Online Appendix to
Incentives for Quality through Endogenous Routing

Lauren Xiaoyuan Lu* Jan A. Van Mieghem† R. Canan Savaskan‡

I. Proofs

PROOF OF LEMMA 1. Notice that when $t \geq \frac{v}{a}$, $V < 0$. Therefore, we maximize a continuous function over a compact set: $[\underline{t}, \frac{v}{a}] \times [0, 1]$, implying an optimum exists. $C'(1) = \infty$ implies $p^{FB} < 1$. Since the IR constraint must bind, substitute $w = \lambda[a(t + p\bar{F}(t)r) - b]$ into the objective function and the Kuhn-Tucker conditions are

$$\begin{aligned} \frac{\partial V(t^{FB}, p^{FB})}{\partial t} &= \lambda[-a + f(t^{FB})(p^{FB}(ar + c_I) + (1 - p^{FB})c_E)] \leq 0, \\ \frac{\partial V(t^{FB}, p^{FB})}{\partial p} &= \lambda[\bar{F}(t^{FB})(c_E - c_I - ar) - C'(p^{FB})] \leq 0. \end{aligned}$$

Because $c_E - c_I > ar$ gives $\frac{\partial V(t,0)}{\partial p} > 0$, $C'(0) = 0$ implies $p^{FB} > 0$. Now we have shown $p^{FB} \in (0, 1)$, which implies $(p^{FB}(ar + c_I) + (1 - p^{FB})c_E) \in (ar + c_I, c_E)$. Applying the condition $\frac{a}{f(\underline{t})} < ar + c_I$, we get $\frac{\partial V(\underline{t},p)}{\partial t} > 0$, which implies $t > \underline{t}$. ■

PROOF OF LEMMA 2, 3, 4. We will prove Lemma 2. Proofs of Lemma 3 and 4 are similar and thus omitted. The Kuhn-Tucker condition is $-1 + prf(t) \leq 0$ with equality at interior solutions. The boundary conditions follow from the fact that f is strictly monotone decreasing. The uniqueness follows from the second-order condition (SOC) $\lambda aprF''(t) < 0$. ■

PROOF OF PROPOSITION 1. From Equation (1) and (3), we get $f(t^S) > f(t^{FB})$ and thus $t^S < t^{FB}$. Therefore self routing cannot achieve the first best. For dedicated and cross routing, set $p = p^{FB}$ and $b^* = \frac{a}{f(t^{FB})p}$ to achieve t^{FB}. The rest follows from the fact that the IR constraints are satisfied at equality. ■

PROOF OF COROLLARY 1. Proposition 1 and Equation (1) together imply

$$f(t^{FB}(p)) = f(t^D(p)) = f(t^C(p)) = \frac{1}{pr + \frac{1}{a}(pc_I + (1 - p)c_E)}.$$

Because we have implicitly assumed the conditions in Proposition 1, i.e., $c_E > c_I + ar > \frac{a}{f(\underline{t})}$ to

*Kenan-Flagler Business School, University of North Carolina, Chapel Hill, NC 27599, lauren_lu@unc.edu.

†Kellogg School of Management, Northwestern University, Evanston, IL 60208, VanMieghem@northwestern.edu.

‡Kellogg School of Management, Northwestern University, Evanston, IL 60208, r-savaskan@kellogg.northwestern.edu.

ensure the existence of an interior first-best solution, we must have

$$f(\underline{t}) > \frac{1}{pr + \frac{1}{a}(pc_I + (1-p)c_E)} \iff p < \frac{c_E - \frac{a}{f(\underline{t})}}{c_E - c_I - ar}.$$

The second inequality is always satisfied because $\frac{c_E - \frac{a}{f(\underline{t})}}{c_E - c_I - ar} > 1$. Hence, $t^{FB}(p) = t^D(p) = t^C(p) > \underline{t}$. In addition, $f(t^S(p)) = \frac{1}{pr} > f(t^{FB}(p))$ when $t^S(p)$ is interior. Therefore, $t^S(p) < t^{FB}(p)$. The rest follows from the fact that Q is strictly increasing in t at any p. ■

