ONLINE APPENDIX

Hyperspecialization and Hyperscaling: A Resource-based Theory of the Digital Firm

(Giustiziero, Kretschmer, Somaya, and Wu, 2021)

Proof of Proposition 1

Given the constraint $\tau \in [\tau_I, 1]$, the Lagrangian of firm *i*'s objective function is

$$VQ_{ia}(\tau r) - \alpha V(Q_{ia}(\tau r) - Q_{ib}((1-\tau)r)) + \lambda_1(\tau - \tau_I) + \lambda_2(1-\tau), \tag{A1}$$

with first order condition

$$\frac{\partial L(\tau,\lambda_1,\lambda_2)}{\partial \tau} = \varsigma(\tau r)r - \alpha(\varsigma(\tau r)r + \varphi((1-\tau)r)r) + \lambda_1 - \lambda_2 = 0, \tag{A2}$$

where $\varsigma(r) = \frac{\partial VQ_{ia}}{\partial r}$ and $\varphi(r) = \frac{\partial VQ_{ib}}{\partial r}$ and complementary slackness conditions:

$$\lambda_1(\tau - \tau_I) = 0, \tag{A3}$$

$$\lambda_2(1-\tau) = 0. \tag{A4}$$

The firm integrates if $\lambda_1 > 0$ and $\lambda_2 = 0$. The first order condition and the complementary slackness conditions imply $\lambda_1 = \alpha(\varsigma(\tau_I r)r + \varphi((1-\tau_I)r)r) - \varsigma(\tau_I r)r > 0$, which is true if $\alpha > \frac{\varsigma(\tau_I r)}{\varsigma(\tau_I r)+\varphi((1-\tau_I)r)} \geq \frac{1}{2}$. Because $Q_{ia} = Q_{ib}$ implies $\varsigma(1)(\tau_I r)^\sigma = \varphi(1)\big((1-\tau_I)r\big)^\sigma \rightarrow \varsigma(\tau_I r)\tau_I r = \varphi\big((1-\tau_I)r\big)(1-\tau_I)r$. Solving for τ_I, we derive $\tau_I = \frac{\varphi((1-\tau_I)r)}{\varsigma(\tau_I r)+\varphi((1-\tau_I)r)}$, so that the critical α can be rewritten as:

$$\alpha > 1 - \tau_I \geq \frac{1}{2}. \tag{A5}$$

For all feasible directions ϵ such that $\tau_I + \epsilon > \tau_I$, the product $\pi'(\tau_I, r)\epsilon$ is negative because (A2), $\lambda_1 > 0$, and $\lambda_2 = 0$ imply $\pi'(\tau_I, r)$ is negative while ϵ is positive by definition. This ensures that $\tau = \tau_I$, $\lambda_1 = \alpha V(\varsigma(\tau_I r)r + \varphi\big((1-\tau_I)r\big) - V\varsigma(\tau_I r)r$, and $\lambda_2 = 0$ is a local maximum because $\pi(\tau, r)$ cannot increase in the proximity of the constraint. When $\sigma < 1$, because π is strictly concave down, the point identifies a global maximum.

When $\sigma > 1, \tau = \tau_I, \lambda_1 = \alpha(\varsigma(\tau_I r)r + \varphi((1-\tau_I)r)r) - \varsigma(\tau_I r)r$, and $\lambda_2 = 0$ identifies a global maximum if:

$$\alpha > 1 - {\tau_I}^\sigma \geq \frac{1}{2}. \tag{A6}$$

The condition in (A6) is derived by comparing profits at the endpoints $\tau = \tau_I$ and $\tau = \tau_S = 1$.

