e-companion: Routing and staffing in customer service chat systems with impatient customers

Tolga Tezcan
University of Rochester, Simon Graduate School of Business
tolga.tezcan@simon.rochester.edu

Jiheng Zhang
The Hong Kong University of Science and Technology
j.zhang@ust.hk

We prove Lemma 1 in §EC1. We present the queueing equations and some preliminary analysis in §EC2. Then we prove Theorem 1 in §EC3 and Theorem 2 in §EC4. Proof of Proposition 1 is placed in §EC5 and that of Theorem 4 in §EC6. The proofs of the results in Section 5 are presented in §EC7. Finally, proofs of supplemental results appear in §EC8.

EC1. Proof of Lemma 1

We first state some preliminary results. We note that

$$P_i^{Ab}\hat{d}_i = i\nu \tag{EC1}$$

hence $P_i^{Ab}\hat{d}_i$ is increasing in i. Also by (3)

$$P_{j'}^{Ab} > P_j^{Ab} \text{ for } j' > j. \tag{EC2}$$

The proof of the following result is elementary hence skipped.

Lemma EC1. *If for a level j, $1 < j < I$, $\hat{d}_j = \hat{d}_{j'}$ for some $j' < j$ then level j cannot be efficient. Also, for any efficient level i_j*

$$\left(1 - P_{i_j}^{Ab}\right)\hat{d}_{i_j} \geq \left(1 - P_i^{Ab}\right)\hat{d}_i, \tag{EC3}$$

for all $i \leq i_j$.

Remark EC1 (Equivalent definitions). It can easily be checked that if a level i is inefficient than the following hold for some $i_1 \le i \le i_2$

$$\left(P_{i_2}^{Ab}\hat{d}_{i_2} - P_{i_1}^{Ab}\hat{d}_{i_1}\right)\hat{d}_i \le \left(P_{i_2}^{Ab}\hat{d}_{i_2} - P_i^{Ab}\hat{d}_i\right)\hat{d}_{i_1} + \left(P_i^{Ab}\hat{d}_i - P_{i_1}^{Ab}\hat{d}_{i_1}\right)\hat{d}_{i_2}, \tag{EC4}$$

$$\left(\hat{d}_{i_2} - \hat{d}_{i_1}\right)\left(P_{i_2}^{Ab}\hat{d}_{i_2} - P_i^{Ab}\hat{d}_i\right) \le \left(\hat{d}_{i_2} - \hat{d}_i\right)\left(P_{i_2}^{Ab}\hat{d}_{i_2} - P_{i_1}^{Ab}\hat{d}_{i_1}\right), \tag{EC5}$$

and

$$\left(\hat{d}_{i_2} - \hat{d}_{i_1}\right)P_i^{Ab}\hat{d}_i \ge \left(\hat{d}_i - \hat{d}_{i_1}\right)P_{i_2}^{Ab}\hat{d}_{i_2} + \left(\hat{d}_{i_2} - \hat{d}_i\right)P_{i_1}^{Ab}\hat{d}_{i_1}. \tag{EC6}$$

Proof of Lemma 1: The routing LP (10) can be written as

$$\min \sum_{i=1}^{I} P_i^{Ab}\lambda_i + \lambda_{I+1} \tag{EC7}$$

$$st.$$

$$\sum_{i=1}^{I} \frac{\lambda_i}{\hat{d}_i} \le N, \tag{EC8}$$

$$\sum_{i=1}^{I} \lambda_i + \lambda_{I+1} \ge \lambda, \tag{EC9}$$

$$\lambda_i \ge 0, \quad i = 1, 2, \ldots, I, I+1. \tag{EC10}$$

Let y_N, y_λ and y_i denote the dual variables associated with the constraints (EC8), (EC9) and (EC10), respectively. We can write the dual of the routing LP, referred to as LP-D, as

$$\begin{aligned}
&\max \lambda y_\lambda - N y_N \\
&st. \\
&y_\lambda - \frac{1}{\hat{d}_i} y_N + y_i \le P_i^{Ab}, \quad i = 1, 2, \ldots, I \\
&y_\lambda + y_{I+1} \le 1, \quad i = 1, 2, \ldots, I \\
&y_\lambda \ge 0, \\
&y_N \ge 0, \\
&y_i \ge 0, \quad i = 1, 2, \ldots, I+1.
\end{aligned}$$

LP-D is equivalent to the following LP, referred to as LP-DE,

$$\max \frac{\lambda}{N} y_\lambda - y_N$$

$$st.$$

$$\hat{d}_i y_\lambda - y_N \le P_i^{Ab}\hat{d}_i, \quad i = 1, 2, \ldots, I, \tag{EC11}$$

$$y_\lambda \le 1, \tag{EC12}$$

$$y_\lambda \ge 0, \tag{EC13}$$

$$y_N \geq 0. \tag{EC14}$$

For the rest of the proof we focus on establishing an optimal solution of LP-DE. Because an optimal solution of LP-D can easily be derived from an optimal solution of LP-DE, we also refer to such a solution as an optimal solution of LP-D, with a slight abuse of terminology.

We divide the proof into a few cases depending on the value of the parameters N and λ. First, assume that $\frac{\lambda}{N} < \hat{d}_1$. Consider the solution $(y_\lambda^*, y_N^*) = (P_1^{Ab}, 0)$ with objective value $\frac{\lambda}{N} P_1^{Ab}$ for LP-DE. We claim that it is an optimal solution of LP-DE. Note first that it is feasible by (EC2). In addition, by (EC11) for $i = 1$, for any feasible solution (y_λ, y_N) the objective function is bounded by

$$\frac{\lambda}{N}\left(P_1^{Ab} + \frac{y_N}{\hat{d}_1}\right) - y_N \leq \frac{\lambda}{N} P_1^{Ab},$$

where the last inequality follows from the fact that $\frac{\lambda}{N} < \hat{d}_1$. Thus, $(y_\lambda^*, y_N^*) = (P_1^{Ab}, 0)$ is an optimal solution of LP-DE and LP-D. Note also that the feasible solution given by $\lambda_1 = \lambda$, and $\lambda_i = 0$, $i = 2, \ldots, I, I+1$ for the routing LP (EC7) attains the same objective value with the optimal objective function value of LP-D, therefore it must be an optimal solution of (10), proving (15).

Next, assume that $\frac{\lambda}{N} = \hat{d}_{i^*}$ for an efficient level i^*. By (EC4) and (EC5) and the fact that level i^* is efficient for any $i_1 < i^* < i_2$ and $i_1, i_2 \in \mathcal{N}^c$

$$\frac{P_{i^*}^{Ab}\hat{d}_{i^*} - P_{i_1}^{Ab}\hat{d}_{i_1}}{\hat{d}_{i^*} - \hat{d}_{i_1}} \leq \frac{P_{i_2}^{Ab}\hat{d}_{i_2} - P_{i_1}^{Ab}\hat{d}_{i_1}}{\hat{d}_{i_2} - \hat{d}_{i_1}} \leq \frac{P_{i_2}^{Ab}\hat{d}_{i_2} - P_{i^*}^{Ab}\hat{d}_{i^*}}{\hat{d}_{i_2} - \hat{d}_{i^*}}, \tag{EC15}$$

where by convention $\cdot/0 = \infty$ and $\hat{d}_{i_2} \geq \hat{d}_{i^*} > \hat{d}_{i_1}$ by Lemma EC1 and the fact that $i_1, i_2 \in \mathcal{N}^c$. Let y_λ^* be such that

$$\frac{P_{i^*}^{Ab}\hat{d}_{i^*} - P_{i_1}^{Ab}\hat{d}_{i_1}}{\hat{d}_{i^*} - \hat{d}_{i_1}} \leq y_\lambda^* \leq \frac{P_{i_2}^{Ab}\hat{d}_{i_2} - P_{i^*}^{Ab}\hat{d}_{i^*}}{\hat{d}_{i_2} - \hat{d}_{i^*}}, \quad \text{for all } i_1, i_2 \in \mathcal{N}^c, \quad i_1 < i^* \text{ and } i_2 > i^*. \tag{EC16}$$

The existence of such y_λ^* is guaranteed by the fact that level i^* is assumed to be efficient and by (EC15). Also, set

$$y_N^* = \hat{d}_{i^*} y_\lambda^* - P_{i^*}^{Ab}\hat{d}_{i^*}. \tag{EC17}$$

First note that by the fact that level I is efficient, $\hat{d}_I > \hat{d}_{i^*}$ and by (EC16) with $i_2 = I$, $y_\lambda^* \leq 1$. Using (EC16) and (EC17) it is easily checked that (y_λ^*, y_N^*) satisfies (EC11) for all $i \in \mathcal{N}^c$. Also, for $i \in \mathcal{N}$, (y_λ^*, y_N^*) satisfies (EC11) by (EC1), the definition of $\mathcal{N}$ and the fact that (y_λ^*, y_N^*) satisfies (EC11) for $i \in \mathcal{N}^c$. Hence, (y_λ^*, y_N^*) is a feasible solution of LP-DE with the objective function value

$P^{Ab}_{i^*}\hat{d}_{i^*}$. The optimality of (y^*_λ, y^*_N) for LP-DE (and so for LP-D) follows from the fact that for any feasible solution (y_λ, y_N) the objective function satisfies

$$\frac{\lambda}{N}y_\lambda - y_N = \hat{d}_{i^*}y_\lambda - y_N \le P^{Ab}_{i^*}\hat{d}_{i^*}$$

by our assumption that $\frac{\lambda}{N} = \hat{d}_{i^*}$. Note also that the feasible solution given by $\lambda_{i^*} = \lambda$, $\lambda_i = 0$ for $i \ne i^*$ for the routing LP (EC7) attains the same objective value as the optimal objective function value of LP-D, therefore it must be an optimal solution of (10), proving (16) in this case.

Now assume that $\hat{d}_I N > \lambda > \hat{d}_1 N$ and that $\frac{\lambda}{N} \ne \hat{d}_{i^*}$ for any $i^* \in \mathcal{F}$. Let i^*_{j+1} be defined as in (14) and define

$$(y^*_\lambda, y^*_N) = \left(\frac{P^{Ab}_{i^*_{j+1}}\hat{d}_{i^*_{j+1}} - P^{Ab}_{i^*_j}\hat{d}_{i^*_j}}{\hat{d}_{i^*_{j+1}} - \hat{d}_{i^*_j}}, \frac{P^{Ab}_{i^*_{j+1}}\hat{d}_{i^*_{j+1}}\hat{d}_{i^*_j} - P^{Ab}_{i^*_j}\hat{d}_{i^*_j}\hat{d}_{i^*_{j+1}}}{\hat{d}_{i^*_{j+1}} - \hat{d}_{i^*_j}} \right). \tag{EC18}$$

Also define

$$\delta = \frac{\hat{d}_{i^*_{j+1}} - \frac{\lambda}{N}}{\hat{d}_{i^*_{j+1}} - \hat{d}_{i^*_j}}.$$

Using (EC11) for i^*_j and i^*_{j+1}, we have for any feasible solution (y_λ, y_N)

$$\frac{\lambda}{N}y_\lambda - y_N = \delta(\hat{d}_{i^*_j}y_\lambda - y_N) + (1-\delta)(\hat{d}_{i^*_{j+1}}y_\lambda - y_N) \le \frac{\hat{d}_{i^*_{j+1}} - \frac{\lambda}{N}}{\hat{d}_{i^*_{j+1}} - \hat{d}_{i^*_j}}P^{Ab}_{i^*_j}\hat{d}_{i^*_j} + \frac{\frac{\lambda}{N} - \hat{d}_{i^*_j}}{\hat{d}_{i^*_{j+1}} - \hat{d}_{i^*_j}}P^{Ab}_{i^*_{j+1}}\hat{d}_{i^*_{j+1}}.$$

On the other hand, it is easy to verify that (y^*_λ, y^*_N) defined in (EC18) achieves this best possible value. Therefore it is enough to show that (y^*_λ, y^*_N) is a feasible solution for LP-DE to conclude that it is an optimal solution for LP-DE.

By the monotonicity property (EC1) and Lemma EC1, (y^*_λ, y^*_N) satisfies the constraints (EC12), (EC13) and (EC14). Also, it satisfies (EC11) for $i = i^*_j$ and $i = i^*_{j+1}$. In fact, these two constraints are tight at this point. We next prove that (y^*_λ, y^*_N) satisfies (EC11) for $i \ne i^*_j, i^*_{j+1}$. For $i < i^*_j$,

$$\begin{aligned}\hat{d}_i y^*_\lambda - y^*_N &= \hat{d}_{i^*_{j+1}}y^*_\lambda - y^*_N - (\hat{d}_{i^*_{j+1}} - \hat{d}_i)y^*_\lambda \\ &= P^{Ab}_{i^*_{j+1}}\hat{d}_{i^*_{j+1}} - (\hat{d}_{i^*_{j+1}} - \hat{d}_i)\left(\frac{P^{Ab}_{i^*_{j+1}}\hat{d}_{i^*_{j+1}} - P^{Ab}_{i^*_j}\hat{d}_{i^*_j}}{\hat{d}_{i^*_{j+1}} - \hat{d}_{i^*_j}}\right).\end{aligned} \tag{EC19}$$

Since level i^*_j is efficient, by (EC5) and Lemma EC1

$$\frac{P^{Ab}_{i^*_{j+1}}\hat{d}_{i^*_{j+1}} - P^{Ab}_{i^*_j}\hat{d}_{i^*_j}}{\hat{d}_{i^*_{j+1}} - \hat{d}_{i^*_j}} \ge \frac{P^{Ab}_{i^*_{j+1}}\hat{d}_{i^*_{j+1}} - P^{Ab}_i\hat{d}_i}{\hat{d}_{i^*_{j+1}} - \hat{d}_i}.$$

This implies by (EC19) that $\hat{d}_i y^*_\lambda - y^*_N \le P^{Ab}_i\hat{d}_i$. So (EC11) is satisfied for any $i < i^*_j$ by (y^*_λ, y^*_N). For $i > i^*_{j+1}$, a similar argument applies by using the efficiency of level i^*_{j+1}. Specifically, by (EC5)

$$(\hat{d}_i - \hat{d}_{i^*_j})P^{Ab}_{i^*_{j+1}}\hat{d}_{i^*_{j+1}} \le (\hat{d}_{i^*_{j+1}} - \hat{d}_{i^*_j})P^{Ab}_i\hat{d}_i + (\hat{d}_i - \hat{d}_{i^*_{j+1}})P^{Ab}_{i^*_j}\hat{d}_{i^*_j}. \tag{EC20}$$

Then

$$\begin{aligned}\hat{d}_i y^*_\lambda - y^*_N &= \hat{d}_{i^*_j} y^*_\lambda - y^*_N + (\hat{d}_i - \hat{d}_{i^*_j}) y^*_\lambda \\ &= P^{Ab}_{i^*_j}\hat{d}_{i^*_j} + (\hat{d}_i - \hat{d}_{i^*_j}) \left(\frac{P^{Ab}_{i^*_{j+1}}\hat{d}_{i^*_{j+1}} - P^{Ab}_{i^*_j}\hat{d}_{i^*_j}}{\hat{d}_{i^*_{j+1}} - \hat{d}_{i^*_j}} \right) \le P^{Ab}_i \hat{d}_i,\end{aligned}$$

where the last inequality follows from (EC20).

If $\mathcal{U}_{i^*_j} = \emptyset$, then the proof is complete. Also, for all $i \in \mathcal{N}$, (y^*_λ, y^*_N) satisfies (EC11), as described above if (y^*_λ, y^*_N) satisfies (EC11) for all $i \in \mathcal{N}^c$. So next we consider $\mathcal{U}_{i_j} \cap \mathcal{N}^c$. For any $i \in \mathcal{U}_{i^*_j} \cap \mathcal{N}^c$, let

$$\delta_{j+1} = \frac{P^{Ab}_{i^*_{j+1}}\hat{d}_{i^*_{j+1}} - P^{Ab}_i \hat{d}_i}{P^{Ab}_{i^*_{j+1}}\hat{d}_{i^*_{j+1}} - P^{Ab}_{i^*_j}\hat{d}_{i^*_j}}$$

and $\delta_j = 1 - \delta_{j+1}$. By (EC1), $\delta_j, \delta_{j+1} \in [0,1]$. It follows from Lemma EC2 below and (EC6) that

$$\hat{d}_i \le \delta_{j+1}\hat{d}_{i^*_j} + \delta_j \hat{d}_{i^*_{j+1}}. \tag{EC21}$$

Then

$$\begin{aligned}\hat{d}_i y^*_\lambda - y^*_N &\le \delta_{j+1}\left(\hat{d}_{i^*_j} y^*_\lambda - y^*_N\right) + \delta_j \left(\hat{d}_{i^*_{j+1}} y^*_\lambda - y^*_N\right) \\ &\le \delta_{j+1} P^{Ab}_{i^*_j}\hat{d}_{i^*_j} + \delta_j P^{Ab}_{i^*_{j+1}}\hat{d}_{i^*_{j+1}} \\ &= P^{Ab}_i \hat{d}_i,\end{aligned}$$

where the first inequality follows from (EC21) and the second inequality follows from the fact that (y^*_λ, y^*_N) satisfies the constraint (EC11) for i^*_j and i^*_{j+1}. So (y^*_λ, y^*_N) satisfies the constraint (EC11) for all $1 \le i \le I$. Therefore, (y^*_λ, y^*_N) is an optimal solution for LP-DE and LP-D. In addition, the feasible solution given by (16) for the routing LP (EC7) attains the same objective value as the optimal objective value of LP-D, therefore it must be an optimal solution of (10), proving (16) in this case.