PROOF OF COROLLARY 2. Because $t^S(p) = \arg\min_{t' \geq \underline{t}} \{t' + p\bar{F}(t')r\} < t^{FB}(p)$, $w^{FB}(p) = \lambda[a(t^{FB}(p) + p\bar{F}(t^{FB}(p))r) - b^{FB}(p)] = w^S(p) = \lambda[a(t^S(p) + p\bar{F}(t^S(p))r) - b^S(p)]$ implies $b^{FB}(p) > b^S(p)$. The rest follows from the fact that $b^{FB}(p) = b^D(p) = b^C(p)$. ■

PROOF OF LEMMA 5. Notice that when $t \geq \frac{v}{a}$, $V < 0$. Therefore, we maximize a continuous function over a compact set: $[\underline{t}, \frac{v}{a}] \times [0,1]$, implying an optimum exists. Assuming an interior optimum exists, the optimal solution $\{t^{FB}, p^{FB}\}$ is then given by the first-order conditions in the lemma. ■

PROOF OF LEMMA 6. Evaluate the second derivative of $U(t)$ at any interior critical point t^S using $f(t^S) = 1/pr$:

$$U''(t^S) = \frac{bprF''(t^S)}{(t^S + p\bar{F}(t^S)r)^2} + \frac{2b(1 - prf(t^S))}{(t^S + p\bar{F}(t^S)r)^3} = \frac{bprF''(t^S)}{(t^S + p\bar{F}(t^S)r)^2} < 0.$$

Because $U(t)$ is strictly concave at any interior critical point, $U(t)$ is strictly pseudoconcave (Avriel, Diewert, Schaible & Zang (1988)) and thus t^S is a unique global maximum. The boundary conditions follow from the fact that f is strictly monotone decreasing. ■

PROOF OF LEMMA 7. Evaluate the second derivative of $U_1(t)$ at any interior critical point t^D using $f\left(t^D\right) = \frac{1 - p\bar{F}(t^D)}{pt^D}$:

$$U_1''(t^D) = b\left[\frac{pF''(t^D)}{t^D} - \frac{2}{(t^D)^3}(pt^D f(t^D) + p\bar{F}(t^D) - 1)\right] = \frac{bpF''(t^D)}{t^D} < 0.$$

Because $U_1(t)$ is strictly concave at any interior critical point, $U_1(t)$ is strictly pseudoconcave (Avriel et al. (1988)) and thus t^D is a unique global maximum. The boundary condition follows from the fact that $tf(t) + \bar{F}(t)$ is strictly monotone decreasing in t and $\bar{F}(\underline{t}) = 1$. ■

PROOF OF LEMMA 8. We first assume an interior solution exists for agent i's problem and derive the equation that determines an interior symmetric Nash equilibrium. We then prove that the equation must have an interior solution under the existence condition. The first-order condition (FOC) for agent i's problem is

$$\frac{\partial U_i(t_i, t_j)}{\partial t_i} = \frac{b(1-\rho_j)}{(1-\rho_i\rho_j)^2}\frac{1}{t_i^2}\left[p(t_i f(t_i) + \bar{F}(t_i))\left(1 + \rho_j\left(1 - \frac{r}{t_i}\right)\right) + \rho_i\rho_j - 1\right] = 0.$$

Let $\hat{t}_i$ be the critical point satisfying the FOC. The second derivative evaluated at $\hat{t}_i$ is

$$\begin{aligned}\frac{\partial^2 U_i(\hat{t}_i, t_j)}{\partial t_i^2} &= \frac{b(1-\rho_j)}{(1-\rho_i\rho_j(\hat{t}_i))^2}\frac{1}{\hat{t}_i^2}\left[p\hat{t}_i F''(\hat{t}_i)\left(1+\rho_j\left(1-\frac{r}{\hat{t}_i}\right)\right) + p(\hat{t}_i f(\hat{t}_i)+\bar{F}(\hat{t}_i))\frac{r}{\hat{t}_i^2}\rho_j + \rho_j\frac{\partial\rho_i(\hat{t}_i)}{\partial t_i}\right] \\ &\quad + b(1-\rho_j)\frac{\partial\left[\frac{1}{(1-\rho_i\rho_j)^2}\frac{1}{t_i^2}\right]}{\partial t_i}\left[p(\hat{t}_i f(\hat{t}_i)+\bar{F}(\hat{t}_i))\left(1+\rho_j\left(1-\frac{r}{\hat{t}_i}\right)\right)+\rho_i(\hat{t}_i)\rho_j - 1\right].\end{aligned}$$