Partial integration corresponds to an interior solution $\tau_C \in (\tau_I, 1)$, requiring $\lambda_1 = 0$ and $\lambda_2 = 0$. The first order condition implies:

$$(1-\alpha)\varsigma(\tau_C r) = \alpha\varphi((1-\tau_C)r) = \eta. \tag{A7}$$

Using Euler's homogenous function theorem and (A7), the second order condition for a maximum is $\tau_C^{-1}(\sigma-1)\,\eta + (1-\tau_C)^{-1}(\sigma-1)\eta < 0$, which holds only if $\sigma < 1$. When $\sigma < 1$, π is strictly concave down and $\tau = \tau_C, \lambda_1 = 0,$ and $\lambda_2 = 0$ identifies a global maximum when $\alpha \leq 1-\tau_I$. Because the characterization $Q_{ia}(\tau_I r) = Q_{ib}((1-\tau_I)\,r)$ implies that $1 > \tau_I > 0$, we have that $1 > 1-\tau_I > 0$.

From the above arguments, it follows $\sigma > 1$ never leads to an interior solution. If $\alpha \leq 1-{\tau_I}^{\sigma}$ and $\sigma > 1$, it must be $\lambda_1 = 0$ and $\lambda_2 > 0$. Then, $\tau_s = 1$ and $\lambda_2 = (1-\alpha)\varsigma(r)r$. Moving along all feasible directions, ϵ', such that $\tau_s + \epsilon' < \tau_s = 1$, $\pi'(\tau_s, r)\epsilon'$ is negative because (A2), $\lambda_1 = 0$, and $\lambda_2 > 0$ imply $\pi'(\tau_s, r)$ is positive while ϵ' is negative by definition. This ensures $\tau = \tau_S, \lambda_1 = 0$, and $\lambda_2 = (1-\alpha)V\varsigma(r)r$ identifies a global maximum when $\sigma > 1$ and $\alpha \leq 1-{\tau_I}^{\sigma}$. Therefore, when resources are scalable, the critical α is $1-{\tau_I}^{\sigma}$. Because $1 > \tau_I > 0$, we have $1 > 1-{\tau_I}^{\sigma} > 0$ since $0 < {\tau_I}^{\sigma} < 1$ for any $\sigma > 0$.

Assume $\sigma < 1$. If $\alpha \leq 1-\tau_I \rightarrow \tau = \tau_C$ and if $\alpha > 1-\tau_I \rightarrow \tau = \tau_I$. Then, τ can be characterized as $\tau = \tau_I + (\tau_C - \tau_I)H(1-\tau_I-\alpha)$, where $H(x) = \begin{cases} 1 & x\geq 0 \\ 0 & x<0 \end{cases}$ is the Heaviside step function. Noting that τ_C is an implicit function of α defined by the first order condition (A7), differentiating τ_C with respect to α gives:

$$\frac{\partial \tau_C}{\partial \alpha} = \frac{1}{\sigma-1}\frac{\tau_C(1-\tau_C)}{\alpha(1-\alpha)}. \tag{A8}$$

Because $\frac{1}{\sigma-1}$ is negative and the other factors are positive, $\frac{\partial \tau_C}{\partial \alpha}$ is negative. We can then express $\frac{\partial \tau}{\partial \alpha}$ as:

$$\frac{\partial \tau}{\partial \alpha} = \frac{\partial \tau_C}{\partial \alpha}H(1-\tau_I-\alpha) + (\tau_C-\tau_I)\delta(1-\tau_I-\alpha). \tag{A9}$$

The function $\delta(x) = \begin{cases} 0 & x\neq 0 \\ +\infty & x=0 \end{cases}$ is Dirac delta function, also called pulse function, which corresponds to the derivative of the Heaviside step function. Since $\tau_C = \tau_I$ when $1-\tau_I = \alpha$, $(\tau_C-\tau_I)\delta(1-\tau_I-\alpha) = 0$ for all $\alpha \in (0,1)$.[1] The derivative $\frac{\partial \tau}{\partial \alpha}$ is then negative and equal to $\frac{\partial \tau_C}{\partial \alpha}$ when $\alpha \leq 1-\tau_I$, and equal to zero when $\alpha > 1-\tau_I$. Because $\frac{\partial \tau}{\partial \alpha}$ is defined for every $\alpha \in (0,1)$, τ continuous in α. We deduce that the nature of the firm's response to (infinitesimal) changes in the parameter α is continuous, with adjustments to vertical scope occurring at the margin.