Finally, we study the case where $\lambda \ge \hat{d}_I N$. Consider the solution $(y^*_\lambda, y^*_N) = (1, \hat{d}_I(1 - P^{Ab}_I))$ with objective value $\frac{\lambda}{N} - \hat{d}_I(1 - P^{Ab}_I)$ for LP-DE. We claim that it is an optimal solution of LP-DE. Note first that it is feasible by (6). In addition, by (EC11) for $i = I$ and (EC12), for any feasible solution (y_λ, y_N) the objective function is bounded by $\frac{\lambda}{N} - \hat{d}_I(1 - P^{Ab}_I)$. Thus, (y^*_λ, y^*_N) is an optimal solution of LP-DE and LP-D. Note also that the feasible solution given by $\lambda_I = \hat{d}_I N$, $\lambda_{I+1} = \lambda - \hat{d}_I N$, and $\lambda_i = 0$, $i = 1, \ldots, I-1$ for the routing LP (EC7) attains the same objective value with the optimal objective function value of LP-D, therefore it must be an optimal solution of (10), proving (17).

□

EC1.1. Proof of (EC21)

We now focus on the inefficient levels in $\mathcal{U}_{i_j^*} \cap \mathcal{N}^c$ (when $\mathcal{U}_{i_j} \neq \emptyset$) for an efficient level i_j. For notational simplicity we assume $\mathcal{N} = \emptyset$ and use only the notation $\mathcal{U}_{i_j^*}$ the rest of this section. (We do not make use of this assumption in the proof.) For an inefficient level i, we denote the set of all pairs (i_1, i_2) with $i_1 < i < i_2$ that satisfy (EC5) by $\mathcal{J}(i)$. For any $i \in \mathcal{U}_{i_j}$, by the definition of inefficiency, there exists a pair $(k, j) \in \mathcal{J}(i)$. However, the definition itself does not require (i_j, i_{j+1}) to be one of the pairs in $\mathcal{J}(i)$. In this section we prove that $(i_j, i_{j+1}) \in \mathcal{J}(i)$.

Let $a_i = P_i^{Ab} \hat{d}_i$ and consider the set of pairs $\{(\hat{d}_i, a_i) : i = \{1, 2, \ldots, I\}\}$ on a two-dimensional plane. Assume that $\hat{d}_k \neq \hat{d}_j$ and let $L^{k,j} : \mathbb{R} \to \mathbb{R}$ denote the (straight) line passing through points $(\hat{d}_k, a_k)$ and $(\hat{d}_j, a_j)$, hence,

$$L^{k,j}(x) = \frac{a_j - a_k}{\hat{d}_j - \hat{d}_k} x - \frac{a_j - a_k}{\hat{d}_j - \hat{d}_k} \hat{d}_k + a_k, \ x \in \mathbb{R}.$$

From here on when we consider $L^{k,j}$ we assume without further mention that $\hat{d}_k \neq \hat{d}_j$. For $k < i < j$, we have

$$L^{k,j}(\hat{d}_i) = a_j \frac{\hat{d}_i - \hat{d}_k}{\hat{d}_j - \hat{d}_k} + a_k \frac{\hat{d}_j - \hat{d}_i}{\hat{d}_j - \hat{d}_k}.$$

If a level i is inefficient then by (EC6) either $\hat{d}_k = \hat{d}_i$ for some $k < i$ or

$$a_i \geq L^{k,j}(\hat{d}_i) \tag{EC22}$$

for some $(k, j) \in J(i)$.

LEMMA EC2. *Let $i_j \in \mathcal{F}$. For any $i \in \mathcal{U}_{i_j}$ (EC21) holds.*

Note that (EC21) is equivalent to

$$a_i \geq L^{i_j, i_{j+1}}(\hat{d}_i). \tag{EC23}$$

We use the following results in the proof that are immediate from algebraic manipulations. Let $k < i$ and $k < j$ and assume that $L^{k,i}(x) \leq L^{k,j}(x)$ for some $x > \hat{d}_k$, then

$$L^{k,i}(x) \leq L^{k,j}(x) \text{ for all } x > \hat{d}_k \tag{EC24}$$

Similarly, if for $k < i$ and $j < i$ $L^{k,i}(x) \leq L^{j,i}(x)$ for some $x < \hat{d}_i$, then

$$L^{k,i}(x) \leq L^{j,i}(x) \text{ for all } x < \hat{d}_i. \tag{EC25}$$

Let $i < j$ and $k < l$ and assume that there exists $x_1 > x_0$ such that $L^{i,j}(x) \leq L^{k,l}(x)$ for some $x < x_0$, and $L^{i,j}(x) \geq L^{k,l}(x)$ for some $x > x_1$, then

$$L^{i,j}(x) \leq L^{k,l}(x) \text{ for all } x < x_0, \quad \text{and} \quad L^{i,j}(x) \geq L^{k,l}(x) \text{ for all } x > x_1. \tag{EC26}$$

Proof of Lemma EC2: We prove the result using an induction argument. Because level i is inefficient, by Lemma EC3 below, $(k_1, i_{j+1}) \in \mathcal{J}(i)$ for some $k_1 < i$. Hence, by (EC22) (note that by Lemma EC1 and the fact that i_{j+1} is efficient we have $\hat{d}_i < \hat{d}_{i_{j+1}}$ for all $i < i_{j+1}$)

$$a_i \geq L^{k_1, i_{j+1}}(\hat{d}_i). \tag{EC27}$$

Now we proceed by induction. Let $i = i_j + 1$. Assume that $k_1 < i_j$ as otherwise the proof is complete. Because level i_j is efficient

$$a_{i_j} < L^{k_1, i_{j+1}}(\hat{d}_{i_j}). \tag{EC28}$$

Since $a_{i_j} = L^{i_j, i_{j+1}}(\hat{d}_{i_j})$, we have $L^{i_j, i_{j+1}}(x) \leq L^{k_1, i_{j+1}}(x)$ for all $x \leq \hat{d}_{i_{j+1}}$ by (EC25). The result then follows from (EC27).

Now assume that for $i \in \mathcal{U}_{i_j}$, $i > i_j + 1$ and

$$a_{i'} \geq L^{i_j, i_{j+1}}(\hat{d}_{i'})$$

for all $i_j + 1 \leq i' \leq i - 1$. By Lemma EC3, there exists $1 \leq k_1 < i$ such that $(k_1, i_{j+1}) \in \mathcal{J}(i)$. If $k_1 \leq i_j$, the result follows as above using (EC28). Now assume that $k_1 > i_j$. By the induction argument

$$a_{k_1} = L^{k_1, i_{j+1}}(\hat{d}_{k_1}) \geq L^{i_j, i_{j+1}}(\hat{d}_{k_1}).$$

Therefore, by (EC25)

$$L^{k_1, i_{j+1}}(x) \geq L^{i_j, i_{j+1}}(x)$$

for $x < \hat{d}_{i_{j+1}}$. Then (EC23) follows from (EC27). □

Lemma EC3. *Let $i_j \in \mathcal{F}$. For any $i \in \mathcal{U}_{i_j}$,*

$$a_i \geq L^{k, i_{j+1}}(\hat{d}_i) \tag{EC29}$$

for some $1 \leq k < i$.

Proof of Lemma EC3: Assume that level i is inefficient. Then by (EC6) either $\hat{d}_i = \hat{d}_j$ for some $j < i$ or

$$a_i \geq L^{j_1, j_2}(\hat{d}_i). \tag{EC30}$$

for some $j_1 < i < j_2$.

First assume that $\hat{d}_i = \hat{d}_{i'}$ for some $i' < i$. Then, by Lemma EC1 $\hat{d}_{i_{j+1}} > \hat{d}_i = \hat{d}_{i'}$. Hence, we have

$$a_i \geq L^{i',i_{j+1}}(\hat{d}_i)$$

by (EC24) completing the proof in this case.

Now assume that $\hat{d}_i > \hat{d}_{i-1}$ and that (EC30) holds. It follows from (EC24) and Lemma EC1 that $\hat{d}_{j_2} > \hat{d}_i$. We prove the result in this case recursively by induction. First suppose $i = i_{j+1} - 1$. In this case, if $j_2 = i_{j+1}$ the result follows. Otherwise, we must have $j_2 > i_{j+1}$. The fact that level i_{j+1} is efficient implies

$$L^{j_1,i_{j+1}}(\hat{d}_{i_{j+1}}) = a_{i_{j+1}} < L^{j_1,j_2}(\hat{d}_{i_{j+1}}),$$

hence by (EC24)

$$L^{j_1,i_{j+1}}(x) \leq L^{j_1,j_2}(x), \text{ for all } x > \hat{d}_{j_1}. \tag{EC31}$$

The result (EC29) with $k = j_1$ follows from (EC30) and (EC31) (with $x = \hat{d}_i$ since $\hat{d}_i > \hat{d}_{j_1}$).

Now suppose $i < i_{j+1} - 1$ and by induction that for all i' with $i < i' < i_{j+1}$

$$a_{i'} \geq L^{k_1(i'),i_{j+1}}(\hat{d}_{i'}) \tag{EC32}$$

for some $k_1(i') < i'$. By the discussion above, if $\hat{d}_i = \hat{d}_{i-1}$ the proof is immediate, therefore we focus on the case when $\hat{d}_i > \hat{d}_{i-1}$. If $\hat{d}_i > \hat{d}_{i-1}$ then it follows from (EC24) and Lemma EC1 that $\hat{d}_{j_2} > \hat{d}_i$. Also, by Lemma EC1 and the fact that i_{j+1} is efficient we have $\hat{d}_{j_2} < \hat{d}_{i_{j+1}}$. Hence for the rest of the proof the following holds

$$\hat{d}_{i-1} < \hat{d}_i < \hat{d}_{j_2} < \hat{d}_{i_{j+1}}. \tag{EC33}$$

If $j_2 \geq i_{j+1}$, then the result (EC29) (with $k = j_1$) follows in a similar way to the discussion leading to (EC31). So we now focus on the case where $j_2 < i_{j+1}$. We can assume the pair (j_1, j_2) we pick is the one such that $j_2 = \min\{j : (j_1, j) \in \mathcal{J}(i)\}$. By the induction assumption (EC32), for this j_2

$$a_{j_2} \geq L^{k_1(j_2),i_{j+1}}(\hat{d}_{j_2}). \tag{EC34}$$

There are three different cases to be analyzed separately.

(i) $k_1(j_2) < j_1 < i < j_2$

(ii) $j_1 \leq k_1(j_2) < i < j_2$

(iii) $j_1 < i \leq k_1(j_2) < j_2$

Case (i): If $a_i \geq L^{k_1(j_2),i_{j+1}}(\hat{d}_i)$, then result (EC29) with $k = k_1(j_2)$ follows immediately. So we focus on the case where

$$a_i < L^{k_1(j_2),i_{j+1}}(\hat{d}_i). \tag{EC35}$$

We show below that in this case,

$$a_{j_1} \leq L^{k_1(j_2),i_{j+1}}(\hat{d}_{j_1}). \tag{EC36}$$

We now show the result (EC29) holds with $k = j_1$ using (EC35) and (EC36). Figure 1 gives a graphic demonstration of the argument. Let $\tilde{L}^{j_1,j_2}$ denote the line that passes through the points $(\hat{d}_{j_1}, a_{j_1})$ and $(\hat{d}_{j_2}, L^{k_1(j_2),i_{j+1}}(\hat{d}_{j_2}))$. Since (EC34) holds, by (EC24), we have that

$$L^{j_1,j_2}(x) \geq \tilde{L}^{j_1,j_2}(x), \text{ for all } x \geq \hat{d}_{j_1}. \tag{EC37}$$

On the other hand, since (EC36) holds, by (EC25), we have that

$$L^{k_1(j_2),i_{j+1}}(x) \geq L^{j_1,i_{j+1}}(x), \text{ for all } x \leq \hat{d}_{i_{j+1}}.$$

Plugging $x = \hat{d}_{j_2}$ in the above yields $L^{k_1(j_2),i_{j+1}}(\hat{d}_{j_2}) = \tilde{L}^{j_1,j_2}(\hat{d}_{j_2}) \geq L^{j_1,i_{j+1}}(\hat{d}_{j_2})$. This, again by (EC24), implies that

$$\tilde{L}^{j_1,j_2}(x) \geq L^{j_1,i_{j+1}}(x), \text{ for all } x \geq \hat{d}_{j_1}. \tag{EC38}$$

The result follows from (EC30), (EC37) and (EC38) (with $x = \hat{d}_i$ since $\hat{d}_i \geq \hat{d}_{j_1}$).

To complete the proof in Case (i), we next prove (EC36). We first show that (EC34) and the opposite of (EC36) implies that

$$L^{j_1,j_2}(x) \geq L^{k_1(j_2),i_{j+1}}(x), \text{ for all } \hat{d}_{j_1} \leq x \leq \hat{d}_{j_2}. \tag{EC39}$$

(In other words, the line segment connecting $(\hat{d}_{j_1}, a_{j_1})$ and $(\hat{d}_{j_2}, a_{j_2})$ lies above the line passing through $(\hat{d}_{k_1(j_2)}, a_{k_1(j_2)})$ and $(\hat{d}_{i_{j+1}}, a_{i_{j+1}})$.) Let $\tilde{L}^{j_1,j_2}$ be defined as above. By (EC34) and (EC24)

$$\tilde{L}^{j_1,j_2}(x) \leq L^{j_1,j_2}(x), \text{ for all } x \geq \hat{d}_{j_1} \tag{EC40}$$

and if (EC36) does not hold then by (EC25)

$$\tilde{L}^{j_1,j_2}(x) \geq L^{k_1(j_2),i_{j+1}}(x), \text{ for all } x \leq \hat{d}_{j_2} \tag{EC41}$$

We have (EC39) by (EC40) and (EC41). By (EC39) and (EC30), we must have that $a_i \geq L^{k_1(j_2),i_{j+1}}(\hat{d}_i)$, contradicting (EC35), hence (EC36) holds.

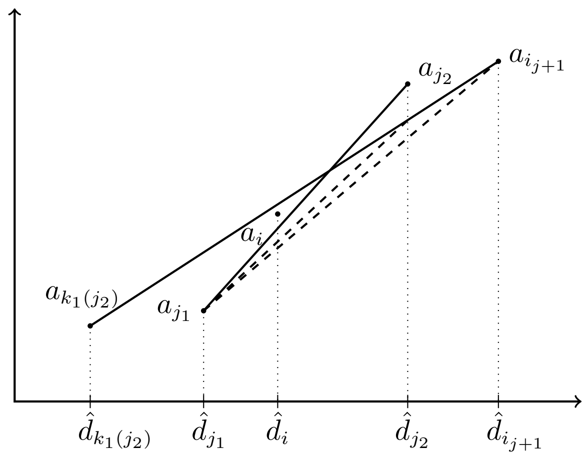

Figure 1 **Graphic Presentation in Case (i).**

Case (ii): If $a_{k_1(j_2)} < L^{j_1,j_2}(\hat{d}_{k_1(j_2)})$, then this and (EC34) implies that

$$L^{j_1,j_2}(x) \ge L^{k_1(j_2),i_{j+1}}(x), \text{ for all } \hat{d}_{k_1(j_2)} \le x \le \hat{d}_{j_2}. \tag{EC42}$$

Since $\hat{d}_{k_1(j_2)} \le \hat{d}_i \le \hat{d}_{j_2}$, the result (EC29) with $k = k_1(j_2)$ follows immediately from (EC30) and from plugging $x = \hat{d}_i$ in (EC42).