Substituting $\frac{\partial\rho_i(\hat{t}_i)}{\partial t_i} = -\frac{pr}{\hat{t}_i^2}[\hat{t}_i f(\hat{t}_i)+\bar{F}(\hat{t}_i)]$ and the FOC into the second derivative gives

$$\frac{\partial^2 U_i(\hat{t}_i, t_j)}{\partial t_i^2} = \frac{b(1-\rho_j)}{(1-\rho_i\rho_j(\hat{t}_i))^2}\frac{1}{\hat{t}_i^2}\left[p\hat{t}_i F''(\hat{t}_i)\left(1+\rho_j\left(1-\frac{r}{\hat{t}_i}\right)\right)\right] < 0.$$

The inequality follows from the fact that $F''(\cdot) < 0$ and that $1+\rho_j\left(1-\frac{r}{\hat{t}_i}\right) > 0$. Because it is strictly concave at any interior critical point, $U_i(t_i, t_j)$ is strictly pseudoconcave in t_i (Avriel et al. (1988)), implying that $\hat{t}_i$ is a unique global maximum. Therefore, the equation that determines a symmetric Nash equilibrium (t^C, t^C) is $g(t^C) = 0$, where $g(t^C)$ is the left hand side of Equation (6).

It remains to show that Equation (6) has an interior solution. When $p > \bar{p}$,

$$g(\underline{t}) > \rho(\underline{t})\left(1-\frac{r}{\underline{t}}\right) + 1 + \rho(\underline{t})^2 - 1 > 0.$$

Because $f = F'$, it is integrable on $[\underline{t}, \infty)$ and $\lim_{t\to\infty} f(t) = 0$. Since $1/t$ is not integrable on $[\underline{t}, \infty)$, $f(t) = o(1/t)$ as $t \to \infty$. Thus, $\lim_{t\to\infty} tf(t) = 0$. Therefore,

$$\begin{aligned}\lim_{t\to\infty} g(t) &= p\lim_{t\to\infty}[tf(t)+\bar{F}(t)] \times \lim_{t\to\infty}\left[\rho(t)\left(1-\frac{r}{t}\right)+1\right] + \lim_{t\to\infty}\rho(t)^2 - 1 \\ &= p\lim_{t\to\infty} tf(t) - 1 < 0.\end{aligned}$$

Hence, there exists a $\bar{t}$ such that $g(\bar{t}) < 0$. Because $g(\cdot)$ is continuous and $g(\underline{t}) > 0$ and $g(\bar{t}) < 0$, applying the Intermediate Value Theorem implies that there exists a $t^C \in (\underline{t}, \bar{t})$ such that $g(t^C) = 0$. ■

PROOF OF PROPOSITION 2. From Equation (4), we get $f(t^S) > f(t^{FB})$ and thus $t^S(p) < t^{FB}(p)$. Therefore self routing cannot achieve the first best. Notice that the agents' optimal effort under dedicated and cross routing only depends on p and the corresponding FOCs are different from the FOC of t^{FB}. Obviously, setting $p = p^{FB}$ in the FOCs of dedicated and cross routing do not give the same solution as t^{FB} (except in an extremely special case where the different FOCs happen to have same the solution, which is a trivial case that we do not consider here). Hence, the first best solution cannot be implemented by any of the routing schemes. ■

PROOF OF COROLLARY 3. To show $t^S(p) < t^D(p)$, substituting $t^S(p)$ into the FOC of $t^D(p)$ yields that

$$\frac{w}{t^S(p)}[pt^S(p)f(t^S(p)) + p\bar{F}(t^S(p)) - 1] = \frac{w}{t^S(p)}[\frac{t^S(p)}{r} + p\bar{F}(t^S(p) - 1] > 0.$$

