Online Appendix

Appendix A: Notation

Conventions. Scalars are set in italic; 2D vectors (per type) use arrows. Superscript k indexes item types; subscript $i \in \{1,2\}$ indexes stations. The vector $e = (1,1)$.

Table 2: Main symbols and notation used in the paper.

Symbol	Type	Meaning / Where Used
Core primitives		
a	scalar	Initial quality score on first visit (single item).
λ	scalar	Exogenous base arrival rate to the network.
l_i	scalar	Threshold (release/transfer target) at Station i.
$l = (l_1, l_2)$	2D vector	Threshold pair for a single item (two-station model).
w_i	scalar	Rate-normalized workload at Station i.
$w = (w_1, w_2)$	2D vector	Workload pair for a single item (two-station model).
θ_i	scalar	Drift (quality improvement rate) at Station i.
σ_i	scalar	Diffusion coefficient at Station i.
σ_{pr}	scalar	Post-release diffusion coefficient.
$\mathcal{B}_t^i$	process	In-station quality BM at Station i (positive drift).
$\mathcal{B}_t^{pr}$	process	Post-release quality BM (drift $\eta(l)$, start $e \cdot l$).
Single-station objects		
l^*	scalar	Optimal threshold in the one-station setting.
$W_i^r(l)$	scalar	Rate-normalized workload at Station i for threshold(s) l.
τ	r.v.	Hitting time to threshold (inverse Gaussian under BM).
Post-release and re-work		
$e \cdot l$	scalar	Total discharge score $l_1 + l_2$ (starting level post-release).
$p(l)$	function $\to$ scalar	Re-work probability $p(l) = \exp\big(-(\nu + \varrho\,\eta(l)\,(e \cdot l))\big)$.
$N(l)$	function $\to$ scalar	Expected number of visits $N(l) = [1 - p(l)]^{-1}$.

Continued on next page

Symbol	Type	Meaning / Where Used
ν	scalar	Baseline log-scale parameter in the re-work probability; $e^{-\nu}$ is the base probability of no re-work.
ϱ	scalar	$\varrho := 2/\sigma_{pr}^2$
Two-station aggregation (CES) and geometry		
$\alpha \in (0,1)$	scalar	Station-importance weight (CES).
$\beta \leq 1$	scalar	CES substitution/complementarity parameter (β=1 subst.; $\beta \downarrow -\infty$ complementarity).
$\eta(l)$	function $\to$ scalar	CES aggregator: $(\alpha l_1^\beta + (1-\alpha) l_2^\beta)^{1/\beta}$.
$\mathcal{W}$	set	Attainable workload region in $\mathbb{R}_+^2$.
$f_{\alpha,\beta}(\cdot)$	function	Boundary of $\mathcal{W}$: $w_2 \geq f_{\alpha,\beta}(w_1)$.
$\mathbb{h}_{\alpha,\beta}(\cdot)$	function	Auxiliary function in boundary characterization.
w_1^0	scalar	Right-endpoint of nontrivial boundary segment of $\mathcal{W}$.
Optimization, costs, and capacity		
c_i	scalar	Holding/effort cost rate at Station i.
$V(l)$ or $V(\vec{l})$	scalar	Total expected cost objective.
$\mathcal{R}^c$	scalar	Relative cost ratio (effective Station 2 vs. Station 1).
$\widetilde{\mathcal{R}}^c$	scalar	Corrected cost ratio (accounts for duals κ).
C_i	scalar	Capacity at Station i.
$\mathcal{L}_i(\vec{l})$	scalar	Total load on Station i: $(\lambda/\theta_i)\, W_i^r(\vec{l})$.
κ_i	scalar	Dual variable for capacity at Station i.
Multi-type extension		
$k \in [K]$	index	Type index, $[K] = \{1, \ldots, K\}$.
$l^k = (l_1^k, l_2^k)$	2D vector	Thresholds for type k.
$\vec{l} = (l^1, \ldots, l^K)$	$2K$-vector	Stacked decision vector over all types.
$\theta_i^k,\ \sigma_i^k$	scalars	Drift and diffusion for type k at Station i.
$\alpha^k,\ \beta^k$	scalars	CES weight and parameter for type k.

Continued on next page

Symbol	Type	Meaning / Where Used
$\vec{w}^k = (w_1^k, w_2^k)$	2D vector	Workload pair for type k.
$\mathbf{w} = (\vec{w}^1, \ldots, \vec{w}^K)$	$2K$-vector	Stacked workload vector across types.
$\mathbf{W}_i^r(\vec{l})$	K-vector	$(W_i^{r,1}(l^1), \ldots, W_i^{r,K}(l^K))$, type workloads at Station i.
$\mathcal{W}_k$	set	Attainable workload set for type k.
$\mathcal{W}^\times$	set	Product set $\prod_{k\in[K]} \mathcal{W}_k$.
$\boldsymbol{\beta} = (\beta^1, \ldots, \beta^K)$	K-vector	Type-wise substitution/complementarity parameters.
$\mathcal{C}_{\boldsymbol{\beta}}$	set	Feasible capacity vectors under $\boldsymbol{\beta}$.

Appendix B: Variants and Extensions

To simplify notation, throughout the extensions we assume that $\nu = 0$.

B.1. Attainable Workload Region for n Stations

Consider the case of n stations and let $(l_1, \ldots, l_n)$ denote the transfer target from each of the n stations. The drift and diffusion coefficient of each Station i, $i = 1, \ldots, n$ are θ_i and σ_i. The constant elasticity of substitution (CES) production function for n stations is:

$$\eta(l) = \left[\sum_{i=1}^{n} \alpha_i l_i^{\beta}\right]^{\frac{1}{\beta}}, \quad 0 < \alpha_i < 1, \quad \beta \leq 1, \tag{17}$$

where $\sum_{i=1}^{n} \alpha_i = 1$; we use henceforth α for the probability vector $\alpha_1, \ldots, \alpha_n$. The rate-normalized workload in Station i is $W_i^r(l) = l_i N(l)$. Consider here the case where the initial state in the first visit is 0 as in subsequent visits.

The final part of Theorem 1 shows that the attainable workload region is given by

$$\mathcal{W} := \{w \in \mathbb{R}_+^n : \eta(w)(e \cdot w) \geq \Gamma^2/\varrho\}.$$

This, as for $n = 2$, is a convex region. This is formally argued within the proof of Theorem 1.

B.2. Processing Complementarity

The analysis so far focused on the case where the two process steps can have *outcome* complementarity, but there is no processing complementarity. That is, the improvement rate in one station does not depend on the quality score reached at the other station.

In this section we consider the other extreme: processing complementarity/substitution. First, we assume that process steps are perfect outcome substitutes and have the same contribution to the return probability ($\alpha = 0.5$ and $\beta = 1$). The number of visits is then

$$N_{PC}(l) = \left[1 - e^{-\frac{1}{2}(e\cdot l)^2}\right]^{-1}.$$

We introduce process dependence by having a non-constant drift in Station 2, a drift $\theta_2(l_1)$ that depends on the final score at completion in Station 1. We use $\theta_2(l_1) = \gamma + y l_1$. This is clearly *not* as general or flexible as what we allowed in outcome complementarity; however, it provides initial tractability. We then explore numerically variations on this structure.

Notice that if $y=0$ and $\gamma>0$ (so that, $\theta_2\equiv\gamma>0$), we restore perfect processing substitution as in our base model. If $\gamma=0$ and $y>0$, then $\theta_2(l_1)$ is strictly increasing, and the two steps are strong complements in the sense that one cannot commence processing in Station 2 unless some processing is done in Station 1 ($l_1>0$).

It will be useful, for clarity, to fix $\theta_1=1$. In that case, l_1 is not only the target but also the mean processing time in Station 1.

The objective function takes the form

$$V(l)=\left(c_1l_1+\frac{c_2}{\theta_2(l_1)}l_2\right)N_{PC}(l)\ .$$

Instead of considering l as the decision, let us define

$$\tau_1=l_1,\ \ \tau_2=l_2/\theta_2(\tau_1).$$

Then, with $\tau=(\tau_1,\tau_2)$, and (re)defining

$$N_{PC}(\tau)=\left[1-e^{-\frac{1}{2}(\tau_1+\theta_2(\tau_1)\tau_2)^2}\right]^{-1},$$

we write

$$V(\tau)=(c_1\tau_1+c_2\tau_2)N(\tau).$$

This makes clear that the optimization problem with processing dependence is the same as one with outcome dependence but where the decisions are the processing times rather than the target, and where the processing times impact returns through the sum $\tau_1+\theta_2(\tau_1)\tau_2$. We define rate-normalized workloads as before

$$W_1(\tau)=\tau_1N_{PC}(\tau),\ \text{and}\ W_2(\tau)=\theta_2(\tau_1)\tau_2N_{PC}(\tau).$$

The following is a characterization of the attainable workload region.

LEMMA 4. *Suppose that $\theta_2(\tau_1)=\gamma+y\tau_1$ with either $\gamma=0$, $y>0$ or $\gamma>0$, $y=0$. In either case, the region*

$$\mathcal{W}=\{(w_1,w_2)\geq 0:\exists\tau\ \text{ s.t. } W_1(\tau)=\tau_1N_{PC}(\tau),\ \ W_2(\tau)=\theta_2(\tau_1)\tau_2N_{PC}(\tau)\},$$

is equivalently given by

$$\mathcal{W}=\{(w_1,w_2)\geq 0:\ \ w_2\geq f_{\alpha,\beta}(w_1)\}$$

where

$$f_{\alpha,\beta}(w_1)=\begin{cases}a-bw_1, & \text{if } w_1\leq b/a,\\ 0, & \text{otherwise},\end{cases}$$

where a,b are constants given explicitly in terms of γ,y and Γ from (6).

Note that $f_{\alpha,\beta}$ here denotes the boundary function in this variant and differs from the function defined in Theorem 1.

Thus, in either extreme case for our model, the attainable workload region is a polyhedron (as in the case of outcome complementarity with $\beta=1$ and $\alpha=1/2$). The difference is in the objective function. Given $w=(w_1,w_2)\in\mathcal{W}$, let $\tau=\tau(w)$ be such that $w_1=\tau_1N_{PC}(\tau)$ and $w_2=\tau_2N_{PC}(\tau)$. Then, the objective function is given by

$$c_1w_1+c_2\frac{w_2}{\theta_2(\tau_1(w))}.$$

In the case that $\gamma > 0$, $y = 0$, we have $\theta_2(\cdot) \equiv \gamma$ and thus, a linear objective function. Therefore, there exists a threshold of cost ratio where the optimal solution is of the form $(0, w_2)$ when the cost ratio is below the threshold and of the form $(w_1, 0)$ beyond the threshold.

If, instead $\theta_2(\tau_1(w)) = y\tau_1(w)$, the objective function is non-linear. Moreover, because $w_1 = \tau_1 N(\tau) \geq \tau_1$, for any $w_2 > 0$

$$c_1 w_1 + c_2 \frac{w_2}{\theta_2(\tau_1(w))} \geq c_1 w_1 + c_2 \frac{w_2}{y w_1} \uparrow \infty, \text{ as } w_1 \downarrow 0.$$

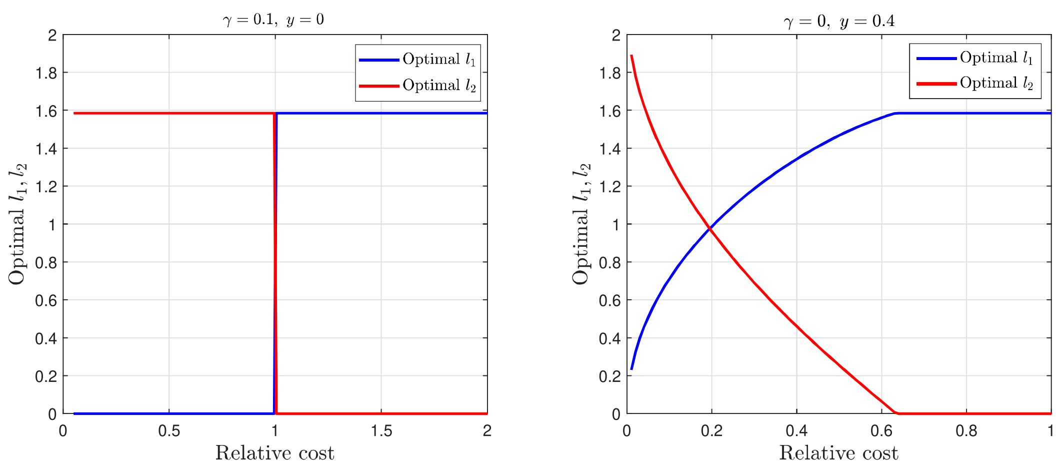
In this scenario, then, we cannot have $(0, w_2)$ (for any $w_2 > 0$) as an optimal solution. Recalling that $w_2 \geq \max\{a - bw_1, 0\}$ for any $w \in \mathcal{W}$, we further have that for any such w, the optimal cost is lower bounded by

$$\begin{cases} c_1 w_1 + c_2 \frac{a}{y w_1} - \frac{b}{y}, & \text{if } w_1 < a/b \\ c_1 \frac{a}{b}, & \text{otherwise} \end{cases}$$

The first line is minimized at $w_1 = \sqrt{\frac{c_2}{c_1}\frac{a}{y}}$ at the value $2\sqrt{c_1 c_2 a/y} - b/y$. For all c_2 large enough, this is greater than $c_1 a/b$. We arrive, therefore, at the conclusion that an optimal solution of the form $(w_1, 0)$ does exist in this case.

