S^{\lambda}_{\pi}(N, \Lambda^{\lambda})] \geq \lambda - N^{\lambda}.$$

(A.5)

(In words, (A.5) states that the expected steady-state rate at which customers abandon or are outsourced must equal or exceed the difference between the mean arrival rate and the service capacity.) From (A.5) and our assumption that $\min(a, p) > c$,

$$E[z_{\pi}(N, \Lambda^{\lambda}(X))] \geq \min(a, p)E[\Lambda^{\lambda}(P_{\pi}(\text{out}; N, \Lambda^{\lambda}) + P_{\pi}(\text{ab}; N, \Lambda^{\lambda})))] \geq c(\lambda - N^{\lambda}).$$

The above inequality then implies

$$\liminf_{\lambda \to \infty} \frac{cN^{\lambda} + E[z_{\pi}(N, \Lambda^{\lambda}(X))]}{\lambda} \geq c\delta + c(1 - \delta) = c.$$

Proof of part (ii): When no customers are outsourced, for every realization $l^{\lambda} = l^{\lambda}(x) = \lambda + x\sqrt{\lambda}$ of $\Lambda^{\lambda}$, the system operates as a $M/M/N^{\lambda} + M$ queue, which was analyzed in [Garnett et al. (2002)]. Since the staffing assumption implies

$$\sqrt{N^{\lambda}} \left(1 - \frac{l^{\lambda}(x)}{N^{\lambda}}\right) \to \beta - x \text{ as } \lambda \to \infty,$$
the condition of Theorem 4 in Garnett et al. (2002) is satisfied, and so
\[ \sqrt{N^{\lambda}} P_{\tau(\infty)}^\lambda(\text{ab}; l^\lambda) \to \triangle \in (0, \infty) \text{ as } \lambda \to \infty. \] (A.6)

Since \( P_{\tau(\infty)}^\lambda(\text{out}; N, l^\lambda) = 0 \),
\[ z_{\tau(\infty)}^\lambda(N, l^\lambda) = a l^\lambda P_{\tau(\infty)}^\lambda(\text{ab}; N, l^\lambda). \] (A.7)

It follows from (A.6) and (A.7) that
\[ \frac{z_{\tau(\infty)}^\lambda(N, l^\lambda)}{\lambda} \to 0 \text{ as } \lambda \to \infty, \]
and so
\[ \frac{z_{\tau(\infty)}^\lambda(N, \Lambda^\lambda)}{\lambda} \to 0 \text{ almost surely, as } \lambda \to \infty. \]

Since \( P_{\tau(\infty)}^\lambda(\text{ab}; l^\lambda) \leq 1 \) for any realization \( l^\lambda \) of \( \Lambda^\lambda \), (A.7) implies
\[ \frac{z_{\tau(\infty)}^\lambda(N, \Lambda^\lambda)}{\lambda} \leq a \left( 1 + \frac{1}{\sqrt{\lambda}} \right) \leq a(1 + |X|) \]
for all \( \lambda \geq 1 \). Since \( E|X| < \infty \) by assumption, the dominated convergence theorem implies that
\[ \frac{E\left[ \frac{z_{\tau(\infty)}^\lambda(N, \Lambda^\lambda)}{\lambda} \right]}{\lambda} \to 0 \text{ as } \lambda \to \infty, \]
which is sufficient to complete the proof.

**Proof of Lemma 1:** This is Theorem 5.2 from Koçağa and Ward (2010) adapted to the setting of this paper, and, to read this proof, the reader is advised to have a copy of that paper on hand. We abbreviate Koçağa and Ward (2010) to KW for the remainder of this proof. First note that Theorem 5.2 in KW holds for any threshold admission policy \( \theta^N \) defined in Theorem 5.1 in that paper, and not only for the policy \( \theta^* \) that appears in Theorem 5.2. Next, when \( X \) in this paper realizes as \( x \), the arrival rate in KW is \( l^\lambda = x \sqrt{\lambda} \), so that
\[ l^\lambda(x) - N^\lambda = -m \sqrt{\lambda} + o\left( \sqrt{\lambda} \right) \text{ for } m = \beta - x. \]

The above display implies that the conditions in Theorem 5.2 in KW are satisfied, and so, noting that \( \lim_{t \to \infty} E[\xi(t, \theta^* N)]/t \) in their notation is \( z_{\tau}(N, l^\lambda(x)) \) in ours, for \( \tau = \{ \tau(T^\lambda) : \lambda \geq 0 \} \) and
\[ T^\lambda = N^\lambda + \tilde{T} \sqrt{l^\lambda(x)}, \]
it follows that
\[ \frac{1}{\sqrt{\lambda}} z_{\tau}(N, l^\lambda(x)) \to \kappa, \text{ as } \lambda \to \infty, \] (A.8)
where \( \kappa \) is as defined in (4.9) in KW.

\footnote{Although \( \triangle \) is explicitly specified in Garnett et al. (2002), we do not provide the expression here, because knowledge that the limit is finite is enough for our purposes.}
Finally, we match the notation and show the algebra that establishes $\kappa$ in (4.9) in KW is exactly $\hat{z}(m, \hat{T})$. Substituting the notation in Table A.1 into $\kappa$ in (4.9) in KW shows

$$
\kappa = \frac{p \exp \left( \frac{-\gamma}{2} \left( \hat{T}^2 + 2m \hat{T} \right) \right) + a \left[ 1 - \exp \left( \frac{-\gamma}{2} \left( \hat{T}^2 + 2m \hat{T} \right) \right) \right] + \sqrt{2\pi} m \exp \left( \frac{m^2}{2\gamma} \right) \left( \Phi \left( \frac{m}{\sqrt{\gamma}} \right) - \Phi \left( \sqrt{\gamma} (\hat{T} + \frac{m}{\gamma}) \right) \right)}{\sqrt{2\pi} \left[ \exp \left( \frac{m^2}{2\gamma} \right) \Phi(m) + \frac{1}{\sqrt{\pi}} \exp \left( \frac{m^2}{2\gamma} \right) \left( \Phi \left( \sqrt{\gamma} (\hat{T} + \frac{m}{\gamma}) \right) - \Phi \left( \frac{m}{\sqrt{\gamma}} \right) \right) \right]},
$$

Next, multiplying by $\phi(m/\sqrt{\gamma})/\phi(m/\sqrt{\gamma})$ and recalling that $\phi(x) = (1/\sqrt{2\pi}) \exp(-x^2/2)$ for $x \in (-\infty, \infty)$ yields

$$
\kappa = \frac{p \phi \left( \sqrt{\gamma} (\hat{T} + \frac{m}{\gamma}) \right) + a \left[ \phi \left( \frac{m}{\sqrt{\gamma}} \right) - \phi \left( \sqrt{\gamma} (\hat{T} + \frac{m}{\gamma}) \right) + \frac{m}{\sqrt{\pi}} \left[ \phi \left( \frac{m}{\sqrt{\gamma}} \right) - \Phi \left( \sqrt{\gamma} (\hat{T} + \frac{m}{\gamma}) \right) \right] \right]}{\phi \left( \phi(m) \right) \Phi(m) + \frac{1}{\sqrt{\gamma}} \left( \Phi \left( \sqrt{\gamma} (\hat{T} + \frac{m}{\gamma}) \right) - \Phi \left( \frac{m}{\sqrt{\gamma}} \right) \right),}
$$

which is exactly $\hat{z}(m, \hat{T})$.

