

OPTIMAL CONTROL OF FLEXIBLE SERVERS IN TWO TANDEM QUEUES WITH OPERATING COSTS

DIMITRIOS G. PANDELIS

*Department of Mechanical and Industrial Engineering
University of Thessaly
38334 Volos, Greece
E-mail: d_pandelis@mie.uth.gr*

We consider two-stage tandem queuing systems with dedicated servers in each station and flexible servers that can serve in both stations. We assume exponential service times, linear holding costs, and operating costs incurred by the servers at rates proportional to their speeds. Under conditions that ensure the optimality of nonidling policies, we show that the optimal allocation of flexible servers is determined by a transition-monotone policy. Moreover, we present conditions under which the optimal policy can be explicitly determined.

1. INTRODUCTION

We study the optimal dynamic assignment of flexible servers in two-stage tandem queuing systems, both without arrivals (clearing systems) and with arrivals. The problem we consider is motivated by applications in manufacturing systems where flexible resources (e.g., cross-trained workers, reconfigurable machines) are used to improve performance under varying demand and operating conditions. Narongwanich, Duenyas, and Birge [16] considered the optimal allocation of capacity investments between dedicated and reconfigurable systems for firms that make products with random demands and introduced new generations of products in random future times. Hopp and Van Oyen [13] provide a framework for justifying workforce cross-training in an organization. They established the connection of cross-training to the organization's objectives and discussed methods of implementation appropriate for the specific application. Models of serial production systems with throughput maximization as the objective have been analyzed by Van Oyen, Gel, and Hopp [22],

Andradottir, Ayhan, and Down [4–6], Gel, Hopp, and Van Oyen [10,11], Hopp, Tekin, and Van Oyen [12], Irvani, Van Oyen, and Sims [15], and Ahn and Righter [3]. These models include collaborative and noncollaborative work disciplines, open systems with external arrivals that are independent of the system state, closed systems (e.g., CONWIP systems where a departure triggers a new arrival), and various forms of cross-training (e.g., full, zone, or hierarchical cross-training).

There is also much research on the optimal use of flexible servers in tandem systems with holding costs. Because the mathematical models involved are quite complex, most results refer to two-stage systems and exponential service times. Rosberg, Varaiya, and Walrand [19] considered a system with Poisson arrivals, a server with a constant service rate in the downstream station, and a server with a controllable service rate in the upstream station. They showed that the optimal service rate is nondecreasing in the length of the first queue and nonincreasing in the length of the second queue. Weber and Stidham [23] considered a system of n stations with arrivals at each station, where, in addition to convex holding costs, servers incur operating costs that are convex functions of their service rates. They showed that the optimal policy is transition-monotone; that is, when a job leaves a queue, the optimal service rate in that queue is not increased and the optimal service rates in the other queues are not decreased. Duenyas, Gupta, and Olsen [7] and Irvani, Posner, and Buzacott [14] characterized optimal policies for models of n and two-stage systems, respectively, with one flexible server and setup costs. Pandelis and Teneketzis [18] studied two-stage clearing systems with multiple flexible servers where jobs join the second queue with probability p after completing service in the upstream station. They derived conditions under which the policy that gives priority to jobs in the upstream station is optimal for general service times. For a two-stage clearing system with two flexible servers, Ahn, Duenyas, and Zhang [2] provided necessary and sufficient conditions under which an exhaustive policy for the upstream or the downstream station is optimal. Similar results were obtained by Ahn, Duenyas, and Lewis [1] for the model with arrivals. The results of Ahn et al. [2] have been extended in two more directions. First, Schiefermayr and Weichbold [20] obtained the optimal policy for all values of holding costs and service times. Second, Weichbold and Schiefermayr [24] derived conditions for the optimality of exhaustive policies when jobs require the second stage of service with a certain probability.

With the exception of Rosberg et al. [19] where there is a dedicated server in the downstream station, the aforementioned articles have assumed no dedicated capacity. Farrar [8,9] studied two versions of a clearing system with dedicated servers in both stations and one flexible server. In the constrained version, the flexible server can only work in the upstream station, whereas in the unconstrained version, the server can work in both stations. For both versions, he showed that the optimal policy is characterized by a switching curve; the flexible server is idled or assigned to the downstream station when the number of jobs there exceeds a certain threshold that depends on the number of jobs in the first queue. He also showed that the slope of the switching curve is greater than or equal to -1 . Pandelis [17] showed the validity of the results obtained by Farrar [8,9] for the case when jobs might leave the system after

completing service in the first station and specified subsets of the state space where the optimal policy can be explicitly determined. The same structure for the optimal policy was derived by Wu, Lewis, and Veatch [26] for a system with a number of flexible servers with the additional feature that all servers are subject to breakdowns. In their model, they assumed that processing requirements are the same in both stations, so actual processing times depend on the servers' speeds. Finally, Wu, Down, and Lewis [25] showed the optimality of a transition-monotone policy for the previous model with arrivals and no dedicated servers in the upstream station.

For clearing systems with no server breakdowns, we extend the results of Wu et al. [26] in two directions. First, we relax the assumption that the two stations are identical in terms of jobs' processing requirements. Second, we assume that all servers incur operating costs, different for each station, during the time they work. To the best of our knowledge, Weber and Stidham [23] is the only previous work that has considered operating costs. Under the additional assumption that operating costs for flexible servers are proportional to their speeds, we get an equivalent problem with one flexible server. Depending on holding costs, operating costs, and service rates, there might be situations when it is optimal to idle the flexible server (e.g., when the number of jobs in the two stations is low enough to yield higher operating than potential holding costs). We derive a condition under which idling the flexible server is not optimal when there are jobs in the downstream station. For this instance of the problem, we show that when the flexible server has higher relative operating cost than the dedicated server in the same station, the optimal policy possesses the switching curve properties proven by Farrar [8]. We also show for the case of no dedicated capacity in one of the two stations that it might be possible to determine the optimal policy explicitly. Specifically, the flexible server is always assigned to the station with no dedicated servers if the reduction in holding cost rate is larger in that station.

Assuming dedicated capacity in only one of the two stations, we obtain similar results for the problem with arrivals under the expected average cost criterion. After specifying the maximal arrival rate for which the problem has a solution and a condition for the optimality of nonidling policies, we show the following: (1) For dedicated capacity in the second stage only, the incentive to use the flexible server at the first stage is nonincreasing in the number of jobs at the second stage and (2) priority should be given to the station with no dedicated servers if the reduction in holding cost rate is larger and the operating cost is smaller there.

The article is organized as follows. In Section 2 the clearing model is formulated as a discrete-time Markov decision process and the structure of optimal nonidling policies is derived. The model with arrivals is analyzed in Section 3. Extensions are discussed as part of our concluding remarks in Section 4.

2. CLEARING SYSTEMS

We consider a tandem queuing system consisting of two stations. There is a number of jobs initially present in the system and no other external arrivals. Jobs that complete service in

the upstream station (station 1) transfer to the downstream station (station 2) where they receive additional service and then leave the system. Every job in station $i, i = 1, 2$, requires an exponentially distributed amount of service with mean Y_i . There are dedicated servers, one for each station, working at speeds s_{d1} and s_{d2} ; these servers are forced to work when there are jobs present in their corresponding station. There are also N additional servers that can serve in both stations; we assume that these servers can move from station to station instantaneously without any cost and work at speed $s_i, i = 1, 2, \dots, N$ —the same for both stations. Let μ_1, μ_2 and μ_{1i}, μ_{2i} denote the processing rates for jobs served by the dedicated servers and flexible server i respectively at stations 1 and 2. We have

$$\mu_1 = \frac{s_{d1}}{Y_1}, \quad \mu_2 = \frac{s_{d2}}{Y_2} \tag{1}$$

and

$$\mu_{1i} = \frac{s_i}{Y_1}, \quad \mu_{2i} = \frac{s_i}{Y_2}, \quad i = 1, 2, \dots, N. \tag{2}$$

Each job in station $i, i = 1, 2$, incurs a linear holding cost at rate h_i . Moreover, dedicated servers incur linear operating costs at rates b_1 and b_2 during the time they work, whereas flexible servers incur costs at rates c_{1i} and $c_{2i}, i = 1, 2, \dots, N$, during the time they work in stations 1 and 2, respectively. Our objective is to determine an allocation strategy for the flexible servers that minimizes the total expected holding and operating costs until the system is cleared of all jobs. Allowing preemptions at times of service completions and assuming that work prior to completions is lost, we formulate the problem as a Markov decision process with state space $\{(x_1, x_2) : x_1, x_2 \geq 0\}$, where x_1 and x_2 are the number of jobs in station 1 and station 2, respectively, including those in service. Starting from state (x_1, x_2) , we let $V(x_1, x_2)$ be the minimum total expected holding and operating cost incurred until the system empties, where $V(0, 0) = 0$. Let also $M = \{1, 2, \dots, N\}$ be the set of flexible servers. To get an expression for the minimum cost function $V(x_1, x_2)$, we minimize over all possible assignments of flexible servers to the two stations. After applying uniformization to convert the continuous-time problem to an equivalent discrete-time problem (see Serfozo [21]), we obtain dynamic programming equations (3)–(5), where terms $V_F^1(x_1, x_2)$ and $V_G^2(x_1, x_2)$ correspond to machines in subsets F and G of M being assigned to stations 1 and 2, respectively. In deriving (3)–(5), we have assumed that the uniformization rate is $\mu_1 + \mu_2 + v_1 + v_2 = 1$, where $v_1 = \sum_i \mu_{1i}$ and $v_2 = \sum_i \mu_{2i}$. We have also assumed that servers in the same station can collaborate to work on the same job, so that processing rates are additive. For $x_1, x_2 \geq 1$, we have

$$\begin{aligned} V(x_1, x_2) = & h_1x_1 + h_2x_2 + b_1 + b_2 + \mu_1V(x_1 - 1, x_2 + 1) \\ & + \mu_2V(x_1, x_2 - 1) + (v_1 + v_2)V(x_1, x_2) \\ & + \min_{\substack{F, G \subset M \\ F \cap G = \emptyset}} \{V_F^1(x_1, x_2) + V_G^2(x_1, x_2)\}, \end{aligned} \tag{3}$$

$$V(x_1, 0) = h_1 x_1 + b_1 + \mu_1 V(x_1 - 1, 1) + (\mu_2 + \nu_1 + \nu_2) V(x_1, 0) + \min_{FCM} \{V_F^1(x_1, 0)\}, \quad (4)$$

$$V(0, x_2) = h_2 x_2 + b_2 + \mu_2 V(0, x_2 - 1) + (\mu_1 + \nu_1 + \nu_2) V(0, x_2) + \min_{GCM} \{V_G^2(0, x_2)\}, \quad (5)$$

where

$$V_F^1(x_1, x_2) = c_{1F} + \mu_{1F} [V(x_1 - 1, x_2 + 1) - V(x_1, x_2)], \quad (6)$$

$$V_G^2(x_1, x_2) = c_{2G} + \mu_{2G} [V(x_1, x_2 - 1) - V(x_1, x_2)], \quad (7)$$

and

$$c_{1F} = \sum_{i \in F} c_{1i}, \quad c_{2G} = \sum_{i \in G} c_{2i}, \quad \mu_{1F} = \sum_{i \in F} \mu_{1i}, \quad \mu_{2G} = \sum_{i \in G} \mu_{2i}. \quad (8)$$

We assume that the following conditions hold.