Figure 9 is a summary of these derivations. It shows that, in both cases, at least one corner solution exists. This suggests that process complementarity of this form is somewhat weaker than outcome complementarity in that it does not force an optimal solution where both stations have non-negligible workload.

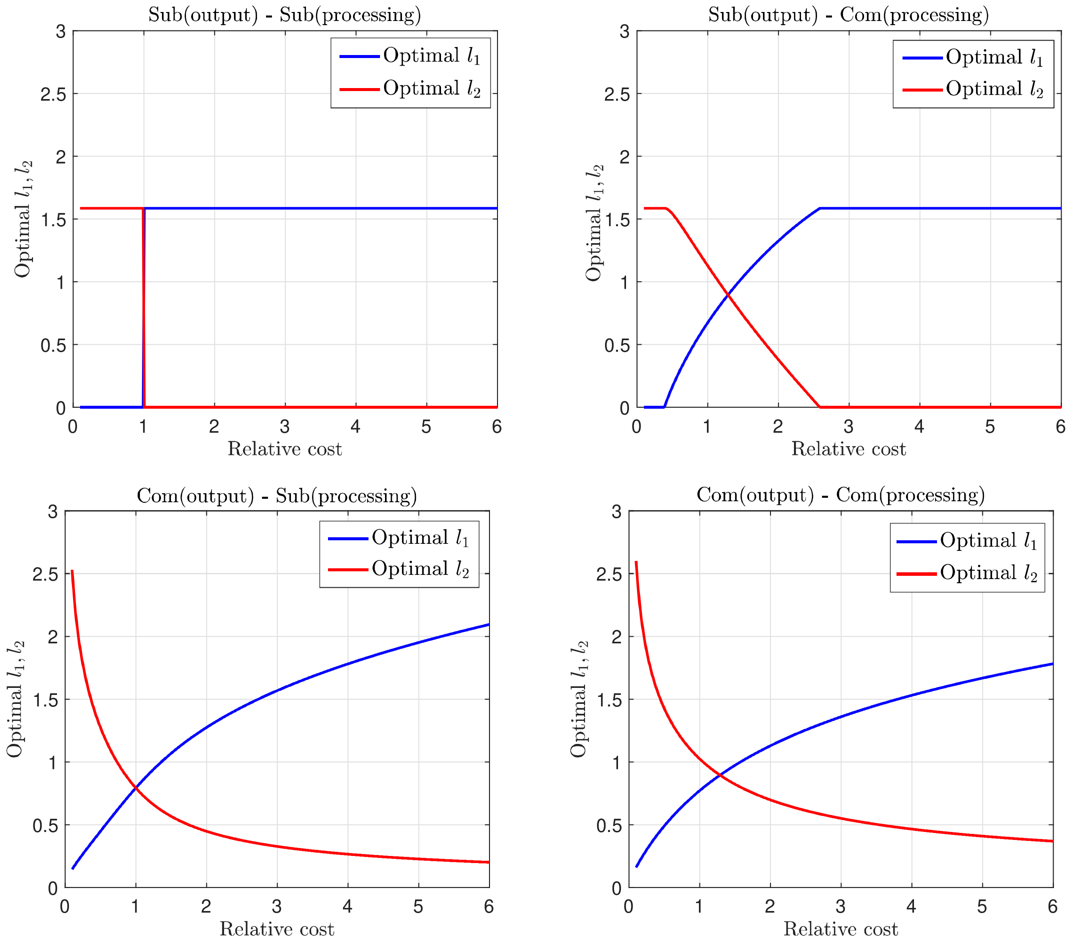
Figure 9 **Optimal solution with $\theta_2(l_1) = \gamma + y l_1$. The parameters are $\varrho = 1$, $\theta_1 = 1$ and $c_1 = 1$.**



We provide numerical examples to substantiate this claim. We restrict our attention here to the case where $\theta_2(l_1) = \gamma + y l_1$. The output complementarity is determined by the value of β, while the processing complementarity is determined by the value of y. We consider four combinations: output and processing substitution; output and processing complementarity; output substitution and processing complementarity; and output complementarity and processing substitution. Figure 10 is a visualization of the optimal decisions.

We observe that the processing substitution/complementary mitigates the effect of the output substitution/complementarity. Under output substitution and processing complementarity (top right plot), the optimal l_1 and l_2 can be zero, while under output complementarity and processing substitution (bottom left plot), these cannot

Figure 10 **Optimal l_1, l_2 for different outcome/processing substitution/complementarity. The parameters are $\varrho = 1$, $\theta_1 = 1$, $\theta_2 = 0.6$, $c_1 = c_2 = 1$; for the outcome substitution/complementarity: $\alpha = 0.5$, $\gamma = 0.1$, $\beta = 1$ and $\beta = 0.25$; for the processing substitution/complementarity: $y = 0$ and $y = 0.1$.**



be zero. This suggests that outcome substitution/complementarity has a stronger effect than processing substitution/complementarity over the total level of substitution/complementarity.

Numerical results for the mixed model $\gamma, y > 0$ with multiple types and finite capacity show that the optimal solution exhibits similar structural properties to the case of outcome substitution/complementarity.

B.3. Incorporating Re-Work Probability Up to a Predefined Time

The re-work probability, as defined in (2), is the probability of a positive-drift BM, $\mathcal{B}_t^{pr}$, starting at $l_1 + l_2$ to hit 0 in finite time. In this section we study the case where the re-work probability is the probability of $\mathcal{B}_t^{pr}$ to hit 0 in a predefined time $T > 0$.

The first time passage distribution at 0 of $\mathcal{B}_t^{pr}$ that starts at $l_1 + l_2$ has the possibly defective PDF (see, e.g., Bhattacharya and Waymire 2009, Chapter 1, page 27):

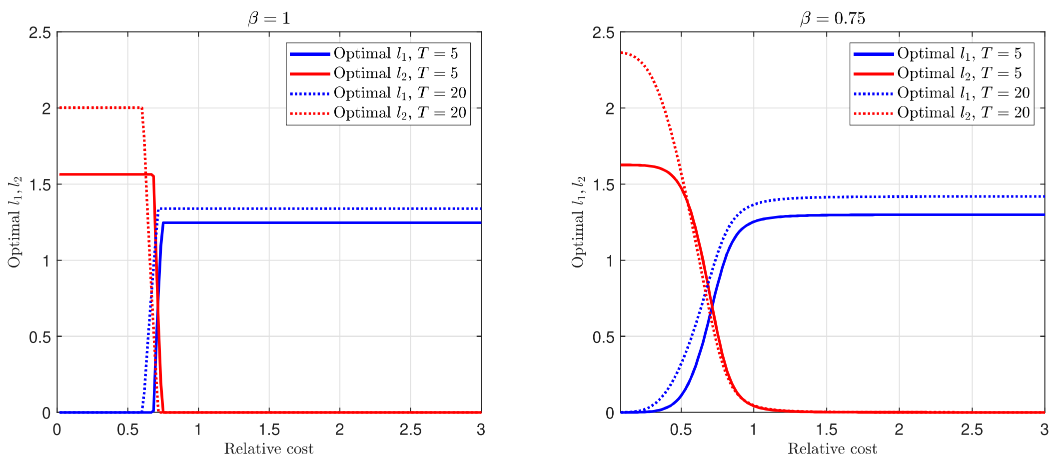
$$f_p(t) = \frac{l_1 + l_2}{\sqrt{2\pi \sigma_{pr}^2 t^3}} \exp\left(-\frac{(-l_1 - l_2 - \eta(l)t)^2}{2\sigma_{pr}^2 t}\right).$$

The re-work probability $p_T(l)$ as a function of T can hence be calculated by

$$p_T(l) = \int_0^T f_p(t)\mathrm{d}t.$$

Figure 11 illustrates the solution structure for different values of T. We derived it by numerically calculating $p_T(l)$, and solving (11) when replacing $p(l)$ with $p_T(l)$. We observe that both stations' efforts increase with T. This is because for a fixed effort mix, the re-work probability increases with T; to overcome this, the discharge threshold $l_1 + l_2$ is increased. The solution structure, however, remains the same as our base model with $p(l)$.

Figure 11 Solution structure when the re-work probability is up to time T. The parameters are $\varrho = 1$, $\theta_1 = 1$, $\theta_2 = 0.6$, $c_1 = 1$, $\alpha = 0.7$.



The results are basically identical to our base case and this can be explained by alluding to the attainable workload region foundation of our earlier results. Specifically, let us define, as before, $N_T(l) = (1 - p_T(l))^{-1}$ and $W_i(l) = l_i N(l)$. Fixing $z \in [0, \infty)$, we have that, with the constraint that $w_2/w_1 = z$,

$$w_1 \geq F_p(z) := \min_{l_1} l_1 N_T(l_1, z l_1). \tag{18}$$

This function plays the role of $\Gamma/\mathbb{h}_{\alpha,\beta}(z)$ in our baseline model. The constraint on w_2 is $w_2 \geq w_1 F_p^{-1}(w_1)$.

Thus, the attainable workload region is given by

$$\mathcal{W} := \{(w_1, w_2) \geq 0 : \;\; w_2 \geq w_1 F_p^{-1}(w_1)\}.$$

This feasible region is depicted in Figure 12 for $\alpha = 0.7$ and two different values of β. As before with $\beta = 1$, the feasible region "hits" the axis at a non-trivial angle; it asymptotes to the axes for $\beta < 1$. As in the baseline model, we end up with three cost regions when $\beta = 1$ and only interior solutions if $\beta < 1$; see Figure 11.

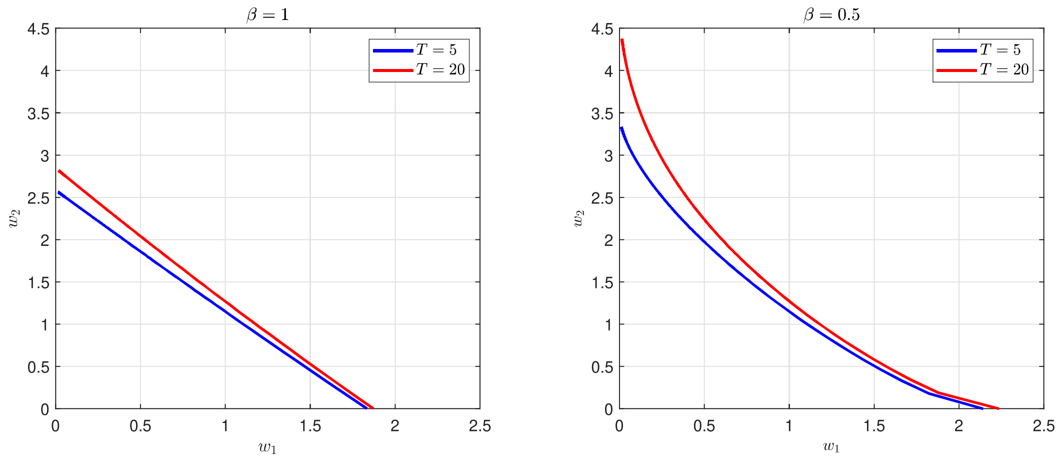
B.4. Limited Capacity with Resource Sharing

In many systems some resources can be shared between the stations. In this case, we face a constraint of the form

$$\mathcal{L}(\vec{l}) = (\lambda/\theta_1) \cdot W_1^r(l) + (\lambda/\theta_2) \cdot W_2^r(l) \leq C$$

We, therefore, want to simultaneously set the stations' effort mix and the resource sharing between the stations.

Figure 12 Feasibility region for different T.



Let C denote the total resource amount for both stations. We consider the following resource sharing problem:

$$\begin{aligned} \min_{\vec{l}\geq 0} \ & (\lambda c_1/\theta_1)\cdot W_1^r(l) + (\lambda c_2/\theta_2)\cdot W_2^r(l) \\ \text{s.t. } & (\lambda/\theta_1)\cdot W_1^r(l) + (\lambda/\theta_2)\cdot W_2^r(l) \leq C. \end{aligned} \tag{19}$$

For a vector (of pairs) $\mathbf{w} = ((w_1^1, w_2^1), \ldots (w_1^K, w_2^K))$, let $\mathbf{w}_1 = (w_1^1, \ldots, w_1^K)$ be the sub-vector that has the first coordinate from each pair and similarly define $\mathbf{w}_2$ for the second coordinate. Then, (19) is equivalently formulated as

$$\begin{aligned} & \min \ (\lambda c_1/\theta_1)\cdot \mathbf{w}_1 + (\lambda c_2/\theta_2)\cdot \mathbf{w}_2 \\ & \text{s.t. } (\lambda/\theta_1)\cdot \mathbf{w}_1 + (\lambda/\theta_2)\cdot \mathbf{w}_2 \leq C, \\ & \mathbf{w} \in \mathcal{W}^{\times}, \end{aligned} \tag{20}$$

Since all input must be processed, some capacity levels C might be infeasible. The set of feasible capacity vectors C is given by

$$\mathcal{C}_\beta = \{C \geq 0 : \exists \vec{l} \geq 0, \text{ s.t. } \mathcal{L}(\vec{l}) \leq C\}.$$

Lemma 5 characterized the optimal effort mix and the optimal resource sharing for a given total capacity C.