<table>
<thead>
<tr>
<th>KW notation</th>
<th>This paper’s notation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma^2$</td>
<td>2</td>
</tr>
<tr>
<td>$\mu$</td>
<td>1</td>
</tr>
<tr>
<td>$m$</td>
<td>$\beta - x$</td>
</tr>
<tr>
<td>$h_I$</td>
<td>0</td>
</tr>
<tr>
<td>$a$</td>
<td>$a$</td>
</tr>
<tr>
<td>$c$</td>
<td>$p$</td>
</tr>
<tr>
<td>$l$</td>
<td>$\hat{T}$</td>
</tr>
</tbody>
</table>

Table A.1: The notation match between Koçağa and Ward (2010) and this paper.

**Proof of Corollary 1:** This follows because a careful reading of the proof of Theorem 5.2 in KW shows that that result holds whenever the term multiplying $\hat{T}$ in (7) is of the same order as the square root of the arrival rate.

**Proof of Theorem 1:** It follows from Lemma 1 that for a given realization $x$ of the random variable $X$

$$
\sqrt{\lambda} \left( \frac{cN^\lambda + z_\tau^\lambda(N, x)}{\lambda} - c \right) \to c\beta + \hat{z}(\beta - x, \hat{T}) \text{ as } \lambda \to \infty.
$$

For this proof, we must argue that the interchange of limit and expectation is valid; in particular, it is enough to show that

$$
\frac{1}{\sqrt{\lambda}} E \left[ z_\tau^\lambda(N, \Lambda^\lambda(X)) \right] \to E \left[ \hat{z}(\beta - X, \hat{T}) \right], \text{ as } \lambda \to \infty. \tag{A.9}
$$

Recall that

$$
z_\tau^\lambda(N, \Lambda^\lambda(X)) = p\Lambda^\lambda P_\tau^\lambda(\text{out}; N, \Lambda^\lambda(X)) + a\gamma Q_\tau^\lambda(N, \Lambda^\lambda(X)).
$$

Suppose we can show

$$
\frac{1}{\sqrt{\lambda}} E \left[ \Lambda^\lambda(X) P_\tau^\lambda(\text{out}, N, \Lambda^\lambda(X)) \right] \to EL_1(X) \text{ as } \lambda \to \infty \tag{A.10}
$$
and
\[ \frac{1}{\sqrt{\lambda}} \gamma E \left[ Q_\tau^\lambda(N, \Lambda^\lambda(X)) \right] \to EL_2(X) \text{ as } \lambda \to \infty, \]  
(A.11)
where, for any \( x \in (-\infty, \infty) \),
\[
L_1(x) := \phi \left( \sqrt{\gamma} \left( \hat{T} + \frac{m}{\gamma} \right) \right) \Phi(m) + \frac{1}{\sqrt{\gamma}} \left( \Phi \left( \sqrt{\gamma} \left( \hat{T} + \frac{m}{\gamma} \right) \right) - \Phi \left( \frac{m}{\sqrt{\gamma}} \right) \right),
\]
\[
L_2(x) := \left[ \phi \left( \frac{m}{\sqrt{\gamma}} \right) - \phi \left( \sqrt{\gamma} \left( \hat{T} + \frac{m}{\gamma} \right) \right) + \frac{m}{\sqrt{\gamma}} \left[ \Phi \left( \frac{m}{\sqrt{\gamma}} \right) - \Phi \left( \sqrt{\gamma} \left( \hat{T} + \frac{m}{\gamma} \right) \right) \right] \right] \phi \left( \frac{m}{\phi(m)} \right) \Phi(m) + \frac{1}{\sqrt{\gamma}} \left( \Phi \left( \sqrt{\gamma} \left( \hat{T} + \frac{m}{\gamma} \right) \right) - \Phi \left( \frac{m}{\sqrt{\gamma}} \right) \right).
\]

Since it is straightforward to see that \( L_1(x) + L_2(x) = z(\beta - x, \hat{T}) \), which implies \( E[L_1(X)] + E[L_2(X)] = E \left[ \hat{z}(\beta - X, \hat{T}) \right] \), to complete the proof, it is enough to show (A.10) and (A.11).

The argument to show (A.10):

We begin by observing that for any realization \( x \) of \( X \), an argument similar to the proof of part (c) of Theorem 5.2 (on page 311 of Koçag˘a and Ward (2010)) shows that
\[
\frac{l^\lambda(x)}{\sqrt{\lambda}} P^\lambda_{\tau}(\text{out}; N, l^\lambda(x)) \to L_1(x),
\]
as \( \lambda \to \infty \), almost surely. Then, to show (A.10), it is sufficient to show that \( \Lambda^\lambda(X) P^\lambda_{\tau}(\text{out}, \Lambda^\lambda(X)) / \sqrt{\lambda} \) is bounded by an integrable random variable. For this, first note that
\[
P^\lambda_{\tau}(\text{out}; N, \Lambda^\lambda(X)) \leq B(N, \Lambda^\lambda(X)),
\]
where \( B(m, \lambda) \) is the Erlang blocking probability in a \( M/M/m/m \) model that has offered load \( \lambda \) (which is exactly equal to the arrival rate in our model since \( \mu = 1 \)). Also,
\[
B(N, \Lambda^\lambda(X)) \leq B(N, \Lambda^\lambda(|X|)),
\]
because the Erlang blocking probability is increasing in the offered load. Furthermore, the Erlang loss function \( L(m, \lambda) := \lambda B(m, \lambda) \) (again, when the offered load equals the arrival rate) is subadditive by Theorem 1 in Smith and Whitt (1981), and so
\[
L(N, \Lambda^\lambda(X)) \leq L(N, \lambda) + L(N, \sqrt{\lambda} |X|),
\]
which implies
\[
\Lambda^\lambda(X) B(N, \Lambda^\lambda(X)) \leq \lambda B(N, \lambda) + \sqrt{\lambda} |X| B(N, \sqrt{\lambda} |X|). \tag{A.12}
\]
Next, also noting that \( B(m, \lambda) \leq 1 \) for any positive integer \( m \) and finite \( \lambda \geq 0 \), we have that
\[
\frac{\Lambda^\lambda(X)}{\sqrt{\lambda}} P^\lambda_{\tau}(\text{out}; N, \Lambda^\lambda(X)) \leq \sqrt{\lambda} B(N, \lambda) + |X|. \tag{A.13}
\]
Since (see for example, Whitt (1984))
\[
\sqrt{\lambda} B(N, \lambda) \to \frac{\phi(\beta)}{\Phi(\beta)} \text{ as } \lambda \to \infty,
\]
5
and recalling that $E|X| < \infty$ by assumption, it follows that the right-hand-side of (A.13) is bounded by an integrable random variable, so that (A.10) is justified.