(A1)

$$\frac{c_{1i}}{s_i} = \tilde{c}_1, \quad \frac{c_{2i}}{s_i} = \tilde{c}_2, \quad i = 1, 2, \dots, N, \text{ for some constants } \tilde{c}_1 \text{ and } \tilde{c}_2.$$

(A2)

$$\frac{c_2}{\nu_2} \leq \frac{h_2 + b_2}{\mu_2}.$$

(A3)

$$\frac{c_1}{\nu_1} \geq \frac{b_1}{\mu_1}, \quad \frac{c_2}{\nu_2} \geq \frac{b_2}{\mu_2}$$

Condition (A1) implies that operating costs incurred by flexible servers are proportional to their speeds. According to condition (A2), the mean operating cost for a job served in station 2 by the flexible server alone is no more than the corresponding mean holding and operating cost incurred during this job's service by the dedicated server alone. According to condition (A3), the operating cost per unit speed is greater for the flexible server compared to the dedicated server in the same station.

We show in the following proposition that under condition (A1), there exists an optimal policy that prescribes the same action for all flexible servers.

PROPOSITION 1: *Assume that condition (A1) holds. Then there exists an optimal policy that either idles all flexible servers or assigns them all to a single station.*

PROOF: ASSUMING $F \neq \emptyset$ and $m \in F$ and letting $F_m = F - \{m\}$, we obtain from condition (A1) and (1), (2), (6), and (8)

$$\begin{aligned} V_F^1(x_1, x_2) - V_{F_m}^1(x_1, x_2) &= c_{1m} + \mu_{1m}[V(x_1 - 1, x_2 + 1) - V(x_1, x_2)] \\ &= \mu_{1m}V_1(x_1, x_2), \end{aligned}$$

where

$$V_1(x_1, x_2) = \tilde{c}_1 Y_1 + V(x_1 - 1, x_2 + 1) - V(x_1, x_2).$$

Similarly, for $G \neq \emptyset$, $n \in G$, and $G_n = G - \{n\}$, we obtain from condition (A1) and (1), (2), (7), and (8)

$$\begin{aligned} V_G^2(x_1, x_2) - V_{G_n}^2(x_1, x_2) &= c_{2n} + \mu_{2n}[V(x_1, x_2 - 1) - V(x_1, x_2)] \\ &= \mu_{2n}V_2(x_1, x_2), \end{aligned}$$

where

$$V_2(x_1, x_2) = \tilde{c}_2 Y_2 + V(x_1, x_2 - 1) - V(x_1, x_2).$$

We have the following cases.

- (i) $V_1(x_1, x_2) \geq 0, V_2(x_1, x_2) \geq 0$: It is optimal to idle all flexible servers.
- (ii) $V_1(x_1, x_2) \geq 0, V_2(x_1, x_2) < 0$: It is optimal to assign all flexible servers to station 2 when there is at least one job present there, otherwise idling is optimal.
- (iii) $V_1(x_1, x_2) < 0, V_2(x_1, x_2) \geq 0$: It is optimal to assign all flexible servers to station 1 when there is at least one job present there, otherwise idling is optimal.
- (iv) $V_1(x_1, x_2) < 0, V_2(x_1, x_2) < 0$: Idling any flexible server is not optimal.

Therefore, it is optimal to either idle or use all flexible servers. It remains to show that when case (iv) holds, it is optimal to assign all servers to a single station. Let F and G be such that $F \cup G = M$ and $G \neq \emptyset$. For $i \in G$, we denote by F_i and G_i the sets of servers assigned to stations 1 and 2, respectively, after moving server i from station 2 to station 1 (i.e., $F_i = F \cup \{i\}, G_i = G - \{i\}$). From condition (A1) and (1), (2), and (6)–(8), we get

$$\begin{aligned} V_{F_i}^1(x_1, x_2) + V_{G_i}^2(x_1, x_2) - [V_F^1(x_1, x_2) + V_G^2(x_1, x_2)] \\ &= c_{1i} + \mu_{1i}[V(x_1 - 1, x_2 + 1) - V(x_1, x_2)] - c_{2i} \\ &\quad - \mu_{2i}[V(x_1, x_2 - 1) - V(x_1, x_2)] \\ &= \mu_{2i}V_{12}(x_1, x_2), \end{aligned} \tag{9}$$

where

$$\begin{aligned} V_{12}(x_1, x_2) &= (\tilde{c}_1 - \tilde{c}_2)Y_2 + \frac{Y_2}{Y_1}[V(x_1 - 1, x_2 + 1) - V(x_1, x_2)] \\ &\quad - [V(x_1, x_2 - 1) - V(x_1, x_2)]. \end{aligned}$$

Equation (9) implies that when $V_{12}(x_1, x_2) \geq 0$, it is optimal to assign all flexible servers to station 2; otherwise, assign them to station 1. ■

As a consequence of Proposition 1, replacing the N flexible servers with one server working at rates v_1 and v_2 and incurring operating costs at rates c_1 and c_2 given by $c_1 = \sum_i c_{1i}$ and $c_2 = \sum_i c_{2i}$ results in an equivalent problem. The optimality equations for this problem take the following form:

$$\begin{aligned} V(x_1, x_2) = & h_1 x_1 + h_2 x_2 + b_1 + b_2 + \mu_1 V(x_1 - 1, x_2 + 1) \\ & + \mu_2 V(x_1, x_2 - 1) + (v_1 + v_2)V(x_1, x_2) \\ & + \min\{0, V_M^1(x_1, x_2), V_M^2(x_1, x_2)\}, \end{aligned} \quad (10)$$

$$\begin{aligned} V(x_1, 0) = & h_1 x_1 + b_1 + \mu_1 V(x_1 - 1, 1) + (\mu_2 + v_1 + v_2)V(x_1, 0) \\ & + \min\{0, V_M^1(x_1, 0)\}, \end{aligned} \quad (11)$$

$$\begin{aligned} V(0, x_2) = & h_2 x_2 + b_2 + \mu_2 V(0, x_2 - 1) + (\mu_1 + v_1 + v_2)V(0, x_2) \\ & + \min\{0, V_M^2(0, x_2)\}. \end{aligned} \quad (12)$$

Next, we show that, under condition (A2), idling the flexible server is not optimal when there is at least one job in station 2. Actually, condition (A2) is the condition under which it is optimal to use the flexible server when there are no jobs in station 1. To see this, note that when there are no jobs present at one of the two stations, we can derive the following expressions for the value function by conditioning on the time of the first service completion:

$$V(x_1, 0) = \min\left\{\frac{h_1 x_1 + b_1}{\mu_1}, \frac{h_1 x_1 + b_1 + c_1}{\mu_1 + v_1}\right\} + V(x_1 - 1, 1), \quad (13)$$

$$V(0, x_2) = \min\left\{\frac{h_2 x_2 + b_2}{\mu_2}, \frac{h_2 x_2 + b_2 + c_2}{\mu_2 + v_2}\right\} + V(0, x_2 - 1). \quad (14)$$

Therefore, when station 1 or station 2 is empty of jobs, idling the flexible server is optimal for $x_2 < X_2$ and $x_1 < X_1$, respectively, where

$$X_2 = \frac{c_2 \mu_2 - b_2 v_2}{h_2 v_2}, \quad X_1 = \frac{c_1 \mu_1 - b_1 v_1}{h_1 v_1}.$$

Therefore, if condition (A2) holds, idling the flexible server is not optimal when $x_1 = 0$. We show in Proposition 2 that this is also true for $x_1 > 0$, provided that there are jobs in station 2. First, we need the following lemma (for the proof, see Appendix A).

LEMMA 1: Assume that condition (A2) holds. Then, for $x_1, x_2 \geq 1$,

$$V(x_1, x_2) - V(x_1, x_2 - 1) \geq \frac{h_2 x_2 + b_2 + c_2}{\mu_2 + v_2}.$$

PROPOSITION 2: Condition (A2) is a sufficient condition under which the optimal policy always uses the flexible server when there are jobs present in the downstream station.

PROOF: Lemma 1 and condition (A2) yield

$$v_2[V(x_1, x_2) - V(x_1, x_2 - 1)] - c_2 \geq 0$$

[i.e., $V_M^2(x_1, x_2) \leq 0$], which by (10) implies that for $x_1, x_2 \geq 1$, the policy that assigns the flexible server to station 2 has no more cost than the one that idles the server. ■

Assuming condition (A2) holds and using (6)–(8), optimality equations (10)–(12) are reduced to the following. For $x_1, x_2 \geq 1$,

$$\begin{aligned} V(x_1, x_2) = & h_1x_1 + h_2x_2 + b_1 + b_2 + \mu_1V(x_1 - 1, x_2 + 1) \\ & + \mu_2V(x_1, x_2 - 1) + \min\{c_2 + v_2V(x_1, x_2 - 1) \\ & + v_1V(x_1, x_2), c_1 + v_1V(x_1 - 1, x_2 + 1) + v_2V(x_1, x_2)\}. \end{aligned} \tag{15}$$

For $x_2 = 0$ or $x_1 = 0$,

$$\begin{aligned} V(x_1, 0) = & h_1x_1 + b_1 + \mu_1V(x_1 - 1, 1) + (\mu_2 + v_2)V(x_1, 0) \\ & + \min\{v_1V(x_1, 0), c_1 + v_1V(x_1 - 1, 1)\}, \end{aligned} \tag{16}$$

$$\begin{aligned} V(0, x_2) = & h_2x_2 + b_2 + \mu_2V(0, x_2 - 1) + (\mu_1 + v_1)V(0, x_2) \\ & + \min\{c_2 + v_2V(0, x_2 - 1), v_2V(0, x_2)\}. \end{aligned} \tag{17}$$

The first (resp. second) argument in the minimization terms corresponds to assigning the flexible server to station 2 (resp. station 1). When there are no jobs present in a station, assigning the additional server to that station is equivalent to idling. Although we have established that idling is not optimal when condition (A2) is true, we have left the corresponding term in (17) to facilitate subsequent analysis.