LEMMA 5 **(optimal resource sharing).** *Fix capacity $C \in \mathcal{C}_\beta$. Let $\bar{l}_k^*(\cdot)$ be the solution for type k items in Theorem 2, as a function of the relative cost. Then, the unique solution $\vec{l}^*$ to (19) is given by*

$$l_k^* = \bar{l}_k^*(\widetilde{\mathcal{R}}_k^c), \ k \in [K],$$

where

$$\widetilde{\mathcal{R}}_k^c = \mathcal{R}_k^c \left(\frac{c_2^k + \kappa}{c_2^k}\right) / \left(\frac{c_1^k + \kappa}{c_1^k}\right),$$

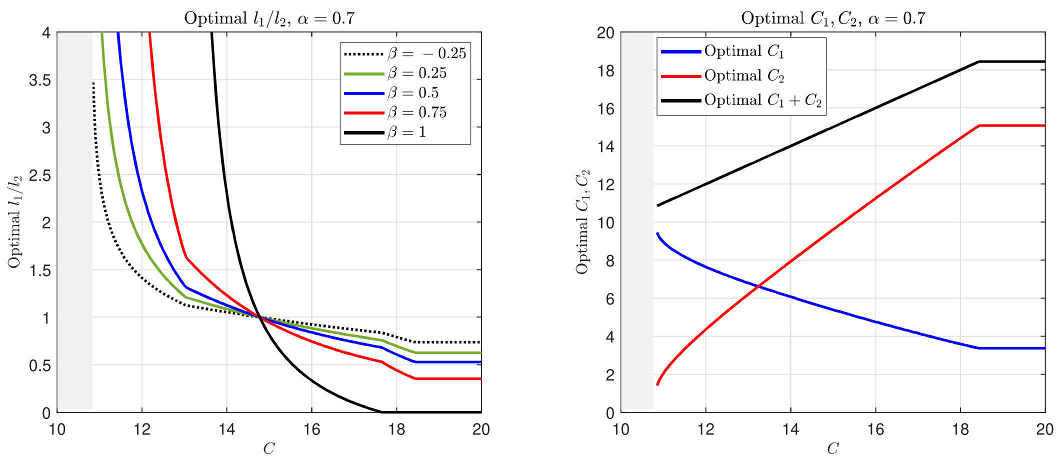
and κ is the minimal non-negative value for which $\mathcal{L}(\vec{l}^*)(\widetilde{\mathcal{R}}^c) \leq C$. The optimal capacity allocation to each station is $C_1^* = (\lambda/\theta_1)\cdot W_1^r(l^*)$, $C_2^* = (\lambda/\theta_2)\cdot W_2^r(l^*)$.

The proof of Lemma 5 is argued identically to Theorem 4.

The dual variable κ for the resource constraint scale up/down the relative cost. If $c_1 < c_2$, the relative cost is shrunk for all item groups. The factor by which it is shrunk depends on the baseline costs c_1^k, c_2^k. This, in turn, diverts more effort to Station 2, the more expensive one. If $c_1^k > c_2^k$, however, the relative cost is scaled up for all item groups. This, in turn, diverts more effort to Station 1, the more expensive one. If $c_1^k = c_2^k$, the relative cost remains intact.

Important is that—because the finite capacity is mapped to a perturbed price $\widetilde{\mathcal{R}}^c$ and because l_1 is increasing and $\widetilde{\mathcal{R}}^c$ and l_2 is decreasing—the capacity allocated to the two stations is not increasing in total capacity. Instead, as the total capacity increases when $c_1 > c_2$, less of it is allocated to Station 1 and more of it is allocated to station 2. This is illustrated in Figure 13. The right plot shows the optimal capacity allocation for Stations 1 and 2 vs. the available capacity C. When there is ample capacity, most of the capacity (90%) is allocated to Station 2 – the less costly one. However, as capacity becomes scarce, more capacity is allocated to Station 1 and less to Station 2. When $c_1 < c_2$, it is the other way around.

Figure 13 Optimal resource sharing as a function of the available capacity C for $k \in [5]$, where each k corresponds to a different value of β. The left plots present l_1^k/l_2^k, and the right plots present the optimal capacity allocation C_1^* and C_2^*. The parameters are $\varrho = 1$, $\theta_1 = 1$, $\theta_2 = 0.6$, $\lambda_k = 1$, $c_1 = 3$, $c_2 = 1$.



B.5. Throughput Maximization

Throughput maximization is a special case of an optimization problem over $\mathcal{W}$. Specifically,

$$\begin{aligned} \lambda^* = &\max_{w_1, w_2, \lambda} \lambda \\ &\text{s.t. } \lambda w_1/\theta_1 \le C_1; \quad \lambda w_2/\theta_2 \le C_2; \\ &w_1, w_2 \in \mathcal{W}. \end{aligned}$$

It is immediate that, at optimality,

$$\lambda^* = \min\left\{\frac{C_1\theta_1}{w_1^*}, \frac{C_2\theta_2}{w_2^*}\right\} = \max_{w \in \mathcal{W}} \min\left\{\frac{C_1\theta_1}{w_1}, \frac{C_2\theta_2}{w_2}\right\}.$$

This is a problem of maximizing the concave function

$$g(w) = \min\left\{\frac{C_1\theta_1}{w_1}, \frac{C_2\theta_2}{w_2}\right\}.$$

over the convex set $\mathcal{W}$; it is, hence, a convex optimization problem. Per Lemma 8, the feasible set shrinks as β decreases. In turn, the throughput is monotonically decreasing in β. Stated verbally, this shows that the model and its math are consistent with prior intuition: the greater the substitution, the more throughput we should be able to handle, because substitution allows flexibility to spread out the load.

To state this formally, with multiple types we make the assumption that the mix is fixed and it is the total volume that changes. That is, $\lambda^k = \Lambda a^k$, where $a^k \geq 0$ and $\sum_{k\in[K]} a^k = 1$.

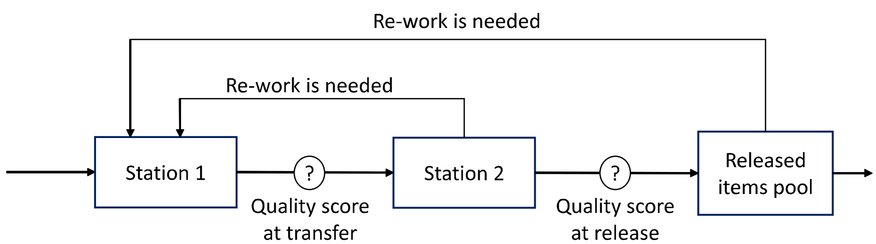
$$\begin{aligned}\lambda^* = &\max_{\mathbf{w},\lambda} \lambda \\ &\text{s.t. } \lambda \sum_{k\in[K]} \frac{a^k}{\theta_1^k} w_1^k \leq C_1; \quad \lambda \sum_{k\in[K]} \frac{a^k}{\theta_2^k} w_2^k \leq C_2, \\ &\mathbf{w} \in \mathcal{W}^\times.\end{aligned}$$

LEMMA 6. *Given $C_1, C_2 \geq 0$, the maximal throughput can be found by solving a convex optimization problem. The optimal solution is increasing in β: the greater the substitution, the greater the throughput.*

B.6. Incorporating Re-Work from Station 2 to Station 1.

Some fallback in Station 2 can be handled within Station 2's environment. Some cases, however, require intensive intervention in Station 1. In this section we analyze the case where quality deterioration and re-work can occur while in Station 2.

Figure 14 Model illustration with re-work from Station 2.



Specifically, if an item's quality score hits some re-work threshold $a \leq 0$ while in Station 2, the item is returned to Station 1. In other words, items stay in Station 2 until their quality score either reaches $l_1 + l_2$ (when they are released) or zero (when they are returned); this takes, in expectation,

$$m_2(l) := \frac{l_2(1 - e^{\varrho l_1 \theta_2}) + l_1(1 - e^{-\varrho l_2 \theta_2})}{\theta_2\left(e^{-\varrho l_2 \theta_2} - e^{\varrho l_1 \theta_2}\right)}.$$

We now distinguish between two types of re-work: The first occurs after Station 2; the expected number of these returns remains $N(l)$.

The second type of re-work occurs while processing in Station 2; the expected number of returns from Station 2 is $N_2(l) := [1 - p_2(l)]^{-1}$, where

$$p_2(l) = \frac{e^{-\varrho l_2 \theta_2} - 1}{e^{-\varrho l_2 \theta_2} - e^{\varrho l_1 \theta_2}}$$

is the re-work probability from Station 2 – the probability of hitting zero before hitting level $l_1 + l_2$.

The total costs of Stations 1 and 2 are then given by

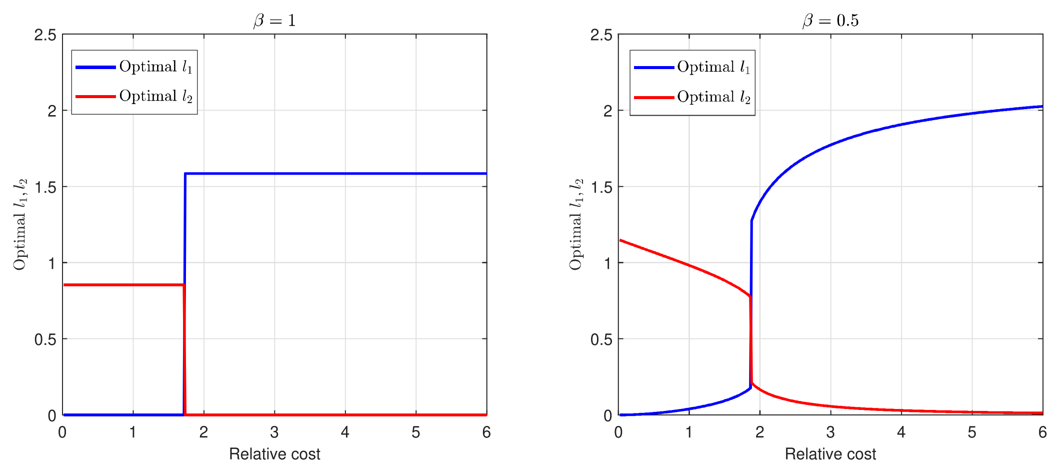
$$\begin{aligned}\text{Station 1 cost } &= c_1\, m_1(l)\,[N(l) + N_2(l) - 1]\,; \\ \text{Station 2 cost } &= c_2\, m_2(l)\,[N(l) + N_2(l)]\,.\end{aligned}$$

We are, therefore, looking to solve

$$V^* = \inf_{l \geq 0} V(l, d) = \inf_{l \geq 0} \text{ Station 1 cost } + \text{Station 2 cost.} \tag{21}$$

Figure 15 illustrates the solution structure when the re-work threshold from Station 2 is $a = -1$. We derived it by numerically solving (21). In this example and many others we examined, the solution structure is consistent with the one derived in §4, demonstrating the solution robustness. Moreover, we see that in contrast to the base model and its variants that we considered, having $\beta < 1$ does not guarantee that (l_1, l_2) changes smoothly with cost. Instead, there is a jump at a cost-threshold cost as in the case where $\beta = 1$ and $\alpha = 0.5$. This is certainly specific to our construction of this variant; nevertheless, it showcases the richness of the type of behavior we would observe as we introduce additional features.

Figure 15 Solution structure when there are returns from Station 2 to Station 1. The parameters are $\varrho = 1$, $\theta_1 = 1$, $\theta_2 = 0.6$, $c_1 = 1$, $\alpha = 0.5$, $a = -1$.



Appendix C: Data Requirements and Parameter Estimation for Model Implementation

Estimating the quality-evolution parameters θ, σ from service-time samples. While our model stipulates that quality evolves as a positive-drift Brownian motion, the drift and variance parameters can be estimated without directly observing the quality path. Instead, we can rely on station-level processing-time data. Under our model, processing times follow an inverse Gaussian distribution—a distribution widely used for modeling and estimating service times (see Whitmore 1975, 1979 and the recent Hashimoto et al. 2023).

The required data include the processing times of items at each station and for each visit. Such data are typically recorded systematically in operational databases. From these data, we can estimate the parameters of the inverse Gaussian distribution, which correspond to l_i/θ_i and $(l_i/\sigma_i)^2$ for each station i. These parameters can be obtained using standard maximum-likelihood estimation methods.