The argument so show (A.11):

We begin by observing that for any realization $x$ of $X$, an argument similar to the proof of part (b) of Theorem 5.2 in Koçaga and Ward (2010) shows that

$$\frac{\gamma Q_\tau(T^\lambda(N, l^\lambda(x)))}{\sqrt{\lambda}} \to L(x),$$

as $\lambda \to \infty$, almost surely. Then, to show (A.11), it is sufficient to show that $\gamma Q_\tau(N, \lambda^\lambda(X))/\sqrt{\lambda}$ is bounded by an integrable random variable. Since $\gamma Q_\tau(N, \lambda^\lambda(x)) = \lambda^\lambda(x) P_\tau^\lambda(ab; N, \lambda^\lambda(x))$ for any realization $x$ of $X$ (as observed in Remark 1), it is sufficient to show that $\Lambda^\lambda(X) P_\tau^\lambda(ab; N, \lambda^\lambda(X))/\sqrt{\lambda}$ is bounded by an integrable random variable. It follows from a coupling argument that for any realization $x$ of $X$,

$$P_\tau^\lambda(ab; N, \lambda^\lambda(X)) \leq B(N^\lambda, \lambda^\lambda(X)),$$

where $B$ is the Erlang blocking probability defined in the argument to show (A.10). Therefore, the bound in (A.12) implies

$$\frac{\Lambda^\lambda(X)}{\sqrt{\lambda}} P_\tau^\lambda(ab; N, \lambda^\lambda(X)) \leq \sqrt{\lambda} B(N, \lambda) + |X|.$$

The right-hand side of the above expression is bounded by an integrable random variable for the exact same reason that (A.13) was.

Proof of Proposition 3: We first observe that

$$E \left[ \hat{\zeta}(\beta - X, \hat{T}^*(\beta - X)) \right] \leq$$

(A.14)

$$E \left[ \hat{\zeta}(\beta - X, \hat{T}^*(\beta - X))|X \leq \beta + 1 \right] + E \left[ \hat{\zeta}(\beta - X, \hat{T}^*(\beta - X))|X > \beta + 1 \right].$$

Next, we use the following two claims. The proofs of the claims can be found immediately following this proof.

Claim 1. For any fixed $\beta \in (-\infty, \infty)$, $\hat{\zeta}(\beta - x, \hat{T}^*(\beta - x))$ is an increasing function of $x$ on $(-\infty, \infty)$.

Claim 2. For any fixed $\beta \in (-\infty, \infty)$ and $x > \beta + 1$,

$$\hat{\zeta}(\beta - x, 0) < p(x - \beta + 1).$$

It follows from Claim 1 that

$$E \left[ \hat{\zeta}(\beta - X, \hat{T}^*(\beta - X))|X \leq \beta + 1 \right] \leq \hat{\zeta}(-1, \hat{T}^*(-1)),$$

(A.15)

and it follows from Claim 2 that

$$E \left[ \hat{\zeta}(\beta - X, \hat{T}^*(\beta - X))|X > \beta + 1 \right] \leq p(E|X| + \beta + 1).$$

(A.16)

Since $\hat{\zeta}(-1, \hat{T}^*(-1)) < \infty$ from the definition of $\hat{\zeta}$ in (8) and $\hat{T}^*$ in (9) (recalling that there exists a unique $T^*$ that solves (9) by Proposition 4.1 in Koçaga and Ward (2010)) and $E|X| < \infty$ by assumption, we conclude from (A.14), (A.15), and (A.16) that

$$E \left[ \hat{\zeta}(\beta - X, \hat{T}^*(\beta - X)) \right] \leq \hat{\zeta}(-1, \hat{T}^*(-1)) + p(E|X| + \beta + 1) < \infty.$$
Then, the inequality (A.19) implies that

From the definition of \( \hat{\pi} \) in (8), the symmetry of the normal distribution, and algebra,

Next, from (10), since \( \hat{T}_2^* \) to satisfy (9) with \( x \) replaced by \( x_2 \), and let \( \tau = \{\tau(T^\lambda): \lambda \geq 0\} \) where

Then, Corollary 1 implies that

Suppose \( (N^\lambda - \lambda)/\sqrt{\lambda} \to \beta \) as \( \lambda \to \infty \), define \( \hat{T}_2^* \) to satisfy (9) with \( x \) replaced by \( x_2 \), and let \( \tau = \{\tau(T^\lambda): \lambda \geq 0\} \) where

It follows from (A.17) and (A.18) that

Next, from (10), since \( \hat{T}_1^* \) is the minimizer of \( \hat{z}(\beta - x_1, T) \) over \( T \in [0, \infty) \),

We conclude from the previous two displays that

which completes the proof, since \( x_1 \) and \( x_2 \) are arbitrary.

Proof of Claim 2 We use the following inequality, that is given in 7.1.13 of Abramowitz and Stegun (1965),

From the definition of \( \hat{z} \) in (8), the symmetry of the normal distribution, and algebra,

Then, the inequality (A.19) implies that

\[
\hat{z}(\beta - x, 0) < \frac{p}{\sqrt{2}} \left( \frac{x - \beta}{\sqrt{2}} + \sqrt{\left( \frac{x - \beta}{\sqrt{2}} \right)^2 + 2} \right) \text{ for all } x \geq \beta + 1.
\]
Finally, since
\[
\frac{x - \beta}{\sqrt{2}} + \sqrt{\left(\frac{x - \beta}{\sqrt{2}}\right)^2 + 2} \leq \frac{x - \beta}{\sqrt{2}} + \sqrt{\left(\frac{x - \beta}{\sqrt{2}} + \sqrt{2}\right)^2} = \sqrt{2}(x - \beta + 1),
\]
the stated claim is established.

**Proof of Claim 3.** Let \( p < a \). Then, for any realization \( x \) of \( X \), \( \hat{T}^* \) defined in (9) is continuous in \( \beta \).

**Proof of Claim 4.** For any realization \( x \) of \( X \), \( \hat{\zeta}(\beta - x, \hat{T}^*(\beta - x)) \) is a decreasing function of \( \beta \) on \((-\infty, \infty)\).

**Proof of (a):** For any realization \( x \) of \( X \), any \( \beta \in \mathbb{R} \), and \( N^\lambda \) that satisfies the conditions of Lemma 1, by Lemma 1, \( \hat{\zeta}(\beta - x, \hat{T}^*(\beta - x)) \) is obtained as the limit of the non-negative cost function \( z_\lambda^\beta(N, \ell^\lambda(x)) \), scaled by \( 1/\sqrt{\lambda} \), for \( \tau^* = \{T_\lambda^\beta : \lambda \geq 0\} \) and \( T_\lambda^\beta \) defined in (15). Hence \( \hat{\zeta}(\beta - x, \hat{T}^*(\beta - x)) \) is a non-negative function of \( \beta \), and so
\[
c\beta + E \left[ \hat{\zeta}(\beta - X, \hat{T}^*(\beta - X)) \right] \to \infty \text{ as } \beta \to \infty.
\]
Next, we handle the case that \( \beta \to -\infty \). When \( p < a \), recall from (9) that \( \hat{T}^*(\beta - x) \) solves
\[
\hat{\zeta}(\beta - x, \hat{T}^*(\beta - x)) = -p(\beta - x) + (a - p)\gamma \hat{T}^*(\beta - x),
\]
for any realization \( x \) of \( X \). Then, adding \( c\beta \) to both sides, taking expectations, and recalling that \( E X = 0 \),
\[
c\beta + E \left[ \hat{\zeta}(\beta - X, \hat{T}^*(\beta - X)) \right] = (c - p)\beta + (a - p)\gamma E \left[ \hat{T}^*(\beta - X) \right].
\]
Since \( c - p < 0 \) by assumption and \( \hat{T}^*(\beta - x) \geq 0 \) for any realization \( x \), it follows that
\[
c\beta + E \left[ \hat{\zeta}(\beta - X, \hat{T}^*(\beta - X)) \right] \to \infty \text{ as } \beta \to -\infty,
\]
when \( p < a \).