Based on (15)–(17), we define the decision function $d(x_1, x_2)$ as follows:

$$\begin{aligned} d(x_1, x_2) = & c_2 - c_1 + v_2[V(x_1, x_2 - 1) - V(x_1, x_2)] \\ & + v_1[V(x_1, x_2) - V(x_1 - 1, x_2 + 1)], \quad x_1, x_2 \geq 1, \\ d(x_1, 0) = & v_1[V(x_1, 0) - V(x_1 - 1, 1)] - c_1, \quad x_1 \geq 1, \\ d(0, x_2) = & v_2[V(0, x_2 - 1) - V(0, x_2)] + c_2, \quad x_2 \geq 1. \end{aligned}$$

The definition of the function $d(x_1, x_2)$ implies that it is optimal to assign the flexible server to station 1 (resp. station 2) when $d(x_1, x_2) > 0$ (resp. $d(x_1, x_2) \leq 0$); $d(x_1, x_2)$ denotes the incentive to have the flexible server work in station 1 instead of station 2 in state (x_1, x_2) .

Remark: We can explicitly determine $d(x_1, 0)$ and $d(0, x_2)$ using (13) and (14). We have

$$d(x_1, 0) = \frac{v_1 h_1 x_1 + v_1 b_1 - \mu_1 c_1}{\mu_1 + v_1 \mathbf{1}(x_1 \geq X_1)}, \quad (18)$$

$$d(0, x_2) = -\frac{v_2 h_2 x_2 + v_2 b_2 - \mu_2 c_2}{\mu_2 + v_2}. \quad (19)$$

It can be easily verified that condition (A2) yields $d(0, x_2) \leq 0$; that is, it does not pay to idle the flexible server. Also, $d(x_1, 0) \leq 0$ for $x_1 \leq X_1$ and $d(x_1, 0) \geq 0$ for $x_1 \geq X_1$.

Using optimality equations (15)–(17) and identities $\min(a, b) = b + (a - b)^-$ and $\min(a, b) = a - (a - b)^+$ for the first and second terms, respectively, of the differences appearing in the definition of $d(x_1, x_2)$, we obtain the following recursive expressions for function $d(x_1, x_2)$, $x_1, x_2 \geq 1$:

$$d(x_1, x_2) = f(x_1, x_2) + v_1 d(x_1, x_2)^- + v_2 d(x_1, x_2)^+, \quad (20)$$

where

$$\begin{aligned} f(x_1, x_2) = & v_1(h_1 - h_2) - v_2 h_2 + \mu_1 d(x_1 - 1, x_2 + 1) + \mu_2 d(x_1, x_2 - 1) \\ & + v_1 d(x_1 - 1, x_2 + 1)^+ + v_2 d(x_1, x_2 - 1)^- \\ & + (v_1 b_1 - \mu_1 c_1) \mathbf{1}(x_1 = 1) + (\mu_2 c_2 - v_2 b_2) \mathbf{1}(x_2 = 1). \end{aligned} \quad (21)$$

The expressions derived for $d(x_1, x_2)$ provide an efficient way for the numerical computation of the optimal policy; starting with the values of $d(x_1, 0)$ and $d(0, x_2)$ obtained from (18) and (19), we can successively use (20) and (21) to compute $d(x_1, x_2)$ for any $x_1, x_2 \geq 1$. Furthermore, they provide the basis for proving qualitative properties of the optimal policy. Specifically, the optimal policy is characterized by a single switching curve (when the number of jobs present in the downstream station exceeds a certain threshold, it is optimal to assign the flexible server to that station), which is the main result of this section given in the following theorem.

THEOREM 1: *Assume conditions (A2) and (A3) are true. Then, for each $x_1 \geq 1$, there exists a positive integer $t(x_1)$ such that it is optimal to assign the flexible server to the downstream station only when $x_2 \geq t(x_1)$. Moreover, $t(x_1) = 1$ for $x_1 \leq X_1$, and the slope of $t(x_1)$ is greater than or equal to -1 .*

Remarks: We believe that the slope of the switching curve is actually nonnegative and the lower bound of -1 is only the most we can prove. To verify our intuition, we used (18)–(21) to determine the optimal policy for several thousand cases, where, for each case, the parameter values were selected randomly. Not even one case exhibited a decreasing switching curve. On the other hand, condition (A3) is not just a technical condition needed for the proof. The following example illustrates that if condition (A3) is not satisfied, the existence of a single switching curve is not guaranteed.

Example: Consider a system with $\mu_1 = 0.2$, $\mu_2 = 0.3$, $\nu_1 = \nu_2 = 0.25$, $h_1 = 2$, $h_2 = 1$, $b_1 = 6$, $b_2 = 10$, $c_1 = 3$, and $c_2 = 4$. It is easy to verify that condition (A2) is satisfied, but not condition (A3). Using (18)–(21) to compute the decision function for $x_1 = 1$, we get $d(1, 0) = 3.1111$, $d(1, 1) = -0.1616$, $d(1, 2) = 0.0875$, $d(1, 3) = 0.1199$, $d(1, 4) = 0.0116$, and $d(1, 5) = -0.1529$. We see that the optimal policy is switching between stations 1 and 2 three times, so it is not characterized by a single switching curve.

The proof of Theorem 1 requires the use of Lemmas 3–5, whose proofs are given in Appendix A. Lemma 2 is an auxiliary result.

LEMMA 2: Assume $A - B = g + \lambda(A^- - B^-) + \mu(A^+ - B^+)$ with $0 < \lambda < 1$ and $0 < \mu < 1$. Then $A - B$ and g have the same sign.

PROOF: Assuming that $A \neq B$, $A \leq B$, and $A \geq B$ for $g = 0$, $g > 0$, and $g < 0$, respectively, we are led to contradictions. ■

LEMMA 3: For all $x_1 \geq 1$ and $x_2 \geq 0$,

$$d(x_1, x_2 + 1) < d(x_1, x_2).$$

LEMMA 4: For all $x_1 > 0$,

$$\lim_{x_2 \rightarrow \infty} d(x_1, x_2) = -\infty.$$

LEMMA 5: For all $x_1, x_2 \geq 1$,

$$d(x_1, x_2) > 0 \Rightarrow d(x_1, x_2) \leq d(x_1 + 1, x_2 - 1).$$

PROOF OF THEOREM 1: For $x_1 \leq X_1$, we have $d(x_1, 0) \leq 0$, which implies that $d(x_1, x_2) < 0$ for all $x_2 \geq 1$ (Lemma 3). Therefore, $t(x_1) = 1$, $x_1 \leq X_1$. For $x_1 > X_1$, we have $d(x_1, 0) > 0$, $d(x_1, x_2)$ decreasing in x_2 (Lemma 3), and $d(x_1, x_2) < 0$ for x_2 sufficiently large (Lemma 4), so there exists an integer $t(x_1)$ such that $d(x_1, x_2) \leq 0$ for $x_2 \geq t(x_1)$. Finally, by Lemma 5, we have for any $x_1, x_2 \geq 1$,

$$d(x_1, x_2) > 0 \Rightarrow d(x_1 + 1, x_2 - 1) > 0.$$

Therefore, the slope of the switching curve $t(x_1)$ defined by the optimal policy is greater than or equal to -1 . ■

We end this section by considering systems with no dedicated capacity at the upstream (resp. downstream) station. We show that the optimal policy assigns the flexible server to the upstream (resp. downstream) station when the reduction in holding cost rate resulting from the use of the flexible server is larger in that station.

THEOREM 2: Let $\mu_1 = 0$ and $v_1(h_1 - h_2) \geq v_2h_2$. Then it is optimal to assign the flexible server at station 1 for all $x_1 > 0$.

PROOF: We will show that $d(x_1, x_2) \geq 0$ for $x_1 > 0$. From (20) and (21), we have for $x_1, x_2 \geq 1$,

$$\begin{aligned} d(x_1, x_2) &= v_1(h_1 - h_2) - v_2h_2 + \mu_2d(x_1, x_2 - 1) \\ &\quad + v_1d(x_1 - 1, x_2 + 1)^+ + v_1d(x_1, x_2 - 1)^- \\ &\quad + v_1b_1\mathbf{1}(x_1 = 1) + (\mu_2c_2 - v_2b_2)\mathbf{1}(x_2 = 1) \\ &\quad + v_1d(x_1, x_2)^- + v_2d(x_1, x_2)^+. \end{aligned}$$

We prove the result by induction on x_2 . We start the induction by noting that $d(x_1, 0) = h_1x_1 > 0$. Then the induction step follows from condition (A3) and Lemma 2. ■

THEOREM 3: Let $\mu_2 = 0$ and $v_1(h_1 - h_2) \leq v_2h_2$. Then it is optimal to assign the flexible server at station 2 for all $x_2 > 0$.

PROOF: We show $d(x_1, x_2) \leq 0$ for $x_2 > 0$ by induction on x_1 . We have $d(0, x_2) = -h_2x_2 < 0$, and for $x_1, x_2 \geq 1$,

$$\begin{aligned} d(x_1, x_2) &= v_1(h_1 - h_2) - v_2h_2 + \mu_1d(x_1 - 1, x_2 + 1) \\ &\quad + v_1d(x_1 - 1, x_2 + 1)^+ + v_2d(x_1, x_2 - 1)^- \\ &\quad + (v_1b_1 - \mu_1c_1)\mathbf{1}(x_1 = 1) - v_2b_2\mathbf{1}(x_2 = 1) \\ &\quad + v_1d(x_1, x_2)^- + v_2d(x_1, x_2)^+. \end{aligned}$$

The result is proven by the induction hypothesis, condition (A3), and Lemma 2. ■

3. SYSTEMS WITH ARRIVALS

In this section we consider the previous model with Poisson arrivals of rate λ . We assume that condition (A1) holds and there are no operating costs for the dedicated servers ($b_1 = b_2 = 0$). Let β , $0 \leq \beta \leq 1$, be a discount factor. We define by $V_{n,\beta}^\theta(x_1, x_2)$ and $V_\beta^\theta(x_1, x_2)$ the expected n -step discounted cost and the expected infinite horizon discounted cost, respectively, under policy θ starting at state (x_1, x_2) . We also define the expected average cost under θ by

$$J^\theta(x_1, x_2) = \limsup_{n \rightarrow \infty} \frac{V_{n,1}^\theta(x_1, x_2)}{n}.$$

Our objective is to characterize an allocation strategy for the flexible servers that minimizes the expected average cost. We were not able to do so for the general model with dedicated servers in both stations, but for special cases with a dedicated server in one of the two stations.