Now assume $\sigma > 1$. If $\alpha \leq 1-{\tau_I}^{\sigma} \rightarrow \tau = \tau_S$ and $\alpha > 1-{\tau_I}^{\sigma} \rightarrow \tau = \tau_I$. Then, $\tau = \tau_I + (\tau_S - \tau_I)H((1-{\tau_I}^{\sigma})-\alpha)$, which is discontinuous because $\lim_{\alpha\rightarrow 1-{\tau_I}^{\sigma}-} \tau = \tau_S \neq \lim_{\alpha\rightarrow 1-{\tau_I}^{\sigma}+} \tau = \tau_I$. By the chain rule, the derivative $\frac{\partial \tau}{\partial \alpha}$ can be expressed as:

$$\frac{\partial \tau}{\partial \alpha} = -(\tau_S-\tau_I)\delta((1-{\tau_I}^{\sigma})-\alpha). \tag{A10}$$

From the properties of Dirac delta function, it follows that $\frac{\partial \tau}{\partial \alpha}$ is zero everywhere except at $\alpha = 1-{\tau_I}^{\sigma}$, where it pulses and spikes to $-\infty$. We infer that, when the resource bundle is scalable, (infinitesimal) positive changes in the parameter α can lead to vertical expansion only in the proximity of the critical α line, altering vertical scope discontinuously from specialization to integration. Q.E.D.

[1] $\delta(0)0 = 0$ because, by the algebraic properties of the Dirac delta function, $\delta(x)x = 0$ for all $x \in \mathbb{R}$.

Proof of Proposition 2

For $\sigma < 1$, the effect of scalability on the critical α is $\frac{\partial(1-\tau_I)}{\partial\sigma} = -\frac{\partial\tau_I}{\partial\sigma}$. Using $Q_{ia}(\tau_I r) = Q_{ib}((1-\tau_I)\, r)$ to implicitly differentiate τ_I with respect to σ gives $\frac{\partial\tau_I}{\partial\sigma} = \frac{(1-\tau_I)\tau_I(ln(1-\tau_I)-ln(\tau_I))}{\sigma}$. Therefore, $\frac{\partial(1-\tau_I)}{\partial\sigma} = -\frac{\partial\tau_I}{\partial\sigma} = -\frac{(1-\tau_I)\tau_I(ln(1-\tau_I)-ln(\tau_I))}{\sigma}$, which is less than or equal to zero because (A5) implies $1-\tau_I \geq \tau_I \rightarrow ln(1-\tau_I) \geq ln(\tau_I)$. For $\sigma > 1$, the effect of scalability on the critical α is $\frac{\partial(1-{\tau_I}^{\sigma})}{\partial\sigma} = -\frac{\partial\tau_I}{\partial\sigma}\sigma{\tau_I}^{\sigma-1} - {\tau_I}^{\sigma}(\tau_I)$. Given that $\frac{\partial\tau_I}{\partial\sigma} = \frac{(1-\tau_I)\tau_I(ln(1-\tau_I)-ln(\tau_I))}{\sigma}$, $\frac{\partial(1-{\tau_I}^{\sigma})}{\partial\sigma}$ can be rewritten as $-{\tau_I}^{\sigma}(\tau_I ln(\tau_I) + (1-\tau_I)ln(1-\tau_I))$, which is positive because $1 > \tau_I > 0$ implies $ln(\tau_I), ln(1-\tau_I) < 0$. Q.E.D.