Now assume that $a_{k_1(j_2)} \ge L^{j_1,j_2}(\hat{d}_{k_1(j_2)})$. This implies that

$$L^{k_1(j_2),i_{j+1}}(\hat{d}_{k_1(j_2)}) \ge L^{j_1,j_2}(\hat{d}_{k_1(j_2)}).$$

It follows from (EC34) that

$$L^{k_1(j_2),i_{j+1}}(\hat{d}_{j_2}) \le L^{j_1,j_2}(\hat{d}_{j_2}).$$

By (EC26),

$$L^{k_1(j_2),i_{j+1}}(x) \ge L^{j_1,j_2}(x), \text{ for all } x \le \hat{d}_{k_1(j_2)}.$$

In particular, the above holds for $x = \hat{d}_{j_1}$ since $\hat{d}_{j_1} \le \hat{d}_{k_1(j_2)}$. Thus, by (EC25),

$$L^{k_1(j_2),i_{j+1}}(x) \ge L^{j_1,i_{j+1}}(x), \text{ for all } x \le \hat{d}_{i_{j+1}}.$$

Since $\hat{d}_{j_2} \le \hat{d}_{i_{j+1}}$, the above inequality holds for $x = \hat{d}_{j_2}$, this together with (EC34) implies that $a_{j_2} \ge L^{j_1,i_{j+1}}(\hat{d}_{j_2})$. By (EC24),

$$L^{j_1,j_2}(x) \ge L^{j_1,i_{j+1}}(x), \text{ for all } x \ge \hat{d}_{j_1}. \tag{EC43}$$

The result (EC29) with $k = j_1$ follows from (EC30) and (EC43) (with $x = \hat{d}_i$ since $\hat{d}_i \ge \hat{d}_{j_1}$).

Case (iii): Since level j_2 is the smallest indexed level such that $(j_1, j_2) \in \mathcal{J}(i)$ we must have

$$a_i \le L^{j_1,k_1(j_2)}(\hat{d}_i). \tag{EC44}$$

Also note that $\hat{d}_{i_{j+1}} > \hat{d}_{k_1(j_2)} \ge \hat{d}_i > \hat{d}_{j_1}$ by (EC33). (EC44) together with (EC30) implies $L^{j_1,k_1(j_2)}(\hat{d}_i) \ge L^{j_1,j_2}(\hat{d}_i)$. Since $\hat{d}_{k_1(j_2)} \ge \hat{d}_i$ by (EC24),

$$a_{k_1(j_2)} = L^{k_1(j_2),i_{j+1}}(\hat{d}_{k_1(j_2)}) = L^{j_1,k_1(j_2)}(\hat{d}_{k_1(j_2)}) \ge L^{j_1,j_2}(\hat{d}_{k_1(j_2)}).$$

On the other hand, (EC34) implies that

$$L^{k_1(j_2),i_{j+1}}(\hat{d}_{j_2}) \le L^{j_1,j_2}(\hat{d}_{j_2}).$$

By (EC26), we have

$$L^{k_1(j_2),i_{j+1}}(x) \le L^{j_1,j_2}(x) \text{ for all } x \ge \hat{d}_{j_2}.$$

In particular, the above inequality holds for $x = \hat{d}_{i_{j+1}}$ since $\hat{d}_{i_{j+1}} \ge \hat{d}_{j_2}$. Thus, by (EC24), (EC43) holds in this case as well. The result (EC29) with $k = j_1$ follows from (EC30) and (EC43) (with $x = \hat{d}_i$ since $\hat{d}_i \ge \hat{d}_{j_1}$). $\square$

EC2. Queueing equations

In this section we provide the details of the queueing equations for the CSC systems and set the notation for the rest of results in the appendix.

EC2.1. Notation

All random variables and processes are defined on a common probability space $(\Omega, \mathcal{G}, P)$ unless specified otherwise. The symbols $\mathbb{N}, \mathbb{R}$ and $\mathbb{R}_+$ are used to denote nonnegative integers, real numbers and nonnegative real numbers, respectively. For $d \in \mathbb{N}$, $\mathbb{R}^d$ denotes the d-dimensional Euclidean space; thus, $\mathbb{R} = \mathbb{R}^1$. The space of functions $f : \mathbb{R}_+ \to \mathbb{R}^d$ that are right-continuous on $[0, \infty)$ and have left limits in $(0, \infty)$ is denoted by $\mathbb{D}(\mathbb{R}_+, \mathbb{R}^d)$ or simply $\mathbb{D}^d$; similarly, with $T > 0$, the space of functions $f : [0, T] \to \mathbb{R}^d$ that are right-continuous on $[0, T)$ and have left limits in $(0, T]$ is denoted by $\mathbb{D}([0, T], \mathbb{R}^d)$. For $f \in \mathbb{D}^d$, $f(t-)$ denotes its left limit at $t > 0$. Each stochastic process whose sample paths are in $\mathbb{D}^d$ is considered to be a $\mathbb{D}^d$-valued random element. The space $\mathbb{D}^d$ is assumed to be endowed with the u.o.c. topology (see Billingsley (1968)). For a function $f : \mathbb{R} \to \mathbb{R}^d$ with d being some positive integer, we say that t is a regular point of f if f is differentiable at t and use $\dot{f}(t)$ to denote its derivative at t. We use $E_\upsilon[\cdot]$ to denote the conditional expectation and $P_\upsilon\{\cdot\}$ to denote the conditional probability given that $(Q(0), Z(0))$ is distributed according to υ. Similarly, we use $E_x[\cdot]$ to denote the conditional expectation and $P_x\{\cdot\}$ to denote the conditional probability given that $(Q(0), Z(0)) = x$.

EC2.2. Queueing Equations

In this section we introduce the queueing equations for the CSC systems. Fix λ and let $Z_i^\lambda(t)$ denote the number of agents serving i customers and $Q^\lambda(t)$ denote the number of customers in queue at time t. Also, let $\{S_i : i=1,\dots,I+1\}$ denote a set of independent rate 1 Poisson processes and $A^\lambda(t)$ denote the number of arrivals at the system by time t which is also assumed to be independent of $\{S_i : i=1,\dots,I+1\}$. We denote by $A_i^\lambda(t)$ the number of customers who are routed to an agent serving i customers for $i=0,1,\dots,I-1$ and by $A_I^\lambda(t)$ the number of customers who are routed to the queue upon arrival by time t in the λth system. We set $Z^\lambda=(Z_i^\lambda(0); i=0,1,\dots,I,\ t\geq 0)$ and $\mathcal{A}^\lambda=(A^\lambda(t), A_i^\lambda(t); i=0,1,\dots,I,\ t\geq 0)$. Let

$$D_i^\lambda(t)=S_i\left(\hat{d}_i\int_0^t Z_i^\lambda(s)ds\right),\ i=1,\dots,I$$

and

$$D_{I+1}^\lambda(t)=S_{I+1}\left(\gamma\int_0^t Q^\lambda(s)ds\right)$$

denote the number of customers who leave the system while receiving service from an agent at level i and the number of abandonments from queue by time t, respectively. The following queueing equations are satisfied under any policy for $t\geq 0$.

$$Z_0^\lambda(t)=Z_0^\lambda(0)-A_0^\lambda(t)+D_1^\lambda(t), \tag{EC45}$$

$$Z_i^\lambda(t)=Z_i^\lambda(0)+A_{i-1}^\lambda(t)-A_i^\lambda(t)+D_{i+1}^\lambda(t)-D_i^\lambda(t),\ i=1,\dots,I-2, \tag{EC46}$$

$$Z_{I-1}^\lambda(t)=Z_{I-1}^\lambda(0)+A_{I-2}^\lambda(t)-A_{I-1}^\lambda(t)+L_q^\lambda(t)-D_{I-1}^\lambda(t), \tag{EC47}$$

$$Q^\lambda(t)=Q^\lambda(0)+A_I^\lambda(t)-D_I^\lambda(t)-D_{I+1}^\lambda(t)+L_q^\lambda(t), \tag{EC48}$$

$$\sum_{i=0}^{I}A_i^\lambda(t)=A^\lambda(t), \tag{EC49}$$

$$Z_i^\lambda(t)\geq 0, A_i^\lambda(0)=0 \text{ and } A_i^\lambda \text{ is nondecreasing},\ i=0,1,\dots,I. \tag{EC50}$$

$$\sum_{i=0}^{I}Z_i^\lambda(t)=N^\lambda, \tag{EC51}$$

where

$$L_q^\lambda(t)=\int_0^t \mathbb{1}\left\{Q^\lambda(s)=0\right\}dD_I^\lambda(s).$$

By (EC51), $Z_I^\lambda(t)=N^\lambda-\sum_{i=0}^{I-1}Z_i^\lambda(t)$, for all $t\geq 0$, hence

$$D_I^\lambda(t)=S_I\left(\hat{d}_I\left(N^\lambda t-\int_0^t\sum_{i=0}^{I-1}Z_i^\lambda(s)ds\right)\right).$$

Also, an arrival can be routed to an agent at level i only when there is at least one agent at that level, hence,

$$\int_0^t \mathbb{1}\left\{Z_i^\lambda(s-)=0\right\} dA_i^\lambda(s)=0,\ t\geq 0.$$

In addition, because we focus on non-idling policies, customers wait in the queue only when all agents are at level I, therefore

$$Q^\lambda(t)\left(\sum_{i=0}^{I-1} Z_i^\lambda(t)\right)=0,\ t\geq 0. \tag{EC52}$$

Additional equations under the proposed policy: Note that the policy proposed in §4.1 is a static priority policy once the basic levels i_j^λ and i_{j+1}^λ are fixed. For levels with indices below i_j^λ, those with lower indices have higher priorities. Levels in $\mathcal{U}_{i_j}^\lambda$ have lower priorities than those levels with indices below i_j^λ but higher priorities than i_j^λ. Also their priorities among those levels in this set are set in the reverse order of their indices. All levels with indices larger than i_{j+1}^λ (including i_{j+1}^λ) have lower priorities than levels with indices lower than i_{j+1}^λ. Also for two levels with indices larger than i_{j+1}^λ, levels in $\mathcal{N}'$ have lower priorities than those levels that are not in $\mathcal{N}'$. Levels that are not in $\mathcal{N}'$ have decreasing priority in their indices and levels that are in $\mathcal{N}'$ have increasing priority in their indices. We can re-index the levels using $\{v_i^\lambda\}$ so that v_i^λ denotes the priority index of level i and $v_i^\lambda < v_{i'}^\lambda$ if level i has priority over i'. We note that under the proposed policy, for $0<t_1<t_2$, we have

$$A_i^\lambda(t_2)-A_i^\lambda(t_1)=0,\quad \text{if} \sum_{\{i':v_{i'}^\lambda<v_i^\lambda\}} Z_{i'}^\lambda(s)>0,\ \text{for all } s\in[t_1,t_2],\ i=1,\ldots,I-1. \tag{EC53}$$

EC3. Proof of Theorem 1

Fix λ and let π^λ denote a non-idling policy. We first prove that under π^λ there exists a unique steady state distribution $\upsilon(\pi^\lambda)$ for the processes (Z^λ, Q^λ). We prove this result using Foster's criteria. Choose $K>0$ large enough so that $\gamma(K-N^\lambda)+I\hat{d}_I N^\lambda-\lambda>-c$, for some $c>0$. We say a state (Z,Q) is feasible if (EC51) and (EC52) are satisfied. (Throughout the proof we use the notation defined in §EC2.) Given a (feasible) state (Z,Q), with $Z=(Z_0,Z_1,\ldots,Z_I)$, where Z_i is the number of agents at level i and Q is the queue length, define $f(Z,Q)=\sum_{i=0}^I Z_i+Q$. Then, for any state (Z,Q) with $f(Z,Q)>K$, for the generator of the underlying Markov chain Γ, we have

$$\Gamma f(Z,Q)=\lambda-\hat{d}_I N^\lambda-\gamma Q<-c$$

because the policy is assumed to be non-idling. Therefore (Z^λ,Q^λ) is positive recurrent for each λ by Foster's criteria (see for example Meyn and Tweedie (2009)).

Because under any non-idling policy there exists a unique stationary distribution, by Meyn and Tweedie (2009) and (EC45)–(EC48), we have for each λ

$$P^{\lambda,\pi^\lambda}(Ab) = \lim_{T\to\infty} \frac{Ab^\lambda(T)}{T\lambda} = \frac{E_{\upsilon(\pi^\lambda)}\left[\sum_{i=1}^I i\nu Z_i^\lambda(0) + \gamma Q^\lambda(0)\right]}{\lambda}. \tag{EC54}$$

Obviously, under $\upsilon(\pi^\lambda)$

$$\sum_{i=1}^I Z_i^\lambda(0) \le N. \tag{EC55}$$

Because the underlying chain is ergodic

$$\lim_{T\to\infty} \frac{D_i^\lambda(T)}{T} = E_{\upsilon(\pi^\lambda)}\left[\hat{d}_i Z_i^\lambda(0)\right] \text{ a.s.,}$$

for $i = 0, 1, \ldots, I$ and

$$\lim_{T\to\infty} \frac{D_{I+1}^\lambda(T)}{T} = E_{\upsilon(\pi^\lambda)}\left[\gamma Q^\lambda(0)\right] \text{ a.s.}$$

In addition, by stationarity

$$\lambda = \sum_{i=0}^I E_{\upsilon(\pi^\lambda)}\left[\hat{d}_i Z_i^\lambda(0)\right] + E_{\upsilon(\pi^\lambda)}\left[\gamma Q^\lambda(0)\right]. \tag{EC56}$$

Consider the following LP

$$\begin{aligned}
&\min_{\{Q, Z_i; i=0,1,\ldots,I\}} P^{Ab} \\
&st.\ \gamma Q + \sum_{i=1}^I i\nu Z_i = P^{Ab}\lambda, \\
&\sum_{i=0}^I Z_i \le N^\lambda, \\
&\gamma Q + \sum_{i=1}^I \hat{d}_i Z_i \ge \lambda, \\
&Q, Z_i \ge 0,\ \ i = 0, 1, \ldots, I.
\end{aligned} \tag{EC57}$$

Denote the optimal solution of this LP by $\tilde{P}^{Ab}(\lambda, N^\lambda)$. By (EC54), (EC55), and (EC56), we have that

$$P^{\lambda,\pi^\lambda}(Ab) \ge \tilde{P}^{Ab}(\lambda, N^\lambda).$$

Also, LP (EC57) is equivalent to LP (10), which can be seen by setting $\lambda_i = \hat{d}_i Z_i$ and $\lambda_{I+1} = \gamma Q$ there and using (2). Therefore,

$$P^{\lambda,\pi^\lambda}(Ab) \ge P^{Ab}(1, \lambda^{-1}N^\lambda). \tag{EC58}$$

Then taking the limit on both sides of (EC58) as λ goes to infinity and using the fact

$$\lim_{\lambda\to\infty} P^{Ab}(1,\lambda^{-1}N^{\lambda}) = P^{Ab}(1,N),$$

which follows from (9), and the fact that the optimal solution of the LP is a continuous function of its constraints gives the desired result. □

EC4. Proof of Theorem 2

In this section we prove Theorem 2. (Throughout the proof we use the notation defined in §EC2.) The proof is based on analyzing the (fluid) limits of the fluid scaled processes $(n^{-1}Z^n, n^{-1}Q^n)$. Specifically we find the steady state of the fluid limits by establishing the fluid model equations. Then we show that the steady state of the fluid scaled processes are tight. Then we combine these results to complete the proof. For the case when there is only one basic level in the limit the proof is slightly different but the idea is similar.

We analyze the fluid limits of the queuing processes under the proposed policies in §EC4.1 and then prove the tightness of the stationary distributions of the fluid scaled processes in §EC4.2. We present the proof in §EC4.3

EC4.1. Analysis of the fluid model

In this section we analyze the fluid limits of the chat systems under the proposed policies. First we establish the fluid model equations that are satisfied by the fluid limits and then find the steady state of the fluid model solutions.