Assuming that the uniformization rate is $\lambda + \mu_1 + \mu_2 + \nu_1 + \nu_2 = 1$, we obtain optimality equations (22)–(24) for the three aforementioned problems. In deriving these equations we have taken into account condition (A1), under which the optimal policy prescribes the same action for all flexible servers, either idling them or assigning them to the same station (the proof is along the lines of that of Proposition 1).

$$V_{n,\beta}(x_1, x_2) = h_1x_1 + h_2x_2 + T_\beta V_{n-1,\beta}(x_1, x_2) \tag{22}$$

$$V_\beta(x_1, x_2) = h_1x_1 + h_2x_2 + T_\beta V_\beta(x_1, x_2) \tag{23}$$

$$J + w(x_1, x_2) = h_1x_1 + h_2x_2 + T_1 w(x_1, x_2), \tag{24}$$

where the operator T_β is defined by

$$\begin{aligned} T_\beta V(x_1, x_2) = & \beta[\lambda V(x_1 + 1, x_2) + \mu_1 V(x_1 - 1, x_2 + 1) \\ & + \mu_2 V(x_1, x_2 - 1)] \\ & + \min\{\beta(\nu_1 + \nu_2)V(x_1, x_2), \\ & c_1 + \beta[\nu_1 V(x_1 - 1, x_2 + 1) + \nu_2 V(x_1, x_2)], \\ & c_2 + \beta[\nu_2 V(x_1, x_2 - 1) + \nu_1 V(x_1, x_2)]\}, \end{aligned} \tag{25}$$

$$\begin{aligned} T_\beta V(x_1, 0) = & \beta[\lambda V(x_1 + 1, 0) + \mu_1 V(x_1 - 1, 1) + \mu_2 V(x_1, 0)] \\ & + \min\{\beta(\nu_1 + \nu_2)V(x_1, 0), \\ & c_1 + \beta[\nu_1 V(x_1 - 1, 1) + \nu_2 V(x_1, 0)]\}, \end{aligned} \tag{26}$$

$$\begin{aligned} T_\beta V(0, x_2) = & \beta[\lambda V(1, x_2) + \mu_2 V(0, x_2 - 1) + \mu_1 V(0, x_2)] \\ & + \min\{\beta(\nu_1 + \nu_2)V(0, x_2), \\ & c_2 + \beta[\nu_2 V(0, x_2 - 1) + \nu_1 V(0, x_2)]\}, \end{aligned} \tag{27}$$

$$T_\beta V(0, 0) = \beta[\lambda V(1, 0) + (\mu_1 + \mu_2 + \nu_1 + \nu_2)V(0, 0)]. \tag{28}$$

The solution of (22) gives the minimum expected n -step discounted cost (with $V_{0,\beta}(x_1, x_2) = 0$), whereas the solution of (23), if it exists, gives the minimum infinite horizon discounted cost. Finally, if a solution (J, w) of (24) exists and π is the policy that attains the minimum on its right-hand side, π is the optimal policy and J is the minimum average cost satisfying

$$J = \inf_{\theta} J^\theta(x_1, x_2) = J^\pi(x_1, x_2) \quad \text{for all } x_1 \text{ and } x_2.$$

Propositions 3 and 4 make statements about the existence of solutions to (23) and (24), respectively, and their properties. For the proofs, we refer readers to Wu et al. [25], where analogous results are proven for the problem with holding costs and unreliable servers. The same arguments remain valid for our problem because linear operating costs make a finite contribution to the infinite horizon discounted cost and the average cost.

PROPOSITION 3: For $0 < \beta < 1$, (23) has a unique solution given by

$$V_\beta(x_1, x_2) = \lim_{n \rightarrow \infty} V_{n,\beta}(x_1, x_2).$$

We define the maximal throughput Λ as the largest value of the arrival rate that is less than or equal to the average service rate at each station. The maximal throughput can be obtained from the solution of the linear program

$$\begin{aligned} \max \quad & \Lambda \\ \text{s.t.} \quad & \Lambda \leq \mu_i + v_i \delta_i, \quad i = 1, 2, \\ & \delta_1 + \delta_2 \leq 1, \\ & \delta_1, \delta_2 \geq 0, \end{aligned}$$

where δ_1 and δ_2 are the proportions of time the flexible server is assigned to the upstream and downstream station, respectively. It is straightforward to show that

$$\Lambda = \begin{cases} \mu_1 + v_1 & \text{if } \mu_1 + v_1 \leq \mu_2 \\ \mu_2 + v_2 & \text{if } \mu_2 + v_2 \leq \mu_1 \\ \frac{\mu_1 v_2 + v_1 \mu_2 + v_1 v_2}{v_1 + v_2} & \text{otherwise.} \end{cases}$$

PROPOSITION 4: For $\lambda < \Lambda$ (24) has a solution (J, w) , where J is given by $\lim_{n \rightarrow \infty} V_{n,1}(x_1, x_2)/n$. Moreover, for a fixed state (x_1^*, x_2^*) , there exists a sequence $\beta_l \rightarrow 1$ for which

$$\lim_{l \rightarrow \infty} [V_{\beta_l}(x_1, x_2) - V_{\beta_l}(x_1^*, x_2^*)] = w(x_1, x_2).$$

We prove that when the reduction in holding cost rate resulting from using the flexible server in the downstream station is larger than the corresponding operating cost, idling the flexible server is not optimal when there are jobs present there.

PROPOSITION 5: Assume $v_2 h_2 > c_2$. Then idling the flexible server is not optimal for $x_2 > 0$.

PROOF: We start by showing that for β sufficiently close to 1, the result is true for the finite horizon problem, except for $n = 1$, in which case, idling the flexible server is optimal because there is nothing to be gained in terms of potential holding costs savings. In particular, we will show that the policy that assigns the flexible server to station 2 has no more cost than the one that idles the server, which by (25) and (27) implies that for $n \geq 1$,

$$\beta v_2 [V_{n,\beta}(x_1, x_2) - V_{n,\beta}(x_1, x_2 - 1)] \geq c_2. \tag{29}$$

A sufficient condition for (29) to be true is

$$V_{n,\beta}(x_1, x_2) - V_{n,\beta}(x_1, x_2 - 1) \geq h_2, \tag{30}$$

because then (29) would be satisfied for $\beta \geq c_2/v_2 h_2$. We prove (30) by induction on n . For $n = 1$, it is satisfied with equality. For $n > 1$, we have from (22),

$$\begin{aligned} V_{n,\beta}(x_1, x_2) - V_{n,\beta}(x_1, x_2 - 1) &= h_2 + T_\beta V_{n-1,\beta}(x_1, x_2) \\ &\quad - T_\beta V_{n-1,\beta}(x_1, x_2 - 1). \end{aligned}$$

Using (25)–(28) and the induction hypothesis, we get by a straightforward term-by-term comparison that $T_\beta V_{n-1,\beta}(x_1, x_2) \geq T_\beta V_{n-1,\beta}(x_1, x_2 - 1)$, which completes the induction. Letting $n \rightarrow \infty$ in (29) and applying Proposition 3, we obtain for $\beta \geq c_2/v_2h_2$,

$$\beta v_2[V_\beta(x_1, x_2) - V_\beta(x_1, x_2 - 1)] \geq c_2. \tag{31}$$

Applying now Proposition 4 to (31), we get

$$v_2[w(x_1, x_2) - w(x_1, x_2 - 1)] \geq c_2,$$

which by (24), (25), and (27) proves the result for the average cost problem. ■

To further characterize the policy that minimizes the average cost, we determine properties of the optimal policy for the finite horizon problem and then use Propositions 3 and 4 to show that the same properties are valid under the average cost criterion. We define decision function $d_{n,\beta}(x_1, x_2)$ as in the previous section for clearing systems:

$$\begin{aligned} d_{n,\beta}(x_1, x_2) &= c_2 - c_1 + \beta\{v_2[V_{n,\beta}(x_1, x_2 - 1) - V_{n,\beta}(x_1, x_2)] + v_1[V_{n,\beta}(x_1, x_2) \\ &\quad - V_{n,\beta}(x_1 - 1, x_2 + 1)]\}, \quad x_1, x_2 \geq 1, \\ d_{n,\beta}(x_1, 0) &= \beta v_1[V_{n,\beta}(x_1, 0) - V_{n,\beta}(x_1 - 1, 1)] - c_1, \quad x_1 \geq 1, \\ d_{n,\beta}(0, x_2) &= \beta v_2[V_{n,\beta}(0, x_2 - 1) - V_{n,\beta}(0, x_2)] + c_2, \quad x_2 \geq 1. \end{aligned}$$

Under the assumption of Proposition 5, $d_{n,\beta}(x_1, x_2)$, $n \geq 1$, determines the optimal policy at stage $n + 1$; the flexible server is assigned to the upstream station if $d_{n,\beta}(x_1, x_2) \geq 0$, and downstream otherwise. The following three lemmas characterize that optimal policy for large values of the discount factor when there is no dedicated capacity in one of the two stations. Their proofs can be found in Appendix B.

LEMMA 6: Let $\mu_1 = 0$, $c_2 < v_2h_2$, and $\beta \geq \beta_1$, where

$$\begin{aligned} \beta_1 &= \max \left\{ \frac{c_2}{v_2h_2}, \frac{2c_1 - c_2}{v_2h_2 - (v_1 + \mu_2)c_2 + 2c_1 - c_2} \right\} \quad \text{if } 2c_1 > c_2, \\ \beta_1 &= \frac{c_2}{v_2h_2}, \quad \text{otherwise.} \end{aligned}$$

Then for $n \geq 1$, $x_1 \geq 1$, and $x_2 \geq 0$,

$$d_{n,\beta}(x_1, x_2 + 1) \leq d_{n,\beta}(x_1, x_2).$$

LEMMA 7: Let $\mu_1 = 0$, $c_1 \leq c_2 < v_2h_2 < v_1(h_1 - h_2)$, and $\beta \geq \beta_2$, where

$$\beta_2 = \max \left\{ \frac{c_2}{v_2h_2}, \frac{c_2 - c_1}{v_1(h_1 - h_2) - v_2h_2 + c_2 - c_1} \right\}.$$

Then for $n \geq 1$, $x_1 \geq 1$, and $x_2 \geq 0$,

$$d_{n,\beta}(x_1, x_2) \geq 0.$$

LEMMA 8: Let $\mu_2 = 0$, $c_2 < v_2 h_2$, $c_2 \leq c_1$, $v_1(h_1 - h_2) < v_2 h_2$, and $\beta \geq \beta_3$, where

$$\beta_3 = \max \left\{ \frac{c_2}{v_2 h_2}, \frac{c_1 - c_2}{v_2 h_2 - v_1(h_1 - h_2) + c_1 - c_2} \right\}.$$

Then for $n \geq 1$, $x_1 \geq 0$, and $x_2 \geq 1$,

$$d_{n,\beta}(x_1, x_2) \leq 0.$$

When there is no dedicated server in station 1, Lemma 6 implies that the gain from having the flexible server work at station 1 is nonincreasing in the number of jobs in station 2. According to Lemma 7, it is always optimal to assign the flexible server at station 1 if the operating cost is smaller and the reduction in holding cost rate is larger in that station. By reversing these conditions, we show in Lemma 8 that the optimal policy always assigns the flexible server at station 2. Theorems 4–6 are the counterparts of Lemmas 6–8 for the average cost criterion.