In practice, the quality thresholds l_i can be obtained either directly or indirectly. They may be directly observed, for example, when quality or performance scores are recorded at transfer or release, or inferred from design-level targets used in engineering or service planning. When l_i is not directly observed, it can be estimated jointly with θ_i and σ_i by fitting the inverse-Gaussian likelihood to the observed processing-time data, treating l_i as a latent parameter inferred from the same data. Because l_i enters both the mean and the shape parameters of the inverse-Gaussian distribution, any measurement error in l_i propagates to the estimated drift and diffusion parameters (θ_i, σ_i). However, since the inverse-Gaussian likelihood depends on l_i only through the ratios l_i/θ_i and l_i/σ_i, a small multiplicative error in l_i can be absorbed by a roughly proportional rescaling of θ_i and σ_i. Consequently, such error primarily affects the overall scale of these parameters rather than their ratio θ_i/σ_i, leaving the process dynamics and identification largely unchanged. Substantial mismeasurement, however, would proportionally rescale predicted processing times and re-work probabilities, as these quantities depend directly on the magnitude of l_i.

Subsequently, we estimate the quality score at transfer or release to build estimates for θ_i and σ_i^2.

Estimating the substitution parameters α, β from realized re-work. The parameters of the CES production function, α and β, can be estimated via maximum likelihood using item-level data. Assume we have a dataset of m items, indexed by j, each possibly generating multiple visits. For item j, let N_j denote the observed number of visits until final completion (no further re-work). Under our model, all visits for the same item share the same re-work probability $p_{\alpha,\beta}(l_1, l_2)$, so that N_j follows a geometric distribution with success probability $1 - p_{\alpha,\beta}(l_1, l_2)$. Hence, the likelihood contribution of item j is

$$L_j(\alpha, \beta) = (1 - p_{\alpha,\beta}(l_1, l_2))\, p_{\alpha,\beta}(l_1, l_2)^{N_j - 1},$$

and the total likelihood is $L(\alpha, \beta) = \prod_{j=1}^{m} L_j(\alpha, \beta)$. The corresponding log-likelihood is

$$\ell(\alpha, \beta) = \sum_{j=1}^{m} \Big[\log(1 - p_{\alpha,\beta}(l_1, l_2)) + (N_j - 1) \log(p_{\alpha,\beta}(l_1, l_2)) \Big].$$

Maximizing this likelihood yields consistent estimates for α and β while properly accounting for the dependence across visits generated by the same item. Synthetic simulation experiments, described below, assess the finite-sample precision of the maximum-likelihood estimator and confirm that it accurately recovers the true parameters even in moderate sample sizes. Note that these results are obtained under the assumed model specification and do not establish robustness to model misspecification.

If, in addition to binary re-work indicators, data on the time elapsed between release and re-work were available, the estimation procedure could be extended to use this richer information. Under our model, post-release quality evolves as a BM with drift $\eta(l_1, l_2)$ and diffusion coefficient σ_{pr}, so the time until re-work corresponds to the first-hitting time of this process, which follows an inverse Gaussian distribution. While our current approach infers these post-release parameters indirectly from re-work probabilities, observing actual return times would allow them to be estimated directly from their likelihood function, thereby providing a dynamic validation of the post-discharge process.

Synthetic Precision Check. To evaluate the accuracy and precision of the item-level maximum likelihood estimator, we conducted a synthetic simulation calibrated to the maintenance setting. Each replication simulated m independent items, where item j's thresholds $(l_{1,j}, l_{2,j})$ were generated with independent variation in their total effort $l_{1,j} + l_{2,j}$ and in the relative mix between Station 1 and Station 2, so that overall scale and allocation effects were varied separately rather than by sampling $l_{1,j}$ and $l_{2,j}$ independently.

For each item j, the number of visits N_j was drawn from a geometric distribution with success probability $1 - p_j$, where

$$p_j = p_{\alpha,\beta}(l_{1,j}, l_{2,j}) = \exp\{-\eta_{\alpha,\beta}(l_{1,j}, l_{2,j})(l_{1,j} + l_{2,j})\}$$

Table 3 Simulation-based evaluation of the item-level MLE under per-item thresholds.

Design (α_0, β_0)	m	Mean visits / item	$\hat{\alpha}$		$\hat{\beta}$	
			Bias	RMSE	Bias	RMSE
(0.60, 1.00)	500	1.230	-0.0085	0.0408	-0.0713	0.1272
(0.60, 1.00)	1000	1.229	-0.0049	0.0319	-0.0502	0.0917
(0.60, 0.50)	500	1.286	0.0022	0.0396	0.0311	0.1568
(0.60, 0.50)	1000	1.288	-0.0013	0.0272	0.0089	0.1027
(0.60, 0.00)	500	1.395	0.0019	0.0297	0.0084	0.1090
(0.60, 0.00)	1000	1.394	-0.0002	0.0209	0.0100	0.0755
(0.55, -0.50)	500	1.542	-0.0006	0.0349	0.0059	0.1408
(0.55, -0.50)	1000	1.540	-0.0006	0.0237	0.0024	0.0904

denotes the re-work probability defined in (2). The true parameter values (α_0, β_0) were fixed at four representative combinations corresponding to strong substitution ($\beta_0 = 1$) and three levels of complementarity $(0.5, 0, -0.5)$. For each design we ran 300 replications with sample sizes $m \in \{500, 1000\}$. The likelihood maximization is performed using a multi-start Nelder–Mead routine (via `fminsearch` in MATLAB), initialized over a grid of starting values, with convergence determined by default tolerance settings; a fixed random seed is used to ensure reproducibility. Table 3 reports the resulting mean number of visits and the bias and root-mean-square error (RMSE) of the estimators $\hat{\alpha}$ and $\hat{\beta}$.

The results indicate that both parameters are estimated with high precision even in moderate samples ($m = 500$), and that bias and RMSE decline further as the sample size increases. Across all designs, the bias in α is negligible ($|\text{bias}| < 0.01$), while the bias in β remains below 0.08 at $m = 500$ and below 0.05 at $m = 1000$, with RMSEs typically under 0.10.

Appendix D: Parameter Calibration for Maintenance Example

To provide an empirical grounding for the stylized *Maintenance Example*, we consider large-scale industrial and facility maintenance tasks such as those performed on power-generation turbines, compressors, and other heavy-plant equipment. These systems involve two sequential activities: an initial *repair stage* followed by *preventive service* performed while the equipment remains offline.

Station 1 (Repair) represents corrective actions that restore functionality— for instance, replacing a failed generator bearing, realigning a gearbox, or repairing electrical components. **Station 2 (Preventive service)** represents inspection and tuning activities— such as lubrication, recalibration, torque checks, or component balancing—that enhance long-term reliability and reduce the probability of future re-work.

Industry studies and field reports on wind-turbine operation and maintenance indicate that the scope and time requirements of these activities vary widely across sites. Facts (2024) notes that regular service campaigns often involve full-day maintenance windows that include inspection, cleaning, calibration, and documentation tasks, while Committee (2017) highlights that corrective interventions—when a specific fault has been identified—typically achieve a faster improvement in the asset's operational condition per hour of work, even though they may target a smaller subset of components.

Anchoring to these observations, we set the quality-improvement rates (Brownian drifts) to

$$\theta_1 = 0.15\ \text{h}^{-1}\ \text{(repair)} \quad \text{and} \quad \theta_2 = 0.10\ \text{h}^{-1}\ \text{(preventive service)},$$

reflecting that corrective actions directly address known issues and thus improve the quality state more rapidly per hour, whereas preventive activities yield slower but more comprehensive gains that sustain reliability over time. We also use the normalized value $\lambda = 1$.

For labor costs, we use U.S. Bureau of Labor Statistics (2024) medians—repair tasks corresponding to *industrial-machinery mechanics* (SOC 49-9041) and preventive-service tasks corresponding to *equipment or quality-control technicians* (SOC 51-9061)— augmented by a conservative 40–60% overhead. This yields $c_1 \in [\$40, \$50]/\text{hr}$ and $c_2 \in [\$35, \$45]/\text{hr}$; therefore, we set $c_1 = 45$ and $c_2 = 40$.

We take $\alpha = 0.4$ and explore $\beta \in \{1,\ 0.5,\ 0.25,\ -4,\ -25\}$ to span the range from substitution to complementarity between repair and preventive service. Here, $\alpha = 0.4$ indicates that greater weight is placed on Station 2 (preventive service) in the aggregated-quality function: while repair is essential to restore basic functionality, preventive service has a larger marginal influence on long-term reliability and re-work reduction. These values reflect realistic maintenance durations for large assets, confirming that the comparative statics are plausible for real operations.

Under these inputs, the optimality conditions in Theorem 2 yield the order-of-magnitude solutions summarized in the first column ($C_1 = \infty$) of Table 4.

Table 4 Calibrated optimal thresholds and mean durations (hours) across capacity settings.

		$C_1 = \infty$		$C_1 = 6$		$C_1 = 10$	
β	Interaction type	(l_1^*, l_2^*)	Durations	(l_1^*, l_2^*)	Durations	(l_1^*, l_2^*)	Durations
1	Full substitution	(1.77,0)	(11.8, 0)	(0.21,1.27)	(1.4, 12.7)	(0.36,1.15)	(2.4, 11.5)
0.5	Mild complementarity	(0.94,0.67)	(6.3, 6.7)	(0.21,1.38)	(1.4, 13.8)	(0.36,1.21)	(2.4, 12.1)
0.25	Moderate complementarity	(0.90,0.77)	(6.0, 7.7)	(0.21,1.46)	(1.4, 14.6)	(0.36,1.24)	(2.4, 12.4)
-4	Strong complementarity	(0.81,0.78)	(5.4, 7.8)	(0.21,4.44)	(1.4, 44.4)	(0.36,2.44)	(2.4, 24.4)
-25	Full complementarity	(0.80,0.79)	(5.3, 7.9)	(0.21,5.43)	(1.4, 54.3)	(0.36,3.03)	(2.4, 30.3)

D.1. Finite-Capacity Calibration

To examine how limited preventive-service capacity affects the optimal design, we extend the calibrated maintenance example to a setting with a finite number of preventive-service workers. We consider two representative capacity levels for the repair stage, $C_1 \in \{6, 10\}$, and five degrees of interaction between preventive service and repair: $\beta \in \{1, 0.5, 0.25, -4, -25\}$. The corresponding optimal thresholds and mean durations are reported in Table 4.

The results illustrate how the capacity of the repair stage shapes the system's overall performance. When repair capacity is tight ($C_1 = 6$), the system relies more heavily on preventive service to sustain quality, as reflected in the longer preventive durations across all interaction levels (β values). Under strong or full complementarity ($\beta = -4$ and -25), limited repair capacity substantially reduces the effectiveness of preventive service, resulting in particularly long processing times at Station 2. When capacity increases ($C_1 = 10$), both stations operate more efficiently, and the preventive durations shorten accordingly. Overall, these patterns confirm that capacity constraints amplify performance losses in complementary systems, whereas their effect is less pronounced when the interaction between stages approaches substitution.

Appendix E: Additional Numerical Experiments

This section presents additional numerical experiments designed to illustrate the robustness and practical implications of the analytical results. All numerical results were obtained using standard MATLAB routines; the solution of a single instance requires under 0.01 seconds on a laptop CPU.

E.1. Sensitivity to Parameter Misspecification

Table 5 presents the results of a sensitivity analysis. Each parameter was independently perturbed by $\pm 10\%$, $\pm 20\%$, and $\pm 50\%$ around the calibrated baseline reported in Appendix D, while all other parameters were held fixed. The columns labeled "Δ Mean Duration (hours,%)" report the absolute and relative changes in the average processing time at each station relative to the baseline solution. Across all scenarios, the variation in mean duration remains modest relative to the size of the perturbation. Even for larger changes of $\pm 20\%$ and $\pm 50\%$, the induced variation is substantially smaller than the magnitude of the parameter shift, indicating that the optimal design and operating regime are robust to parameter misspecification.

E.2. Comparison with Other Benchmark Policies

Although our proposed policies are analytically derived, proved to be optimal, and admit closed-form expressions, it is instructive to compare their performance with other intuitive benchmark policies that are often used in practice. We consider two such benchmarks that reflect simple, decentralized decision rules.

Cost-Priority Policy. This benchmark prioritizes effort allocation to the more cost-effective stage—that is, the station with the smaller effective cost ratio c_i/θ_i. Such a rule captures cost-driven decision making guided by immediate efficiency considerations rather than joint optimization.

Decentralized Policy. In this benchmark, each station determines its own quality threshold independently, ignoring the interaction captured by the complementarity or substitution parameter β. This represents an uncoordinated design in which decisions are made from a local perspective rather than a system-wide one. Specifically, station $i \in \{1,2\}$ solves

$$V_i^* = \min_{l_i \in \mathbb{R}_+} \frac{c_i}{\theta_i} l_i N_i(l_i),$$

where

$$N_i(l_i) = \left[1 - \exp\left(-\nu - \varrho\,(\alpha l_i^{\beta})^{1/\beta} l_i\right)\right]^{-1}.$$

In practice, this benchmark represents departments or teams setting their own operating targets without accounting for cross-stage effects.