We show $t^D(p) < t^C(p)$ by contradiction. Suppose $t^D(p) \geq t^C(p)$ and it follows from the FOC of $t^D(p)$ that

$$pt^C(p)f(t^C(p)) + p\bar{F}(t^C(p)) - 1 \geq 0.$$

Then,

$$\begin{aligned} & p[t^C(p)f(t^C(p)) + \bar{F}(t^C(p))][1 + (1 - \frac{r}{t^C(p)})\rho(t^C(p))] + \rho(t^C(p))^2 - 1 \\ \geq \; & 1 + (1 - \frac{r}{t^C(p)})\rho(t^C(p)) + \rho(t^C(p))^2 - 1 \\ = \; & \frac{\rho(t^C(p))}{t^C(p)}(t^C(p) + p\bar{F}(t^C(p)r - r) > 0, \end{aligned}$$

contradicting the FOC of $t^C(p)$. The rest follows from the fact that Q is strictly increasing in t at any p. ■

PROOF OF COROLLARY 4. Because $t^S = \arg\min_{t' \geq t} \{t' + p\bar{F}(t')r\}$, $w^C(p) = a - \frac{b^C(p)}{t^C(p) + p\bar{F}(t^C(p))r} = w^S(p) = a - \frac{b^S(p)}{t^S(p) + p\bar{F}(t^S(p))r}$ implies that $b^C(p) > b^S(p)$. Because $w^D(p) = \frac{a[t^D(p) + p\bar{F}(t^D(p))r] - b^D(p)}{2t^D(p)}$, then zero wage rate implies $b^D(p) = a[t^D(p) + p\bar{F}(t^D(p))r]$. The inequality follows from Lemma 3. ■

PROOF OF PROPOSITION 3. First we compare the principal's profit rate at a given p,

$$\begin{aligned} V^C(p) - V^S(p) \; = \; & \frac{1}{[t^C(p) + p\bar{F}(t^C(p))r][t^S(p) + p\bar{F}(t^S(p))r]} \\ & \cdot \Big\{ (v - C(p)) \left(t^C(p) - t^S(p)\right) \left(pr\frac{F(t^C(p)) - F(t^S(p))}{t^C(p) - t^S(p)} - 1 \right) \\ & + [pc_I + (1-p)c_E] \, [t^C \bar{F}(t^S(p)) - t^S \bar{F}(t^C(p))] \Big\}. \end{aligned}$$

(i) Since $t^C(p) > t^S(p)$, it follows that $\frac{F(t^C(p)) - F(t^S(p))}{t^C(p) - t^S(p)} < f(t^S(p)) = \frac{1}{pr}$ and $t^C\bar{F}(t^S(p)) - t^S\bar{F}(t^C(p)) > 0$. Therefore, $V^C(p) - V^S(p) > 0$ if c_I, c_E, and C'' are sufficiently large. Thus under the conditions, $V^C = V^C(p^C) \geq V^C(p^S) > V^S(p^S) = V^S$. The first inequality follows from the optimality of V^C. (ii) Similarly, $V^C(p) - V^S(p) < 0$ if v is sufficiently large. Thus under the condition, $V^C = V^C(p^C) < V^S(p^C) \leq V^S(p^S) = V^S$. The second inequality follows from the optimality of V^S. Comparing V^D with V^S is similar and thus omitted. Now we compare V^D with V^C. Because $2t^D(p) > t^D + p\bar{F}(t^D)r$,

$$V^C(p) - V^D(p) > \frac{v + \bar{F}(t^C(p))(pc_I + (1-p)c_E) - C(p)}{t + p\bar{F}(t^C(p))r} - \frac{v + \bar{F}(t^D(p))(pc_I + (1-p)c_E) - C(p)}{t + p\bar{F}(t^D(p))r}.$$

The rest is similar to comparing V^S with V^C.■

PROOF OF LEMMA 9.

$$\frac{\partial U_i(t_i,t_{-i})}{\partial t_i} = \frac{b(1-\rho_i)}{\left[1+(1-\frac{\rho_i}{N-1})\sum_{j\neq i}\frac{\rho_j}{1-\rho_j}\right]^2}\frac{1}{t_i^2}\left\{p[t_if(t_i)+\bar{F}(t_i)]\left[1+\frac{Nt_i-r}{(N-1)t_i}\sum_{j\neq i}\frac{\rho_j}{N-1-\rho_j}\right]\right.$$
$$\left. -1-\left(1-\frac{\rho_i}{N-1}\right)\sum_{j\neq i}\frac{\rho_j}{N-1-\rho_j}\right\}$$