THEOREM 4: If $\mu_1 = 0$ and $c_2 < v_2 h_2$, the incentive to use the flexible server at station 1 does not increase as the number of jobs in station 2 increases.

PROOF: Applying successively Propositions 3 and 4 to the result of Lemma 6, we get

$$\begin{aligned} c_2 + v_2[w(x_1, 0) - w(x_1, 1)] + v_1[w(x_1, 1) - w(x_1 - 1, 2)] \\ \leq v_1[w(x_1, 0) - w(x_1 - 1, 1)], \end{aligned}$$

and for $x_2 \geq 1$,

$$\begin{aligned} v_2[w(x_1, x_2) - w(x_1, x_2 + 1)] + v_1[w(x_1, x_2 + 1) - w(x_1 - 1, x_2 + 2)] \\ \leq v_2[w(x_1, x_2 - 1) - w(x_1, x_2)] + v_1[w(x_1, x_2) \\ - w(x_1 - 1, x_2 + 1)], \end{aligned}$$

which in view of (24)–(26) proves the statement of the theorem. ■

THEOREM 5: If $\mu_1 = 0$ and $c_1 \leq c_2 < v_2 h_2 < v_1(h_1 - h_2)$, the optimal policy gives priority to the upstream station.

THEOREM 6: If $\mu_2 = 0$ and $c_2 < v_2 h_2$, $c_2 \leq c_1$, $v_1(h_1 - h_2) < v_2 h_2$, the optimal policy gives priority to the downstream station.

The proofs of Theorems 5 and 6 are based on Lemmas 7 and 8, respectively, and are derived by identical arguments to that of Theorem 4.

4. DISCUSSION

We considered scheduling a number of flexible servers in a two-stage tandem queuing system with dedicated servers at each stage, where operating costs are incurred by the servers during the time they work. Assuming a collaborative work discipline and exponential service times, we showed that when operating cost rates due to flexible servers are proportional to their speeds, we get an equivalent problem with one flexible server. Furthermore, we identified conditions on holding costs, operating costs, and service times (see Propositions 2 and 5) under which there exist optimal nonidling policies for the flexible server when there are jobs in station 2. The analysis in the article focused on deriving properties of the optimal policy when the conditions that ensure the optimality of nonidling policies are satisfied.

For clearing systems, we showed that under condition (A2), allocating the flexible server to the downstream station becomes optimal as the number of jobs there increases. Moreover, the slope of the switching curve defined by the optimal policy is greater than or equal to -1 . When condition (A2) is not satisfied, in which case idling is optimal for certain states of the problem, we have not been able to derive any structural properties of the optimal policy. Recall that when there are no jobs present in station 1 or 2, idling the flexible server is optimal when $x_2 < X_2$ and $x_1 < X_1$, respectively, where X_1 and X_2 have been defined in Section 2. We conjecture that when $x_2 \geq X_2$, idling is not optimal for any number of jobs in station 1. We believe this is true for the following reason. If idling is not optimal when station 1 is empty, there must be no issue with high operating costs in station 2, which is a fact independent of jobs present in station 1. The reverse is not true, that is, when $x_1 \geq X_1$, idling might be optimal when there are jobs in station 2. This is the case when holding and operating costs in station 2 are so large that it would not be beneficial to send jobs to station 2 (a process that would be facilitated by the use of the flexible server in station 1) or have the flexible server work in station 2.

For systems with arrivals, we showed the optimality of nonidling policies when the reduction in holding cost rate is larger than the operating cost in station 2. Moreover, we showed that the transition-monotonicity property of the optimal policy for clearing systems remains valid when there is no dedicated capacity in the upstream station. It is not clear whether or under what conditions this property holds for the general model as well. However, some numerical experimentation with the finite horizon problem seems to indicate that this might be the case.

References

1. Ahn, H.S., Duenyas, I., & Lewis, M.E. (2002). The optimal control of a two-stage tandem queuing system with flexible servers. *Probability in the Engineering and Informational Sciences* 16: 453–469.
2. Ahn, H.S., Duenyas, I., & Zhang, R. (1999). Optimal stochastic scheduling of a two-stage tandem queue with parallel servers. *Advances in Applied Probability* 31: 1095–1117.
3. Ahn, H.S. & Righter, R. (2006). Dynamic load balancing with flexible workers. *Advances in Applied Probability* 38: 621–642.

4. Andradottir, S., Ayhan, H., & Down, D.G. (2001). Server assignment policies for maximizing the steady-state throughput of finite queueing systems. *Management Science* 47: 1421–1439.
5. Andradottir, S., Ayhan, H., & Down, D.G. (2003). Dynamic server allocation for queueing networks with flexible servers. *Operations Research* 51: 952–968.
6. Andradottir, S., Ayhan, H., & Down, D.G. (2007). Compensating for failures with flexible servers. *Operations Research* 55: 753–768.
7. Duenyas, I., Gupta, D., & Olsen, T. (1998). Control of a single server tandem queueing system with setups. *Operations Research* 46: 218–230.
8. Farrar, T.M. (1992). Resource allocation in systems of queues. Ph.D. dissertation, Cambridge University, Cambridge.
9. Farrar, T.M. (1993). Optimal use of an extra server in a two station tandem queueing network. *IEEE Transactions on Automatic Control* 38: 1296–1299.
10. Gel, E.S., Hopp, W.J., & Van Oyen, M.P. (2002). Factors affecting opportunity of worksharing as a dynamic load balancing mechanism. *IIE Transactions* 34: 847–863.
11. Gel, E.S., Hopp, W.J., & Van Oyen, M.P. (2007). Hierarchical cross-training in work-in-process-constrained systems. *IIE Transactions* 39: 125–143.
12. Hopp, W.J., Tekin, E., & Van Oyen, M.P. (2004). Benefits of skill-chaining in production lines with cross-trained workers. *Management Science* 50: 83–98.
13. Hopp, W.J. & Van Oyen, M.P. (2004). Agile workforce evaluation: A framework for cross-training and coordination. *IIE Transactions* 36: 919–940.
14. Iravani, S.M., Posner, M.J., & Buzacott, J.A. (1997). A two-stage tandem queue attended by a moving server with holding and switching costs. *Queueing Systems* 26: 203–228.
15. Iravani, S.M., Van Oyen, M.P., & Sims, K.T. (2005). Structural flexibility: A new perspective on the design of manufacturing and service operations. *Management Science* 51: 151–166.
16. Narongwanich, W., Duenyas, I., & Birge, J. (2003). Optimal portfolio of reconfigurable and dedicated capacity under uncertainty. <http://users.iems.northwestern.edu/~jrbirge/Public/html/new.html>
17. Pandelis, D.G. (2007). Optimal use of excess capacity in two interconnected queues. *Mathematical Methods of Operations Research* 65: 179–192.
18. Pandelis, D.G. & Teneketzis, D. (1994). Optimal multiserver stochastic scheduling of two interconnected priority queues. *Advances in Applied Probability* 26: 258–279.
19. Rosberg, Z., Varaiya, P., & Walrand, J. (1982). Optimal control of service in tandem queues. *IEEE Transactions on Automatic Control* 27: 600–609.
20. Schiefermayr, K. & Weichbold, J. (2005). A complete solution for the optimal stochastic scheduling of a two-stage tandem queue with two flexible servers. *Journal of Applied Probability* 42: 778–796.
21. Serfozo, R. (1978). An equivalence between continuous and discrete time Markov decision processes. *Operations Research* 27: 616–620.
22. Van Oyen, M.P., Gel, E.S., & Hopp, W.J. (2001). Performance opportunity for workforce agility in collaborative and noncollaborative work systems. *IIE Transactions* 33: 761–777.
23. Weber, R.R. & Stidham, S. (1987). Optimal control of service rates in networks of queues. *Advances in Applied Probability* 19: 202–218.
24. Weichbold, J. & Schiefermayr, K. (2006). The optimal control of a general tandem queue. *Probability in the Engineering and Informational Sciences* 20: 307–327.
25. Wu, C.H., Down, D.G., & Lewis, M.E. (2007). Heuristics for allocation of reconfigurable resources in a serial line with reliability considerations. *IEEE Transactions* (In Press).
26. Wu, C.H., Lewis, M.E., & Veatch, M. (2006). Dynamic allocation of reconfigurable resources in a two-stage tandem queueing system with reliability considerations. *IEEE Transactions on Automatic Control* 51: 309–314.