Consider the asymptotic regime where (9) holds. Consider the following fluid model equations.

$$\bar{Z}_0(t) = \bar{Z}_0(0) - \bar{A}_0(t) + \hat{d}_1 \int_0^t \bar{Z}_1(s)ds, \tag{EC59}$$

$$\bar{Z}_i(t) = \bar{Z}_i(0) + \bar{A}_{i-1}(t) - \bar{A}_i(t) + \hat{d}_{i+1} \int_0^t \bar{Z}_{i+1}(s)ds - \hat{d}_i \int_0^t \bar{Z}_i(s)ds,\ i = 1,\dots,I-2, \tag{EC60}$$

$$\bar{Z}_{I-1}(t) = \bar{Z}_{I-1}(0) + \bar{A}_{I-2}(t) - \bar{A}_{I-1}(t) + \bar{L}_q(t) - \hat{d}_{I-1} \int_0^t \bar{Z}_{I-1}(s)ds, \tag{EC61}$$

$$\bar{Q}(t) = \bar{Q}(0) + \bar{A}_I(t) - \hat{d}_I \int_0^t \bar{Z}_I(s)ds - \gamma \int_0^t \bar{Q}(s)ds + \bar{L}_q(t), \tag{EC62}$$

$$\sum_{i=0}^{I} \dot{\bar{A}}_i(t) = \dot{\bar{A}}(t) = 1, \tag{EC63}$$

$$\bar{Z}_I(t) = N - \sum_{i=0}^{I-1} \bar{Z}_i(t), \tag{EC64}$$

$$\sum_{\{k: v_k \le v_i\}} \dot{\bar{A}}_k(t) = 1 \text{ if } \bar{Z}_{i'}(t) > 0, \text{ for some } i' \text{ with } v_i \ge v_{i'},\ i = 0,1,\dots,I-1, \tag{EC65}$$

$$\dot{\bar{L}}_q(t) = 0 \text{ if } \bar{Q}(t) > 0, \tag{EC66}$$

$$\bar{Q}(t)\sum_{i=0}^{I-1}\bar{Z}_i(t)=0, \tag{EC67}$$

$$\bar{Z}_i(t)\ge 0, \bar{Q}(t)\ge 0,\ i=0,1,\ldots,I, \tag{EC68}$$

$$\bar{A}_i, \bar{L}_q \text{ are non-decreasing},\ i=0,1,\ldots,I, \tag{EC69}$$

where the priorities of levels are found based on $i^*(1,N)$ as explained in §EC2.2. We refer to (EC59)–(EC69) as the fluid model and to $(\bar{Q},\bar{Z},\bar{\mathcal{A}},\bar{L}_q)$ with $\bar{Z}=\big(\bar{Z}_i(t); i=0,1,\ldots,I,\ t\ge 0\big)$ and $\bar{\mathcal{A}}=(\bar{A}(t),\bar{A}_i(t); i=0,1,\ldots,I,\ t\ge 0)$ which satisfies these equations as a fluid model solution. It can be shown as in Dai and Tezcan (2011) that every fluid model solution is Lipschitz and so differentiable a.e.

Because the proposed policy depends on the values of λ and N^λ, if $i^*(1,N)$ has only one element, there may be a "discontinuity" in the proposed policies along the sequence of systems. Specifically, the proposed policy $\pi^{\lambda,*}$ may be different for each λ and may fluctuate between different policies as $\lambda\to\infty$. We prove Theorem 2 separately for the cases when such discontinuity is present and when it is not. Under the following assumption we will show that the proposed policy is independent of λ for large λ.

Assumption 1. *One of the following conditions holds*

i) The set $i^*(1,N)$ has two elements;

ii) The set $i^*(1,N)$ has only one element and all the levels are efficient;

iii) $N^\lambda=\lambda N$ for some $N\ge 0$ and for λ large enough.

If Assumption 1 holds, one can work with the fluid model to complete the proof of Theorem 2. Otherwise, one needs to work with the fluid limits as described in Proposition 2. We next show that the fluid model is obtained from the underlying queueing system equations. The fluid scaling of the queueing processes are defined as $\bar{Z}^\lambda(t)=\lambda^{-1}Z^\lambda(t)$, $\bar{Q}^\lambda(t)=\lambda^{-1}Q^\lambda(t)$, $\bar{\mathcal{A}}^\lambda(t)=\lambda^{-1}\mathcal{A}^\lambda(t)$ and $\bar{L}_q^\lambda(t)=\lambda^{-1}L_q^\lambda(t)$. Assume that

$$\big(\bar{Z}^\lambda(0),\bar{Q}^\lambda(0)\big)\to\big(\bar{Z}(0),\bar{Q}(0)\big)\ \text{ as } \lambda\to\infty \text{ a.s.} \tag{EC70}$$

with $\bar{Z}(0)=\big(\bar{Z}_i(0); i=0,1,\ldots,I\big)$.

Proposition 1. *Consider a sequence of chat service systems and assume that* (9), (EC70) *and Assumption 1 hold. Then the sequence $\{\big(\bar{Z}^\lambda,\bar{Q}^\lambda,\bar{\mathcal{A}}^\lambda,\bar{L}_q^\lambda\big)\}$ is tight a.s. in the Skorohod space endowed with the u.o.c. topology and every limit $\big(\bar{Z},\bar{Q},\bar{\mathcal{A}},\bar{L}_q\big)$ is a fluid model solution.*

Proof of Proposition 1: Consider a sequence of chat service systems and assume that (9), (EC70) and Assumption 1 hold. The proof of the tightness of the sequence $\{(\bar{Z}^\lambda, \bar{Q}^\lambda, \bar{\mathcal{A}}^\lambda, \bar{L}_q^\lambda)\}$ is standard (see Dai and Tezcan (2011)) hence we skip the details. Also, the fact that fluid limits satisfy the fluid model equations can be proved as in there. We also note that all the fluid model equations (EC59)–(EC69) except (EC65) are satisfied by the fluid limits under any sequence of non-idling policies. We next show that the fluid limits under the proposed policy satisfy the policy specific equation (EC65).

Fix a fluid limit $(\bar{Z}, \bar{Q}, \bar{\mathcal{A}}, \bar{L}_q)$ and assume that $\bar{Z}_i(t) > 0$ for some $i \leq I-1$. Then, there exists a subsequence, denoted again by λ and $\omega \in \Omega$ such that

$$(\bar{Z}^\lambda, \bar{Q}^\lambda, \bar{\mathcal{A}}^\lambda, \bar{L}_q^\lambda) \to (\bar{Z}, \bar{Q}, \bar{\mathcal{A}}, \bar{L}_q), \text{ u.o.c. as } \lambda \to \infty. \tag{EC71}$$

By the continuity of the fluid model solutions, (EC71) implies that

$$\bar{Z}_i^\lambda(s) > 0 \tag{EC72}$$

for $s \in [t-\delta, t+\delta]$, for some $\delta > 0$ small enough and for all λ large enough. Under Assumption 1(i), by the continuity of an LP on its constraints and because there can be at most two basic levels, for λ large enough we have $i^*(\lambda, N^\lambda) = i^*(1, N)$. Under Assumption 1(iii) we have $i^*(\lambda, N^\lambda) = i^*(1, N)$ for all λ large enough. Under Assumption 1(ii) the set of basic levels may be different for two different λ's, however, because all the levels are efficient the policy is a strict priority rule giving priority to levels with a lower index for each λ. Condition (EC65) then follows from (EC49), (EC53), (EC71) and (EC72). $\square$

Steady state of the fluid limits: Next we establish the steady state of the fluid model. Fix N and consider the optimal solution $\lambda^*(1,N) = (\lambda_i^*(1,N); i = 0, 1, \ldots, I, I+1)$ of (10). Let z^* be defined as in (18) and

$$q^* = \begin{cases} 0, & \text{if } 1 \leq \hat{d}_I N \\ \frac{1-\hat{d}_I N}{\gamma}, & \text{if } 1 > \hat{d}_I N. \end{cases} \tag{EC73}$$

Theorem EC1. *Let $M > N$ be such that $q^* < M$. For any fluid model solution $(\bar{Z}, \bar{Q}, \bar{\mathcal{A}}, \bar{L}_q)$ with $\|(\bar{Q}(0), \bar{Z}(0))\| < 2M$ and for any $\epsilon > 0$ there exists $T(M,\epsilon) > 0$ such that*

$$\|(\bar{Q}(t), \bar{Z}(t)) - (q^*, z^*)\| < \epsilon \tag{EC74}$$

for all $t \geq T(M,\epsilon)$.

We present the proof of this result in §EC8.1.

When Assumption 1 does not hold: Next we analyze the fluid limits when Assumption 1 does not hold and prove a result similar to Theorem EC1. In this case we work with the fluid limits instead of the fluid model solutions as the fluid limits may not satisfy (EC65) because of the aforementioned "discontinuity" in the proposed policy in the sequence of CSC systems.

PROPOSITION 2. *Consider a sequence of chat service systems and assume that* (9) *and* (EC70) *hold. Also assume that* $i^*(1,N)$ *has only one element,*

$$\left(\bar{Z}^\lambda, \bar{Q}^\lambda, \bar{\mathcal{A}}^\lambda, \bar{L}_q^\lambda\right) \to \left(\bar{Z}, \bar{Q}, \bar{\mathcal{A}}, \bar{L}_q\right), \text{ a.s. u.o.c. as } \lambda \to \infty \tag{EC75}$$

and that

$$\limsup_{\lambda\to\infty} P\left\{\bar{Q}^\lambda(0) > M\right\} < \epsilon$$

for some M *and* $\epsilon > 0$*. Then, there exists* $T > 0$ *large enough such that*

$$\limsup_{\lambda\to\infty} P\left\{\left\|(\bar{Q}^\lambda(t), \bar{Z}^\lambda(t)) - (q^*, z^*)\right\| > \epsilon\right\} < 2\epsilon \tag{EC76}$$

for all $t \geq T$.

We present the proof in §EC8.2.

EC4.2. Convergence of steady state quantities

Consider the asymptotic regime where (9) holds. Let $\bar{\upsilon}(\pi^\lambda)$ denote the stationary distribution of $(\bar{Z}^\lambda(t), \bar{Q}^\lambda(t))$ under a non-idling policy π^λ. The existence and uniqueness of $\bar{\upsilon}(\pi^\lambda)$ follow from the proof of Theorem 1. We next show that the sequence of fluid scaled stationary distributions is tight.

THEOREM EC2. *Consider a sequence of chat service systems that satisfies* (9) *under a sequence of non-idling policies* $\{\pi^\lambda\}$ *and let* $\left(\bar{Z}^\lambda(0), \bar{Q}^\lambda(0)\right)$ *be distributed according to* $\bar{\upsilon}(\pi^\lambda)$*. Then the sequence* $\left\{(\bar{Z}^\lambda(0), \bar{Q}^\lambda(0))\right\}$ *is tight and the sequence* $\left\{\bar{Q}^\lambda(0)\right\}$ *is uniformly integrable.*

A proof is presented in §EC8.3.

EC4.3. Proof of Theorem 2

Consider a sequence of chat service systems that satisfies (9). Let υ^λ denote the stationary distribution in the λth system. The first claim follows from the second one, Theorem EC2, and the fact that

$$P^\lambda_{\upsilon^\lambda}(Ab) = \lim_{T\to\infty} \frac{Ab^\lambda(T)}{T\lambda} = \frac{E_{\upsilon^\lambda}\left[\gamma Q^\lambda(0) + \sum_{i=1}^I i\nu Z_i^\lambda(0)\right]}{\lambda}. \tag{EC77}$$

We focus on proving the second part for the rest of the proof. Assume that the initial state $(Q^\lambda(0), Z^\lambda(0))$ has distribution υ^λ. By Theorem EC2, the sequence $\{(\bar{Q}^\lambda(0), \bar{Z}^\lambda(0))\}$ is tight. Therefore, it is sufficient to show that every convergent subsequence of $(\bar{Q}^\lambda(0), \bar{Z}^\lambda(0))$ converges to the same limit (q^*, z^*). To this end consider a convergent subsequence, denoted again by $\{\lambda\}$ for notational simplicity. We show that

$$(\bar{Q}^\lambda(0), \bar{Z}^\lambda(0)) \to (q^*, z^*) \text{ as } n \to \infty \text{ in probability.} \tag{EC78}$$

To prove (EC78) it is enough to show that

$$\limsup_{\lambda\to\infty} P\left\{\left\|(\bar{Q}^\lambda(0), \bar{Z}^\lambda(0)) - (q^*, z^*)\right\| > \epsilon\right\} < \epsilon' \tag{EC79}$$

for any $\epsilon, \epsilon' > 0$. Therefore, fix $\epsilon > 0$ and $\epsilon' > 0$. By Theorem EC2, there exists $M > 0$ such that

$$\limsup_{\lambda\to\infty} P\left\{\left\|(\bar{Q}^\lambda(0), \bar{Z}^\lambda(0))\right\| > M\right\} < \epsilon'/2. \tag{EC80}$$

Using an argument similar to that in Theorem A.1 in Dai and Tezcan (2011) and using Theorem 4.4 in Billingsley (1968), the sequence

$$\left\{(\bar{Q}^\lambda(0), \bar{Z}^\lambda(0), \bar{Q}^\lambda, \bar{Z}^\lambda, \bar{\mathcal{A}}^\lambda, \bar{L}_q^\lambda, \bar{S}^\lambda\right\} \tag{EC81}$$

is tight, where $\bar{S}^\lambda = (\bar{S}_1^\lambda, \bar{S}_2^\lambda, \ldots, \bar{S}_{I+1}^\lambda)$ and $\bar{S}_i^\lambda(t) = \lambda^{-1} S_i(\lambda t)$, $i = 1, 2, \ldots, I+1$. Consider a convergent subsequence, again denoted by λ, and let the limit of this subsequence be denoted by

$$\left\{(\bar{Q}(0), \bar{Z}(0), \bar{Q}, \bar{Z}, \bar{\mathcal{A}}, \bar{L}_q, \bar{S}\right\}.$$

Also observe that $\bar{S}_i(t) = 1$, for $\bar{S} = \left(\bar{S}_1(t), \ldots, S_{I+1}; t \geq 0\right)$ and $\bar{A}(t) = t$, $t \geq 0$ a.s. Obviously, it is enough to prove (EC79) for any such (further) subsequence.

By appealing to the Skorohod representation theorem, we may choose an equivalent distributional representation (which we will denote by putting a " ˜ " above the symbols) such that the sequence of random processes

$$\left\{\left(\tilde{\bar{Q}}^\lambda(0), \tilde{\bar{Z}}^\lambda(0), \tilde{\bar{Q}}^\lambda, \tilde{\bar{Z}}^\lambda, \tilde{\bar{A}}^\lambda, \tilde{\bar{L}}_q^\lambda, \tilde{\bar{S}}^\lambda\right)\right\}$$

as well as the limit

$$\left(\tilde{\bar{Q}}(0), \tilde{\bar{Z}}(0), \tilde{\bar{Q}}, \tilde{\bar{Z}}, \tilde{\bar{A}}, \tilde{\bar{L}}_q, \tilde{\bar{S}}\right)$$

are defined on a new probability space, say $(\tilde{\Omega}, \tilde{\mathcal{G}}, \tilde{P})$, so that $\tilde{P}$-a.s.

$$\left(\tilde{\bar{Q}}^\lambda(0), \tilde{\bar{Z}}^\lambda(0), \tilde{\bar{Q}}^\lambda, \tilde{\bar{Z}}^\lambda, \tilde{\bar{A}}^\lambda, \tilde{\bar{L}}_q^\lambda, \tilde{\bar{S}}^\lambda\right) \to \left(\tilde{\bar{Q}}(0), \tilde{\bar{Z}}(0), \tilde{\bar{Q}}, \tilde{\bar{Z}}, \tilde{\bar{A}}, \tilde{\bar{L}}_q, \tilde{\bar{S}}\right) \tag{EC82}$$

u.o.c. as $\lambda \to \infty$. Clearly $\tilde{\bar{S}}_i(t) = 1$, $t \geq 0$ a.s. for $\tilde{\bar{S}} = \left(\tilde{\bar{S}}_1(t), \ldots, \tilde{\bar{S}}_{I+1}; t \geq 0\right)$, hence we drop it from the notation for the rest of the proof. Also, $\tilde{\bar{A}}(t) = t$ a.s. for $t \geq 0$.