Proof of Proposition 3

When the resource bundle is non-scalable, the critical α is $1-\tau_I$. The effect of fungibility on the critical α is then $\frac{\partial(1-\tau_I)}{\partial\,\varphi(1)} = -\frac{\partial\tau_I}{\partial\varphi(1)}$. Using $Q_{ia}(\tau_I r) = Q_{ib}((1-\tau_I)\, r)$, implicitly differentiating τ_I with respect to $\varphi(1)$ gives $\frac{\partial\tau_I}{\partial\varphi(1)} = \frac{(1-\tau_I)\tau_I}{\sigma}$, which is positive because both τ_I and $(1-\tau_I)$ are positive. Therefore $-\frac{\partial\tau_I}{\partial\varphi(1)}$ is negative. When the resource bundle is scalable, the critical α is $1-{\tau_I}^{\sigma}$. The effect of fungibility of the critical α is then given by the derivative $\frac{\partial(1-{\tau_I}^{\sigma})}{\partial\,\varphi(1)} = -\frac{\partial\tau_I}{\partial\varphi(1)}\sigma{\tau_I}^{\sigma-1}$. Since $\frac{\partial\tau_I}{\partial\varphi(1)} = \frac{(1-\tau_I)\tau_I}{\sigma}$, $\frac{\partial(1-{\tau_I}^{\sigma})}{\partial\,\varphi(1)} = -(1-\tau_I){\tau_I}^{\sigma}$, which is negative because $(1-\tau_I), {\tau_I}^{\sigma} > 0$. Q.E.D.

Proof of Proposition 4

Because the optimal scaling rule satisfies $\frac{\partial\pi(\tau_j,r)}{\partial r} = \frac{\partial C(r)}{\partial r}$ for $j \in \{S, C, I\}$ and $C(\cdot)$ is monotonically increasing, a firm opting for sourcing regime j is as large as or larger than a firm opting for sourcing regime $i \neq j$ if $\frac{\partial\pi(\tau_j,r)}{\partial r} \geq \frac{\partial\pi(\tau_i,r)}{\partial r}$ for all $r > 0$.

Consider a digital firm whose scaling exponent is $\sigma > 1$. By (A6), the firm will specialize if $\alpha \leq 1-{\tau_I}^{\sigma}$, or else it will integrate. When it specializes, the marginal productivity of its resources is $\frac{\partial\pi(\tau_S,r)}{\partial r} = (1-\alpha)\varsigma(r)$, when it integrates, $\varsigma(\tau_I r)\tau_I$. By Euler's theorem, $\varsigma(\tau_I r)\tau_I$ is equivalent to ${\tau_I}^{\sigma}\varsigma(r)$. Using (A6), we have $(1-\alpha)\varsigma(r) \geq \left(1-(1-{\tau_I}^{\sigma})\right)\varsigma(r) = {\tau_I}^{\sigma}\varsigma(r)$.

However, because $\frac{\partial({\tau_I}^{\sigma})}{\partial\,\varphi(1)} = \frac{\partial\tau_I}{\partial\varphi(1)}\sigma{\tau_I}^{\sigma-1} = (1-\tau_I){\tau_I}^{\sigma} > 0$ (with $\frac{\partial\tau_I}{\partial\varphi(1)} = \frac{(1-\tau_I)\tau_I}{\sigma}$ being the derivative of τ_I with respect to $\varphi(1)$ implied by $Q_{ia}(\tau_I r) = Q_{ib}((1-\tau_I)\, r)$), the difference $\frac{\partial\pi(\tau_S,r)}{\partial r} - \frac{\partial\pi(\tau_I,r)}{\partial r} = (1-\alpha)\varsigma(r) - {\tau_I}^{\sigma}\varsigma(r)$ is decreasing in the fungibility of firm *i*'s resources.

Next, we compare a specialized digital firm with scaling exponent $\sigma > 1$ to a partially integrated industrial firm with scaling exponent $\sigma' < 1$ so that $\frac{\partial VQ_{ia}(1)}{\partial r} = \varsigma'(1) \geq \frac{\partial VQ_{ib}(1)}{\partial r} = \varphi'(1) > 0$. The marginal productivity of the specialized digital firm's resources is $(1-\alpha)\varsigma(r)$. The marginal productivity of the resources of the partially integrated firm is $(1-\alpha)\varsigma'(\tau'_C r)\tau'_C + \alpha\varphi'\left((1-\tau'_C)r\right)(1-\tau'_C)$, which, by (A7), can be expressed as $\frac{\partial\pi(\tau'_C,r)}{\partial r} = (1-\alpha)\varsigma'(\tau'_C r)$. The marginal productivity of the specialized firm is greater than that of the partially integrated firm because $\varsigma(r) > \varsigma'(\tau'_C r)$ if $r \geq r'$. Whether this threshold is met depends on the specifics of the cost function. The cut-off value for r identifies a point whose surpassing can lead to sustained growth and can be interpreted as a tipping point or critical mass that must be attained in order to trigger hyperscaling.