APPENDIX A
Proofs of Lemmas 1 and 3–5

PROOF OF LEMMA 1: Let $V^{\pi_i}(x_1, x_2)$, $i = 0, 1, 2$, be the expected cost of a policy that at state (x_1, x_2) idles the extra server, assigns it to station 1, or assigns it to station 2, respectively, and proceeds optimally after the first service completion. Therefore,

$$V(x_1, x_2) - V(x_1, x_2 - 1) = \min_{i=0,1,2} \{V^{\pi_i}(x_1, x_2) - V(x_1, x_2 - 1)\}, \tag{A.1}$$

$$V(x_1, x_2 - 1) \leq V^{\pi_i}(x_1, x_2 - 1),$$

implying that

$$V(x_1, x_2) - V(x_1, x_2 - 1) \geq \min_{i=0,1,2} \{V^{\pi_i}(x_1, x_2) - V^{\pi_i}(x_1, x_2 - 1)\}. \tag{A.2}$$

For $x_1, x_2 \geq 1$, by conditioning on the time of the first service completion, we derive the following expressions:

$$\begin{aligned} V^{\pi_0}(x_1, x_2) &= \frac{h_1x_1 + h_2x_2 + b_1 + b_2}{\mu_1 + \mu_2} + V(x_1 - 1, x_2 + 1) \frac{\mu_1}{\mu_1 + \mu_2} \\ &\quad + V(x_1, x_2 - 1) \frac{\mu_2}{\mu_1 + \mu_2}, \end{aligned} \tag{A.3a}$$

$$\begin{aligned} V^{\pi_1}(x_1, x_2) &= \frac{h_1x_1 + h_2x_2 + b_1 + b_2 + c_1}{\mu_1 + \mu_2 + \nu_1} + V(x_1 - 1, x_2 + 1) \frac{\mu_1 + \nu_1}{\mu_1 + \mu_2 + \nu_1} \\ &\quad + V(x_1, x_2 - 1) \frac{\mu_2}{\mu_1 + \mu_2 + \nu_1}, \end{aligned} \tag{A.3b}$$

$$\begin{aligned} V^{\pi_2}(x_1, x_2) &= \frac{h_1x_1 + h_2x_2 + b_1 + b_2 + c_2}{\mu_1 + \mu_2 + \nu_2} + V(x_1 - 1, x_2 + 1) \frac{\mu_1}{\mu_1 + \mu_2 + \nu_2} \\ &\quad + V(x_1, x_2 - 1) \frac{\mu_2 + \nu_2}{\mu_1 + \mu_2 + \nu_2}. \end{aligned} \tag{A.3c}$$

We will show that

$$V^{\pi_i}(x_1, 1) - V(x_1, 0) \geq \frac{h_2 + b_2 + c_2}{\mu_2 + \nu_2}, \quad i = 0, 1, 2, \tag{A.4}$$

$$V^{\pi_i}(x_1, x_2) - V^{\pi_i}(x_1, x_2 - 1) \geq \frac{h_2x_2 + b_2 + c_2}{\mu_2 + \nu_2}, \quad i = 0, 1, 2, x_2 > 1, \tag{A.5}$$

which by (A.1) and (A.2) suffice to prove the statement of the lemma. From (A.3a)–(A.3c), we have

$$V^{\pi_0}(x_1, 1) - V(x_1, 0) = \frac{h_1x_1 + h_2 + b_1 + b_2}{\mu_1 + \mu_2} + [V(x_1 - 1, 2) - V(x_1, 0)] \frac{\mu_1}{\mu_1 + \mu_2}, \tag{A.6a}$$

$$\begin{aligned} V^{\pi_1}(x_1, 1) - V(x_1, 0) &= \frac{h_1x_1 + h_2 + b_1 + b_2 + c_1}{\mu_1 + \mu_2 + \nu_1} \\ &\quad + [V(x_1 - 1, 2) - V(x_1, 0)] \frac{\mu_1 + \nu_1}{\mu_1 + \mu_2 + \nu_1}, \end{aligned} \tag{A.6b}$$

$$\begin{aligned}
 V^{\pi_2}(x_1, 1) - V(x_1, 0) &= \frac{h_1 x_1 + h_2 + b_1 + b_2 + c_2}{\mu_1 + \mu_2 + v_2} \\
 &+ [V(x_1 - 1, 2) - V(x_1, 0)] \frac{\mu_1}{\mu_1 + \mu_2 + v_2}, \quad (\text{A.6c})
 \end{aligned}$$

and for $x_2 > 1$,

$$\begin{aligned}
 V^{\pi_0}(x_1, x_2) - V^{\pi_0}(x_1, x_2 - 1) &= \frac{h_2}{\mu_1 + \mu_2} + [V(x_1 - 1, x_2 + 1) - V(x_1 - 1, x_2)] \frac{\mu_1}{\mu_1 + \mu_2} \\
 &+ [V(x_1, x_2 - 1) - V(x_1, x_2 - 2)] \frac{\mu_2}{\mu_1 + \mu_2}, \quad (\text{A.7a})
 \end{aligned}$$

$$\begin{aligned}
 V^{\pi_1}(x_1, x_2) - V^{\pi_1}(x_1, x_2 - 1) &= \frac{h_2}{\mu_1 + \mu_2 + v_1} + [V(x_1 - 1, x_2 + 1) \\
 &- V(x_1 - 1, x_2)] \frac{\mu_1 + v_1}{\mu_1 + \mu_2 + v_1} + [V(x_1, x_2 - 1) \\
 &- V(x_1, x_2 - 2)] \frac{\mu_2}{\mu_1 + \mu_2 + v_1}, \quad (\text{A.7b})
 \end{aligned}$$

$$\begin{aligned}
 V^{\pi_2}(x_1, x_2) - V^{\pi_2}(x_1, x_2 - 1) &= \frac{h_2}{\mu_1 + \mu_2 + v_2} + [V(x_1 - 1, x_2 + 1) \\
 &- V(x_1 - 1, x_2)] \frac{\mu_1}{\mu_1 + \mu_2 + v_2} + [V(x_1, x_2 - 1) \\
 &- V(x_1, x_2 - 2)] \frac{\mu_2 + v_2}{\mu_1 + \mu_2 + v_2}. \quad (\text{A.7c})
 \end{aligned}$$

The proof of (A.4) and (A.5) is by means of induction on x_1 and x_2 . First, we prove (A.4) for $x_1 = 1$. From (13) and (14) and condition (A2), we get

$$\begin{aligned}
 V(0, 2) - V(1, 0) &= \frac{2h_2 + b_2 + c_2}{\mu_2 + v_2} - \min \left\{ \frac{h_1 + b_1}{\mu_1}, \frac{h_1 + b_1 + c_1}{\mu_1 + v_1} \right\} \Rightarrow \\
 V(0, 2) - V(1, 0) &\geq \frac{2h_2 + b_2 + c_2}{\mu_2 + v_2} - \frac{h_1 + b_1}{\mu_1}, \quad (\text{A.8a})
 \end{aligned}$$

$$V(0, 2) - V(1, 0) \geq \frac{2h_2 + b_2 + c_2}{\mu_2 + v_2} - \frac{h_1 + b_1 + c_1}{\mu_1 + v_1}. \quad (\text{A.8b})$$

Substituting (A.8a) into (A.6a) and (A.6c), (A.8b) into (A.6b), and taking into account condition (A2) yields the result. To prove (A.5) for $x_1 = 1$, we use induction on x_2 . From (14) and condition (A2), we get

$$V(0, x_2 + 1) - V(0, x_2) = \frac{h_2(x_2 + 1) + b_2 + c_2}{\mu_2 + v_2}. \quad (\text{A.9})$$

Applying the induction hypothesis (with the induction base, $x_2 = 2$, having been established by the fact that the statement of the lemma has been proven to hold for $x_1 = x_2 = 1$) we also get

$$V(1, x_2 - 1) - V(1, x_2 - 2) \geq \frac{h_2(x_2 - 1) + b_2 + c_2}{\mu_2 + v_2}. \quad (\text{A.10})$$

Then the result follows from (A.7a)–(A.7c), (A.9), and (A.10). Next we prove (A.4) and (A.5) for $x_1 > 1$ by induction on x_1 . Starting with (A.4), we have from (13),

$$\begin{aligned} V(x_1 - 1, 2) - V(x_1, 0) &= V(x_1 - 1, 2) \\ &\quad - \min\left\{\frac{h_1x_1 + b_1}{\mu_1}, \frac{h_1x_1 + b_1 + c_1}{\mu_1 + \nu_1}\right\} - V(x_1 - 1, 1) \\ &\geq \frac{2h_2 + b_2 + c_2}{\mu_2 + \nu_2} - \min\left\{\frac{h_1x_1 + b_1}{\mu_1}, \frac{h_1x_1 + b_1 + c_1}{\mu_1 + \nu_1}\right\} \\ &\Rightarrow V(x_1 - 1, 2) - V(x_1, 0) \geq \frac{2h_2 + b_2 + c_2}{\mu_2 + \nu_2} - \frac{h_1x_1 + b_1}{\mu_1}, \\ V(x_1 - 1, 2) - V(x_1, 0) &\geq \frac{2h_2 + b_2 + c_2}{\mu_2 + \nu_2} - \frac{h_1x_1 + b_1 + c_1}{\mu_1 + \nu_1}, \end{aligned}$$

where the first inequality is a result of the induction hypothesis. The rest of the proof is identical to the one for $x_1 = 1$ with h_1 replaced by h_1x_1 . Finally, we prove (A.5) by induction on x_2 . Applying the induction hypothesis, we get

$$\begin{aligned} V(x_1 - 1, x_2 + 1) - V(x_1 - 1, x_2) &\geq \frac{h_2(x_2 + 1) + b_2 + c_2}{\mu_2 + \nu_2}, \\ V(x_1, x_2 - 1) - V(x_1, x_2 - 2) &\geq \frac{h_2(x_2 - 1) + b_2 + c_2}{\mu_2 + \nu_2}, \end{aligned}$$

so the proof is essentially identical to the proof for $x_1 = 1$. ■

PROOF OF LEMMA 3: The proof is by induction on x_1 . First, we prove the result for $x_1 = 1$ by induction on x_2 . We establish the induction base by showing that $d(1, 1) < d(1, 0)$.

CASE 1: $X_1 > 1 \Rightarrow d(1, 0) < 0$.

From (21) and (18), we get

$$\begin{aligned} f(1, 1) &= -\nu_1h_2 - (\nu_2h_2 + \nu_2b_2 - \mu_2c_2) + (\mu_1 + \mu_2 + \nu_2)d(1, 0) \\ &\quad + \mu_1d(0, 2) < 0, \end{aligned} \tag{A.11}$$

because of condition (A2) and $d(0, 2) < 0$. Because $f(1, 1) < 0$, we get from (20) and Lemma 2 that $d(1, 1) < 0$, so by (20) and (A.11),

$$d(1, 1) = \frac{f(1, 1)}{1 - \nu_1} = \frac{-\nu_1h_2 - (\nu_2h_2 + \nu_2b_2 - \mu_2c_2) + \mu_1d(0, 2)}{1 - \nu_1} + d(1, 0) < d(1, 0),$$

where the inequality follows from condition (A2) and $d(0, 2) < 0$.

CASE 2: $X_1 \leq 1 \Rightarrow d(1, 0) \geq 0$.