Table 6 compares the optimal integrated design with the two benchmark policies across different levels of substitutability and complementarity, as captured by the parameter β. Recall that $\beta \in (-\infty, 1]$ spans the full range from

Table 5 Sensitivity of the optimal design to parameter misspecification. Each parameter is varied by $\pm 10\%$, $\pm 20\%$, and $\pm 50\%$, holding the other parameters fixed.

				Δ Mean Duration (hours, %)	
Parameter	**Change**	l_1^*	l_2^*	**Station 1**	**Station 2**
α	$\pm 10\%$	0.80–0.87	0.73–0.79	0.24 (4.3%)	0.29 (3.9%)
	$\pm 20\%$	0.76–0.90	0.70–0.82	0.48 (8.7%)	0.59 (7.7%)
	$\pm 50\%$	0.61–0.99	0.61–0.91	1.24 (22.4%)	1.50 (19.7%)
β	$\pm 10\%$	0.83–0.83	0.76–0.76	0.01 (0.2%)	0.02 (0.2%)
	$\pm 20\%$	0.83–0.84	0.76–0.76	0.03 (0.5%)	0.03 (0.4%)
	$\pm 50\%$	0.82–0.85	0.75–0.77	0.07 (1.3%)	0.08 (1.1%)
c_1/θ_1	$\pm 10\%$	0.79–0.88	0.72–0.80	0.32 (5.7%)	0.35 (4.6%)
	$\pm 20\%$	0.75–0.95	0.69–0.83	0.64 (11.7%)	0.71 (9.3%)
	$\pm 50\%$	0.66–1.24	0.55–0.92	1.92 (34.7%)	1.85 (24.3%)
c_2/θ_2	$\pm 10\%$	0.78–0.88	0.73–0.80	0.31 (5.7%)	0.36 (4.7%)
	$\pm 20\%$	0.73–0.92	0.70–0.84	0.63 (11.3%)	0.72 (9.5%)
	$\pm 50\%$	0.57–1.05	0.63–1.04	1.60 (28.9%)	2.05 (26.9%)

strong complementarity to full substitution: $\beta = 1$ corresponds to perfect substitution, while $\beta < 0$ captures complementarity, with increasingly negative values reflecting stronger complementarity. Corner solutions with $l_i = 0$ arise under substitution regimes, whereas under complementarity the optimal solution is typically interior.

When $\beta < 0$, the benchmark policies that deactivate one stage become *degenerate*, as effective quality collapses to zero when only a single station operates. In contrast, the integrated policy consistently delivers superior outcomes across all interaction regimes, achieving substantial cost savings relative to both benchmarks. These results underscore the operational value of jointly optimizing both thresholds and highlight the importance of coordinated decision-making within an integrated framework.

Appendix F: Proofs of Main Results

This section contains the proof of the main results (propositions and theorems). Proofs of all lemmas appear in Appendix G.

Proof of Theorem 1. We will prove this through explicit characterization of the constraints that define the set $\mathcal{W}$. We break the space into "rays":

$$\mathcal{W}_z = \left\{ w_1, w_2 \in \mathcal{W},\ \frac{w_2}{w_1} = z \right\}.$$

For $w_1, w_2 \in \mathcal{W}_z$,

$$w_1 \geq \min_{l_1 \geq 0} l_1 N(l_1, z l_1); \tag{22}$$

Table 6 Comparison with benchmark policies. "Degenerate" indicates that the benchmark collapses under $\beta < 0$, yielding $\eta = 0$.

	Objective value			Cost saving (%)	
β	Optimal	Benchmark 1	Benchmark 2	Benchmark 1	Benchmark 2
1	743.3	743.3	1,117.5	0.0%	33.5%
0.5	772.8	1,346	1,588	42.6%	51.3%
0.25	773.7	6881.5	3347.2	88.8%	76.9%
-4	775.3	Degenerate	819.8	–	5.4%
-25	775.3	Degenerate	847.6	–	8.5%

$$w_2 = (w_1)\, z \geq z \min_{l_1 \geq 0}\; l_1 N(l_1, z l_1).$$

Let us expand the right-hand side of (22)

$$\begin{aligned} l_1 N(l_1, z l_1) &= l_1 \left[1 - e^{-\nu - \varrho \eta(l_1, z l_1) l_1 (z+1)}\right]^{-1} \\ &= l_1 \left[1 - e^{-\nu - \varrho l_1^2 \left(\alpha + (1-\alpha) z^{\beta}\right)^{\frac{1}{\beta}} (z+1)}\right]^{-1} \\ &= \frac{1}{\sqrt{\mathbb{h}_{\alpha,\beta}(z)}} \sqrt{\mathbb{h}_{\alpha,\beta}(z)}\, l_1 \left[1 - e^{-\nu - l_1^2 \mathbb{h}_{\alpha,\beta}(z)}\right]^{-1}, \end{aligned}$$

where

$$\mathbb{h}_{\alpha,\beta}(z) := \varrho \left(\alpha + (1-\alpha) z^{\beta}\right)^{\frac{1}{\beta}} (z+1).$$

Changing variables by defining $y = \sqrt{\mathbb{h}_{\alpha,\beta}(z)}\, l_1$, yields the following

$$w_1 = \min_{l_1 \geq 0}\; l_1 N(l_1, z l_1) = \frac{1}{\sqrt{\mathbb{h}_{\alpha,\beta}(z)}} \min_{y \geq 0} y \left[1 - e^{-\nu - y^2}\right]^{-1} = \frac{\Gamma}{\sqrt{\mathbb{h}_{\alpha,\beta}(z)}},$$

where

$$\Gamma = \min_{y \geq 0} y \left[1 - e^{-\nu - y^2}\right]^{-1}.$$

Notice that Γ does not depend on α, β, or z.

Transformation then gives us

$$w_1^2 = \frac{\Gamma^2}{\mathbb{h}_{\alpha,\beta}(z)} \;\Rightarrow\; \mathbb{h}_{\alpha,\beta}(z) = \frac{\Gamma^2}{w_1^2} \;\Rightarrow\; z = \mathbb{h}_{\alpha,\beta}^{-1}\left(\frac{\Gamma^2}{w_1^2}\right),$$

and thus,

$$w_2 = z\,(w_1) = (w_1)\, \mathbb{h}_{\alpha,\beta}^{-1}\left(\frac{\Gamma^2}{w_1^2}\right).$$

Since $w_2 \geq 0$, the set $\mathcal{W}$ can now be rewritten as

$$\mathcal{W} = \left\{ w_1, w_2 \geq 0 :\; w_2 \geq w_1 \mathbb{h}_{\alpha,\beta}^{-1}\left(\frac{\Gamma^2}{w_1^2}\right) \right\}.$$

Since Γ is independent in z, $\mathcal{W}$ is a convex set provided that the function

$$f_{\alpha,\beta}(w_1) = (w_1)\, \mathbb{h}_{\alpha,\beta}^{-1}\left(\frac{\Gamma^2}{w_1^2}\right)$$

is convex.

We will prove this convexity by showing that $f''_{\alpha,\beta}(w_1) \geq 0$. The following representation will be useful

$$g(z) = \frac{\Gamma}{\sqrt{\mathbb{h}_{\alpha,\beta}(z)}}, \quad m(z) = \frac{z}{\sqrt{\mathbb{h}_{\alpha,\beta}(z)}},$$

so that

$$f_{\alpha,\beta}(w_1) = m\left(g^{-1}(w_1)\right).$$

First, we introduce the following auxiliary lemma.

LEMMA 7. *The function* $g(z) = \Gamma/\sqrt{\mathbb{h}_{\alpha,\beta}(z)}$ *is monotone decreasing in* z *for* $\alpha \geq 0.5$.

The first and second derivatives of $f_{\alpha,\beta}$ are

$$f'_{\alpha,\beta}(w_1) = m'\left(g^{-1}(w_1)\right) \frac{1}{g'\left(g^{-1}(w_1)\right)},$$

and

$$\begin{aligned} f''_{\alpha,\beta}(w_1) &= m''\left(g^{-1}(w_1)\right)\left(\frac{1}{g'\left(g^{-1}(w_1)\right)}\right)^2 + m'\left(g^{-1}(w_1)\right)\left(-\frac{g''\left(g^{-1}(w_1)\right)}{g'\left(g^{-1}(w_1)\right)^3}\right) \\ &= \left(g'\left(g^{-1}(w_1)\right)\right)^{-3}\left[m''\left(g^{-1}(w_1)\right)g'\left(g^{-1}(w_1)\right) - m'\left(g^{-1}(w_1)\right)g''\left(g^{-1}(w_1)\right)\right]. \end{aligned}$$

Per Lemma 7, $g'(z) < 0$; therefore, it suffices to prove that the second-line term in brackets is negative; namely,

$$m''(z)g'(z) - m'(z)g''(z) \leq 0, \quad \forall z \geq 0.$$

This is what we do next.

Direct differentiation yields that

$$m''(z)g'(z) - m'(z)g''(z) = \frac{\mathcal{L}(z)}{4z^2(1+z)^3\left(\alpha + (1-\alpha)z^\beta\right)^{2+\frac{1}{\beta}}}, \tag{23}$$

where

$$\mathcal{L}(z) = -z^{2\beta}(1-\alpha)^2 - z^2\alpha^2 + 2z^\beta(1-\alpha)\alpha\left(-1 - z(1+z) + \beta + z(2+z)\beta\right).$$

Since the denominator in (23) is positive, it suffices to show that $\mathcal{L}(z)$ is negative for all $z \geq 0$. For all $\beta \leq 1$: since $\lim_{z\downarrow 0}\mathcal{L}(z) \leq 0$, it suffices to show that $\mathcal{L}'(z) \leq 0$ for all $z > 0$ to conclude that $\mathcal{L}(z) \leq 0$ for all $z \geq 0$.

Direct differentiation gives

$$\mathcal{L}'(z) = -2\beta z^{-1+2\beta}(1-\alpha)^2 + 2z^{-1+\beta}\left(\beta^2(1+z)^2 - z(1+2z) + \beta(-1+z+z^2)\right)\alpha(1-\alpha) - 2z\alpha^2.$$

In turn,

$$\mathcal{L}'(z) \leq a(z) := -2\beta z^{-1+2\beta}(1-\alpha)^2 + 2z^{-1+\beta}\left(\beta^2(1+z)^2 - z(1+2z) + \beta(-1+z+z^2)\right)\alpha(1-\alpha) - 2z\alpha^2.$$

Since $a(z) \leq 0$, if and only if, $z^{1-\beta}a(z) \leq 0$, for all $z \geq 0$, we proceed with

$$z^{1-\beta}a(z) = -2\beta z^\beta(1-\alpha)^2 + 2\left[(\beta^2+\beta-2)z^2 + (2\beta^2+\beta-1)z + \beta^2 - \beta\right]\alpha(1-\alpha) - 2z^{2-\beta}\alpha^2.$$

Since $\beta \leq 1$, we get that $\beta^2 + \beta - 2 \leq 0$, and $\beta^2 - \beta \leq 0$, so that

$$z^{1-\beta}a(z) = -2\beta z^\beta(1-\alpha)^2 + 2\left[(\beta^2+\beta-2)z^2 + (2\beta^2+\beta-1)z + \beta^2 - \beta\right]\alpha(1-\alpha) - 2z^{2-\beta}\alpha^2.$$

Now, if $2\beta^2+\beta-1\le 0$ (equivalently, $\beta\le 0.5$), then $z^{1-\beta}a(z)\le 0$, and we are done. It remains to consider the case where $\beta>0.5$. In this case,

$$z^{1-\beta}a(z)\le -2\left(\beta z^{\beta}(1-\alpha)^2-(2\beta^2+\beta-1)z\alpha(1-\alpha)+z^{2-\beta}\alpha^2\right).$$

It can easily be verified by differentiation that the right-hand side is decreasing in $\alpha\in[0.5,1]$, when $2\beta^2+\beta-1\ge 0$. We can then take $\alpha=0.5$ (recall that $\alpha\ge 0.5$ by assumption), and get that

$$z^{1-\beta}a(z)\le -0.5\left(\beta z^{\beta}-(2\beta^2+\beta-1)z+z^{2-\beta}\right).$$

Dividing further by z, we seek to prove that

$$-0.5\left(\beta z^{\beta-1}-(2\beta^2+\beta-1)+z^{1-\beta}\right)\le 0,$$

or equivalently,

$$\beta z^{\beta-1}+z^{1-\beta}-2\beta^2-\beta+1\ge 0. \tag{24}$$

Define

$$h(z)=\beta z^{\beta-1}+z^{1-\beta}.$$

Then, $h(z)$ has a unique minimum at

$$z^*=\beta^{-\frac{1}{2(\beta-1)}},\ \text{ and }\ h(z^*)=2\sqrt{\beta}.$$

Thus, to prove (24), it suffices to show that

$$2\sqrt{\beta}-2\beta^2-\beta+1\ge 0.$$

Since this is a decreasing function in β in the relevant range $(0.5,1]$, which equals 0 at the upper boundary, $\beta=1$, we are done.