By the equivalent distributional representation, the limit $\left(\tilde{\bar{Q}}(0), \tilde{\bar{Z}}(0), \tilde{\bar{Q}}, \tilde{\bar{Z}}, \tilde{\bar{A}}, \tilde{\bar{L}}_q\right)$ satisfies the fluid model equations (EC59)–(EC69) if Assumption 1 holds. Hence, it follows from (EC80) and Theorem EC1 that there exists $T > 0$ such that

$$P\left\{\left\|(\tilde{\bar{Q}}(t), \tilde{\bar{Z}}(t)) - (q^*, z^*)\right\| > \epsilon\right\} < \epsilon'/2 \tag{EC83}$$

for all $t \geq T$. By (EC82) and the equivalent distributional representation, this implies that

$$P\left\{\left\|(\bar{Q}^\lambda(t), \bar{Z}^\lambda(t)) - (q^*, z^*)\right\| > \epsilon\right\} < \epsilon'$$

for all $t \geq T$ and λ large enough. We obtain the desired result (EC80) by the fact that the initial state (in the original probability space) $(\bar{Q}^\lambda(0), \bar{Z}^\lambda(0))$ has a stationary distribution for each λ. If Assumption 1 does not hold, (EC83) (hence the result) follows from Proposition 2. $\square$

EC4.4. Proof of Theorem 3

The proof is similar to the proof of Theorem 2 hence we only present a sketch. The main difference in the proof is that we need to show that Theorem EC1 is still valid under π' when (23) holds. The proof is similar to that of Theorem EC1. It is easily checked that (EC110) in Step 1 and (EC111) in Step 2 of Theorem EC1 are still valid when (23) holds. Then Step 3 is identical since π' takes the same actions with $\pi^{\lambda,*}$ when (EC113) and (EC114) hold. $\square$

EC5. Proof of Proposition 1

Assume that (9), (26) and (27). holds. (Throughout the proof we use the notation defined in §EC2.) Recall that $z^*(1,N) = (z_0^*, z_1^*, \ldots, z_I^*)$ denotes the optimal solution of (10) (with staffing level N and arrival rate $\lambda = 1$), see (18).

As in the proof of Theorem 2, we focus on the fluid limits. First we give an explanation of how the queueing equations are obtained under π_P. Let levels i_j and i_{j+1} be efficient and recall that $\mathcal{U}_{i_j}$ denote the set of inefficient levels whose indices are in between the indices of these two levels (i_j and i_{j+1}). With a slight abuse of notation, let $A_i^\lambda(t)$ denote the number of customers who are assigned to an agent at levels $\{i_j\} \cup \mathcal{U}_{i_j}$ upon arrival. By construction of our policy there may be at most one agent at one of the levels in $\mathcal{U}_{i_j}$. Let

$$D^\lambda_{\mathcal{U}_{i_j}}(t) = \sum_{i \in \mathcal{U}_{i_j}} D_i^\lambda(t) \text{ and } \bar{D}^\lambda_{\mathcal{U}_{i_j}}(t) = \frac{D^\lambda_{\mathcal{U}_{i_j}}(t)}{\lambda}.$$

Obviously

$$\sum_{i \in \mathcal{U}_{i_j}} \bar{Z}_i^{\lambda}(\cdot) \to 0 \text{ and } \bar{D}^{\lambda}_{\mathcal{U}_{i_j}}(\cdot) \to 0 \tag{EC84}$$

u.o.c. a.s. as $\lambda \to \infty$.

Now because of the preemption procedure

$$\left| Z^{\lambda}_{i_{j+1}}(t) - Z^{\lambda}_{i_{j+1}}(0) - \left(\frac{A^{\lambda}_{i_j}(t) - D^{\lambda}_{i_{j+1}}(t) - D^{\lambda}_{\mathcal{U}_{i_j}}(t)}{i_{j+1} - i_j} - \frac{\left(A^{\lambda}_{i_{j+1}}(t) - D^{\lambda}_{i_{j+2}}(t) - D^{\lambda}_{\mathcal{U}_{i_{j+1}}}(t)\right)}{i_{j+2} - i_{j+1}} \right) \right| \le 2.$$

Using this condition and (EC84), one can show that the fluid limits of π_P satisfy the following equations in addition to (EC59) and (EC64)–(EC69); for $i \notin \mathcal{F}$

$$\bar{Z}_i(t) = 0, \text{ for all } t \ge 0. \tag{EC85}$$

For $i_j \in \mathcal{F}$,

$$\begin{aligned}
\bar{Z}_{i_j}(t) &= \bar{Z}_{i_j}(0) + \frac{1}{i_j - i_{j-1}} \bar{A}_{i_{j-1}}(t) - \frac{1}{i_{j+1} - i_j} \bar{A}_{i_j}(t) \\
&\quad + \frac{\hat{d}_{i_{j+1}}}{i_{j+1} - i_j} \int_0^t \bar{Z}_{i_{j+1}}(s)ds - \frac{\hat{d}_{i_j}}{i_j - i_{j-1}} \int_0^t \bar{Z}_{i_j}(s)ds, \; i_j \in \mathcal{F} \setminus \{0, I\}, && \text{(EC86)} \\
\bar{Q}(t) &= \bar{Q}(0) + \bar{A}_I(t) - \hat{d}_I \int_0^t \bar{Z}_I(s)ds - \gamma \int_0^t \bar{Q}(s)ds + \bar{L}_q(t), && \text{(EC87)} \\
&\sum_{i_j \in \mathcal{F}} \dot{\bar{A}}_{i_j}(t) = 1, && \text{(EC88)} \\
&\dot{\bar{A}}_{i_{j'}}(t) = 0, \text{ if } \bar{Z}_{i_j}(t) > 0 \text{ for some } j < j', && \text{(EC89)}
\end{aligned}$$

where J is the number of efficient levels. The fluid model equations (EC59), (EC64)–(EC69) and (EC85)–(EC89) are similar to the fluid model equations (EC59)–(EC69), with one difference; the number of agents transitioning from one level to another as explained above. We note again that a solution $\{\bar{Z}, \bar{Q}, \bar{\mathcal{A}}\}$ to equations (EC59), (EC64)–(EC69) and (EC85)–(EC89), referred to as a fluid model solution, is differentiable almost everywhere and we refer to a point where it is differentiable as a regular point.

As in the proof of Theorem 2, it is enough to show that for any fluid model solution $\{\bar{Z}, \bar{Q}, \bar{\mathcal{A}}\}$, given $\epsilon > 0$, there exists T large enough such that

$$\left| \bar{Z}_i(t) - z_i^* \right| < \epsilon. \tag{EC90}$$

for $t \ge T$ and $i = 0, \ldots, I$. We also note that for $i \notin \mathcal{F}$, (EC90) follows from Lemma 1 and (EC85). We prove the result in three steps. We mainly focus on the case when there are two basic levels, whose indices denoted by i_j and i_{j+1}, $0<i_j<i_{j+1}<I$ and $\hat{d}_1N<1<\hat{d}_IN$ and comment on other cases at the end of the proof. Throughout the proof we only consider regular points of the fluid model solution.

Step1 (when $i<i_j$): We first show that (EC90) holds for $i<i_j$. The proof is similar to the proof of Step 1 of Theorem EC1 using an induction argument. For $i=0$, the argument in Step 1 of Theorem EC1 can be repeated verbatim. Assume that $i_k\in\mathcal{F}$ with $0<i_k<i_j$ and there exits t' such that $\bar{Z}_{i'}(t)=0$ for all $t\geq t'$ and $i'<i_k$. We show that there exists $T>t'$ such that $\bar{Z}_{i_k}(t)=0$ for $t\geq T$, completing the induction argument. Because level i_k is the lowest indexed nonempty level then at any regular point $t>t'$

$$\dot{\bar{Z}}_{i'}(t)=0 \text{ for all } i'<i_k.$$

Therefore by (EC59) and (EC86), $\dot{\bar{A}}_{i_{k-1}}(t)=(i_k-i_{k-1})\hat{d}_{i_k}\bar{Z}_{i_k}(t)$. Also by (EC86)–(EC89)

$$\begin{aligned}\dot{\bar{Z}}_{i_k}(t)&=\frac{-1+\hat{d}_{i_k}\bar{Z}_{i_k}(t)+\hat{d}_{i_{k+1}}\bar{Z}_{i_{k+1}}(t)}{i_{k+1}-i_k}\\&\leq\frac{1}{I}\left(-1+\hat{d}_{i_{k+1}}(N-\bar{Z}_{i_k}(t))\right),\end{aligned}\tag{EC91}$$

where the inequality follows from the fact that $i_{k+1}-i_k\leq I$ and $i_k,i_{k+1}\in\mathcal{F}$. Because $i_k<i_j$ and so $i_{k+1}\leq i_j$, (EC91) implies that, when there are two basic levels,

$$\dot{\bar{Z}}_{i_k}(t)\leq-\frac{1}{I}(1-\hat{d}_{i_j}N)<0,$$

whenever $\bar{Z}_{i_k}(t)>0$ for a regular point t. Therefore, there exists T large enough such that

$$\bar{Z}_i(t)=0 \text{ for all } t\geq T \text{ and } i<i_j,$$

completing the proof of (EC90) for $i<i_j$.

Step 2 (When $i>i_{j+1}$): Next we focus on $i>i_{j+1}$ and prove by induction that (EC90) holds for $i>i_{j+1}$. As in the second step of proof of Theorem EC1, $\bar{Q}(t)=0$ for $t>T':=M/(1-\hat{d}_IN)$. Assume that $\bar{Z}_I(t)>0$ for a regular point $t>T'$. If $\bar{Z}_i(t)>0$ for some $i<i_{J-1}$, then by (EC87) and (EC89)

$$\dot{\bar{Z}}_I(t)\leq-\frac{\hat{d}_I\bar{Z}_I(t)}{I-i_{J-1}}.\tag{EC92}$$

If $\bar{Z}_i(t)=0$ for all $i<i_{J-1}$, then $\dot{\bar{Z}}_i(t)=0$ for all $i<i_{J-1}$ for any regular point t by (EC86)–(EC89), hence

$$\dot{\bar{Z}}_I(t)\leq\frac{1-\hat{d}_{i_{J-1}}\bar{Z}_{i_{J-1}}(t)-\hat{d}_I\bar{Z}_I(t)}{I}\leq\frac{1-\hat{d}_{i_{J-1}}N}{I}.\tag{EC93}$$

By (EC92) and (EC93) there exists $T_I > 0$ such that $\bar{Z}_I(t) < \epsilon$ for $t \geq T_I$.

Now consider $J > k > j+1$ and assume that there exits $T_{i_{k+1}}$ such that, for $t \geq T_{i_{k+1}}$,

$$\bar{Z}_i(t) \leq \frac{1}{4I \max_j\{\hat{d}_j\}} \epsilon \tag{EC94}$$

for $i > i_k$ and that

$$\epsilon > \frac{\hat{d}_{i_{k+1}} N}{4I \hat{d}_{i_{j+1}}}. \tag{EC95}$$

Assume also that $\bar{Z}_{i_k}(t) > I\epsilon/\hat{d}_{i_k}$ for a regular point t. If $\bar{Z}_i(t) > 0$ for some $i < i_k - 1$, then by (EC86), (EC89), (EC94) and (EC95)

$$\dot{\bar{Z}}_{i_k}(t) \leq -\frac{\hat{d}_{i_k} \bar{Z}_{i_k}(t)}{i_k - i_{k-1}} + \frac{\hat{d}_{i_{k+1}} \bar{Z}_{i_{k+1}}(t)}{i_{k+1} - i_k} < -\epsilon/2. \tag{EC96}$$

If $\bar{Z}_i(t) = 0$ for all $i < i_{k-1}$, then $\dot{\bar{Z}}_i(t) = 0$ for all $i < i_{k-1}$ for any regular point t, hence by (EC86), (EC89), (EC94) and (EC95)

$$\dot{\bar{Z}}_{i_k}(t) \leq \frac{1 - \hat{d}_{i_{k-1}} \bar{Z}_{i_{k-1}}(t) - \hat{d}_{i_k} \bar{Z}_{i_k}(t)}{(i_k - i_{k-1})} + \frac{\hat{d}_{i_{k+1}} \bar{Z}_{i_{k+1}}(t)}{i_{k+1} - i_k} \leq -\epsilon/4. \tag{EC97}$$

By (EC96) and (EC97) there exists $T_{i_k} > 0$ such that $\bar{Z}_{i_k}(t) < \epsilon$ for $t \geq T_{i_k}$ completing the induction argument.

Step 3 (i_j and i_{j+1}): Now we are ready to finish the proof by focusing on i_j and i_{j+1}. We show that given ϵ there exists T large enough such that

$$\left|\bar{Z}_{i_k}(t) - z^*_{i_k}\right| < \epsilon, \text{ for } k = j, j+1.$$

The proof is similar to the case with $\mathcal{U}_{i_j} = \emptyset$ in the proof of Theorem EC1. Fix $\epsilon > 0$ and given $\epsilon' > 0$ choose $T(\epsilon')$ so that (EC113) and (EC114) hold. Existence of such $T(\epsilon')$ is guaranteed by the first two steps. Then by (EC86), (EC113) and (EC114), if $\bar{Z}_{i_j}(t) = z^*_{i_j} + \tilde{\epsilon}$, for $\tilde{\epsilon} > \epsilon$ and $t > T(\epsilon')$

$$\dot{\bar{Z}}_{i_j}(t) \leq -\frac{1}{i_{j+1} - i_j} \left((\hat{d}_{i_{j+1}} - \hat{d}_{i_j})\epsilon + \hat{d}_{i_{j+1}} \right) \epsilon'$$

Similarly, if $\bar{Z}_{i_j}(t) = z^*_{i_j} - \tilde{\epsilon}$

$$\dot{\bar{Z}}_{i_j}(t) \geq \frac{1}{i_{j+1} - i_j} \left((\hat{d}_{i_{j+1}} - \hat{d}_{i_j})\epsilon - \hat{d}_{i_{j+1}} \epsilon' \right)$$

For $\epsilon' < \left(\hat{d}_{i_{j+1}} - \hat{d}_{i_j}\right) / (4\hat{d}_{i_{j+1}})\epsilon$, this gives the desired.

Other cases: The other cases are handled in a way similar to that described at the end of the proof of Theorem EC1. $\square$

EC6. Proof of Theorem 4

Fix $\epsilon > 0$ that satisfies

$$\epsilon < \frac{\delta}{2Id_m(P_I^{Ab}\hat{d}_I \vee 1)}, \tag{EC98}$$

where $d_m = \max_{1\le i_1, i_2, i\le I}\left(\frac{\hat{d}_i - \hat{d}_{i_2}}{\hat{d}_{i_1} - \hat{d}_{i_2}}\right) \ge 1$ and consider the policy $\hat{\pi}^*(\epsilon)$. By (6), (EC2), Theorem EC2 and (EC77) it is enough to show that

$$\limsup_{T\to\infty}\lim_{\lambda\to\infty}\left\|\bar{Z}^\lambda(T) - z^*(1,N)\right\| < 2d_m\epsilon. \tag{EC99}$$

Note that by Proposition 1, there exist λ and T large enough such that for the virtual system

$$\left\|\bar{Z}^\lambda(T) - z^*(1,N)\right\| < \epsilon.$$

Therefore, if $z^*_{i_{j_k}}(1,N) > \epsilon$, for $k = 1,2$, then the result follows from Theorem 2.

Now assume that $z^*_{i_{j_k}}(1,N) < \epsilon$ for $k = 1$ or $k = 2$. If $i_{j_1} = I$ then we have that all the agents are in states $I-1$ and I in the fluid model as we argue next. Because $\hat{d}_I > \hat{d}_{I-1}$, using the first step of the proof of Theorem EC8.1, we have for large enough t, $\bar{Z}_{I-1}(t) \le d_m\epsilon$ by definition of d_m and $\bar{Z}_I(t) = N - \bar{Z}_{I-1}(t)$. Hence (EC99) follows.

Now assume that $z^*_{i_{j_2}}(1,N) < \epsilon$ and let j_1 denote index of the lowest indexed level whose index is greater than i_{j_1} that is not in $\mathcal{N}$. Note that as in the proof of Theorem EC8.1, $\bar{Z}_{j_1}(t) \le d_m\epsilon$ (by definition of d_m) and $\bar{Z}_{i_{j_1}}(t) \ge N - \bar{Z}_{j_1}(t) - d_m\epsilon$, $\bar{Z}_i(t) = 0$ for $i \le i_{j_1}$ and $\sum_{i>j_1}\bar{Z}_i(t) < d_m\epsilon$ for t large enough. Hence (EC99) follows.