From (21) and (18), we get

$$f(1, 1) = -\nu_1h_2 - (\nu_2h_2 + \nu_2b_2 - \mu_2c_2) + (\mu_1 + \nu_1 + \mu_2)d(1, 0) + \mu_1d(0, 2) \tag{A.12}$$

The result is trivially true when $f(1, 1) < 0$ because in this case (20) and Lemma 2 imply that $d(1, 1) < 0$. In the case $f(1, 1) \geq 0$, we get from (20) and Lemma 2 that $d(1, 1) \geq 0$, which by

(20) and (A.12) yields

$$d(1, 1) = \frac{f(1, 1)}{1 - v_2} = \frac{-v_1 h_2 - (v_2 h_2 + v_2 b_2 - \mu_2 c_2) + \mu_1 d(0, 2)}{1 - v_2} + d(1, 0) < d(1, 0)$$

because of condition (A2) and $d(0, 2) < 0$.

The induction on x_2 is completed by proving that the result holds for $x_2 = l + 1$ provided that it holds for $x_2 = l$. We have

$$d(1, l + 1) - d(1, l + 2) = f(1, l + 1) - f(1, l + 2) + v_1[d(1, l + 1)^- - d(1, l + 2)^-] + v_2[d(1, l + 1)^+ - d(1, l + 2)^+]. \quad (\text{A.13})$$

Using the induction hypothesis, the fact that $d(0, x_2)$ is decreasing in x_2 [(19)] and condition (A3), we get $f(1, l + 1) - f(1, l + 2) > 0$. Therefore, the result for $x_2 = l + 1$ follows from (A.13) and Lemma 2. Next, we show that the result is true for $x_1 = k + 1$ under the assumption that it is true for $x_1 = k$ and any x_2 . Again, we use induction on x_2 . To prove that $d(k + 1, 1) < d(k + 1, 0)$, we consider two cases.

CASE 1: $k + 1 < X_1 \Rightarrow d(k + 1, 0) < 0, d(k, 0) < 0$.

From (18) and the induction hypothesis, we have

$$v_1 h_1 = [d(k + 1, 0) - d(k, 0)]\mu_1, \quad (\text{A.14})$$

$$d(k, 2) < d(k, 0) < 0, \quad (\text{A.15})$$

respectively. Plugging (A.14) and (A.15) into (21), we get

$$f(k + 1, 1) = -v_1 h_2 - (v_2 h_2 + v_2 b_2 - \mu_2 c_2) + (\mu_1 + \mu_2 + v_2)d(k + 1, 0) + \mu_1[d(k, 2) - d(k, 0)] < 0, \quad (\text{A.16})$$

because of condition (A2) and (A.15). Because $f(k + 1, 1) < 0$, we get from (20) and Lemma 2 that $d(k + 1, 1) < 0$, so from (20) and (A.16),

$$\begin{aligned} d(k + 1, 1) &= \frac{f(k + 1, 1)}{1 - v_1} \\ &= \frac{-v_1 h_2 - (v_2 h_2 + v_2 b_2 - \mu_2 c_2) + \mu_1[d(k, 2) - d(k, 0)]}{1 - v_1} \\ &\quad + d(k + 1, 0) < d(k + 1, 0), \end{aligned}$$

where the inequality follows from condition (A2) and (A.15).

CASE 2: $k + 1 \geq X_1 \Rightarrow d(k + 1, 0) \geq 0$.

To prove that $d(k + 1, 1) < d(k + 1, 0)$, we restrict attention to the case $f(k + 1, 1) \geq 0$ as the case $f(k + 1, 1) < 0$ is trivial ($f(k + 1, 1) < 0 \Rightarrow d(k + 1, 1) < 0 \leq d(k + 1, 0)$). For $f(k + 1, 1) \geq 0$, we have $d(k + 1, 1) \geq 0$, and by (20),

$$d(k + 1, 1) = \frac{f(k + 1, 1)}{1 - v_2}. \quad (\text{A.17})$$

In the case $d(k, 0) \geq 0$, (18) yields

$$v_1 h_1 = [d(k + 1, 0) - d(k, 0)](\mu_1 + v_1). \quad (\text{A.18a})$$

When $d(k, 0) < 0$, we have

$$v_1 h_1 = d(k + 1, 0)(\mu_1 + v_1) - d(k, 0)\mu_1. \tag{A.18b}$$

Then, we get from (A.17), (21), and (A.18a) and (A.18b)

$$d(k + 1, 1) = \frac{\tilde{f}(k + 1, 1)}{1 - v_2} + d(k + 1, 0),$$

where

$$\begin{aligned} \tilde{f}(k + 1, 1) = & -v_1 h_2 - (v_2 h_2 + v_2 b_2 - \mu_2 c_2) + \mu_1 [d(k, 2) - d(k, 0)] \\ & + v_1 [d(k, 2)^+ - d(k, 0)] \mathbf{1}(d(k, 0) \geq 0). \end{aligned}$$

The result follows from $\tilde{f}(k + 1, 1) < 0$, which is true because of condition (A2) and the induction hypothesis for $x_1(d(k, 2) < d(k, 0))$. Finally, to prove the result for $x_2 = l + 1$ assuming it is true for $x_2 = l$, we obtain an equation analogous to (A.13) and use the induction hypotheses for $x_1 = k, x_2 = l$, and condition (A3). ■

PROOF OF LEMMA 4: The proof is by induction on x_1 . We begin by noting that $\lim_{x_2 \rightarrow \infty} d(0, x_2) = -\infty$ (see (19)). To complete the induction, we need to show that the result is true for $x_1 = x_1^*$ under the assumption that it is true for $x_1 < x_1^*$.

CASE 1. $x_1^* < X_1 \Rightarrow d(x_1^*, 0) < 0, d(x_1^* - 1, 0) < 0$.

Because $d(x_1, x_2)$ is a decreasing sequence in x_2 (Lemma 3), we have $d(x_1^*, x_2) < 0$ and $d(x_1^* - 1, x_2) < 0$ for any x_2 , so (20) and (21) yield

$$\begin{aligned} d(x_1^*, x_2) = & v_1 h_1 - (v_1 + v_2)h_2 + \mu_1 d(x_1^* - 1, x_2 + 1) + \mu_2 d(x_1^*, x_2 - 1) \\ & + v_2 d(x_1^*, x_2 - 1) + v_1 d(x_1^*, x_2) + (v_1 b_1 - \mu_1 c_1) \mathbf{1}(x_1 = 1) \\ & + (\mu_2 c_2 - v_2 b_2) \mathbf{1}(x_2 = 1). \end{aligned}$$

Taking limits on both sides and rearranging terms, we get

$$\mu_1 \lim_{x_2 \rightarrow \infty} d(x_1^*, x_2) = C + \mu_1 \lim_{x_2 \rightarrow \infty} d(x_1^* - 1, x_2),$$

where C is a constant. The result is a consequence of the induction hypothesis.

CASE 2: $x_1^* \geq X_1 \Rightarrow d(x_1^*, 0) \geq 0$.

We claim that $d(x_1^*, x_2) < 0$ for x_2 sufficiently large; if this were not the case [i.e., if $d(x_1^*, x_2)$ were bounded below by a nonnegative number], we would get from (20) and (21) that for $x_2 > 1$,

$$\begin{aligned} d(x_1^*, x_2) = & v_1 h_1 - (v_1 + v_2)h_2 + \mu_1 d(x_1^* - 1, x_2 + 1) + \mu_2 d(x_1^*, x_2 - 1) \\ & + v_1 d(x_1^* - 1, x_2 + 1)^+ + v_2 d(x_1^*, x_2) + (v_1 b_1 - \mu_1 c_1) \mathbf{1}(x_1^* = 1). \end{aligned}$$

Taking limits and using the induction hypothesis, we obtain

$$(\mu_1 + v_1) \lim_{x_2 \rightarrow \infty} d(x_1^*, x_2) = C + \mu_1 \lim_{x_2 \rightarrow \infty} d(x_1^* - 1, x_2),$$

which, by the induction hypothesis, yields

$$\lim_{x_2 \rightarrow \infty} d(x_1^*, x_2) = -\infty,$$

clearly a contradiction. Therefore, for x_2 sufficiently large,

$$\begin{aligned} d(x_1^*, x_2) &= v_1 h_1 - (v_1 + v_2) h_2 + \mu_1 d(x_1^* - 1, x_2 + 1) + \mu_2 d(x_1^*, x_2 - 1) \\ &\quad + v_2 d(x_1^*, x_2 - 1) + v_1 d(x_1^*, x_2) + (v_1 b_1 - \mu_1 c_1) \mathbf{1}(x_1^* = 1) \\ &\Rightarrow \mu_1 \lim_{x_2 \rightarrow \infty} d(x_1^*, x_2) = C + \mu_1 \lim_{x_2 \rightarrow \infty} d(x_1^* - 1, x_2), \end{aligned}$$

and $\lim_{x_2 \rightarrow \infty} d(x_1^*, x_2) = -\infty$ follows from the induction hypothesis. ■

PROOF OF LEMMA 5: For each integer $K \geq 2$, we introduce the function

$$d_K(x_1) = d(x_1, K - x_1), \quad x_1 = 0, 1, \dots, K.$$

Therefore, we need to show that for all $K \geq 2$,

$$d_K(x_1) > 0 \Rightarrow d_K(x_1) \leq d_K(x_1 + 1).$$

Let

$$\underline{K} = \min \{K : d_K(x_1) > 0 \text{ for some } 1 \leq x_1 \leq K - 1\}$$

and

$$\overline{K} = \min \{K \geq \underline{K} : d_{K+1}(x_1) \leq 0 \text{ for all } x_1 \leq K\}.$$

It suffices to prove the result for $\underline{K} \leq K \leq \overline{K}$, as the same arguments can be repeated for $K > \overline{K}$. The proof is by induction on K (the inductive scheme is the one used by Wu et al. [26]). First, note that when $d(x_1 - 1, 1) > 0$, it follows from Lemma 3 that $d(x_1 - 1, 0) > d(x_1 - 1, 1) > 0$, implying $x_1 - 1 > X_1$, which by (18) yields $d(x_1 - 1, 0) < d(x_1, 0)$. Therefore,

$$d_{x_1}(x_1 - 1) > 0 \Rightarrow d_{x_1}(x_1 - 1) < d_{x_1}(x_1). \tag{A.19}$$

To start the induction, we prove the result for $K = \underline{K}$. If $\underline{K} = 2$, in which case $d_{\underline{K}}(\underline{K} - 1) > 0$, the result is a special case of (A.19). Consider now $\underline{K} > 2$. Because $d_{\underline{K}-1}(x_1) \leq 0$ for $x_1 \leq \underline{K} - 2$, we have by Lemma 3 that $d_{\underline{K}}(x_1) \leq 0$ for $x_1 \leq \underline{K} - 2$. Therefore, $d_{\underline{K}}(\underline{K} - 1) > 0$ and it suffices to prove $d_{\underline{K}}(\underline{K} - 1) \leq d_{\underline{K}}(\underline{K})$, which follows from (A.19).