Finally, we compare a specialized digital firm with scaling exponent $\sigma > 1$ to an integrated industrial firm with scaling exponent $\sigma' < 1$. The productivity of the integrated firm is given by $\tau'_I \varsigma'(\tau'_I r)$. The productivity of the specialized firm is $(1-\alpha)\varsigma(r) \geq \big(1-(1-\tau_I{}^{\sigma})\big)\varsigma(r) = \tau_I \varsigma(\tau_I r)$. We have that $\tau_I \varsigma(\tau_I r) > \tau'_I \varsigma'(\tau'_I r)$ if $r \geq r''$(note that $\frac{\partial \tau_I}{\partial \sigma} > 0$—see proof of Proposition 2—and, therefore, $\tau_I > \tau'_I$). Also in this case, the cut-off value for r can be interpreted as a tipping point or critical mass.

It is interesting to note that, if the specialized firm's resource bundle is non-scalable with scaling exponent $\sigma' < 1$, the productivity under integration would have dominated the productivity under specialization for $\alpha > 1 - \tau'_I{}^{\sigma'}$. Given that integration requires $\alpha > 1 - \tau'_I > 1 - \tau'_I{}^{\sigma'}$, the integrated firm would have been larger than the specialized firm. Q.E.D**.**

Model Extension: N-firm Game-theoretic Model and Proof of Proposition 5

The "threat points" within which α leads to trade between firm i and firm *j*s are fully determined by the second-stage maximization programs delineating the optimal allocation of resources. For firm i, this corresponds to the Lagrangian in (A1). Therefore, when firm *i*'s resource bundle is scalable, firm *i* will specialize if α is below the threat point $(1-\tau_I{}^{\sigma})$, or else it will integrate. When the resource bundle is non-scalable, firm *i* will opt for concurrent sourcing if α is below the threat point $1-\tau_I$, or else it will integrate.

For any firm *j*, the second-stage maximization program can be converted to the Lagrangian:

$$(1-\theta\alpha)VQ_{ja}(\tau_j r_j) + \theta\alpha V Q_{jb}((1-\tau_j)r_j) + \lambda_{j1}(\tau_{jI}-\tau_j) + \lambda_{j2}(\tau_j - 0). \qquad \text{(A11)}$$

This maximization problem mirrors firm *i*'s. When the complementors' resources are scalable, the complementors will specialize in *b* if α is above the "threat point" $\theta^{-1}(1-\tau_{jI})^{\sigma_j}$, or else they will integrate (where $\theta \in (0,1)$ is the transaction cost parameter).[2] When their resource bundle is non-scalable, the complementors will perform concurrent sourcing if α is above the threat point $\theta^{-1}(1-\tau_{jI})$, or else they will integrate.[3]

The equilibrium value of α is determined by the market clearing constraint requiring that *a*s and *b*s are produced in one-to-one proportions,

$$g = Q_{ia}(\tau r) - Q_{ib}((1-\tau)r) - N\left(Q_{jb}\left((1-\tau_j)r_j\right) - Q_{ja}(\tau_j r_j)\right) = 0, \qquad \text{(A12)}$$

where $\tau, r, \tau_j,$ and r_j are a function of α. The market clearing constraint must always be satisfied in equilibrium. If it were not, because, for instance, firm *i* produced an excess supply of *a*s, then firm *i* would deviate by either reducing its resource stock, r, or by integrating more, increasing τ, so as to match the complementors' supply of *b*s. In doing so, firm *i* would reduce its costs (by reducing r) or increase its revenues (by reducing τ), ultimately increasing its profits. We also note that when $N > 1$, none of the complementors can profitably undercut the "realized α." If a complementor deviated by offering a lower α to firm *i,* such complementor would not capture the whole market. On the contrary, it would scale less

[2] As in the case of firm *i*, when resource bundle is scalable, the threat point is determined by comparing profits at the corner solutions $\tau_j = \tau_{jI}$ and $\tau_j = 0$.