Let $\hat{t}_i$ be the critical point satisfying the FOC. To simplify notation, let $\hat{\rho}_i = \rho_i(\hat{t}_i)$. The second derivative evaluated at $\hat{t}_i$ is

$$\frac{\partial^2 U_i(\hat{t}_i,t_{-i})}{\partial t_i^2} = \frac{b}{\left[1+\left(1-\frac{\hat{\rho}_i}{N-1}\right)\sum_{j\neq i}\frac{\rho_j}{N-1-\rho_j}\right]^2}\frac{1}{\hat{t}_i^2}\left\{p\hat{t}_iF''(\hat{t}_i)\left[1+\frac{Nt_i-r}{(N-1)t_i}\sum_{j\neq i}\frac{\rho_j}{N-1-\rho_j}\right]\right.$$
$$\left.+p[\hat{t}_if(\hat{t}_i)+\bar{F}(\hat{t}_i)]\frac{1}{N-1}\frac{r}{\hat{t}_i^2}\sum_{j\neq i}\frac{\rho_j}{N-1-\rho_j}+\frac{1}{N-1}\frac{\partial\hat{\rho}_i}{\partial t_i}\sum_{j\neq i}\frac{\rho_j}{N-1-\rho_j}\right\}$$
$$+b\frac{\partial\left(\frac{1}{\left[1+\left(1-\frac{\hat{\rho}_i}{N-1}\right)\sum_{j\neq i}\frac{\rho_j}{N-1-\rho_j}\right]^2}\frac{1}{\hat{t}_i^2}\right)}{\partial t_i}\left\{[p(\hat{t}_if(\hat{t}_i)+\bar{F}(\hat{t}_i)]\left[1+\frac{Nt_i-r}{(N-1)t_i}\sum_{j\neq i}\frac{\rho_j}{N-1-\rho_j}\right]\right.$$
$$\left.-1-\left(1-\frac{\hat{\rho}_i}{N-1}\right)\sum_{j\neq i}\frac{\rho_j}{N-1-\rho_j}\right\}$$
$$= \frac{b}{\left[1+(1-\frac{\hat{\rho}_i}{N-1})\sum_{j\neq i}\frac{\rho_j}{1-\rho_j}\right]^2}\frac{1}{\hat{t}_i^2}p\hat{t}_iF''(\hat{t}_i)\left[1+\frac{Nt_i-r}{(N-1)t_i}\sum_{j\neq i}\frac{\rho_j}{N-1-\rho_j}\right] < 0.$$

The last equality follows from the FOC above and that $\frac{\partial\hat{\rho}_i}{\partial t_i} = -\frac{r}{\hat{t}_i^2}p[\hat{t}_if(\hat{t}_i)+\bar{F}(\hat{t}_i)]$. The inequality follows from the fact that $F''(\cdot)<0$ and that $\rho_j<1$. Because it is strictly concave at any interior critical point, $U_i(t_i,t_j)$ is strictly pseudoconcave in t_i (Avriel et al. (1988)), implying that $\hat{t}_i$ is a unique global maximum. Assuming a symmetric equilibrium gives Equation (7). The proof for the existence condition is similar to that of Lemma 8 and thus omitted. ■

PROOF OF LEMMA 10. If $t^D(p)=\underline{t}$, we are done because $t_N^C(p)$ is interior. Now consider interior $t^D(p)$. Suppose to the contrary $t^D(p)\geq t_N^C(p)$ and it follows from the FOC of $t^D(p)$ that $p[t_N^Cf(t_N^C)+\bar{F}(t_N^C)]\geq 1$ (for simplicity, we use t_N^C to denote $t_N^C(p)$.) Let $g_N(t)$ denote the FOC of the symmetric equilibrium of the N-agent system.