When \( a \leq p \), we first observe that \( T^*(\beta - x) = \infty \) for any realization \( x \) of \( X \) and so
\[
\hat{\zeta}(\beta - x, T^*(\beta - x)) = \hat{\zeta}(\beta - x, \infty).
\]
Next, we let \( \hat{\zeta}(\beta - x, \hat{T}^*(\beta - x)) \) be the minimum cost and \( \hat{T}^*(\beta - x) \) the optimal threshold of the diffusion control problem with costs \( a \) and \( p \) such that \( p < a = a \leq p \). It follows from the optimality of \( \hat{T}^* \) that, for any realization \( X = x \),
\[
\hat{\zeta}(\beta - x, \hat{T}^*(\beta - x)) \leq \hat{\zeta}(\beta - x, \infty) = \hat{\zeta}(\beta - x, \infty),
\]
For any fixed point \((\beta - x, \infty)\) and \(\hat{z}(\beta - x, \infty)\) do not depend on \(p\) or \(\hat{p}\). Hence
\[
c\beta + E \left[ \hat{z}(\beta - X, \hat{T}^*(\beta - X)) \right] \leq c\beta + E [\hat{z}(\beta - X, \infty)].
\]
Since \(p < a = a\), we can also repeat the same arguments as in the case \(p < a\) to get
\[
c\beta + E \left[ \hat{z}(\beta - X, \hat{T}^*(\beta - X)) \right] \to \infty \text{ as } \beta \to -\infty.
\]
Combining the previous two displays we get
\[
c\beta + E \left[ \hat{z}(\beta - X, \hat{T}^*(\beta - X)) \right] = c\beta + E [\hat{z}(\beta - X, \infty)] \to \infty \text{ as } \beta \to -\infty,
\]
when \(a \leq p\).

**Proof of (b):** Set \(\beta_0 < 0\) and \(|\beta_0|\) arbitrarily large. From the definition of continuity, it is enough to show that if \(\{\beta_n\}\) is a sequence in \((\beta_0, \infty)\) that converges to \(\beta \in (\beta_0, \infty)\) as \(n \to \infty\), then
\[
c\beta_n + E \left[ \hat{z} \left( \beta_n - X, \hat{T}^*(\beta_n - X) \right) \right] \to c\beta + E \left[ \hat{z} \left( \beta - X, \hat{T}^*(\beta - X) \right) \right] \text{ as } n \to \infty. \tag{A.20}
\]
Since \(\hat{z}\) is a continuous function of its arguments (as can be seen immediately from its definition in (8)) and \(\hat{T}^*\) is a continuous function of \(\beta\) by Claim 3, it follows that for any realization \(x\) of \(X\),
\[
\hat{z}(\beta_n - x, \hat{T}^*(\beta_n - x)) \to \hat{z}(\beta - x, \hat{T}^*(\beta - x)) \text{ as } n \to \infty.
\]
For every \(n\), by Claim 4,
\[
\hat{z}(\beta_n - x, \hat{T}^*(\beta_n - x)) \leq \hat{z}(\beta_0 - x, \hat{T}^*(\beta_0 - x)), \text{ for any } x \in (-\infty, \infty).
\]
Since by Proposition 3,
\[
E \left[ \hat{z} \left( \beta_0 - X, \hat{T}^*(\beta_0 - X) \right) \right] < \infty,
\]
the dominated convergence theorem implies that
\[
E \left[ \hat{z} \left( \beta_n - X, \hat{T}^*(\beta_n - X) \right) \right] \to E \left[ \hat{z} \left( \beta - X, \hat{T}^*(\beta - X) \right) \right] \text{ as } n \to \infty,
\]
from which (A.20) follows.

**Proof of Claim 3** Let \(p < a\), and for a fixed realization of \(X\) as \(x\), define
\[
g(\beta, \hat{T}) := (a - p)\gamma\hat{T} - \hat{z}(\beta - x, \hat{T}) - p(\beta - x).
\]
For any fixed point \((\beta_1, \hat{T}^*) \in \mathbb{R} \times [0, \infty)\), (8) shows that \(\hat{z}(\beta - x, \hat{T})\) is differentiable. Hence
\[
\frac{\partial g}{\partial \hat{T}} \bigg|_{\hat{T}=\hat{T}^*} = (a - p)\gamma - \frac{d\hat{z}(\beta - x, \hat{T})}{d\hat{T}} \bigg|_{\hat{T}=\hat{T}^*}.
\]
Recall from (9) and (10) that when \(g(\beta_1, \hat{T}^*) = 0\), then
\[
\hat{z}(\beta_1 - x, \hat{T}^*) \leq \hat{z}(\beta_1 - x, \hat{T}) \text{ for any other } \hat{T} \geq 0.
\]
Hence the first-order conditions imply that

\[
\frac{d\hat{z}(\beta - x, \hat{T})}{dT} \bigg|_{\hat{T}=\hat{T}^*} = 0,
\]

and so

\[
\frac{\partial g}{\partial \hat{T}} \bigg|_{\hat{T}=\hat{T}^*} = (a - p)\gamma.
\]

Because \(a < p\) by assumption, the conditions of the implicit function theorem are satisfied. We use the implicit function theorem to conclude that there exists an open set \(S_{\beta_1}\) containing \(\beta_1\) and an open set \(S_{\hat{T}^*}\), containing \(\hat{T}^*\), and a unique continuously invertible function \(h : S_{\beta_1} \to S_{\hat{T}^*}\) such that

\[
\{(\beta, h(\beta)) : \beta \in S_{\beta_1}\} = \{(\beta, \hat{T}^*) \in S_{\beta_1} \times S_{\hat{T}^*}\}.
\]

The existence of a unique continuously invertible function \(h\) holds for any \(\beta_1 \in \mathbb{R}\), and so the proof is complete.

**Proof of Claim 4:** As in the proof of Claim 1, it follows from a coupling argument that the expected steady-state number of customers waiting in a \(M/M/N/B + M\) queue, as well as the expected steady-state loss proportion, is decreasing as \(N\) increases but \(B\) remains fixed. Next, recall from Section 3 that for any fixed realization \(x\) of \(X\), the long-run average operating cost associated with an admissible routing policy \(\pi\) in the system with mean arrival rate \(\lambda\) is

\[
z_\lambda^*(N, l^\lambda(x)) = pl^\lambda(x)P_{\pi}^\lambda(\text{out}; N, l^\lambda(x)) + a\gamma Q_{\pi}^\lambda(N, l^\lambda(x)).
\]

Therefore, for two staffing policies

\[N_1 = \{\lambda + \beta_1\sqrt{\lambda} : \lambda \geq 0\}\] and \(N_2 = \{\lambda + \beta_2\sqrt{\lambda} : \lambda \geq 0\}\) with \(\beta_1 \leq \beta_2\)

and the identical threshold routing policy (equivalently, the identical \(B\) because the threshold is on the total number of customers in the system) \(\tau = \{T^\lambda : \lambda \geq 0\}\) with

\[T^\lambda = N_1^\lambda + \hat{T}_1^* \sqrt{l^\lambda(x)},\]

for \(\hat{T}_1^*\) that satisfies (9) with \(\beta_1\) replacing \(\beta\), it follows that

\[z_\lambda^*(N_2, l^\lambda(x)) < z_\tau(N_1, l^\lambda(x)).\]