We conclude that $\mathcal{L}(z)\le 0$, $\forall z\ge 0$, and that the function $f_{\alpha,\beta}(w_1)$ is convex. Note that in case where $\beta<1$ (strictly), $\beta^2+\beta-2<0$, and $\beta^2-\beta<0$, so that $z^{1-\beta}a(z)<0$ (strictly), and therefore, $f_{\alpha,\beta}(w_1)$ is strictly convex.

Lastly, we prove (9) for n stations; $n=2$ is then a special case.

Specifically, let $(l_1,\ldots,l_n)$ denote the transfer target from each of the n stations. The service cost, drift and diffusion coefficient of each Station i, $i=1,\ldots,n$ are c_i, θ_i and σ_i.

The constant elasticity of substitution (CES) production function for n stations is:

$$\eta(l)=\left[\sum_{i=1}^{n}\alpha_i l_i^{\beta}\right]^{\frac{1}{\beta}},\quad 0<\alpha_i<1,\quad \beta\le 1, \tag{25}$$

where $\sum_{i=1}^n\alpha_i=1$; we use henceforth α for the probability vector $\alpha_1,\ldots,\alpha_n$.

The workload in Station i is

$$W_i(l)=\lambda\frac{l_i}{\theta_i}N(l),\ \text{and}\ W_i^r(l)=l_iN(l).$$

As before, we first break the space into “rays”:

$$\mathcal{W}_z=\left\{w\in\mathcal{W},\ \ \frac{w_n}{w_1}=z_n\right\}.$$

where $z = (1, z_2, \ldots, z_n)$.

For $w \in \mathcal{W}_z$,

$$w_1 \geq \min_{l_1 \geq 0} \; l_1 N(l_1 z);$$
$$w_n = w_1 z_n \geq z_n \min_{l_1 \geq 0} \; l_1 N(l_1 z),$$

where $l_1 z$ is the vector $l_1 z = (l_1, l_2 z_2, \ldots, l_n z_n)$.

We then have

$$w_1 = \min_{l_1 \geq 0} \; l_1 N(l_1, z l_1) = \frac{1}{\sqrt{\mathbb{h}_{\alpha,\beta}(z)}} \min_{y \geq 0} y \left[1 - e^{-\nu - y^2}\right]^{-1} = \frac{\Gamma}{\sqrt{\mathbb{h}_{\alpha,\beta}(z)}},$$

where

$$\Gamma = \min_{y \geq 0} y \left[1 - e^{-\nu - y^2}\right]^{-1}.$$

and

$$\mathbb{h}_{\alpha,\beta}(z) := \varrho \left(\alpha_1 + \sum_{i=2}^{n} \alpha_i z_i^{\beta}\right)^{\frac{1}{\beta}} \left(1 + \sum_{i=2}^{n} z_i\right).$$

The attainable workload set in terms of w_1 and the multipliers z is then given by those $w_1, z_2, \ldots, z_n \geq 0$ such that

$$\frac{\Gamma}{w_1} - \sqrt{\mathbb{h}_{\alpha,\beta}(z_2, \ldots, z_n)} \leq 0,$$

Substituting $w_n / w_1 = z_n$, we get

$$\mathcal{W} := \left\{ w \in \mathbb{R}^n_+ : \frac{\Gamma}{w_1} - \sqrt{\mathbb{h}_{\alpha,\beta}(w_2/w_1, \ldots, w_n/w_1)} \leq 0 \right\}.$$

Basic manipulation (extracting w_1) gives

$$\sqrt{\mathbb{h}_{\alpha,\beta}(w_2/w_1, \ldots, w_n/w_1)} = \frac{1}{w_1} \sqrt{\varrho \eta(w)(e \cdot w)},$$

which then gives

$$\mathcal{W} = \left\{ w \in \mathbb{R}^n_+ : \frac{\Gamma}{w_1} - \frac{1}{w_1} \sqrt{\varrho \eta(w)(e \cdot w)} \leq 0 \right\} = \left\{ w \in \mathbb{R}^n_+ : \eta(w)(e \cdot w) \geq \Gamma^2 / \varrho \right\},$$

as stated.

Finally, we prove that $\mathcal{W}$ is a convex subset of $\mathbb{R}^n_+$.

The production function $\eta(w)$ is concave in $w \in \mathbb{R}^n_+$ for any probability vector α and $\beta \leq 1$. It is, in particular, log-concave. The function $(e \cdot w)$ is trivially concave and hence log-concave. The function $\eta(w)(e \cdot w)$, being the product of two log-concave functions, is itself log-concave; it is, in particular, quasi-concave. The contour set $\mathcal{W} = \{w \in \mathbb{R}^n_+ : \eta(w)(e \cdot w) \geq \Gamma^2 / \varrho\}$ is then a convex set. Q.E.D.

Proof of Proposition 1. Direct differentiation gives

$$f'_{\alpha,\beta}(w_1) = \mathbb{h}^{-1}_{\alpha,\beta}\left(\frac{\Gamma^2}{w_1^2}\right) - \frac{2\Gamma^2}{w_1^2 \mathbb{h}'_{\alpha,\beta}\left(\mathbb{h}^{-1}_{\alpha,\beta}\left(\frac{\Gamma^2}{w_1^2}\right)\right)}.$$

As $w_1 \uparrow w_1^0$, $z(w_1) = \mathbb{h}^{-1}_{\alpha,\beta}(\Gamma^2/w_1^2) \downarrow 0$, by definition of w_1^0.

Since $w_1^0 = \Gamma/\sqrt{\mathbb{h}_{\alpha,\beta}(0)} > 0$, we have

$$\lim_{w_1 \uparrow w_1^0} f'_{\alpha,\beta}(w_1) = -\frac{2\Gamma^2}{(w_1^0)^2} \lim_{z\downarrow 0} \frac{1}{\mathbb{h}'_{\alpha,\beta}(z)}.$$

Therefore, we proceed by focusing on $\lim_{z\downarrow 0} \mathbb{h}'_{\alpha,\beta}(z)$.

Recall that,

$$\begin{aligned}
\mathbb{h}_{\alpha,\beta}(z) &:= \varrho\left(\alpha + (1-\alpha)z^\beta\right)^{\frac{1}{\beta}}(z+1);\\
\mathbb{h}'_{\alpha,\beta}(z) &= \varrho\left[\left(\alpha + (1-\alpha)z^\beta\right)^{\frac{1}{\beta}} + \left(\alpha + (1-\alpha)z^\beta\right)^{\frac{1}{\beta}-1} z^{\beta-1})(z+1)(1-\alpha)\right].
\end{aligned}$$

Therefore,

$$\lim_{z\downarrow 0} \mathbb{h}'_{\alpha,\beta}(z) = \varrho\alpha^{\frac{1}{\beta}} + \lim_{z\downarrow 0} \varrho\alpha^{\frac{1}{\beta}-1}(1-\alpha)z^{\beta-1} = \begin{cases} \infty, & \text{if } \beta < 1;\\ \varrho, & \text{if } \beta = 1,\end{cases}$$

where the first equality if because $\frac{1}{\beta} - 1 \geq 0$, and the second equality is because $z^{\beta-1} \to \infty$ when $\beta < 1$, and $z^{\beta-1} = 1$ when $\beta = 1$.

We conclude that

$$\lim_{w_1 \uparrow w_1^0} f'_{\alpha,\beta}(w_1) = \begin{cases} 0, & \text{if } \beta < 1;\\ -\frac{2\Gamma^2}{(w_1^0)^2\varrho} := -\gamma_1, & \text{if } \beta = 1.\end{cases}$$

Finally, because $w_1^0 = \Gamma/\sqrt{\mathbb{h}_{\alpha,\beta}(0)}$, $-\gamma_1 = -2\alpha$.

Next, we similarly consider the limit as $y \downarrow 0$. In this case, $z(y) \uparrow \infty$. It can be verified that

$$\frac{\mathbb{h}'_{\alpha,\beta}(z)}{z} \to 2\varrho(1-\alpha)^{\frac{1}{\beta}}, \text{ as } z \uparrow \infty.$$

Since $\mathbb{h}_{\alpha,\beta}(z(w_1)) = \Gamma^2/w_1^2$, we get that

$$f'_{\alpha,\beta}(w_1) = \mathbb{h}^{-1}_{\alpha,\beta}\left(\frac{\Gamma^2}{(w_1)^2}\right) - \frac{2}{(w_1)^2 \mathbb{h}'_{\alpha,\beta}\left(\mathbb{h}^{-1}_{\alpha,\beta}\left(\frac{\Gamma^2}{w_1^2}\right)\right)} = z(w_1) - \frac{2\mathbb{h}_{\alpha,\beta}(z(w_1))}{\mathbb{h}'_{\alpha,\beta}(z(w_1))}.$$

Let $A(z) := (\alpha + (1-\alpha)z^\beta)^{\frac{1}{\beta}}$. Then,

$$\frac{\mathbb{h}'_{\alpha,\beta}(z)}{\mathbb{h}_{\alpha,\beta}(z)} = \frac{A'(z)}{A(z)} + \frac{1}{z+1}.$$

Since

$$\begin{aligned}
\frac{A'(z)}{A(z)} &= \frac{(1-\alpha)z^{\beta-1}}{\alpha + (1-\alpha)z^\beta} = \frac{1}{z}\cdot\frac{1}{1+\frac{\alpha}{(1-\alpha)z^\beta}}\\
&= \frac{1}{z}\left(1 - \frac{\alpha}{(1-\alpha)}z^{-\beta} + o(z^{-\beta})\right),
\end{aligned}$$

we get

$$\frac{\mathbb{h}'_{\alpha,\beta}(z)}{\mathbb{h}_{\alpha,\beta}(z)} = \frac{1}{z}\left(1 - \frac{\alpha}{(1-\alpha)}z^{-\beta}\right) + \frac{1}{z+1} + o(z^{-1-\beta})$$
$$= \frac{2}{z} - \frac{1}{z^2} - \frac{\alpha}{(1-\alpha)}z^{-1-\beta} + o(z^{-1-\beta})$$
$$\frac{2}{z}(1+\epsilon(z)),$$

where $\epsilon(z) = -\frac{1}{2z} - \frac{\alpha}{2(1-\alpha)}z^{-\beta} + o(z^{-\beta})$ is such that $\epsilon(z) \downarrow 0$ as $z \uparrow \infty$. In turn,

$$\frac{2\mathbb{h}_{\alpha,\beta}(z)}{\mathbb{h}'_{\alpha,\beta}(z)} = 2\frac{z}{2}(1 - \epsilon(z) + o(\epsilon(z))$$
$$= z\left(1 + \frac{1}{2z} + \frac{\alpha}{2(1-\alpha)}z^{-\beta} + o(z^{-\beta})\right)$$
$$= z + \frac{1}{2} + \frac{\alpha}{2(1-\alpha)}z^{1-\beta} + o(z^{1-\beta}).$$

Hence

$$f'_{\alpha,\beta}(w_1) = z(w_1) - \frac{2\mathbb{h}_{\alpha,\beta}(z(w_1))}{\mathbb{h}'_{\alpha,\beta}(z(w_1))} = -\frac{1}{2} - \frac{\alpha}{2(1-\alpha)}z^{1-\beta}(w_1) + o\big(z^{1-\beta}(w_1)\big),$$

as $w_1 \downarrow 0$ (hence $z(w_1) \to \infty$).

We then have that

$$\lim_{w_1 \downarrow 0} f'_{\alpha,\beta}(w_1) = \begin{cases} -\frac{1}{2(1-\alpha)}, & \text{if } \beta = 1; \\ -\infty, & \text{if } \beta < 1, \end{cases}$$

as stated. Q.E.D.

Proof of Theorem 2. Recall our optimization problem

$$\min_{w_1, w_2 \geq 0} w_1 + \mathcal{R}^c w_2$$
$$\text{s.t. } w \in \mathcal{W}.$$

Then (see for example, Bertsekas 1997, Proposition 2.1.2), w^* is an optimal point if and only if

$$(1, \mathcal{R}^c)'(w - w^*) \geq 0, \quad \forall w \in \mathcal{W}.$$

We will prove two claims:

1. If $\beta = 1$, there exists $\mathcal{R}^c$ small enough, for which $w^* = (0, w_2^0)$. If $\beta = 1$, there also exists $\mathcal{R}^c$ large enough, for which $w^* = (w_1^0, 0)$.
2. If $\beta < 1$, the optimal solution always has $w_1, w_2 > 0$.