Now assume that $z^*_{i_{j_1}}(1,N) < \epsilon$ and let j_2 denote the index of the lowest indexed level whose index is greater than i_{j_2} that is not in $\mathcal{N}$. By steps 1 and 2 of Theorem EC8.1, $\bar{Z}_i(t) = 0$ for $i < i_{j_2} - 1$, and $\bar{Z}_{i_{j_2}-1}(t) < d_m\epsilon$ by definition of d_m and $\sum_{i>j_2}\bar{Z}_i(t) < d_m\epsilon/I$, for $t \ge T$ and T large enough. Note that because j_2 denotes the index of the smallest indexed level whose index is larger than i_{j_2} that is not in $\mathcal{N}$, all levels in $\{i_{j_2}+1, \ldots, j_2 - 1\}$ (if there are any) are in $\mathcal{N}$. In this case (EC99) follows similarly to the third step of the proof of Theorem EC8.1 as we explain next. Set $T = 0$ for notational simplicity. If $\bar{Z}_{i_{j_2}-1}(t) + \bar{Z}_{i_{j_2}}(t) > N - d_m\epsilon$ for any t the result readily follows. Therefore assume that $\bar{Z}_{i_{j_2}-1}(0) + \bar{Z}_{i_{j_2}}(0) < N - d_m\epsilon$. Then, it is easy to show using the fluid model equations that either $\bar{Z}_{i_{j_2}-1}(s) = 0$ or $\bar{Z}_{i_{j_2}-1}(s) + \bar{Z}_{i_{j_2}}(s) > N - d_m\epsilon$ for s large enough. In the latter case the desired result follows from the preceding argument so assume that the former holds. Also assume that $\bar{Z}_{i_{j_2}}(t) < N - d_m\epsilon$ for all $t \ge s$ as otherwise the result follows again from the preceding argument. Then it is easy to check that $\bar{Z}_{i_{j_2}-1}(t)=0$ and similar to (EC119) and (EC120) we have for $t\geq s$ that

$$\bar{Z}_{i_{j_2}}(t)\geq \bar{Z}_{i_{j_2}}(s) \tag{EC100}$$

and for any $\epsilon_b>0$ there exists finite $T(\epsilon_b)$ large enough such that

$$\hat{d}_b(t)\geq (1-\epsilon_b)\vee\left(\hat{d}_b(s)\wedge 1\right) \tag{EC101}$$

for $\hat{d}_b(t)=\hat{d}_{i_{j_2}}\bar{Z}_{i_{j_2}}(t)+\hat{d}_{j_2}\bar{Z}_{j_2}(t)$ and $t\geq T(\epsilon_b)$. The rest of the proof follows similarly to the third step of the proof of Theorem EC8.1. □

EC7. Proofs of the results in Section 5

Proof of Theorem 5: The proof follows from Theorems 1 and 2. First it is easily checked that

$$\frac{N^*(\lambda^n,p^{Ab})}{n}=\lambda^n N^*(1,p^{Ab}).$$

Therefore,

$$\lim_{n\to\infty}\frac{N^*(\lambda^n,p^{Ab})}{n}=N^*(\lambda,p^{Ab}). \tag{EC102}$$

Assume that there exists an asymptotically feasible sequence N^n such that

$$\limsup_{n\to\infty}\frac{N^*(\lambda^n,p^{Ab})}{N^n}>1. \tag{EC103}$$

Then there exists a subsequence (for simplicity of notation, we still use index n), along which

$$\lim_{n\to\infty}\frac{N^*(\lambda^n,p^{Ab})}{N^n}>1 \tag{EC104}$$

and

$$\lim_{n\to\infty}\frac{N^n}{n}=N \tag{EC105}$$

for some N. Also, by (EC103) and (EC104), we have

$$N<N^*(\lambda,p^{Ab}). \tag{EC106}$$

Note that $P^{Ab}(\lambda,N)>P^{Ab}(\lambda,N^*(\lambda,p^{Ab}))$, otherwise $N^*(\lambda,p^{Ab})=N$ and (EC106) cannot hold. By Theorems 1 and 2, for any non-idling policy π^n, if the staffing level is equal to N^n in the nth system, we have

$$\liminf_{n\to\infty}P^{Ab,\pi^n}(\lambda^n,N^n)\geq\limsup_{n\to\infty}P^{Ab,\pi^*}(\lambda^n,N^n)=P^{Ab}(\lambda,N)>P^{Ab}\left(\lambda,N^*(\lambda,p^{Ab})\right). \tag{EC107}$$

Hence, N^n cannot be feasible and in turn (EC103) cannot hold. Also,

$$\lim_{n\to\infty} P^{Ab,\pi^*}(\lambda^n, N^*(\lambda^n, p^{Ab})) = P^{Ab}(\lambda, N^*(\lambda, p^{Ab})) \le p^{Ab}, \tag{EC108}$$

where the equality follows from Theorem 2, (38) and (EC102) and the inequality follows from the fact that $N^*(\lambda, p^{Ab})$ is an optimal solution (hence a feasible solution) of the staffing LP. □

Proof of Theorem 6: Fix $\delta > 0$. If $p^{Ab} = P_1^{Ab}$ the result follows from Lemma 1 and (EC2) and Theorem 4 by setting $\epsilon = \delta/2$. Note that if $p^{Ab} > P_1^{Ab}$ then $\lambda \ge (1+\tilde{\delta})N^*(\lambda, p^{Ab})d_1$ for some $\delta \ge \tilde{\delta} > 0$ by Lemma 2 since both levels 1 and I are assumed to be efficient. For notational simplicity we set $N^* = N^*(\lambda, p^{Ab})$. Define

$$\theta = \inf_{\lambda \ge (1+\tilde{\delta})N^*\hat{d}_1} \left\{ P^{Ab}(\lambda, N^*) - P^{Ab}\left(\lambda, (1+\tilde{\delta})N^*\right)\right\}. \tag{EC109}$$

By Lemma 1 and (EC2), $\theta > 0$. Choose $\epsilon > 0$ as in Theorem 4 so that (29) holds with θ. The result then follows from Theorem 4. □

EC8. Proofs of supplemental results

In this section we prove Theorem EC1 in §EC8.1, Proposition 2 in §EC8.2 and Theorem EC2 in §EC8.3. (Throughout this section we use the notation defined in §EC2.)

EC8.1. Proof of Theorem EC1

Fix N and let $i^*(1,N) = \{i_j, i_{j+1}\}$ denote the set of the indices of the basic levels. We mainly focus on the case when $0 < i_j < i_{j+1} < I$ and $\hat{d}_1 N < 1 < \hat{d}_I N$. We describe how the proof can be extended to other cases at the end. In addition we focus mainly on the case when $\mathcal{N} = \emptyset$ and extend the proof using Lemma EC4 below when it is not. Let $(\bar{Q}, \bar{Z}, \bar{\mathcal{A}}, \bar{L}_q)$ denote a fluid model solution. Throughout the proof we use the following fact that if $\bar{Z}_i(t) = 0$ at a regular point t then $\dot{\bar{Z}}_i(t) = 0$, since $\bar{Z}_i$ attains its minimum at t.

Fix $\epsilon > 0$. We prove (EC74) in three steps.

1. We first show that there exists $T_1 \ge 0$ such that $\bar{Z}_i(t) = 0$ for $t \ge T_1$ and $i \le i_j - 1$.
2. Then we show that given $\epsilon > 0$ there exists $T_2 \ge T_1$ such that $\bar{Z}_i(t) < \epsilon$ for $t \ge T_2$ and $i > i_{j+1}$.
3. In the last step, we show that given $\epsilon > 0$ there exists $T_3 \ge T_2$ such that $\left|\bar{Z}_i(t) - z_i^*\right| < \epsilon$ for $t \ge T_3$ and for $i = i_j, i_j + 1 \ldots, i_{j+1}$, completing the proof.

Because $(\bar{Q}, \bar{Z}, \bar{\mathcal{A}}, \bar{L}_q)$ is differentiable almost everywhere, when we take the derivatives with respect to t we only consider the regular points of $(\bar{Q}, \bar{Z}, \bar{\mathcal{A}}, \bar{L}_q)$ throughout the proof without loss of generality.

<u>*Step 1.*</u> Let

$$f(t) = \sum_{i=0}^{i_j-1} (i_j - i)\bar{Z}_i(t).$$

Assume that $f(t) > 0$. Note that by (EC59)–(EC69), for any regular point t we have

$$\dot{f}(t) = -1 + \sum_{i=1}^{i_j} \hat{d}_i \bar{Z}_i(t) < -1 + \hat{d}_{i_j} N < -\delta, \tag{EC110}$$

where $\delta = (\hat{d}_{i_{j+1}} - \hat{d}_{i_j}) z^*_{i_{j+1}}$ and the last inequality follows from (18), and the fact that there are two basic levels. Hence $\bar{Z}_i(t) = 0$, $i \le i_j - 1$, for $t \ge T_1 := N/\delta$.

<u>*Step 2.*</u> Now we consider $i > i_{j+1}$. By (EC62), (EC66) and (EC67), if $\bar{Q}(t) > 0$, we have

$$\dot{\bar{Q}}(t) = 1 - \hat{d}_I N - \gamma \bar{Q}(t).$$

Hence, $\bar{Q}(t) = 0$ for $t > T' := M/(1 - \hat{d}_I N)$. Let

$$f(t) = \sum_{i=i_{j+1}+1}^{I} (i - i_{j+1}) \bar{Z}_i(t).$$

By our assumption $\mathcal{N} = \emptyset$ and (EC59)–(EC69), if $f(t) > \epsilon$ for $t > T'$, then

$$\dot{f}(t) \le -\delta, \tag{EC111}$$

where $\delta = \min\left\{ \hat{d}_{i_{j+1}+1}\epsilon/I, -1 + \hat{d}_{i_{j+1}+1} N \right\}$, which is positive by (14). By our assumption $\mathcal{N} = \emptyset$, we conclude that

$$\bar{Z}_i(t) < \epsilon \tag{EC112}$$

for $t \ge T_2 := T' + N/\delta$ and $i > i_{j+1}$.

<u>*Step 3.*</u> To conclude the proof we need to show that (EC74) is satisfied (for a reselected ϵ) by $\bar{Z}_{i_j}$, $\bar{Z}_i$ for $i \in \mathcal{U}_{i_j}$ (if $\mathcal{U}_{i_j} \ne \emptyset$) and $\bar{Z}_{i_{j+1}}$. By the first two steps, we can assume that for any $\epsilon' > 0$ there exists a finite $T(\epsilon') > 0$ such that for $t \ge T(\epsilon')$, $\bar{Q}(t) = 0$,

$$\bar{Z}_i(t) = 0 \text{ for } i \le i_j - 1 \tag{EC113}$$

and

$$\bar{Z}_i(t) \le \epsilon' \text{ for } i \ge i_{j+1} + 1. \tag{EC114}$$

Assume that $\mathcal{U}_{i_j} = \emptyset$. Fix $\epsilon, \epsilon' > 0$ so that

$$\epsilon' < \left(\hat{d}_{i_{j+1}} - \hat{d}_{i_j} \right) / (4\hat{d}_{i_{j+1}})\epsilon. \tag{EC115}$$

Recall that by Lemma EC1

$$\hat{d}_{i_{j+1}} > \hat{d}_i \tag{EC116}$$

for all $i < i_{j+1}$ otherwise level i_{j+1} cannot be an efficient level. By (EC60), (EC65), (EC67) and (EC113)

$$\dot{\bar{Z}}_{i_j}(t) = -(1 - \hat{d}_{i_j}\bar{Z}_{i_j}(t)) + \hat{d}_{i_{j+1}}\bar{Z}_{i_{j+1}}(t)$$

Fix $\epsilon > 0$. Then by (EC65), (EC113) and (EC114), if $\bar{Z}_{i_j}(t) = z^*_{i_j} + \tilde{\epsilon}$, for $\tilde{\epsilon} > \epsilon$,

$$\dot{\bar{Z}}_{i_j}(t) \leq -(\hat{d}_{i_{j+1}} - \hat{d}_{i_j})\epsilon.$$

Similarly, if $\bar{Z}_{i_j}(t) = z^*_{i_j} - \tilde{\epsilon}$

$$\dot{\bar{Z}}_{i_j}(t) \geq (\hat{d}_{i_{j+1}} - \hat{d}_{i_j})\epsilon - \hat{d}_{i_{j+1}}\epsilon'.$$

This gives the desired result (EC74) for i_j, i_{j+1} with $T > T(\epsilon') + 4N/\left((\hat{d}_{i_{j+1}} - \hat{d}_{i_j})\epsilon\right)$ by (EC115) and (EC116).

Now assume that $\mathcal{U}_{i_j} \neq \emptyset$. Define

$$\hat{d}_b(t) = \hat{d}_{i_j}\bar{Z}_{i_j}(t) + \hat{d}_{i_{j+1}}\bar{Z}_{i_{j+1}}(t).$$

By (EC113), (EC114), (EC116) and (EC64), given $\delta > 0$, there exists $\epsilon', \epsilon > 0$ such that if $\hat{d}_b(t) \geq 1 - 2\epsilon$, and $\bar{Z}_{i_j}(t) \geq z^*_{i_j} - 2\epsilon$ for $t \geq T(\epsilon')$, then

$$\bar{Z}_{i_{j+1}}(t) \in [z^*_{i_{j+1}} - \delta, z^*_{i_{j+1}} + \delta].$$

Therefore, it is enough to prove that for $\epsilon > 0$ small, there exists T such that for $t \geq T \geq T(\epsilon')$,

$$\hat{d}_b(t) \geq 1 - 2\epsilon \text{ and } \bar{Z}_{i_j}(t) \geq z^*_{i_j} - 2\epsilon. \tag{EC117}$$

For the rest of the proof fix $\epsilon > 0$ and $\epsilon' > 0$ small so that

$$(\hat{d}_1 + 1)\left(1 - \frac{\hat{d}_{i_j}}{\hat{d}_{i_{j+1}}}\right)\epsilon \geq 4\left(|\mathcal{U}_{i_j}| + 1\right)\hat{d}_{i_j+1}\epsilon'. \tag{EC118}$$

From here on we only consider $t \geq T(\epsilon')$ so take $T(\epsilon') = 0$ for notational simplicity and assume that (EC113) and (EC114) hold for $t \geq 0$.

We begin the proof in this case with the following preliminary results. For all $t \geq 0$ and $\epsilon > 0$

$$\bar{Z}_{i_j}(t) \geq \bar{Z}_{i_j}(0) \wedge \left(z^*_{i_j} - \epsilon\right) \tag{EC119}$$

and for any $\epsilon_b > 0$ there exists finite $T(\epsilon_b)$ large enough such that

$$\hat{d}_b(t) \geq (1-\epsilon_b) \vee \left(\hat{d}_b(0) \wedge 1\right) \tag{EC120}$$

for $t \geq T(\epsilon_b)$.

Proof of (EC119)**:** By (EC118) we have

$$\epsilon \geq \frac{2\hat{d}_{i_{j+1}}}{\hat{d}_{i_{j+1}} - \hat{d}_{i_j}} \epsilon' \tag{EC121}$$

Note that because level $i_j - 1$ has priority over levels $i \geq i_j$ and level i_j has lower priority than those levels in $\mathcal{U}_{i_j}$, if $\bar{Z}_{i_j}(t) \leq z^*_{i_j} - \epsilon$, either there exists $i \in \mathcal{U}_{i_j}$ with $\bar{Z}_i(t) > 0$ or by (EC118), $\hat{d}_b(t) \geq 1$ and $\bar{Z}_i(t) = 0$ for all $i \in \mathcal{U}_{i_j}$. It is easy to show that if $\hat{d}_b(t) \geq 1$ and $\bar{Z}_i(t) = 0$ for all $i \in \mathcal{U}_{i_j}$, then t cannot be a regular point. Therefore, by (EC60) and (EC65)

$$\dot{\bar{Z}}_{i_j}(t) \geq 0 \tag{EC122}$$

whenever $\bar{Z}_{i_j}(t) \leq z^*_{i_j} - \epsilon$. This gives (EC119).

Proof of (EC120)**:** Now assume that

$$\hat{d}_b(t) \leq 1. \tag{EC123}$$

Let i be the highest indexed non-empty level in $\mathcal{U}_{i_j}$. Then, for all $i < i' < i_{j+1}$,

$$\dot{\bar{Z}}_{i'}(t) = 0,$$

by (EC60)–(EC65) since $\bar{Z}_{i'}$ attains a minimum at t. Therefore, if (EC123) holds, by (EC60)–(EC65)

$$\dot{\bar{Z}}_{i_{j+1}}(t) \geq \frac{1-\hat{d}_b(t)}{|\mathcal{U}_{i_j}|} \text{ and } \dot{\bar{Z}}_{i_j}(t) \geq 0,$$

where $|\mathcal{U}_{i_j}|$ is the cardinality of the set $\mathcal{U}_{i_j}$.