Dividing both sides of the above inequality by \(\sqrt{\lambda}\), taking the limit as \(\lambda \to \infty\), and applying Claim 3 shows that

\[\hat{z} \left(\beta_2 - x, \hat{T}_1^*\right) < \hat{z} \left(\beta_1 - x, \hat{T}_1^*\right),\] (A.21)

Since from (10), \(\hat{T}_2^*\) is a minimizer when \(\beta_2\) replaces \(\beta\) (not \(\beta_1\)),

\[\hat{z} \left(\beta_2 - x, \hat{T}_2^*\right) < \hat{z} \left(\beta_2 - x, \hat{T}_1^*\right),\] (A.22)

The proof is complete from (A.21), (A.22), and the fact that \(\beta_1\) and \(\beta_2\) are arbitrary.

**Proof of Theorem 2:** It follows from Theorem 1 that

\[\hat{C}^\lambda(U) \to \hat{C}^* < \infty, \text{ as } \lambda \to \infty.\]
Therefore, to show asymptotic optimality (see Definition 1), it is enough to show

$$\liminf_{\lambda \to \infty} \hat{C}^\lambda(u) \geq \hat{C}^*$$

(A.23)

under any arbitrary admissible policy

$$u = (N, \pi) = \{(N^\lambda, \pi^\lambda) : \lambda \geq 0\}.$$ 

We first establish (A.23) and then prove

$$cN^\lambda \star + E \left[ \frac{z_{\pi^\lambda}(N^\lambda, \Lambda^\lambda(X))}{\sqrt{\lambda}} - \mathcal{C}^{\lambda, \text{opt}}(\lambda) \right] \to 0, \text{ as } \lambda \to \infty.$$ 

(A.24)

The argument that our proposed policy is asymptotically optimal (that (A.23) holds). We first argue that we need only consider admissible policies under which

$$\liminf_{\lambda \to \infty} \frac{N^\lambda - \lambda}{\sqrt{\lambda}} > -\infty$$

(A.25)

holds. To see this, assume the bound (established at the end of this proof)

$$E \left[ \frac{z_{\pi^\lambda}(N, \Lambda^\lambda(X))}{\sqrt{\lambda}} \right] \geq \min(a, p)(\lambda - N^\lambda)$$

(A.26)

is valid. Then, from the definition of $\hat{C}^\lambda(u)$ and (A.26),

$$\hat{C}^\lambda(u) = c \frac{N^\lambda - \lambda}{\sqrt{\lambda}} + E \left[ \frac{z_{\pi}(N, \Lambda^\lambda(X))}{\sqrt{\lambda}} \right] \geq (\min(a, p) - c) \frac{\lambda - N^\lambda}{\sqrt{\lambda}}.$$ 

If (A.25) does not hold, then it follows from the assumption $\min(a, p) > c$ and the above display that

$$\liminf_{\lambda \to \infty} \hat{C}^\lambda(u) = \infty,$$

which trivially satisfies (A.23). In summary, it is enough to show (A.23) holds for the subset of admissible policies that satisfy (A.25).

Consider any subsequence $\lambda_i$ on which (A.25) holds and also on which the lim inf in (A.23) is attained, so that

$$\lim_{\lambda_i \to \infty} \hat{C}^\lambda_i(u) = \liminf_{\lambda \to \infty} \hat{C}^\lambda(u) < \infty.$$ 

We may assume the limit is finite because otherwise (A.23) holds trivially. On this subsequence, since the limit is finite and

$$\hat{C}^\lambda_i(u) = c \frac{N^\lambda_i - \lambda_i}{\sqrt{\lambda_i}} + E \left[ \frac{z_{\pi}(N, \Lambda^\lambda_i(X))}{\sqrt{\lambda_i}} \right]$$

has its second term positive, it must be the case that

$$\limsup_{\lambda_i \to \infty} \frac{N^\lambda_i - \lambda_i}{\sqrt{\lambda_i}} < \infty.$$
Since

\[-\infty < \liminf_{\lambda_i \to \infty} \frac{N^\lambda_i - \lambda_i}{\sqrt{\lambda_i}} < \limsup_{\lambda_i \to \infty} \frac{N^\lambda_i - \lambda_i}{\sqrt{\lambda_i}} < \infty,\]

the Bolzano-Weierstrass theorem guarantees that any subsequence has a further convergent subsequence \( \lambda_{ij} \) on which

\[
\frac{N^\lambda_{ij} - \lambda_{ij}}{\sqrt{\lambda_{ij}}} \to \beta \in \mathbb{R} \text{ as } \lambda_{ij} \to \infty.
\]  

(A.27)

Since from the properties of the limit

\[
\lim_{\lambda_{ij} \to \infty} \hat{C}(u) \geq c \lim_{\lambda_{ij} \to \infty} \frac{N^\lambda_{ij} - \lambda_{ij}}{\sqrt{\lambda_{ij}}} + \liminf_{\lambda_{ij} \to \infty} \frac{E \left[ z_{\pi}^{\lambda_{ij}}(N, \Lambda^\lambda_{ij}(X)) \right]}{\sqrt{\lambda_{ij}}},
\]

and Fatou’s lemma guarantees

\[
\liminf_{\lambda_{ij} \to \infty} \frac{E \left[ z_{\pi}^{\lambda_{ij}}(N, \Lambda^\lambda_{ij}(X)) \right]}{\sqrt{\lambda_{ij}}} \geq E \left[ \liminf_{\lambda_{ij} \to \infty} \frac{z_{\pi}^{\lambda_{ij}}(N, \Lambda^\lambda_{ij}(X))}{\sqrt{\lambda_{ij}}} \right],
\]

it follows that

\[
\lim_{\lambda_{ij} \to \infty} \hat{C}(u) \geq c\beta + E \left[ \liminf_{\lambda_{ij} \to \infty} \frac{z_{\pi}^{\lambda_{ij}}(N, \Lambda^\lambda_{ij}(X))}{\sqrt{\lambda_{ij}}} \right].
\]  

(A.28)

Next, for any realization \( x \) of \( X \), it is straightforward to see Theorem 5.2 part (ii) in Koçağa and Ward (2010) can be used to conclude

\[
\liminf_{\lambda_{ij} \to \infty} \frac{z_{\pi}^{\lambda_{ij}}(N, I^{\lambda_{ij}}(x))}{\sqrt{\lambda_{ij}}} \geq \hat{z}(\beta - x, \hat{T}^*(\beta - x)),
\]  

(A.29)

because (A.27) implies the conditions of that theorem are satisfied since

\[
\frac{N^\lambda_{ij} - I^{\lambda_{ij}}(x)}{\sqrt{\lambda_{ij}}} = \frac{N^\lambda_{ij} - \lambda_{ij}}{\sqrt{\lambda_{ij}}} - x \to \beta - x, \text{ as } \lambda_{ij} \to \infty.
\]