We start with $\beta = 1$. Take as a candidate for optimality $w^* = (0, w_2^0)$; then, for each $w \in \mathcal{W}$,

$$w_1 + \mathcal{R}^c(w_2 - w_2^0) \geq 0. \tag{26}$$

If $w_2 \geq w_2^0$, there is nothing to verify, so let us assume $w_2 < w_2^0$. Recall that

$$w_2 > f_{\alpha,\beta}(w_1) = w_1^2 \mathbb{h}^{-1}_{\alpha,\beta}\left(\frac{\Gamma^2}{w_1^2}\right).$$

Hence, (26) holds, in particular, if

$$w_1 + \mathcal{R}^c \left(f_{\alpha,\beta}(w_1) - w_2^0\right) = w_1 - \mathcal{R}^c \left(w_2^0 - f_{\alpha,\beta}(w_1)\right) \geq 0.$$

Recall that for $\beta = 1$, the derivative is in $[-\gamma_1, -\gamma_2]$, where $\gamma_1, \gamma_2 \in (0, \infty)$, hence,

$$w_1 - \mathcal{R}^c \left(w_2^0 - f_{\alpha,\beta}(w_1)\right) \geq w_1 - \mathcal{R}^c \gamma_2 w_1.$$

This expression is, in turn, positive for all $w_1 \leq w_1^0$ provided that $\mathcal{R}^c \gamma_2 \leq 1$.

For $w_1 > w_1^0$,

$$w_1 + \mathcal{R}^c \left(w_2 - w_2^0\right) \geq w_1^0 - \mathcal{R}^c w_2^0,$$

which holds as long as $\mathcal{R}^c \leq w_1^0 / w_2^0$.

Combining the above, we have that $w^* = (0, w_2^0)$ is optimal for small enough $\mathcal{R}^c$. The argument for large $\mathcal{R}^c$ is similar and omitted.

For $\beta < 1$, let us again consider $w^* = (0, w_2^0)$; we show that it cannot be optimal. Proceeding in the same way as before, notice that given w_1, $w_1 + \mathcal{R}^c(w_2 - w_2^*)$ is minimized at $f_{\alpha,\beta}(w_1)$. We will show that for any $\mathcal{R}^c > 0$, there exists w_1 small enough, such that

$$w_1 + \mathcal{R}^c \left(f_{\alpha,\beta}(w_1) - w_2^0\right) < 0.$$

Recall that $w_2^0 = f_{\alpha,\beta}(0)$ and that $f'_{\alpha,\beta}(w_1) \to -\infty$, as $w_1 \downarrow 0$ because $\beta < 1$. Thus, given $\mathcal{R}^c$ small enough, such that $f'_{\alpha,\beta}(w_1) \leq -2/\mathcal{R}^c$, then

$$w_1 + \mathcal{R}^c \left(f_{\alpha,\beta}(w_1) - w_2^0\right) \leq w_1 - \mathcal{R}^c \frac{2}{\mathcal{R}^c} w_1 = -w_1 < 0.$$

The argument is similar for $y = (w_1^0, 0)$, and is omitted. We, therefore, showed that there will never be a corner solution when $\beta < 1$.

Note that with the exception of the case where $\beta = 1$ and $\alpha = 0.5$, the optimal solution is unique. This follows from the strict convexity at the end of Theorem 1's proof. Q.E.D.

Proof of Proposition 2. Let us start from item (iii). Per Lemma 8, the function f is increasing in β for each w_1. In particular, the set $\mathcal{W}_{\beta_1} \subseteq \mathcal{W}_{\beta_2}$ if $\beta_1 < \beta_2$. In turn,

$$\bar{V}(\beta_2, \mathcal{R}^c) := \min_{w \in \mathcal{W}_{\beta_2}} (w_1 + \mathcal{R}^c w_2) \leq \min_{w \in \mathcal{W}_{\beta_1}} (w_1 + \mathcal{R}^c w_2) =: \bar{V}_{\beta_1}.$$

We turn to (i). Per Lemma 2, the optimal ratio z^* is achieve at the point where

$$-\left(z - \frac{2\mathbb{h}_{\alpha,\beta}(z)}{\mathbb{h}'_{\alpha,\beta}(z)}\right) = \frac{1}{\mathcal{R}^c}.$$

At $z = 1$ (where $w_1 = w_2$), the left hand-side evaluates to

$$-\left(z - \frac{2\mathbb{h}_{\alpha,\beta}(z)}{\mathbb{h}'_{\alpha,\beta}(z)}\right)\Big|_{z=1} = \frac{2\alpha + 1}{2(1-\alpha) + 1},$$

which does not depend on β.

At $z = 1$, $w_1 = \Gamma / \sqrt{\mathbb{h}_{\alpha,\beta}(1)} = \Gamma / \sqrt{2\varrho}$ and $l_2^* = l_1^*$. Thus, $\eta(l^*) = l_1^* = l_2^*$, so that l^* solves

$$W_1^r(l) = l_1 N(l_1, l_1) = l_1 \frac{1}{1 - e^{-2\varrho l_1^2}} = \frac{\Gamma}{\sqrt{2\varrho}},$$

which is an equation that does not depend on β.

Finally, we turn to (ii). Again, we use the fact that the optimal ratio z^* uniquely solves

$$g_\beta(z) := -\left(z - \frac{2\mathbb{h}_{\alpha,\beta}(z)}{\mathbb{h}'_{\alpha,\beta}(z)}\right) = \frac{1}{\mathcal{R}^c}.$$

First, we note that because f is convex, the right-hand side (which is $-f'$) is decreasing in z. Thus, the inverse g^{-1} is also monotone decreasing in its argument, hence monotone increasing in $\mathcal{R}^c$: the larger $\mathcal{R}^c$ is, the larger z^* is. In particular, $z^* \geq 1$ for $\mathcal{R}^c \geq \mathcal{R}^c_0$ and $z^* \leq 1$ for $\mathcal{R}^c \leq \mathcal{R}^c_0$.

All we need to show now is that z^* is increasing in β when $\mathcal{R}^c \leq \mathcal{R}^c_0$ (which by the above means $z^* \leq 1$) and decreasing for $\mathcal{R}^c \geq \mathcal{R}^c_0$.

Differentiation gives

$$\frac{\partial}{\partial\beta} g_\beta(z) = \frac{-2(1-\alpha)\alpha(z+1)^2 z^{\beta+1}\log(z)}{\left((1-\alpha)(2z+1)z^\beta - \alpha z\right)^2} = \begin{cases} > 0, & \text{if } z < 1 \\ < 0, & \text{if } z > 1, \\ = 0, & \text{if } z = 1. \end{cases}$$

We conclude, as stated, that $g_\beta(z)$ is increasing in β for $z < 1$ (for $\mathcal{R}^c \leq \mathcal{R}^c_0$) and decreasing otherwise. Q.E.D.

Proof of Theorem 3. The existence of an optimal solution w^* is guaranteed by the linearity of the objective function and the convexity of the set $\mathcal{W} \cup \{(w_1, w_2) : (\lambda/\theta_1)w_1 + (\lambda/\theta_2)w_2\}$. With the exception of the case $\beta = 1$, $\alpha = 1/2$, $\mathcal{R}^c = 1$, this solution is unique.

Notice that for $w_1 \leq w_1^0$, the optimal solution must be of the form $(w_1, f_{\alpha,\beta}(w_1))$. The total cost is $w_1 + \mathcal{R}^c f_{\alpha,\beta}(w_1)$ and has the derivative $1 + \mathcal{R}^c f'_{\alpha,\beta}(w_1)$. If $\mathcal{R}^c f'_{\alpha,\beta}(w_1) \leq -1$–equivalently $\mathcal{R}^c |f'_{\alpha,\beta}(w_1)| - 1 \geq 0$— the cost is decreasing in w_1, so that the optimal solution is $(C_1^r, f_{\alpha,\beta}(C_1^r))$. Notice that this is the optimal unconstrained solution at the cost $\widetilde{\mathcal{R}}^c$, where $|f'_{\alpha,\beta}(C_1)| = \frac{1}{\widetilde{\mathcal{R}}^c}$. The solution κ_1 to $\mathcal{R}^c \frac{c_1}{c_1+\kappa_1} = \frac{1}{|f'_{\alpha,\beta}(C_1^r)|}$ is as stated. It is now by construction that with this choice of κ_1 and κ_2, the optimal solution is $(C_1^r, f_{\alpha,\beta}(C_1^r))$. A similar argument is repeated for κ_2.

It is immediate now that κ_1, κ_2 are infinite if $f'_{\alpha,\beta}(C_1^r) = f'_{\alpha,\beta}(f^{-1}(C_2^r)) = 1/\mathcal{R}^c$. Notice that if $\kappa_1 > 0$, then $|f'_{\alpha,\beta}(C_1^r)| > 1/\mathcal{R}^c$; hence, it must be also that $|f'_{\alpha,\beta}(f^{-1}_{\alpha,\beta}(C_2^r))| > 1/os$ and, in turn, that $\kappa_2 = 0$. Indeed, if $|f'_{\alpha,\beta}(C_1^r)| > 1/\mathcal{R}^c$, then $f'_{\alpha,\beta}(C_1^r) < -1/\mathcal{R}^c$ and $(f'_{\alpha,\beta})^{-1}(1/\mathcal{R}^c) > C_1^r$. Moreover, $|f'_{\alpha,\beta}(f^{-1}_{\alpha,\beta}(C_2^r))| \leq 1/\mathcal{R}^c$ implies similarly that $(f'_{\alpha,\beta})^{-1}(1/\mathcal{R}^c) \leq f^{-1}_{\alpha,\beta}(C_2^r)$. Combining these two yields that $C_1^r < f^{-1}_{\alpha,\beta}(C_2^r)$, or equivalently—because $f_{\alpha,\beta}$ and $f^{-1}_{\alpha,\beta}$ are decreasing functions—that $f_{\alpha,\beta}(C_1^r) > C_2^r$, which means that $(C_1^r, C_2^r) \notin \mathcal{W}$ and contradicts the feasibility of this capacity levels. Q.E.D.

Proof of Theorem 4. The existence of an optimal solution follows from the linearity of the objective function and the convexity of the set $\mathcal{W}^\times$. The existence of Lagrange multipliers as stated then follows from standard necessary conditions for convex optimization with inequality constraints; see, e.g., (Bertsekas 1997, Proposition 3.1).

By definition, the optimal constrained value is the same as the optimal unconstrained value with the corrected relative cost $\widetilde{\mathcal{R}}^c_k$. If $C_1 = \infty$, as C_2 increases, the optimal solution places more (or the same, if C_2 is not binding) work in Station 2. Because the corrected cost $\widetilde{\mathcal{R}}^c_k$ either goes up for all types or down for all types, if it goes up for all types then the consumption in Station 2 goes down, contradicting the increased usage. This is argued identically for C_1. Q.E.D.

Appendix G: Proof of Lemmas

Proof of Lemma 1. We prove the lemma for the case that $\nu = 0$. The proof readily extends.

$$W(l) = \frac{\lambda}{\theta} l \left[1 - e^{-\varrho\gamma l^2}\right]^{-1} - \lambda \frac{a}{\theta}.$$

Direct differentiation gives

$$W''(l) = \frac{\lambda}{\theta} \frac{l\gamma\varrho\left(-3 + 2l^2\gamma\varrho Coth\left(l^2\gamma\varrho/2\right)\right)}{-1 + Cosh\left(l^2\varrho\gamma\right)} > 0.$$

The inequality follows because $Cosh(u)$ (the hyperbolic cosine) satisfies $Cosh(u) \geq 1$, for all $u \geq 0$ and because $Coth(u)$ (the hyperbolic cotangent of u) satisfies that $2uCoth(u/2) \geq 4$, for all $u > 0$. We, therefore, have that $W(l)$ is convex in l. The optimal solution then satisfies the first order condition, $W'(l) = 0$, which corresponds to $e^{l^2\gamma\varrho} = 1 + 2l^2\gamma\rho$, as stated.

We turn to the explicit expression of the optimal workload. Changing variable $y = \sqrt{\varrho\gamma}l$, we write

$$W^* = \min_l W(l) = \frac{\lambda}{\theta}\left(\min_{y\geq 0}\left\{\frac{y}{\sqrt{\varrho\gamma}}\left[1 - e^{-y^2}\right]^{-1}\right\} - a\right).$$

Defining $\Gamma = \min_{y\geq 0} y\left[1 - e^{-y^2}\right]^{-1}$ we have the stated result. Q.E.D.

Proof of Lemma 2. If there is an interior optimal solution, it is on the boundary where $w_2 = f_{\alpha,\beta}(w_1)$. The objective function takes the form $w_1 + \mathcal{R}^c f_{\alpha,\beta}(w_1)$ and, by the first order conditions, the optimum satisfies

$$f'_{\alpha,\beta}(w_1^*) = -\frac{1}{\mathcal{R}^c}, \tag{27}$$

which has a unique solution. This is because f is strictly convex for $w_1 \leq w_1^0$ if $\beta < 1$ or if $\beta = 1$ and $\alpha > 1/2$; see the proof of Theorem 1. The strict convexity then implies that $f'' > 0$ and, in turn, that f' is a strictly monotone increasing function.