Note also that if for all the levels $i \in \mathcal{U}_{i_j}$, $\bar{Z}_i(t) = 0$ (implying $\dot{\bar{Z}}_i(t) = 0$) then when (EC123) holds, by (EC60)–(EC65),

$$\dot{\bar{Z}}_{i_{j+1}}(t) \geq \frac{1-\hat{d}_b(t)}{|\mathcal{U}_{i_j}|} \text{ and } \dot{\bar{Z}}_{i_j}(t) = -\frac{1-\hat{d}_b(t)}{|\mathcal{U}_{i_j}|}.$$

Therefore, $\hat{d}_b(t)$ is increasing if $\hat{d}_b(t) \leq 1$ and so (EC120) holds.

We consider the following three cases separately to complete the proof.

(A) $\bar{Z}_{i_j}(0) \geq z^*_{i_j} + \epsilon$

(B) $\bar{Z}_{i_j}(0) \le z^*_{i_j} + \epsilon$, $\hat{d}_b(0) \ge 1$

(C) $\bar{Z}_{i_j}(0) \le z^*_{i_j} + \epsilon$, $\hat{d}_b(0) \le 1$

Case (A): Assume that $\bar{Z}_{i_j}(0) \ge z^*_{i_j} + \epsilon$. Then, by (EC120) for t large enough $\hat{d}_b(t) \ge 1 - \epsilon$ for $t \ge T$. This with (EC119) gives (EC117).

Case (B): Assume that $\bar{Z}_{i_j}(0) \le z^*_{i_j} + \epsilon$ and $\hat{d}_b(0) \ge 1$. By (EC120), $\hat{d}_b(t) \ge 1$, for all $t \ge 0$. Thus, to prove (EC117), it is enough to show that there exists T large enough such that $\bar{Z}_{i_j}(t) \ge z^*_{i_j} - 2\epsilon$ for $t \ge T$. But by (EC122), this can be proved by showing that there exists a $t > 0$ such that $\bar{Z}_{i_j}(t) \ge z^*_{i_j} - 2\epsilon$.

Assume that no such t exists, that is,

$$\bar{Z}_{i_j}(t) \le z^*_{i_j} - 2\epsilon \text{ for all } t \ge 0. \tag{EC124}$$

Let

$$\tilde{\epsilon}_1(t) = z^*_{i_j} - \bar{Z}_{i_j}(t) \tag{EC125}$$

and

$$\tilde{\epsilon}_2(t) = \frac{1 - \hat{d}_{i_j}\bar{Z}_{i_j}(t) - \hat{d}_{i_{j+1}} z^*_{i_{j+1}}}{\hat{d}_{i_{j+1}}} = \frac{\hat{d}_{i_j}}{\hat{d}_{i_{j+1}}}\tilde{\epsilon}_1(t). \tag{EC126}$$

Because $\hat{d}_b(t) \ge 1$ for $t \ge 0$, $\bar{Z}_{i_{j+1}}(t) \ge z^*_{i_{j+1}} + \tilde{\epsilon}_2(t)$.

Let $\delta(t) = \bar{Z}_{i_{j+1}}(t) - z^*_{i_{j+1}} - \tilde{\epsilon}_2(t)$. By (EC121), (EC125), and (EC126),

$$\delta(t) + \sum_{i\in\mathcal{U}_{i_j}} \bar{Z}_i(t) \ge N - \epsilon' - \bar{Z}_{i_j}(t) - z^*_{i_{j+1}} - \tilde{\epsilon}_2(t) = \tilde{\epsilon}_1(t) - \tilde{\epsilon}_2(t) - \epsilon' > 0.5\left(1 - \frac{\hat{d}_{i_j}}{\hat{d}_{i_{j+1}}}\right)\epsilon \tag{EC127}$$

Now define the Lyapunov function

$$f(t) = (i_{j+1} - i_j)\,\bar{Z}_{i_{j+1}}(t) + \sum_{i\in\mathcal{U}_{i_j}} (i - i_j)\,\bar{Z}_i(t).$$

Note that $f(0) < IN$ and that $f(t) = 0$ implies

$$\bar{Z}_{i_{j+1}}(t) + \sum_{i\in\mathcal{U}_{i_j}} \bar{Z}_i(t) = 0. \tag{EC128}$$

If $f(t) > 0$, we have (assuming (EC124) and $\hat{d}_b(t) \ge 1$ for all $t \ge 0$)

$$\dot{f}(t) = (i_{j+1} - i_j)\,\dot{\bar{Z}}_{i_{j+1}}(t) + \sum_{i\in\mathcal{U}_{i_j}} (i - i_j)\,\dot{\bar{Z}}_i(t)$$

$$\leq 1 - \hat{d}_{i_j} \bar{Z}_{i_j}(t) - \hat{d}_{i_{j+1}} \bar{Z}_{i_{j+1}}(t) - \sum_{i \in U_{i_j}} \hat{d}_i \bar{Z}_i(t) + (|\mathcal{U}_{i_j}| + 1)\hat{d}_{i_j+1}\epsilon' \tag{EC129}$$

$$\leq -\hat{d}_{i_{j+1}} \delta(t) - \sum_{i \in \mathcal{U}_{i_j}} \hat{d}_i \bar{Z}_i(t) + (|\mathcal{U}_{i_j}| + 1)\hat{d}_{i_j+1}\epsilon' \tag{EC130}$$

$$\leq -0.5\hat{d}_1 \left(1 - \frac{\hat{d}_{i_j}}{\hat{d}_{i_{j+1}}}\right) \epsilon + (|\mathcal{U}_{i_j}| + 1)\hat{d}_{i_j+1}\epsilon' \tag{EC131}$$

$$\leq -0.25\hat{d}_1 \left(1 - \frac{\hat{d}_{i_j}}{\hat{d}_{i_{j+1}}}\right) \epsilon := -\Delta, \tag{EC132}$$

where (EC129) follows from (EC60)–(EC65) and the fact that level $i_j - 1$ has priority over all the other levels $i \geq i_j - 1$, (EC130) follows from the definition of $\delta(t)$, (EC131) follows from (EC127) and (EC132) follows from (EC118). Therefore, for T large enough $f(t) = 0$ for $t \geq T$ and so (EC128) holds for all $t \geq T$. However, this contradicts the fact that $\hat{d}_b(t) \geq 1$ for all $t \geq 0$. Hence, $\bar{Z}_{i_j}(t) \geq z^*_{i_j} - 2\epsilon$ for some $t \in [0, (IN)/\Delta]$, giving (EC117) in this case.

Case (C): Assume that, $\bar{Z}_{i_j}(0) \leq z^*_{i_j} + \epsilon$ and $\hat{d}_b(0) \leq 1$. By (EC120), we can assume without loss of generality that $\hat{d}_b(t) \geq 1 - \epsilon$ for t large enough. If $\hat{d}_b(t) \geq 1$, the proof is complete by Case (B) so assume that $1 - \epsilon \leq \hat{d}_b(t) < 1$.

Also, if $\bar{Z}_{i_j}(t) > z^*_{i_j} - 2\epsilon$ for some t, (EC117) follows from (EC119). Therefore, assume that $\bar{Z}_{i_j}(t) < z^*_{i_j} - 2\epsilon$ for all $t \geq 0$. Then, because $\hat{d}_b(t) < 1$, $\bar{Z}_{i_{j+1}}(t) < z^*_{i_{j+1}} + \tilde{\epsilon}_2(t)$ for $\tilde{\epsilon}_2$ defined as in (EC126). Hence

$$\sum_{i \in \mathcal{U}_{i_j}} \bar{Z}_i(t) \geq \tilde{\epsilon}_1(t) - \tilde{\epsilon}_2(t) - \epsilon' > 0.5 \left(1 - \frac{\hat{d}_{i_j}}{\hat{d}_{i_{j+1}}}\right). \tag{EC133}$$

Because $\hat{d}_b(t) < 1$, by (EC60)–(EC65), (EC133), and the fact that level $i_{j+1} - 1$ have higher priority than level i_j and level i_{j+1},

$$\sum_{i \in \mathcal{U}_{i_j}} (i - i_j)\, \dot{\bar{Z}}_i(t) \leq -\hat{d}_1 \sum_{i \in \mathcal{U}_{i_j}} \bar{Z}_i(t) \leq -0.5\hat{d}_1 \left(1 - \frac{\hat{d}_{i_j}}{\hat{d}_{i_{j+1}}}\right) \epsilon := -\Delta_0.$$

Therefore, for $T \geq N/\Delta_0$

$$\sum_{i \in \mathcal{U}_{i_j}} \bar{Z}_i(T) = 0.$$

Hence,

$$\bar{Z}_{i_{j+1}}(T) = N - \bar{Z}_{i_j}(T) - \epsilon'.$$

However, because $\bar{Z}_{i_j}(t) < z^*_{i_j} - 2\epsilon$ for all $t \geq 0$, this implies $\hat{d}_b(T) > 1$ for ϵ' small, contradicting our earlier assumption and completing the proof by Case (B).

Other cases: If there is only one basic level i_j with $1 < i_j < I$ then we have

$$\bar{Z}_i(t) = 0, \text{ for } i < i_j - 1$$

and $t \geq T_1 := N/\delta$, where $\delta = (\hat{d}_{i_{j+1}} - \hat{d}_{i_j}) z^*_{i_{j+1}}$, using an argument similar to that in Step 1. Therefore, for $t \geq T_1$, by (EC60) and (EC65)

$$\dot{\bar{Z}}_{i_j-1}(t) \leq -\left(\hat{d}_{i_j} - \hat{d}_{i_j-1}\right) \dot{\bar{Z}}_{i_j-1}(t).$$

Hence

$$\bar{Z}_{i_j-1}(t) < \epsilon \tag{EC134}$$

for $t \geq T_2 := N/\left(\left(\hat{d}_{i_j} - \hat{d}_{i_j-1}\right)\epsilon\right) + T_1$. If $1 < \hat{d}_I N$, the result then follows from Step 2.

If $1 = \hat{d}_I N$, then there is only one basic level I. By (EC60) and (EC62), $t \geq T' := M/(\gamma\epsilon)$, we have $\bar{Q}(t) < \epsilon$ and by (EC134), $\bar{Z}_I(t) > N - \epsilon$, proving the result for $t \geq T'$ large enough.

If $1 > \hat{d}_I N$, we have by (EC110) that $\bar{Z}_I(t) = N$, for all $t > T' := N/(1 - \hat{d}_I N)$. By (EC73) and (EC59)–(EC69), this implies that $\left|\bar{Q}(t) - q^*\right| < \epsilon$ for all $t > T_1 := T' + M/(\gamma\epsilon) + q^*/(\gamma\epsilon)$, proving (EC74) if $1 > \hat{d}_I N$.

Finally we consider the cases when there are two basic levels i_j and i_{j+1} but either $i_j = 0$ or $i_{j+1} = I$. If $i_j = 0$, then $i_{j+1} = 1$, since level 1 is an efficient level by definition. Then the result is proved using (EC112) as in the case with $\mathcal{U}_{i_j} = \emptyset$ in Step 3. If $i_{j+1} = I$ then the result is proved in the same way as above except that we do not need Step 2 and do not require (EC114) in Step 3 anymore. □

If $\mathcal{N} \neq \emptyset$, the second step in the proof needs to be modified. The result in that case follows from the following lemma.

LEMMA EC4. *Assume that the conditions of Theorem EC1 hold and that $i_{j+1} < I$. Then for any $\epsilon > 0$ there exists $t_2 > 0$ such that*

$$\bar{Z}_i(t) < \epsilon \text{ for } t \geq t_2 \text{ and } i > i_{j+1}.$$

Proof of Lemma EC4: Assume that the conditions of Theorem EC1 hold and that $i_{j+1} < I$. Let $\{\nu_i\}$ denote a sequence of nonnegative finite constants $\{\nu_i\}$ for $i \geq i_{j+1} + 1$ such that

- $\nu_i = 1$ if $i \notin \mathcal{N}$,
- If $i \in \mathcal{N}$

$$\nu_i \geq \left(\frac{d^*}{d_i} \vee 1\right) \nu_{i'}, \quad \text{for all } i' > i, \tag{EC135}$$

where $d^* = \min\{d_{i'} : i' \geq i_{j+1} + 1, i' \notin \mathcal{N}\}$.

Also, let $\{\rho_i\}$ be defined for $i \geq i_{j+1}+1$ recursively as follows;

$$\rho_i = \nu_i + \rho_{i-1},$$

where $\rho_{i_{j+1}} = 0$. Define

$$f(t) = \sum_{i=i_{j+1}+1}^{I} \rho_i \bar{Z}_i(t).$$

Observe that, given $\epsilon > 0$, it is enough to show that there exists $T > 0$ such that

$$f(t) < \epsilon, \text{ for all } t \geq T.$$

Fix ϵ and assume that $f(t) > \epsilon$. This implies $\sum_{i=i_{j+1}+1}^{I} \bar{Z}_i(t) > \epsilon/\rho_I$. Recall that we assume $I \notin \mathcal{N}$.

Case 1: Assume that $\bar{Z}_n(t) > 0$ for some $n < i_{j+1}$, then by (EC60) and (EC65)

$$\begin{aligned} \dot{f}(t) &= \sum_{i=i_{j+1}+1}^{I} \rho_i \dot{\bar{Z}}_i(t) \leq \sum_{i=i_{j+1}+1}^{I} \rho_i \left(-d_i \bar{Z}_i(t) + d_{i+1}\bar{Z}_{i+1}(t)\right) \\ &\leq -d^* \sum_{i=i_{j+1}+1}^{I} \bar{Z}_i(t) < -d^* \epsilon/\rho_I. \end{aligned}$$

Case 2: Now assume that $\bar{Z}_n(t) = 0$ for all $n < i_{j+1}$, then for a regular point t,

$$\dot{\bar{Z}}_{i_{j+1}-1}(t) = 0, \tag{EC136}$$

because $\bar{Z}_{i_{j+1}-1}(t) = 0$ (note that this is only possible if $1 \geq d_{i_{j+1}} \bar{Z}_{i_{j+1}}(t)$, which we assume is true), and $\sum_{i=i_{j+1}}^{I} \bar{Z}_i(t) = N$. Note that (EC136) implies

$$\dot{\bar{A}}_{i_{j+1}-1}(t) = d_{i_{j+1}} \bar{Z}_{i_{j+1}}(t). \tag{EC137}$$

Because $1 < d_I N$, we can assume without loss of generality that $\bar{Z}_I(t) < N$, for all $t \geq 0$, and so there exists $i < I$ with $\bar{Z}_i(t) > 0$, by (EC64).