It follows from (A.28) and (A.29) that

\[
\lim_{\lambda_{ij} \to \infty} \hat{C}^{\lambda_{ij}}(u) \geq c\beta + E \left[ \hat{z}(\beta - X, \hat{T}^*(\beta - X)) \right].
\]

Since the definitions of \( \beta^* \) and \( \hat{C}^* \) imply that

\[
c\beta + E \left[ \hat{z}(\beta - X, \hat{T}^*(\beta - X)) \right] \geq \hat{C}^* = c\beta^* + E \left[ \hat{z}(\beta^* - X, \hat{T}^*(\beta^* - X)) \right],
\]

we conclude that (A.23) is satisfied. Since the subsequence \( \lambda_{ij} \) was arbitrary, the proof is complete once we establish the earlier assumed bound (A.26).

\[
2 \text{ Please see Table A.1 to see how to match the notation between Koçağa and Ward (2010) and this paper.}
\]
To see the bound \( (A.26) \) holds, let \( S^\lambda_{\pi}(N, l^\lambda(x)) \) denote the expected number of busy servers when the realized arrival rate is \( l^\lambda(x) = \lambda + x \sqrt{\lambda} \). Since the arrival rate into the system must equal the departure rate from the system (due to both abandonments and service completions),

\[
l^\lambda(x) \left( 1 - P^\lambda_{\pi}(\text{out}; N, l^\lambda(x)) \right) = S^\lambda_{\pi}(N, l^\lambda(x)) + l^\lambda(x) P^\lambda_{\pi}(\text{ab}; N, l^\lambda(x)),
\]

or, equivalently,

\[
l^\lambda(x) \left( P^\lambda_{\pi}(\text{out}; N, l^\lambda(x)) + P^\lambda_{\pi}(\text{ab}; N, l^\lambda(x)) \right) = l^\lambda(x) - S^\lambda_{\pi}(N, l^\lambda(x)).
\]

Since \( S^\lambda_{\pi}(N, l^\lambda(x)) \leq N^\lambda \), it follows that

\[
l^\lambda(x) \left( P^\lambda_{\pi}(\text{out}; N, l^\lambda(x)) + P^\lambda_{\pi}(\text{ab}; N, l^\lambda(x)) \right) \geq l^\lambda(x) - N^\lambda.
\]

From the definition of \( z^\lambda_{\pi}(N, l^\lambda(x)) \), it follows that

\[
z^\lambda_{\pi}(N, l^\lambda(x)) \geq \min(a, p) l^\lambda(x) \left( P^\lambda_{\pi}(\text{out}; N, l^\lambda(x)) + P^\lambda_{\pi}(\text{ab}; N, l^\lambda(x)) \right).
\]

Hence

\[
z^\lambda_{\pi}(N, l^\lambda(x)) \geq \min(a, p) \left( l^\lambda(x) - N^\lambda \right),
\]

and \( (A.26) \) follows by taking expectations and recalling \( EX = 0 \).

The argument that our proposed policy achieves cost \( o(\sqrt{\lambda}) \) higher than the minimum cost (that \( (A.24) \) holds).

Recall that \( C^\lambda_{\text{opt}} \) is the minimum achievable cost defined in (3) for the system with mean arrival rate \( \lambda \). Since

\[
\frac{c N^\lambda_{\ast} + E \left[ z^\lambda_{\pi^*} \left( N^\ast, \Lambda^\lambda(X) \right) \right]}{\sqrt{\lambda}} - C^\lambda_{\text{opt}} = \hat{C}^\lambda(u^*) - \frac{C^\lambda_{\text{opt}} - c\lambda}{\sqrt{\lambda}},
\]

and from Theorem 1

\[
\hat{C}^\lambda(u^*) \rightarrow \hat{C}^* \text{ as } \lambda \rightarrow \infty,
\]

it is enough to establish

\[
\frac{C^\lambda_{\text{opt}} - c\lambda}{\sqrt{\lambda}} \rightarrow \hat{C}^* \text{ as } \lambda \rightarrow \infty.
\]

(A.30)

It follows from Theorem 2 that

\[
\liminf_{\lambda \rightarrow \infty} \frac{C^\lambda_{\text{opt}} - c\lambda}{\sqrt{\lambda}} \geq \hat{C}^*.
\]

(A.31)

Also, a policy \( u^\text{opt} \) that consists of a sequence \( (N^\lambda_{\text{opt}}, \pi^\lambda_{\text{opt}}) \) in which each element of the sequence is exactly optimal for each \( \lambda \), and so achieves the minimum cost \( C^\text{opt} \), must be asymptotically optimal. Hence, because \( u^\text{opt} \) is an admissible policy, from the definition of asymptotic optimality,

\[
\limsup_{\lambda \rightarrow \infty} \frac{C^\lambda_{\text{opt}} - c\lambda}{\sqrt{\lambda}} = \limsup_{\lambda \rightarrow \infty} \hat{C}^\lambda(u^\text{opt}) \leq \lim_{\lambda \rightarrow \infty} \hat{C}^\lambda(u^*) = \hat{C}^*.
\]

(A.32)

The limit \( (A.30) \) follows from \( (A.31) \) and \( (A.32) \).
B. Supporting Numerical Tables

In this section we provide detailed numerical results in tabular format which support our findings in Section 6. We set the mean service time and the mean patience time equal to 1 and fix the cost parameters at $c = 0.1$, $p = 1$ and $a = 5$ unless specified otherwise.

Table B.1 provides the details of the numerical study shown in Figure 1. The first column shows the Uniform arrival rate distribution which has its mean $\lambda$ fixed at 100 and has an increasing CV as we go down in the table. The second column shows the optimal staffing level. Columns three for and five show the staffing level, associated cost and cost percentage error of $U$, respectively. Columns six, seven, and eight report the same numbers for the first alternative policy, $D$, while columns nine, ten and eleven report the same numbers for the second alternative policy $NV$.

Tables B.2, B.3 and B.4 provide details for the numerical studies associated with Figures 2a and 3a, Figures 2b and 3b, and Figures 2c and 3c, respectively. The first column shows the varying staffing cost while the other columns are the same as in Table B.1. Table B.5 provides the $\beta^*$ values used in Tables B.2-B.4.

Tables B.6, B.7 and B.8 provide details for the numerical studies associated with Figures 4a and 5a, Figures 4b and 5b, and Figures 4c and 5c, respectively. The first column shows the parameters of the Beta distribution that $\Lambda$ follows while the other columns are the same as before.