To complete the induction, we assume that the result holds for $K = k - 1$, where $\underline{K} < k \leq \overline{K}$, and prove it for $K = k$. Let $x_1^* = \min\{x_1 : d_k(x_1) > 0\}$, so we need to show that $d_k(x_1^*) \leq d_k(x_1^* + 1) \leq \dots \leq d_k(k - 1) \leq d_k(k)$. The last inequality has already been proven [(A.19)], so we only need to consider $1 \leq x_1^* \leq k - 2$. We start by proving the first inequality $d_k(x_1^*) \leq d_k(x_1^* + 1)$. From (20) and (21), we have

$$\begin{aligned} d_k(x_1^*) - d_k(x_1^* + 1) &= \mu_1 [d_k(x_1^* - 1) - d_k(x_1^*)] + v_1 [d_k(x_1^* - 1)^+ - d_k(x_1^*)^+] \\ &\quad + \mu_2 [d_{k-1}(x_1^*) - d_{k-1}(x_1^* + 1)] + v_2 [d_{k-1}(x_1^*)^- - d_{k-1}(x_1^* + 1)^-] \\ &\quad + (v_1 b_1 - \mu_1 c_1) \mathbf{1}(x_1^* = 1) - (\mu_2 c_2 - v_2 b_2) \mathbf{1}(x_1^* = k - 2) \\ &\quad + v_1 [d_k(x_1^*)^- - d_k(x_1^* + 1)^-] + v_2 [d_k(x_1^*)^+ - d_k(x_1^* + 1)^+]. \end{aligned} \tag{A.20}$$

By the definition of x_1^* and Lemma 3, we get

$$d_k(x_1^* - 1) \leq 0 < d_k(x_1^*) < d_{k-1}(x_1^*). \tag{A.21}$$

Because $d_{k-1}(x_1^*) > 0$, we can apply the induction hypothesis to get $d_{k-1}(x_1^*) \leq d_{k-1}(x_1^* + 1)$, so from (A.20) and (A.21), condition (A3), and Lemma 2, we get $d_k(x_1^*) \leq d_k(x_1^* + 1)$. Having

proven that $d_k(x_1^* + 1) > 0$, we can show that $d_k(x_1^* + 1) \leq d_k(x_1^* + 2) \leq \dots \leq d_k(k - 1)$ by successive application of the same arguments. ■

APPENDIX B Proofs of Lemmas 6–8

Before we proceed to the proofs, we derive expressions for $d_{n,\beta}(x_1, x_2)$. Because $V_{1,\beta}(x_1, x_2) = h_1x_1 + h_2x_2$, we get, for $n = 1$,

$$d_{1,\beta}(x_1, x_2) = (c_2 - c_1) + \beta[v_1(h_1 - h_2) - v_2h_2], \tag{B.1}$$

$$d_{1,\beta}(x_1, 0) = -c_1 + \beta v_1(h_1 - h_2), \tag{B.2}$$

$$d_{1,\beta}(0, x_2) = c_2 - \beta v_2h_2. \tag{B.3}$$

For $n \geq 2$, we obtain the following recursive expressions from (22) and (25)–(28). For $x_1, x_2 \geq 1$,

$$\begin{aligned} d_{n,\beta}(x_1, x_2) &= (1 - \beta)(c_1 - c_2) + \beta[v_1(h_1 - h_2) - v_2h_2 - \mu_1c_1\mathbf{1}(x_1 = 1) + \mu_2c_2\mathbf{1}(x_2 = 1)] \\ &\quad + \beta[\lambda d_{n-1,\beta}(x_1 + 1, x_2) + \mu_1d_{n-1,\beta}(x_1 - 1, x_2 + 1) + \mu_2d_{n-1,\beta}(x_1, x_2 - 1) \\ &\quad + v_1d_{n-1,\beta}(x_1 - 1, x_2 + 1)^+ + v_2d_{n-1,\beta}(x_1, x_2 - 1)^- \\ &\quad + v_1d_{n-1,\beta}(x_1, x_2)^- + v_2d_{n-1,\beta}(x_1, x_2)^+], \end{aligned} \tag{B.4}$$

$$\begin{aligned} d_{n,\beta}(x_1, 0) &= -(1 - \beta)c_1 + \beta[v_1(h_1 - h_2) - v_1c_2 - \mu_1c_1\mathbf{1}(x_1 = 1) - \mu_1c_2\mathbf{1}(x_1 > 1)] \\ &\quad + \beta[(\mu_2 + v_2)\beta v_1 + \mu_1\beta v_2\mathbf{1}(x_1 > 1)][V_{n-1,\beta}(x_1 - 1, 1) - V_{n-1,\beta}(x_1 - 1, 0)] \\ &\quad + \lambda d_{n-1,\beta}(x_1 + 1, 0) + \mu_1d_{n-1,\beta}(x_1 - 1, 1)\mathbf{1}(x_1 > 1) + (\mu_2 + v_2)d_{n-1,\beta}(x_1, 0) \\ &\quad + v_1d_{n-1,\beta}(x_1 - 1, 1)^+ + v_1d_{n-1,\beta}(x_1, 0)^-], \end{aligned} \tag{B.5}$$

$$\begin{aligned} d_{n,\beta}(0, x_2) &= (1 - \beta)c_2 + \beta[-v_2h_2 + \lambda c_2 + \mu_2c_2\mathbf{1}(x_2 = 1)] \\ &\quad + \beta[\beta\lambda v_2[V_{n-1,\beta}(1, x_2 - 1) - V_{n-1,\beta}(1, x_2)] + (\mu_1 + v_1)d_{n-1,\beta}(0, x_2) \\ &\quad + [\mu_2d_{n-1,\beta}(0, x_2 - 1) + v_2d_{n-1,\beta}(0, x_2 - 1)^-]\mathbf{1}(x_2 > 1) + v_2d_{n-1,\beta}(0, x_2)^+]. \end{aligned}$$

The last equation is not needed for the proofs that follow, as we have already shown in Proposition 5 that $d_{n,\beta}(0, x_2) \leq 0$ for $\beta \geq c_2/v_2h_2$. However, we decided to include it in the article in order to have a complete set of expressions for $d_{n,\beta}(x_1, x_2)$.

PROOF OF LEMMA 6: The proof is by induction on n . For any $x_2 \geq 1$, we get from (B.1) and (B.2) that $d_{1,\beta}(x_1, x_2)$ is constant for $x_2 \geq 1$ and

$$d_{1,\beta}(x_1, 0) - d_{1,\beta}(x_1, x_2) = \beta v_2h_2 - c_2,$$

which is nonnegative because $\beta \geq \beta_1 \geq c_2/v_2h_2$. Having established the induction base, we assume that the result holds for $n = k$ and prove it for $n = k + 1$. Equations (B.4) and (B.5)

and the fact that $d_{k,\beta}(0, x_2) \leq 0$ yield

$$\begin{aligned}
 d_{k+1,\beta}(x_1, 0) - d_{k+1,\beta}(x_1, 1) = & -(1 - \beta)(2c_1 - c_2) + \beta[v_2h_2 - (v_1 + \mu_2)c_2] \\
 & + \beta\{\lambda[d_{k,\beta}(x_1 + 1, 0) - d_{k,\beta}(x_1 + 1, 1)] \\
 & + (\mu_2 + v_2)\beta v_1[V_{k,\beta}(x_1 - 1, 1) - V_{k,\beta}(x_1 - 1, 0)] \\
 & + v_1[d_{k,\beta}(x_1 - 1, 1)^+ - d_{k,\beta}(x_1 - 1, 2)^+] \mathbf{1}(x_1 > 1) \\
 & + v_1[d_{k,\beta}(x_1, 0)^- - d_{k,\beta}(x_1, 1)^-] \\
 & + v_2[d_{k,\beta}(x_1, 0) - d_{k,\beta}(x_1, 0)^- - d_{k,\beta}(x_1, 1)^+]\}.
 \end{aligned}$$

The combination of the first two terms of the right-hand side is nonnegative; it is positive when $2c_1 \leq c_2$ and nonnegative when $2c_1 > c_2$ because

$$\beta \geq \beta_1 \geq \frac{2c_1 - c_2}{v_2h_2 - (v_1 + \mu_2)c_2 + 2c_1 - c_2}.$$

Noting that $d_{k,\beta}(x_1, 0) - d_{k,\beta}(x_1, 0)^- = d_{k,\beta}(x_1, 0)^+$, we get from the induction hypothesis and (30) that the remaining term is also nonnegative when $\beta \geq \beta_1$. For $x_2 \geq 1$, $d_{k+1,\beta}(x_1, x_2) - d_{k+1,\beta}(x_1, x_2 + 1) \geq 0$ follows in a straightforward manner from (B.4), $d_{k,\beta}(0, x_2) \leq 0$, and the induction hypothesis. ■

PROOF OF LEMMA 7: By induction on n . From (B.2), we get $d_{1,\beta}(x_1, 0) \geq 0$ because

$$\beta \geq \beta_2 \geq \frac{c_2}{v_2h_2} > \frac{c_1}{v_1(h_1 - h_2)}$$

and from (B.1), we get $d_{1,\beta}(x_1, x_2) \geq 0$ for $x_2 \geq 1$. Assuming that $d_{k,\beta}(x_1, x_2) \geq 0$, we use (B.5) and (B.4) to show that $d_{k+1,\beta}(x_1, x_2) \geq 0$ for $x_2 = 0$ and $x_2 \geq 1$, respectively. The sum of the first two terms of the right-hand side of (B.5) is nonnegative because

$$\beta \geq \beta_2 \geq \frac{c_2}{v_2h_2} > \frac{c_1}{v_1(h_1 - h_2) - v_1c_2 + c_1}.$$

The last term is nonnegative by (30) and the induction hypothesis. Turning to (B.4), the sum of its first two terms is nonnegative because

$$\beta \geq \beta_2 \geq \frac{c_2 - c_1}{v_1(h_1 - h_2) - v_2h_2 + c_2 - c_1},$$

and the remaining term is nonnegative by the induction hypothesis. ■

PROOF OF LEMMA 8: The result has already been proven for $x_1 = 0$ (proof of Proposition 5). For $x_1 > 0$, we use induction on n . Equation (B.1) yields $d_{1,\beta}(x_1, x_2) < 0$. By applying the induction hypothesis to (B.4), we get $d_{n,\beta}(x_1, x_2) \leq 0$ because ■

$$\beta \geq \beta_3 \geq \frac{c_1 - c_2}{v_2h_2 - v_1(h_1 - h_2) + c_1 - c_2}.$$