$$g_N(t_N^C)\geq\frac{1}{N-1}\frac{pr\bar{F}(t_N^C)}{t_N^C}\left[1+\frac{pr\bar{F}(t_N^C)}{t_N^C}-\frac{r}{t_N^C}\right]>0,$$

contradicting the optimality condition of t_N^C. ■

PROOF OF LEMMA 11. To show $t_N^C > t_{N+1}^C$, it suffices to show g_N strictly decreases in both N and t. To simplify notation, we treat N as a real number. Taking the derivative w.r.t. N

$$\frac{\partial g_N(t)}{\partial N} = \frac{1}{(N-1)^2}\frac{pr\bar{F}(t)(prf(t)-1)}{t} < 0$$

for any $t > t^D$ because $f(t) < f(t^D) \leq f(t^S) = 1/pr$. Lemma 10 says that $t^D < t_N^C$ for all $N \geq 2$. Therefore, for the set of equilibrium solutions, g_N strictly decreases in N. Hence $g_N(t_{N+1}^C) > g_{N+1}(t_{N+1}^C) = g_N(t_N^C) = 0$. It remains to show that g_N strictly decreases in t.

$$\begin{aligned}
\frac{\partial g_N(t)}{\partial t} &= p^2rF''(t)\bar{F}(t)\left(1-\frac{r}{(N-1)t}\right) - \frac{p^2[tf(t)+\bar{F}(t)]^2r}{t^2} \\
&\quad + ptF''(t) + \frac{pr[tf(t)+\bar{F}(t)]}{t^2}\frac{prf(t)+N-2}{N-1} \\
&< p^2rF''(t)\bar{F}(t)\left(1-\frac{r}{(N-1)t}\right) - \frac{p^2[tf(t)+\bar{F}(t)]^2r}{t^2} + p\left[tF''(t)+\frac{tf(t)+\bar{F}(t)}{t}\right]
\end{aligned}$$

for any $t > t^D$ (because $f(t) < 1/pr$ and $r \leq t$). Since the first two terms of the RHS are negative, the third term being negative is a sufficient condition for $\frac{\partial g_N(t)}{\partial t} < 0$, which is equivalent to $-\frac{F''(t)}{f(t)} \geq \frac{1}{t} + \frac{\bar{F}(t)}{t^2 f(t)}$. Now let us invoke the DFR assumption and from the definition of DFR we have $-\frac{F''(t)}{f(t)} \geq \frac{f(t)}{\bar{F}(t)}$. If $\frac{f(t)}{\bar{F}(t)} \geq \frac{1}{t} + \frac{\bar{F}(t)}{t^2 f(t)}$, equivalently, if $\frac{tf(t)}{\bar{F}(t)}\left(\frac{tf(t)}{\bar{F}(t)}-1\right) \geq 0$, the third term of $\frac{\partial g_N(t)}{\partial t}$ will be negative. Satisfying the condition calls for the IGFR property and $\underline{t}f(\underline{t}) \geq 1$ so that $\frac{tf(t)}{\bar{F}(t)} \geq \frac{\underline{t}f(\underline{t})}{\bar{F}(\underline{t})} \geq 1$. ■

PROOF OF PROPOSITION 4. Let t_i^* denote the optimal effort when all other agents choose t^D. It suffices to show $U_i(t_i^*, t_{-i}^D) - U_i(t_i^D, t_{-i}^D) \leq \varepsilon$. Claim. $t_i^* \geq t^D$. To show this, substitute t^D into the first derivative $\frac{\partial U_i(t_i, t_{-i}^D)}{\partial t_i}$. Because $p[t^D f(t^D) + \bar{F}(t^D)] = 1$ (Otherwise $t^D = \underline{t}$, we are done.), $\frac{\partial U_i(t_i^D, t_{-i}^D)}{\partial t_i} = g_N(t^D) = \frac{1}{N-1}\frac{pr\bar{F}(t^D)}{t^D}\left[1+\frac{pr\bar{F}(t^D)}{t^D}-\frac{r}{t^D}\right] > 0$. Because the agent's problem is strictly pseudoconcave as shown in Lemma 9, $t_i^* \geq t^D$. Now

$$\begin{aligned}
& U_i(t_i^*, t_{-i}^D) - U_i(t_i^D, t_{-i}^D) \\
= \; & \frac{b}{1+\frac{1-\frac{pr}{N-1}\frac{\bar{F}(t_i^*)}{t_i^*}}{1-\frac{pr}{N-1}\frac{\bar{F}(t^D)}{t^D}}\frac{pr\bar{F}(t^D)}{t^D}}\left[\frac{1-p\bar{F}(t_i^*)}{t_i^*} + \frac{1-\frac{pr}{N-1}\frac{\bar{F}(t_i^*)}{t_i^*}}{1-\frac{pr}{N-1}\frac{\bar{F}(t^D)}{t^D}}\frac{p\bar{F}(t^D)}{t^D}\right] - \frac{b}{t^D+pr\bar{F}(t^D)} \\
\leq \; & \frac{b}{1+\frac{p\bar{F}(t^D)}{t^D}}\left[\frac{1-p\bar{F}(t^D)}{t^D} + \frac{1}{1-\frac{pr}{N-1}\frac{\bar{F}(t^D)}{t^D}}\frac{p\bar{F}(t^D)}{t^D}\right] - \frac{b}{t^D+pr\bar{F}(t^D)}
\end{aligned}$$