<table>
<thead>
<tr>
<th>Arrival rate distribution</th>
<th>$N_{\text{opt}}$</th>
<th>$U$</th>
<th>$D$</th>
<th>$NV$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Lambda = 100$</td>
<td>119</td>
<td>119</td>
<td>12.41</td>
<td>0.01%</td>
</tr>
<tr>
<td>$U[99, 101]$</td>
<td>119</td>
<td>119</td>
<td>12.41</td>
<td>0.01%</td>
</tr>
<tr>
<td>$U[90, 110]$</td>
<td>121</td>
<td>121</td>
<td>12.71</td>
<td>0.01%</td>
</tr>
<tr>
<td>$U[80, 120]$</td>
<td>127</td>
<td>126</td>
<td>13.38</td>
<td>0.03%</td>
</tr>
<tr>
<td>$U[70, 130]$</td>
<td>133</td>
<td>132</td>
<td>14.16</td>
<td>0.07%</td>
</tr>
<tr>
<td>$U[60, 140]$</td>
<td>140</td>
<td>139</td>
<td>14.97</td>
<td>0.05%</td>
</tr>
<tr>
<td>$U[50, 150]$</td>
<td>147</td>
<td>146</td>
<td>15.82</td>
<td>0.06%</td>
</tr>
<tr>
<td>$U[40, 160]$</td>
<td>155</td>
<td>154</td>
<td>16.67</td>
<td>0.03%</td>
</tr>
<tr>
<td>$U[30, 170]$</td>
<td>162</td>
<td>161</td>
<td>17.54</td>
<td>0.05%</td>
</tr>
<tr>
<td>$U[20, 180]$</td>
<td>170</td>
<td>169</td>
<td>18.42</td>
<td>0.04%</td>
</tr>
<tr>
<td>$U[10, 190]$</td>
<td>178</td>
<td>176</td>
<td>19.30</td>
<td>0.06%</td>
</tr>
</tbody>
</table>

Table B.1: Performance of $U$ vs other policies: Varying CV (Figure 1 in the main body)
### Table B.2: Performance of $U$ vs other policies: Varying staffing cost ($c$) under low arrival rate variability (Figures 2a and 3a in the main body)

<table>
<thead>
<tr>
<th>Staffing cost ($c$)</th>
<th>$N_{opt}^{U}$</th>
<th>$U$</th>
<th>$D$</th>
<th>$NV$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_U$</td>
<td>$C(N_U)$</td>
<td>% error</td>
<td>$N_D$</td>
</tr>
<tr>
<td>0.01</td>
<td>134</td>
<td>132</td>
<td>1.38</td>
<td>0.31%</td>
</tr>
<tr>
<td>0.05</td>
<td>126</td>
<td>125</td>
<td>6.54</td>
<td>0.06%</td>
</tr>
<tr>
<td>0.1</td>
<td>121</td>
<td>121</td>
<td>12.71</td>
<td>0.01%</td>
</tr>
<tr>
<td>0.2</td>
<td>116</td>
<td>116</td>
<td>24.56</td>
<td>0.02%</td>
</tr>
<tr>
<td>0.3</td>
<td>112</td>
<td>112</td>
<td>35.93</td>
<td>0.02%</td>
</tr>
<tr>
<td>0.4</td>
<td>108</td>
<td>108</td>
<td>46.91</td>
<td>0.03%</td>
</tr>
<tr>
<td>0.5</td>
<td>104</td>
<td>105</td>
<td>57.51</td>
<td>0.08%</td>
</tr>
<tr>
<td>0.6</td>
<td>100</td>
<td>101</td>
<td>67.72</td>
<td>0.08%</td>
</tr>
<tr>
<td>0.7</td>
<td>95</td>
<td>96</td>
<td>77.48</td>
<td>0.05%</td>
</tr>
<tr>
<td>0.8</td>
<td>89</td>
<td>90</td>
<td>86.71</td>
<td>0.07%</td>
</tr>
<tr>
<td>0.9</td>
<td>75</td>
<td>78</td>
<td>94.99</td>
<td>0.35%</td>
</tr>
<tr>
<td>0.95</td>
<td>59</td>
<td>63</td>
<td>98.38</td>
<td>0.75%</td>
</tr>
<tr>
<td>0.99</td>
<td>1</td>
<td>8</td>
<td>101.01</td>
<td>0.01%</td>
</tr>
</tbody>
</table>

### Table B.3: Performance of $U$ vs other policies: Varying staffing cost ($c$) under moderate arrival rate variability (Figures 2b and 3b in the main body)

<table>
<thead>
<tr>
<th>Staffing cost ($c$)</th>
<th>$N_{opt}^{U}$</th>
<th>$U$</th>
<th>$D$</th>
<th>$NV$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_U$</td>
<td>$C(N_U)$</td>
<td>% error</td>
<td>$N_D$</td>
</tr>
<tr>
<td>0.01</td>
<td>170</td>
<td>165</td>
<td>1.77</td>
<td>1.15%</td>
</tr>
<tr>
<td>0.05</td>
<td>156</td>
<td>154</td>
<td>8.24</td>
<td>0.18%</td>
</tr>
<tr>
<td>0.1</td>
<td>147</td>
<td>146</td>
<td>15.82</td>
<td>0.06%</td>
</tr>
<tr>
<td>0.2</td>
<td>134</td>
<td>134</td>
<td>29.85</td>
<td>0.04%</td>
</tr>
<tr>
<td>0.3</td>
<td>122</td>
<td>123</td>
<td>42.64</td>
<td>0.04%</td>
</tr>
<tr>
<td>0.4</td>
<td>111</td>
<td>112</td>
<td>54.28</td>
<td>0.10%</td>
</tr>
<tr>
<td>0.5</td>
<td>100</td>
<td>102</td>
<td>64.83</td>
<td>0.21%</td>
</tr>
<tr>
<td>0.6</td>
<td>89</td>
<td>91</td>
<td>74.28</td>
<td>0.35%</td>
</tr>
<tr>
<td>0.7</td>
<td>79</td>
<td>81</td>
<td>82.69</td>
<td>0.37%</td>
</tr>
<tr>
<td>0.8</td>
<td>68</td>
<td>71</td>
<td>90.05</td>
<td>0.31%</td>
</tr>
<tr>
<td>0.9</td>
<td>56</td>
<td>58</td>
<td>96.27</td>
<td>0.21%</td>
</tr>
<tr>
<td>0.95</td>
<td>44</td>
<td>45</td>
<td>98.81</td>
<td>0.14%</td>
</tr>
<tr>
<td>0.99</td>
<td>0</td>
<td>0</td>
<td>100.00</td>
<td>0.00%</td>
</tr>
</tbody>
</table>
Table B.4: Performance of $U$ vs other policies: Varying staffing cost ($c$) under high arrival rate variability (Figures 2c and 3c in the main body)

<table>
<thead>
<tr>
<th>Staffing cost ($c$)</th>
<th>$N_{opt}$</th>
<th>$U$</th>
<th>$D$</th>
<th>$NV$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_U$</td>
<td>$C(N_U)$</td>
<td>% error</td>
<td>$N_D$</td>
</tr>
<tr>
<td>0.01</td>
<td>209</td>
<td>202</td>
<td>2.18</td>
<td>1.49%</td>
</tr>
<tr>
<td>0.05</td>
<td>191</td>
<td>188</td>
<td>10.10</td>
<td>0.16%</td>
</tr>
<tr>
<td>0.1</td>
<td>178</td>
<td>176</td>
<td>19.30</td>
<td>0.06%</td>
</tr>
<tr>
<td>0.2</td>
<td>156</td>
<td>156</td>
<td>35.97</td>
<td>0.03%</td>
</tr>
<tr>
<td>0.3</td>
<td>136</td>
<td>138</td>
<td>50.60</td>
<td>0.09%</td>
</tr>
<tr>
<td>0.4</td>
<td>117</td>
<td>119</td>
<td>63.29</td>
<td>0.23%</td>
</tr>
<tr>
<td>0.5</td>
<td>99</td>
<td>101</td>
<td>74.10</td>
<td>0.29%</td>
</tr>
<tr>
<td>0.6</td>
<td>80</td>
<td>83</td>
<td>83.03</td>
<td>0.22%</td>
</tr>
<tr>
<td>0.7</td>
<td>61</td>
<td>65</td>
<td>90.12</td>
<td>0.17%</td>
</tr>
<tr>
<td>0.8</td>
<td>43</td>
<td>47</td>
<td>95.35</td>
<td>0.13%</td>
</tr>
<tr>
<td>0.9</td>
<td>24</td>
<td>28</td>
<td>98.71</td>
<td>0.08%</td>
</tr>
<tr>
<td>0.95</td>
<td>15</td>
<td>15</td>
<td>99.67</td>
<td>0.02%</td>
</tr>
<tr>
<td>0.99</td>
<td>0</td>
<td>0</td>
<td>100.00</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