[3] When the resource bundle is non-scalable, the threat point is reached when the shadow price of integration, captured by the Lagrange multiplier λ_{j1}, becomes positive.

because a lower α would reduce the marginal revenue product of its resources and, ultimately, its incentives to invest in r_j.[4]

Then, for α clearing the market, the profile of actions $\left(\{1, r_S(\alpha)\}, \{0, r_{jS}(\alpha)\}\right)$ is a Nash equilibrium if $\sigma, \sigma_j > 1$ and $1 - \tau_I{}^{\sigma} \geq \alpha \geq \theta^{-1}\left(1 - \tau_{jI}\right)^{\sigma_j}$; $\left(\{\tau_C(\alpha), r_C(\alpha)\}, \{\tau_{jC}(\alpha), r_{jC}(\alpha)\}\right)$ is a Nash equilibrium if $\sigma, \sigma_j < 1$ and $1 - \tau_I \geq \alpha \geq \theta^{-1}\left(1 - \tau_{jI}\right)$; $\left(\{\tau_C(\alpha), r_C(\alpha)\}, \{0, r_{jS}(\alpha)\}\right)$ is a Nash equilibrium if $\sigma < 1$, $\sigma_j > 1$ and $1 - \tau_I \geq \alpha \geq \theta^{-1}\left(1 - \tau_{jI}\right)^{\sigma_j}$; and $\left(\{1, r_S(\alpha)\}, \{\tau_{jC}(\alpha), r_{jC}(\alpha)\}\right)$ is a Nash equilibrium if $\sigma > 1$, $\sigma_j < 1$ and $1 - \tau_I{}^{\sigma} \geq \alpha \geq \theta^{-1}\left(1 - \tau_{jI}\right)$.

Using (A12) to implicitly differentiate α with respect to $-\varphi_j(1), -N,$ and $-\theta$, it follows:

$$\frac{\partial \alpha}{\partial(-\varphi_j(1))} = -\left(\frac{\partial g}{\partial \tau_i}\frac{\partial \tau_i}{\partial \alpha} + \frac{\partial g}{\partial r_i}\left(\frac{\partial r_i}{\partial \tau_i}\frac{\partial \tau_i}{\partial \alpha} + \frac{\partial r_i}{\partial \alpha}\right) + \frac{\partial g}{\partial \tau_j}\frac{\partial \tau_j}{\partial \alpha} + \frac{\partial g}{\partial r_j}\left(\frac{\partial r_j}{\partial \tau_j}\frac{\partial \tau_j}{\partial \alpha} + \frac{\partial r_j}{\partial \alpha}\right)\right)^{-1}\left(\frac{\partial g}{\partial(-\varphi_j(1))}\right) > 0, \quad \text{(A13)}$$

$$\frac{\partial \alpha}{\partial(-N)} = -\left(\frac{\partial g}{\partial \tau_i}\frac{\partial \tau_i}{\partial \alpha} + \frac{\partial g}{\partial r_i}\left(\frac{\partial r_i}{\partial \tau_i}\frac{\partial \tau_i}{\partial \alpha} + \frac{\partial r_i}{\partial \alpha}\right) + \frac{\partial g}{\partial \tau_j}\frac{\partial \tau_j}{\partial \alpha} + \frac{\partial g}{\partial r_j}\left(\frac{\partial r_j}{\partial \tau_j}\frac{\partial \tau_j}{\partial \alpha} + \frac{\partial r_j}{\partial \alpha}\right)\right)^{-1}\left(\frac{\partial g}{\partial(-N)}\right) > 0, \quad \text{(A14)}$$