Case 2(a): Assume that there exists a level n such that $i_{j+1} \leq n < I$, $\bar{Z}_n(t) > 0$ and $n \notin \mathcal{N}'$ and let $i \notin \mathcal{N}'$ be the index of the lowest indexed such level at time t. Note that $i+1 \notin \mathcal{N}$. By the definition of the policy and (EC65), if $i' \in \mathcal{N}'$ and $i' \leq i$

$$\dot{\bar{A}}_{i'}(t) = 0. \tag{EC138}$$

Also, if $i' \notin \mathcal{N}'$ and $i' \leq i$, we claim that

$$\dot{\bar{A}}_{i'}(t) = d_{i'+1} \bar{Z}_{i'+1}(t). \tag{EC139}$$

To prove this first note that $\dot{\bar{Z}}_{i'}(t)=0$, for a regular point t. Therefore, by (EC60), it is enough to show that $\dot{\bar{A}}_{i'-1}(t)=0$. If $\bar{Z}_{i'-1}(t)>0$, then $i'-1\in\mathcal{N}'$ and by (EC138), $\dot{\bar{A}}_{i'-1}(t)=0$. Assume that $\bar{Z}_{i'-1}(t)=0$ and so $\dot{\bar{Z}}_{i'-1}(t)=0$. If $\bar{Z}_{i'-2}(t)>0$, therefore $i'-2\in\mathcal{N}'$, then $\dot{\bar{A}}_{i'-2}(t)=0$ by (EC60). Because $\dot{\bar{Z}}_{i'-1}(t)=0$, this implies $\dot{\bar{A}}_{i'-1}(t)=0$ by (EC60). If $\bar{Z}_{i'-2}(t)=0$, we repeat the same argument until we reach to level i_{j+1} or the highest indexed level $i'-k$ before i' with $\bar{Z}_{i'-k}(t)>0$. Also,

$$\dot{\bar{A}}_{i'}(t)=0,\ \text{for } i'\geq i+1 \tag{EC140}$$

by (EC65). Therefore by (EC60) and (EC63)

$$\dot{\bar{Z}}_{i+k}(t)=-d_{i+k}\bar{Z}_{i+k}(t)+d_{i+k+1}\bar{Z}_{i+k+1}(t),\ \text{for } k=2,\ldots,I-i-1, \tag{EC141}$$

$$\dot{\bar{Z}}_{I}(t)=-d_I\bar{Z}_I(t). \tag{EC142}$$

For $i_{j+1}\leq k\leq i$, by (EC60)–(EC64),

$$\dot{\bar{Z}}_k(t)=\dot{\bar{A}}_{k-1}(t)+d_{k+1}\bar{Z}_{k+1}(t)-\dot{\bar{A}}_k(t)-d_k\bar{Z}_k(t) \tag{EC143}$$

and

$$\dot{\bar{Z}}_{i+1}(t)=\dot{\bar{A}}_i(t)+d_{i+2}\bar{Z}_{i+2}(t)-d_{i+1}\bar{Z}_{i+1}(t). \tag{EC144}$$

By (EC138)–(EC140) and (EC63)

$$\dot{\bar{A}}_i(t)=1-\sum_{\substack{k\in[i_{j+1}+1,i]\\ k\notin\mathcal{N}}} d_k\bar{Z}_k(t)-d_{i_{j+1}}\bar{Z}_{i_{j+1}}(t). \tag{EC145}$$

Then, by (EC141)–(EC144)

$$\begin{aligned}
\dot{f}(t) &= \sum_{k=i_{j+1}+1}^{I-1}\rho_k\left(d_{k+1}\bar{Z}_{k+1}(t)-d_k\bar{Z}_k(t)\right)-\rho_I d_I\bar{Z}_I(t)+\sum_{k=i_{j+1}+1}^{i}\rho_k\left(\dot{\bar{A}}_{k-1}(t)-\dot{\bar{A}}_k(t)\right)+\rho_{i+1}\dot{\bar{A}}_i(t)\\
&\overset{(a)}{\leq} -\sum_{k=i_{j+1}+1}^{I}\nu_k d_k\bar{Z}_k(t)+\sum_{k=i_{j+1}+1}^{i}\nu_k\dot{\bar{A}}_{k-1}(t)+\nu_{i+1}\left(1-\sum_{\substack{k\in[i_{j+1}+1,i]\\ k\notin\mathcal{N}}} d_k\bar{Z}_k(t)-d_{i_{j+1}}\bar{Z}_{i_{j+1}}(t)\right)\\
&\overset{(b)}{=} -\sum_{k=i_{j+1}+1}^{I}\nu_k d_k\bar{Z}_k(t)+\sum_{\substack{k\in[i_{j+1}+1,i]\\ k\notin\mathcal{N}}}\nu_k d_k\bar{Z}_k(t)+1-\sum_{\substack{k\in[i_{j+1}+1,i]\\ k\notin\mathcal{N}}} d_k\bar{Z}_k(t)-d_{i_{j+1}}\bar{Z}_{i_{j+1}}(t)\\
&\overset{(c)}{=} 1-d_{i_{j+1}}\bar{Z}_{i_{j+1}}(t)-\sum_{k=i_{j+1}+1}^{I}\nu_k d_k\bar{Z}_k(t)\\
&\overset{(d)}{\leq} 1-d_{i_{j+1}}\bar{Z}_{i_{j+1}}(t)-d^*(N-\bar{Z}_{i_{j+1}}(t))
\end{aligned} \tag{EC146}$$

$$\stackrel{(e)}{\leq} -c\epsilon,$$

for some constant $c>0$, where (a) follows from (EC140), (EC145) and algebraic manipulations, (b) follows from the fact that $i+1\notin\mathcal{N}$, so $\nu_{i+1}=1$, (EC138) and (EC139), (c) follows from simple algebraic manipulations, (d) follows from the definition of ν_k's and the fact that $\sum_{i=i_{j+1}}^{I}\bar{Z}_i(t)=N$, (e) follows from the definition of d^*.

Case 2(b): Now assume that $\bar{Z}_n(t)=0$ for all $n<i_{j+1}$ and $\bar{Z}_k(t)=0$ for all $k\geq i_{j+1}$ and $k\notin\mathcal{N}'\setminus\{I\}$. Let $i\in\mathcal{N}'$ denote the index of the highest indexed level with $\bar{Z}_i(t)>0$. We note that (EC138) and (EC139) still hold. Also, (EC143) holds by (EC60).

By definition of the policy $\bar{Z}_k(t)=0$ for all $i+1\leq k\leq I-1$ and so

$$\dot{\bar{Z}}_k(t)=0, \text{ for } i+1\leq k\leq I-1, \tag{EC147}$$

because t is a regular point. Therefore, by (EC60)–(EC64)

$$\dot{\bar{Z}}_i(t)=\dot{\bar{A}}_{i-1}(t)-\dot{\bar{A}}_i(t)-d_i\bar{Z}_i(t), \tag{EC148}$$

and

$$\dot{\bar{Z}}_I(t)=\dot{\bar{A}}_i(t)-d_I\bar{Z}_I(t), \tag{EC149}$$

where by (EC137)-(EC139) and (EC147)

$$\dot{\bar{A}}_i(t)=\frac{1-d_{i_{j+1}}\bar{Z}_{i_{j+1}}(t)-d_I\bar{Z}_I(t)-\sum_{\substack{k\in[i_{j+1}+1,i]\\ k\notin\mathcal{N}}}d_k\bar{Z}_k(t)}{I-i}. \tag{EC150}$$

We note that $\dot{\bar{A}}_i(t)\geq 0$ otherwise t cannot be a regular point.

Then, similar to case Case 2(a),

$$\begin{aligned}
\dot{f}(t) &\stackrel{(a)}{\leq} \sum_{k=i_{j+1}+1}^{i}\rho_k\left(d_{k+1}\bar{Z}_{k+1}(t)-d_k\bar{Z}_k(t)\right)-\rho_I d_I\bar{Z}_I(t)+\sum_{k=i_{j+1}+1}^{i-1}\rho_k\left(\dot{\bar{A}}_{k-1}(t)-\dot{\bar{A}}_k(t)\right)\\
&\qquad +\rho_i\dot{\bar{A}}_{i-1}(t)+(\rho_I-\rho_i)\dot{\bar{A}}_i(t)\\
&\stackrel{(b)}{\leq} -\sum_{\substack{k\in[i_{j+1}+1,i]\\ k\in\mathcal{N}}}\nu_k d_k\bar{Z}_k(t)-\rho_I d_I\bar{Z}_I(t)+\frac{\sum_{k=i+1}^{I}\nu_k}{I-i}\left(1-d_{i_{j+1}}\bar{Z}_{i_{j+1}}(t)-\sum_{\substack{k\in[i_{j+1}+1,i]\\ k\notin\mathcal{N}}}d_k\bar{Z}_k(t)\right)\\
&\stackrel{(c)}{\leq} -\sum_{\substack{k\in[i_{j+1}+1,i]\\ k\in\mathcal{N}}}\nu_k d_k\bar{Z}_k(t)-\rho_I d_I\bar{Z}_I(t)+\nu_{i+1}\left(1-d_{i_{j+1}}\bar{Z}_{i_{j+1}}(t)-\sum_{\substack{k\in[i_{j+1}+1,i]\\ k\notin\mathcal{N}}}d_k\bar{Z}_k(t)\right)\\
&\stackrel{(d)}{\leq} \nu_{i+1}\left(1-d_{i_{j+1}}\bar{Z}_{i_{j+1}}(t)-d^*(N-\bar{Z}_{i_{j+1}}(t))\right)
\end{aligned}$$

$$\stackrel{(e)}{\leq} -c\epsilon,$$

for some constant $c > 0$, where (a) follows from (EC139), (EC143), (EC147), (EC148), and (EC149), (b) follows from (EC150) and algebraic manipulations similar to that in (EC146), (c) and (d) follow from (EC135), (e) follows from the fact that $\sum_{k=i_{j+1}}^{I} \bar{Z}_k(t) = N$ and from the definition of d^*. □

EC8.2. Proof of Proposition 2

Assume that the conditions of the proposition hold. We prove the result by contradiction. Assume that (EC76) does not hold. Then we can find a subsequence, denoted again by λ, such that

$$\lim_{n\to\infty} P\left\{\left\|(\bar{Q}^\lambda(T), \bar{Z}^\lambda(T)) - (q^*, z^*)\right\| > \epsilon\right\} > 2\epsilon.$$

We next prove that no such subsequence exists. Let i_j denote the index of the single basic level in $i^*(1,N)$. For simplicity we only consider the case when $i_j < i_{J-1}$, but the proof is similar for other cases. We note that $\left(\bar{Z}, \bar{Q}, \bar{\mathcal{A}}, \bar{L}_q\right)$ satisfies the fluid model equations (EC59)–(EC64) and (EC66)–(EC69). It may not satisfy (EC65) because $i^*(1,N)$ has only one element. First assume that

$$\liminf_{n\to\infty} \left|i^*(\lambda, N^\lambda)\right| = 1. \tag{EC151}$$

Then the result follows from Theorem EC1, because along a subsequence of $\{n\}$ Assumption 1(iii) holds.

If (EC151) does not hold, then at least one the following does:

$$\limsup_{\lambda\to\infty} \mathbb{1}\left(\lambda^*_{i_{j-1}}(\lambda, N^\lambda) > 0\right) = 1 \text{ or } \limsup_{\lambda\to\infty} \mathbb{1}\left(\lambda^*_{i_{j+1}}(\lambda, N^\lambda) > 0\right) = 1 \tag{EC152}$$

by Lemma 1. We focus on the case when the former holds; the proof for the case when only the latter holds is similar.

If the first equation in (EC152) holds, then along a subsequence λ' the limit is attained, hence the proposed policy is $\pi^*(i_{j-1}, i_j)$ (although $i^*(1,N)$ has a single element) for λ' large enough. Fix a sample path $\omega \in \Omega$ such that (EC75) holds and $\bar{Q}(0) < M$.

From the first two steps of the proof of Theorem EC1, there exists $T > 0$ such that

$$\bar{Z}_i(t) = 0, \text{ for } i = 0, 1, \ldots, i_{j-1} - 1 \tag{EC153}$$

and

$$\bar{Z}_i(t) < \epsilon/I, \text{ for } i = i_j + 1, \ldots, I. \tag{EC154}$$

We prove below that there exists $T_1 > T$ such that

$$\bar{Z}_i(t) < \epsilon/I, \text{ for } i = i_{j-1}, \ldots, i_j - 1 \tag{EC155}$$

for $t > T_1$. By (EC64) and (EC153)–(EC155), for $t > T_1$

$$\bar{Z}_{i_j}(t) > N - \epsilon. \tag{EC156}$$

Note that (EC156) holds for a set of sample paths with probability greater than $1-\epsilon$, giving the desired result.

We prove (EC155) to complete the proof. Let

$$f(t) = \sum_{i=i_{j-1}}^{i_j-1} (i_j - i)\bar{Z}_i(t).$$

If $f(t) > \delta$, then by (EC65) (recall that $\pi^*(i_{j-1}, i_j)$ is the policy for the case with two basic levels i_{j-1} and i_j along this sequence), (EC60), (EC63), (EC64), (EC65), (EC153) and (EC154)

$$\dot{f}(t) = -1 + \sum_{i=i_{j-1}}^{i_j-1} \hat{d}_i \bar{Z}_i(t) < -\left(\hat{d}_{i_j} - \hat{d}_{i_{j-1}}\right)\delta/I^2$$

giving (EC156). □

EC8.3. Proof of Theorem EC2

Consider a sequence of chat service systems that satisfies (9) under a sequence of non-idling policies $\{\pi^\lambda\}$. For $x \in \mathbb{R}_+^{I+2}$, define $\Phi(x) = \sum_{i=0}^{I+1} x_i$. Next we show that, for λ large enough, for some $t_0, K > 0$ and $0 < \delta < 1$,

$$\sup_{x \in \mathbb{R}_+^{I+2} : \Phi(x) > K} \frac{E_x\left[\Phi\left(\bar{Z}^\lambda(t_0), \bar{Q}^\lambda(t_0)\right)\right]}{\Phi(x)} < \delta. \tag{EC157}$$

In other words, Φ is a geometric Lyapunov function with a geometric drift size $0 < \delta < 1$, drift time t_0 and exception parameter K (see Gamarnik and Zeevi (2006)). By Theorem 5 in Gamarnik and Zeevi (2006), (EC157) implies that

$$E_{\upsilon(\pi^\lambda)}\left[\Phi\left(\bar{Z}^\lambda(0), \bar{Q}^\lambda(0)\right)\right] < \frac{\phi^\lambda(t_0)K}{1-\gamma}, \tag{EC158}$$

where

$$\phi^\lambda(t_0) = \sup_x \frac{E_x\left[\Phi\left(\bar{Z}^\lambda(t_0), \bar{Q}^\lambda(t_0)\right)\right]}{\Phi(x)}.$$

Because $\sum_{i=0}^{I} \bar{Z}_i^\lambda(t) + \bar{Q}_i^\lambda(t) \le \sum_{i=0}^{I} \bar{Z}_i^\lambda(t) + \bar{Q}^\lambda(0) + \bar{A}^\lambda(t)$,

$$\phi^\lambda(t_0) \le E\left[\exp(\bar{A}^\lambda(t_0))\right].$$

By (52) in Gamarnik and Zeevi (2006) and the fact that arrivals follow a Poisson process,

$$\limsup_{\lambda\to\infty} E\left[\exp(\bar{A}^\lambda(t_0))\right] < \infty$$

for any finite $t_0 > 0$. Hence, (EC157) and (EC158) give tightness. To prove uniform integrability, note that by Markov's inequality and (EC158) we have

$$P_{\upsilon(\pi^\lambda)}\left(\bar{Q}^\lambda(0) > u\right) \le B_0 \exp(-u),$$

for some constant B_0.

We prove (EC157) to complete the proof by choosing appropriate K, t_0 and δ. Let $\bar{N}^\lambda = \lambda^{-1} N^\lambda$

$$\kappa = \max\left(1/\gamma, \sup_\lambda \bar{N}^\lambda\right). \tag{EC159}$$

Choose $K = 7\kappa$ and $t_0 = 1/\gamma$. We now compute an upper bound of $\mathbb{E}_x\left[\Psi\left(\bar{Z}^\lambda(t_0), \bar{Q}^\lambda(t_0)\right)\right]$. Let $Z^\lambda(\cdot)$ denote the total number of customers in an $M/M/\infty$ queue with arrival rate λ and service rate γ. Then by a coupling argument one can show that $E_x\left[\Psi\left(\bar{Z}^\lambda(t_0), \bar{Q}^\lambda(t_0)\right)\right] \le \bar{N}^\lambda + E\left(\bar{Z}^\lambda(t)|\bar{Z}(0) = x_{I+1}\right)$. It follows from the Kolmogorov equation (see Example 2 in Chapter 4.6 of Karlin and Taylor (1975)) that for $t > 0$

$$E(\bar{Z}^\lambda(t)|\bar{Z}(0) = x_{I+1}) = \frac{1}{\gamma}(1 - e^{-\gamma t}) + x_{I+1}e^{-\gamma t}.$$

When $\Phi(x) > K$, we must have $x_{I+1} \ge 6\kappa$ due to the choice of K and the fact that $\sum_{i=0}^{I} x_i < \sup_n \bar{N}^\lambda$. So by (EC159) and the choice t_0,

$$E(\bar{Z}^\lambda(t_0)|\bar{Z}(0) = x_{I+1}) \le \kappa + x_{I+1}e^{-\gamma t} \le \kappa + x_{I+1}/2 \le \frac{2}{3}x_{I+1}.$$

Again, by (EC159)

$$\sup_{x\in\mathbb{R}_+^{I+2}, \Phi(x)>K} \frac{E_x\left[\Psi\left(\bar{Z}^\lambda(t_0), \bar{Q}^\lambda(t_0)\right)\right]}{\Psi(x)} \le \frac{\bar{N}^\lambda + \frac{2}{3}x_{I+1}}{\bar{N}^\lambda + x_{I+1}} = \frac{5}{7}.$$

Thus, (EC157) holds for $\delta = 5/7$. □

References

Billingsley, P. 1968. *Convergence of probability measures*. Wiley, New York.

Dai, J. G., T. Tezcan. 2011. State space collapse in many-server diffusion limits of parallel server systems. *Mathematics of Operations Research* **36**(2) 271–320.

Gamarnik, D., A. Zeevi. 2006. Validity of heavy traffic steady-state approximation in generalized Jackson networks. *Annals of Applied Probability* **16**(1) 56–90.

Karlin, S., H. M. Taylor. 1975. *A first course in stochastic processes*. 2nd ed. Academic Press Inc., New York.

Meyn, S., R. L. Tweedie. 2009. *Markov chains and stochastic stability*. 2nd ed. Cambridge University Press, Cambridge.