Recall that

$$f'_{\alpha,\beta}(w_1) = z(w_1) - \frac{2\mathbb{h}_{\alpha,\beta}(z(w_1))}{\mathbb{h}'_{\alpha,\beta}(z(w_1))}.$$

Hence, $z(w_1)$ is the unique solution z

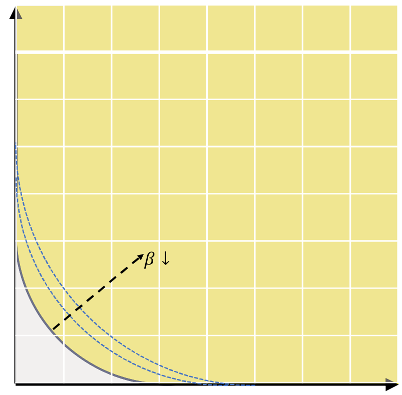
$$z - \frac{2\mathbb{h}_{\alpha,\beta}(z)}{\mathbb{h}'_{\alpha,\beta}(z)} = -\frac{1}{\mathcal{R}^c}.$$

Q.E.D.

Lemma 8, depicted in Figure 16, proves invaluable in our optimization analysis. It establishes that stronger complementarity (smaller β) shrinks the feasibility region.

LEMMA 8 **(complementarity shrinks the feasibility region)**. *The function $f_{\alpha,\beta}(\cdot)$ is decreasing in β; that is, for each $w_1 \geq 0$, $f_{\alpha,\beta_1}(w_1) \geq f_{\alpha,\beta_2}(w_1)$ if $\beta_2 \geq \beta_1$. In turn, given a convex function $g(\cdot): \mathbb{R}_+^2 \to \mathbb{R}$, the optimal value $g^* = \min_{w\in\mathcal{W}} g(w)$ is larger for smaller β (greater complementarity).*

Figure 16 **The attainable workload region shrinks as complementarity increases (β decreases).**



Proof of Lemma 8. Recall that $\mathbb{h}_{\alpha,\beta}(z)=\eta(1,z)(1+z)$, where $\eta(l_1,l_2)=(\alpha l_1^\beta+(1-\alpha)l_2^\beta)^{\frac{1}{\beta}}$. We will first show that $\eta(l_1,l_2)$ and, in turn, $\mathbb{h}_{\alpha,\beta}(z)$ are increasing in β. From direct differentiation it follows that

$$\begin{aligned}\frac{\partial}{\partial\beta}\eta(l_1,l_2)&=\frac{1}{\beta}\eta(l_1,l_2)\left[\frac{(1-\alpha)l_2^\beta\ln(l_2)+\alpha l_1^\beta\ln(l_1)}{\eta^\beta(l_1,l_2)}-\ln\left(\eta(l_1,l_2)\right)\right]\\&=\frac{1}{\beta}\eta^{1-\beta}(l_1,l_2)\left[\alpha l_1^\beta\ln(l_1)+(1-\alpha)l_2^\beta\ln(l_2)-\eta^\beta(l_1,l_2)\ln\left(\eta(l_1,l_2)\right)\right]\\&=\frac{1}{\beta^2}\eta^{1-\beta}(l_1,l_2)\left[\alpha l_1^\beta\ln(l_1^\beta)+(1-\alpha)l_2^\beta\ln(l_2^\beta)-\left[\alpha l_1^\beta+(1-\alpha)l_2^\beta\right]\ln\left(\alpha l_1^\beta+(1-\alpha)l_2^\beta\right)\right]>0,\end{aligned}$$

where the inequality comes from the convexity of the function $x\ln(x)$.

Thus, $\mathbb{h}_{\alpha,\beta}(z)=\eta(1,z)(1+z)$ is increasing in β. In particular, $\mathbb{h}_{\alpha,\beta}^{-1}(\cdot)$ is decreasing in β and we may conclude that, for each w_1, $f_{\alpha,\beta}(w_1)=w_1\mathbb{h}_{\alpha,\beta}^{-1}(\Gamma^2/w_1^2)$ is decreasing in β as stated.

Lastly, since $f_{\alpha,\beta}(w_1)$ is decreasing in β, the feasibility set is increasing in β. Minimizing any convex function $g(\cdot)$ over a smaller set, yields a larger optimal value. Thus, the minimum is larger for smaller values of β. Q.E.D.

Proof of Lemma 3. Notice that if $C_1^r<C^0$, it must be the case that $\widetilde{\mathcal{R}}^c<\mathcal{R}_0^c$ where $\mathcal{R}_0^c$ is as in Proposition 2 and does not depend on β. In turn, by item (ii) there, the ratio is decreasing in β (the optimal ratio $z^*=l_2^*/l_1^*$ is increasing). Moreover, the ratio z^* is greater than 1 in this range. Since $w_2^*(z)$ is increasing in z, and because $f_{\alpha,\beta}(w_1)$ is increasing in β, we then have that w_2^* is increasing in β. In turn, the corresponding cost must be decreasing in β. From here we can conclude that κ_1 must be increasing in β. The argument is the same for the other parts of the lemma. Q.E.D.

Proof of Lemma 6. First, let us rewrite this problem equivalently as a minimization one:

$$\begin{aligned}&\min_{\mathbf{w},\lambda} 1/\lambda\\&\text{s.t. } \frac{1}{\lambda}\geq\frac{\sum_{k\in[K]}\frac{a^k}{\theta_1^k}w_1^k}{C_1};\quad \frac{1}{\lambda}\geq\frac{\sum_{k\in[K]}\frac{a^k}{\theta_2^k}w_2^k}{C_2},\\&\mathbf{w}\in\mathcal{W}^\times.\end{aligned}$$

An optimal solution must satisfy

$$\frac{1}{\lambda}=\max\left\{\frac{\sum_{k\in[K]}\frac{a^k}{\theta_1^k}w_1^k}{C_1},\frac{\sum_{k\in[K]}\frac{a^k}{\theta_2^k}w_2^k}{C_2}\right\}=:\mathcal{G}(w). \tag{28}$$

Notice that because the maximum operations preserves convexity, $\mathcal{G}(w)$ is a convex function. Since $\mathcal{W}^{\times}$ is a convex set, w^* can be found by solving the convex problem

$$\min_{\mathbf{w}} \mathcal{G}(w)$$
$$\text{s.t. } \mathbf{w} \in \mathcal{W}^{\times}.$$

Once solved, λ^* is then given by (28). Q.E.D.

Proof of Lemma 4. Let us consider first the case that $\gamma = 0$, $y > 0$. In this case,

$$w_1 \geq \min \tau_1 N_{PC}(\tau) = \frac{1}{1+y\tau_2} \min_{\tau_1} \tau_1(1+y\tau_2)\left[1-e^{-\frac{1}{2}\tau_1^2(1+y\tau_2)^2}\right]^{-1} = \frac{\Gamma}{1+y\tau_2},$$

where $\Gamma := \min_{x\geq 0} x\left[1-e^{-\frac{1}{2}x^2}\right]^{-1}$. By construction, $W_2(\tau)/W_1(\tau) = y\tau_2$; therefore, we have the constraint $w_1 \geq \frac{\Gamma}{1+w_2/w_1}$, from which follows that $w_2 \geq (\Gamma - w_1) =: f_{\alpha,\beta}(w_1)$, as stated.

For sufficiency, suppose that $w_2, w_1 \geq 0$ satisfy the inequality $w_1 + w_2 \geq \Gamma$. We need to show that there exists τ_1, τ_2 such that $w_1 = \tau_1 N_{PC}(\tau)$ and $w_2 = y\tau_1\tau_2 N_{PC}(\tau)$. To this end, define

$$\tau_1 = \frac{w_1}{N}, \quad \tau_2 = \frac{w_2}{yw_1},$$

and let N be a solution (if one exists) to

$$N = \left[1-e^{-\frac{1}{2}(\tau_1+\theta_2(\tau_1)\tau_2)^2}\right]^{-1} = \left[1-e^{-\frac{1}{2}\left(\frac{w_1+w_2}{N}\right)^2}\right]^{-1}. \quad (29)$$

If such, a solution N that we have shown for which w_1, w_2 are indeed feasible and induced by the above choice of τ_1, τ_2. To see that (29) has indeed a solution for $w_1, w_2 \geq 0$ with $w_1 + w_2 \geq \Gamma$, let us rewrite this equation as

$$\frac{1}{N}\left[1-e^{-\frac{1}{2}\left(\frac{w_1+w_2}{N}\right)^2}\right]^{-1} = 1. \quad (30)$$

Since the function $g(x) = x(1-e^{-x^2}) \to 0$ as $x \downarrow 0$, the left-hand side of (30) diverges as $N \uparrow \infty$. This left-hand side is a continuous function of N. To show that a solution exists, we then only have to show that there exists a choice of N (for this w_1, w_2) such that the left-hand side is smaller or equal to 1.

Multiplying and dividing this left-hand side we have

$$\frac{1}{w_1+w_2}\frac{w_1+w_2}{N}\left[1-e^{-\frac{1}{2}\left(\frac{w_1+w_2}{N}\right)^2}\right]^{-1}.$$

Let

$$y^* := \arg\min_{y\geq 0} y\left[1-e^{-\frac{1}{2}y^2}\right]^{-1};$$

This y^* is unique follows the proof of Lemma 1; set $N^0 = \frac{w_1+w_2}{y^*}$. Recalling that $\Gamma := \min \arg\min_{y\geq 0} y\left[1-e^{-\frac{1}{2}y^2}\right]^{-1}$, we then have with this choice of N, that the left-hand side equals

$$\frac{\Gamma}{w_1+w_2} \leq 1,$$

where the last inequality follows from our assumption that $w_1 + w_2 \geq \Gamma$. We conclude then that at $N = N^0$,

$$\frac{1}{N}\left[1-e^{-\frac{1}{2}\left(\frac{w_1+w_2}{N}\right)^2}\right]^{-1} \leq 1.$$

The fact that $\frac{1}{N}\left[1-e^{-\frac{1}{2}\left(\frac{w_1+w_2}{N}\right)^2}\right]^{-1} \uparrow \infty$ as $N \uparrow \infty$, establishes the existence of N^* that solves (29), as required. This concludes the proof of sufficiency.

The proof for the case where $\gamma > 0$, $y = 0$ is similar and omitted. Q.E.D.

Proof of Lemma 7. We start by proving that $\mathbb{h}_{\alpha,\beta}(z)$ is increasing in z. We have

$$\mathbb{h}'_{\alpha,\beta}(z) = \left(-z^{\beta}(\alpha-1)+\alpha\right)^{\frac{1}{\beta}}\left(1+\frac{(1+z)(-1+\alpha)}{z(-1+\alpha-z^{-\beta}\alpha)}\right),$$

where the first term is positive since $0.5 \le \alpha \le 1$ implies that

$$\alpha \le -z^{\beta}(\alpha-1) \le 0.5z^{\beta}+\alpha.$$

The second term is positive since both denominator and numerator in the quotient are negative for $\alpha \ge 0.5$; specifically,

$$-z^{\beta}+\alpha-1 \le -z^{\beta}\alpha+\alpha-1 \le -0.5z^{\beta}+\alpha-1 \le 0.$$

Finally, because $\mathbb{h}_{\alpha,\beta}(z)$ is increasing in z, $\sqrt{\mathbb{h}_{\alpha,\beta}(z)}$ is also increasing in z, and $1/\sqrt{\mathbb{h}_{\alpha,\beta}(z)}$ is decreasing in z. Q.E.D.

References

Bertsekas, D.P. 1997. *Nonlinear Programming*. Taylor & Francis.

Bhattacharya, R.N., E.C. Waymire. 2009. *Stochastic Processes with Applications*. SIAM.

Committee, American Wind Energy Association (AWEA) Operations & Maintenance. 2017. Chapter 8: Condition-based maintenance. Operations & Maintenance Recommended Practices RP 801 et seq., American Wind Energy Association. URL https://cleanpower.org/wp-content/uploads/2024/06/AWEA-OM-RP-Chapter-8_Condition-Based-Maintenance.pdf. Accessed November 2025.

Facts, Wind Energy – The. 2024. Commissioning, operation and maintenance for wind turbines. https://wind-energy-the-facts.org/commissioning-operation-and-maintenance.html. Accessed November 2025.

Hashimoto, E.M., E.M.M. Ortega, G.M. Cordeiro, V.G. Cancho, I. Silva. 2023. The re-parameterized inverse gaussian regression to model length of stay of COVID-19 patients in the public health care system of Piracicaba, Brazil. *Journal of Applied Statistics* **50**(8) 1665–1685.

U.S. Bureau of Labor Statistics. 2024. Occupational employment and wage statistics (oews), 2024 tables for soc 51-9061 and 49-9041. https://www.bls.gov/oes/. Accessed November 2025.

Whitmore, G.A. 1975. The inverse gaussian distribution as a model of hospital stay. *Health Services Research* **10**(3) 297.

Whitmore, G.A. 1979. An inverse gaussian model for labour turnover. *Journal of the Royal Statistical Society: Series A (General)* **142**(4) 468–478.