$$\frac{\partial \alpha}{\partial(-\theta)} = -\left(\frac{\partial g}{\partial \tau_i}\frac{\partial \tau_i}{\partial \alpha} + \frac{\partial g}{\partial r_i}\left(\frac{\partial r_i}{\partial \tau_i}\frac{\partial \tau_i}{\partial \alpha} + \frac{\partial r_i}{\partial \alpha}\right) + \frac{\partial g}{\partial \tau_j}\frac{\partial \tau_j}{\partial \alpha} + \frac{\partial g}{\partial r_j}\left(\frac{\partial r_j}{\partial \tau_j}\frac{\partial \tau_j}{\partial \alpha} + \frac{\partial r_j}{\partial \alpha}\right)\right)^{-1}\left(\frac{\partial g}{\partial(-\theta)}\right) > 0. \quad \text{(A15)}$$

The sign of the above derivatives is fully determined by their numerators, since the denominator is always negative (intuitively, a positive change in α, which is the share of the pie apportioned to activity *b*, leads to a migration of resources toward that activity, thus having a negative impact on the difference between *a*s and *b*s measured by *g*). In (A13), the numerator is positive because a negative change in the complementors' baseline capability reduces the output of the complementors' main activity, depleting the *b*s in the market. For the market to clear, this reduction in supply needs to be counterbalanced by an increase in α. In (A14), the numerator is negative because a decrease in the number of complementors in the market results, ceteris paribus, in a reduction in the supply of *b*s, which must be met by a greater α. In (A15), an increase in transaction costs reduces the complementors' willingness to trade, diluting the supply of *b*s in the market. This effect must then be offset by an increase in α so as to rebalance the supply of *b*s.

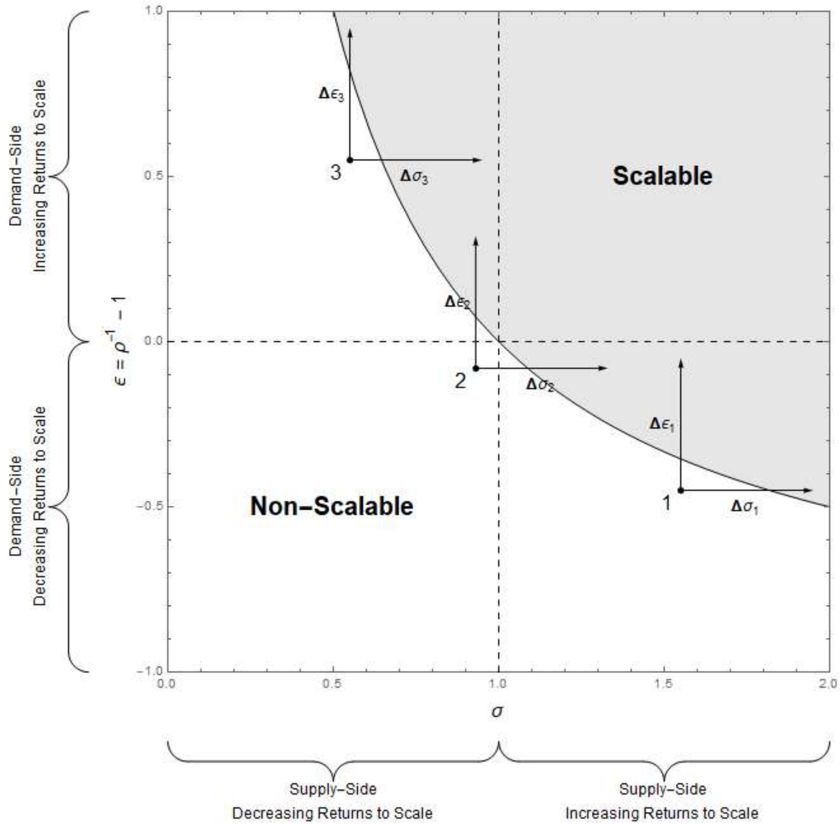
Noting that *g* in (A12) is continuous in α, that for every action profile, that α can get arbitrarily close to one (e.g., for θ arbitrarily small) or arbitrarily close to zero (e.g., for $\varphi_j(1)$ arbitrarily large), we can deduce that $\alpha(\cdot)$ is a function with image $(0,1)$, is differentiable, and monotonic in each variable $\varphi_j(1), N,$ and θ. Because α can fall outside the threat points (e.g., for θ arbitrarily small), if $\sigma < 1$ there exist initial values $\varphi_j(1)$, N, and θ, and increments $\Delta\varphi_j(1) < 0$, $\Delta N < 0$, and $\Delta\theta < 0$ such that $\alpha(\varphi_j(1), N, \theta) \leq 1 - \tau_I \rightarrow \tau = \tau_C$ and $\alpha\left(\varphi_j(1) + \Delta\varphi_j(1), N, \theta\right), \alpha\left(\varphi_j(1), N + \Delta N, \theta\right), \alpha\left(\varphi_j(1), N, \theta + \Delta\theta\right) > 1 - \tau_I \rightarrow \tau = \tau_I$. If $\sigma > 1$, there are values $\varphi_j(1)'$, N', and θ', and increments $\Delta N' < 0$, $\Delta\varphi_j(1)' < 0$, and $\Delta\theta' < 0$ such that $\alpha(\varphi_j(1)', N', \theta') \leq 1 - \tau_I{}^{\sigma} \rightarrow \tau = \tau_S$ and $\alpha\left(\varphi_j(1)' + \Delta\varphi_j(1)', N', \theta'\right)$, $\alpha\left(\varphi_j(1)', N' + \Delta N', \theta'\right)$, $\alpha\left(\varphi_j(1)', N, \theta' + \Delta\theta'\right) > 1 - \tau_I{}^{\sigma} \rightarrow \tau = \tau_I$. Q.E.D.