because $1-\frac{pr}{N-1}\frac{\bar{F}(t_i^*)}{t_i^*} \geq 1-\frac{pr}{N-1}\frac{\bar{F}(t^D)}{t^D}$ and $\frac{1-p\bar{F}(t_i^*)}{t_i^*} \leq \frac{1-p\bar{F}(t^D)}{t^D}$. Choose N_1 large enough s.t.

$\frac{pr}{N_1-1}\frac{\bar{F}(t^D)}{t^D} \leq \frac{1}{2}$. Then $\frac{1}{1-\frac{pr}{N-1}\frac{\bar{F}(t^D)}{t^D}} \leq 1 + 2\frac{pr}{N-1}\frac{\bar{F}(t^D)}{t^D}$. It follows that

$$\begin{aligned} LHS &\leq \frac{b}{1+\frac{p\bar{F}(t^D)}{t^D}}\left[\frac{1-p\bar{F}(t^D)}{t^D} + \frac{p\bar{F}(t^D)}{t^D} + 2\frac{pr}{N-1}\frac{\bar{F}(t^D)}{t^D}\frac{p\bar{F}(t^D)}{t^D}\right] - \frac{b}{t^D + pr\bar{F}(t^D)} \\ &= \frac{1}{N-1}\frac{2rb\left(\frac{p\bar{F}(t^D)}{t^D}\right)^2}{1+\frac{p\bar{F}(t^D)}{t^D}}. \end{aligned}$$

Now choose N_2 large enough s.t. LHS $\leq \varepsilon$. Let $N_\varepsilon = \max(N_1, N_2)$. ■

II. Alternative Incentive Schemes: Penalty and Bonus

Assigning rework to a different agent implicitly punishes the agent for quality failure. In dedicated and cross routing, the agents are punished because they cannot recoup the cost of effort spent on a job that fails quality inspection. Such punishment could be replicated by a modified self routing scheme where the principal executes a monetary punishment whenever a defect is identified. Consider the case of limited demand. Suppose the principal specifies a penalty x for each defect identified, the agents' problem becomes

$$\max_{t \geq \underline{t}} \lambda[b - a(t + p\bar{F}(t)r) - p\bar{F}(t)x].$$

The first-order condition is equivalent to $f(t) = \frac{1}{pr+\frac{1}{a}px}$. Recalling Equation (1), we set

$$x = c_I + \frac{1-p^{FB}}{p^{FB}}c_E$$

to allow the principal to achieve the first-best effort level. Similarly, we can derive the penalty for the case of unlimited demand

$$x = \frac{b\left(c_I + \frac{1-p^{FB}}{p^{FB}}c_E\right)}{\left[t^{FB} + p^{FB}\bar{F}(t^{FB})r\right]A(t^{FB}, p^{FB}) + \bar{F}(t^{FB})\left[p^{FB}c_I + (1-p^{FB})c_E\right]},$$

where $A(t^{FB}, p^{FB})$ is defined as in Lemma 5.

If instead we suppose the principal specifies a bonus y for each first-pass success, we derive the bonus that induces the first-best outcome under limited demand:

$$y = c_I + \frac{1-p^{FB}}{p^{FB}}c_E,$$

and under unlimited demand:

$$y = \frac{b\left(c_I + \frac{1-p^{FB}}{p^{FB}}c_E\right)}{\left[t^{FB} + p^{FB}\bar{F}(t^{FB})r\right]A(t^{FB}, p^{FB}) - \left[\frac{1}{p^{FB}} - \bar{F}(t^{FB})\right]\left[p^{FB}c_I + (1-p^{FB})c_E\right]}.$$