Table B.5: $\beta^*$ values for changing staffing costs and arrival rate variabilities given in Tables B.2-B.4

<table>
<thead>
<tr>
<th>Staffing cost ($c$)</th>
<th>Arrival Rate Variability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low (Table B.2)</td>
</tr>
<tr>
<td>0.01</td>
<td>3.2164</td>
</tr>
<tr>
<td>0.05</td>
<td>2.5108</td>
</tr>
<tr>
<td>0.10</td>
<td>2.1109</td>
</tr>
<tr>
<td>0.20</td>
<td>1.5948</td>
</tr>
<tr>
<td>0.30</td>
<td>1.1972</td>
</tr>
<tr>
<td>0.40</td>
<td>0.8368</td>
</tr>
<tr>
<td>0.50</td>
<td>0.4777</td>
</tr>
<tr>
<td>0.60</td>
<td>0.0881</td>
</tr>
<tr>
<td>0.70</td>
<td>-0.3778</td>
</tr>
<tr>
<td>0.80</td>
<td>-1.0220</td>
</tr>
<tr>
<td>0.90</td>
<td>-2.2158</td>
</tr>
<tr>
<td>0.95</td>
<td>-3.6768</td>
</tr>
<tr>
<td>0.99</td>
<td>-9.2008</td>
</tr>
<tr>
<td>Arrival rate distribution</td>
<td>$N^\text{opt}$</td>
</tr>
<tr>
<td>---------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td></td>
<td>$N_U$</td>
</tr>
<tr>
<td>Beta(1.5, 0.5)</td>
<td>121</td>
</tr>
<tr>
<td>Beta(1.4, 0.6)</td>
<td>121</td>
</tr>
<tr>
<td>Beta(1.3, 0.7)</td>
<td>121</td>
</tr>
<tr>
<td>Beta(1.2, 0.8)</td>
<td>121</td>
</tr>
<tr>
<td>Beta(1.1, 0.9)</td>
<td>121</td>
</tr>
<tr>
<td>Beta(1.0, 1.0)</td>
<td>121</td>
</tr>
<tr>
<td>Beta(0.9, 1.1)</td>
<td>122</td>
</tr>
<tr>
<td>Beta(0.8, 1.2)</td>
<td>122</td>
</tr>
<tr>
<td>Beta(0.7, 1.3)</td>
<td>121</td>
</tr>
<tr>
<td>Beta(0.6, 1.4)</td>
<td>121</td>
</tr>
<tr>
<td>Beta(0.5, 1.5)</td>
<td>121</td>
</tr>
</tbody>
</table>

Table B.6: Performance of $U$ vs other policies: Varying skewness under low arrival rate variability (Figures 4a and 5a in the main body)

<table>
<thead>
<tr>
<th>Arrival rate distribution</th>
<th>$N^\text{opt}$</th>
<th>$U$</th>
<th>$D$</th>
<th>$NV$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_U$</td>
<td>$C(N_U)$</td>
<td>% error</td>
<td>$N_D$</td>
</tr>
<tr>
<td>Beta(1.5, 0.5)</td>
<td>140</td>
<td>139</td>
<td>14.73</td>
<td>0.06%</td>
</tr>
<tr>
<td>Beta(1.4, 0.6)</td>
<td>142</td>
<td>140</td>
<td>14.97</td>
<td>0.14%</td>
</tr>
<tr>
<td>Beta(1.3, 0.7)</td>
<td>144</td>
<td>142</td>
<td>15.18</td>
<td>0.07%</td>
</tr>
<tr>
<td>Beta(1.2, 0.8)</td>
<td>145</td>
<td>144</td>
<td>15.38</td>
<td>0.05%</td>
</tr>
<tr>
<td>Beta(1.1, 0.9)</td>
<td>146</td>
<td>145</td>
<td>15.60</td>
<td>0.05%</td>
</tr>
<tr>
<td>Beta(1.0, 1.0)</td>
<td>147</td>
<td>146</td>
<td>15.82</td>
<td>0.06%</td>
</tr>
<tr>
<td>Beta(0.9, 1.1)</td>
<td>149</td>
<td>147</td>
<td>16.04</td>
<td>0.06%</td>
</tr>
<tr>
<td>Beta(0.8, 1.2)</td>
<td>150</td>
<td>149</td>
<td>16.27</td>
<td>0.04%</td>
</tr>
<tr>
<td>Beta(0.7, 1.3)</td>
<td>150</td>
<td>149</td>
<td>16.52</td>
<td>0.06%</td>
</tr>
<tr>
<td>Beta(0.6, 1.4)</td>
<td>151</td>
<td>150</td>
<td>16.78</td>
<td>0.04%</td>
</tr>
<tr>
<td>Beta(0.5, 1.5)</td>
<td>151</td>
<td>151</td>
<td>17.06</td>
<td>0.03%</td>
</tr>
</tbody>
</table>

Table B.7: Performance of $U$ vs other policies: Varying skewness under moderate arrival rate variability (Figures 4b and 5b in the main body)

<table>
<thead>
<tr>
<th>Arrival rate distribution</th>
<th>$N^\text{opt}$</th>
<th>$U$</th>
<th>$D$</th>
<th>$NV$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$N_U$</td>
<td>$C(N_U)$</td>
<td>% error</td>
<td>$N_D$</td>
</tr>
<tr>
<td>Beta(1.1, 0.9)</td>
<td>175</td>
<td>173</td>
<td>18.85</td>
<td>0.07%</td>
</tr>
<tr>
<td>Beta(1.0, 1.0)</td>
<td>178</td>
<td>176</td>
<td>19.30</td>
<td>0.06%</td>
</tr>
<tr>
<td>Beta(0.9, 1.1)</td>
<td>180</td>
<td>179</td>
<td>19.76</td>
<td>0.05%</td>
</tr>
<tr>
<td>Beta(0.8, 1.2)</td>
<td>182</td>
<td>181</td>
<td>20.24</td>
<td>0.04%</td>
</tr>
<tr>
<td>Beta(0.7, 1.3)</td>
<td>184</td>
<td>183</td>
<td>20.75</td>
<td>0.04%</td>
</tr>
<tr>
<td>Beta(0.6, 1.4)</td>
<td>186</td>
<td>185</td>
<td>21.29</td>
<td>0.03%</td>
</tr>
<tr>
<td>Beta(0.5, 1.5)</td>
<td>187</td>
<td>186</td>
<td>21.87</td>
<td>0.04%</td>
</tr>
</tbody>
</table>

Table B.8: Performance of $U$ vs other policies: Varying skewness under high arrival rate variability (Figures 4c and 5c in the main body)
Electronic Companion References