[4] The marginal revenue product of firm *j*'s resources is $\frac{\partial \pi(\tau_j, r_j)}{\partial r_j} = \frac{\partial\left(VQ_{jb}((1-\tau_j)r_j) - (1-\theta\alpha)V(Q_{jb}(\tau_j r_j) - Q_{ja}((1-\tau_j)r_j))\right)}{\partial r_j}$. This expression determines firm *j*'s scale via the first order condition $\frac{\partial \pi(\tau_j, r_j)}{\partial r_j} = \frac{\partial C(r_j)}{\partial r_j}$.

Asymmetric Effects of Supply-side and Demand-side Returns to Scale on Scalability

Figure A1 below illustrates how similar increases in returns to scale in supply and demand due to digitalization might affect overall scalability, depending on preexisting levels of supply and demand-side returns to scale. In the figure, the horizontal axis measures supply-side returns to scale (or elasticity), σ, and the vertical axis, the elasticity of demand, $\epsilon(\rho) = \rho^{-1} - 1$. (Note that ρ determines the elasticity of demand, but it is not the elasticity parameter itself.) The solid black curve separating the two regions corresponds to the function $\epsilon(\sigma)$, the locus of points for which $\sigma/\rho = 1$. Thus, the curve partitions the parameter space into non-scalable (white) and scalable (shaded) regions.

Figure A1: Asymmetric Effects of Supply-side and Demand-side Returns to Scale



Overall, the impacts of the two parameters can best be described as complementary, because augmenting the returns to scale on one side (either demand or supply) has a larger impact on overall scalability if returns to scale on the other side are also augmented. However, there may be asymmetries in the degree to which increases in supply or demand side returns to scale increase overall scalability and take firms into the scalable region. Case "1" illustrates a situation in which increases in demand-side returns to scale are more consequential for the scalability of a firm's resource bundle, and Case "3" illustrates the opposite situation in which supply-side changes in returns to scale are more consequential. In addition, it is important to consider the degree to which digitalization affects σ and ϵ, as the ultimate change in scalability also depends on the magnitude of the changes on the supply and demand sides as a result of digitalization. In some contexts, changes on the supply side might have larger impacts on scalability (e.g., firms with highly digital resource bundles), whereas in other contexts changes on the demand side might have bigger impacts (e.g., through digital sales and distribution and platform-driven network effects).