Business Networks, Production Chains, and Productivity: A Theory of Input-Output Architecture*

Ezra Oberfield

Federal Reserve Bank of Chicago

ezraoberfield@gmail.com

January 20, 2013

Abstract

This paper develops a theory in which the network structure of production—the set of plant-level input-output relationships—is the endogenous outcome of individual choices, and these choices shape productivity at the micro and macro levels. Entrepreneurs search for production techniques that use alternative sets of inputs, which are in turn produced by other entrepreneurs. The value of a technique depends both on its inherent productivity and on the cost of associated inputs; when producing, each entrepreneur selects the technique that delivers the best combination. These choices collectively determine the economy's equilibrium input-output structure. As new techniques are discovered, entrepreneurs substitute across suppliers in response to changing input prices. Despite the network structure, the model is analytically tractable, allowing for sharp characterizations of aggregate productivity and various micro-level characteristics. When the share of intermediate goods relative to labor in production is high, star suppliers emerge endogenously as small cost differences are amplified. Aggregate productivity depends on how frequently the lower-cost entrepreneurs are selected as suppliers. Larger firms tend to experience smaller reductions in cost but play an important role in the diffusion of cost savings through the network.

Keywords: Networks, Productivity, Supply Chains, Ideas, Diffusion

JEL Codes: O31, O33, O47

*I am grateful for comments from Amanda Agan, Fernando Alvarez, Enghin Atalay, Gadi Barlevy, Marco Bassetto, Jarda Borovicka, Jeff Campbell, Vasco Carvalho, Thomas Chaney, Meredith Crowley, Cristina De Nardi, Aspen Gorry, Joe Kaboski, Sam Kortum, Alejandro Justiniano, Robert Lucas, Devesh Raval, Rob Shimer, Nancy Stokey, Nico Trachter, Mark Wright, and Andy Zuppann, as well as various seminar participants. I also thank Todd Messer and Matt Olson for excellent research assistance. The views expressed herein are those of the author and not necessarily those of the Federal Reserve Bank of Chicago or the Federal Reserve System. All mistakes are my own.

Economists have long understood that productivity is at the heart economic progress. Aggregate productivity reflects both the productivity of individual producers and the economy's network structure of production—who buys inputs from whom—as this determines how changes in one producer's cost of production impacts other producers.[1]

This paper develops a theory in which the network structure of production is the endogenous outcome of individual choices, and studies how these choices determine both micro- and macro-level productivity. The theory is based on the premise that there may be multiple ways to produce a good, each with a different set of inputs. In the model, each entrepreneur sells a particular good and searches for cost-effective techniques to produce that good. Occasionally an entrepreneur discovers a new technique to produce her good that uses another entrepreneur's good as an input, with productivity specific to that input. When producing, the entrepreneur considers both the productivity and cost of the input associated with each technique available to her, and chooses to use the one that delivers the lowest cost of production. These choices collectively determine the equilibrium input-output architecture of the economy, shaping productivity at the micro and macro levels.

In the model, aggregate production technology consists of an evolving set of relationships among producers.[2] The collection of known production techniques form a dynamic network comprising each entrepreneur's potential suppliers and potential customers – others who might use the entrepreneur's good as an input. These jointly determine a firm's size and its contribution to ag-

[1]This point has been stressed by researchers in several areas. The literature on aggregate fluctuations has focused on how sectoral shocks propagate through an economy – see Long and Plosser (1983), Horvath (1998), Dupor (1999), Acemoglu et al. (2012), and Foerster et al. (2011). In the growth literature, Jones (2008) argues that misallocation in one sector raises input prices in other sectors, and the magnitude of the overall effect depends on the input-output structure. In the trade literature, Caliendo and Parro (2011) argue that the welfare impact of changes in tariffs in particular industries depends on the input-output structure.

[2]The theory emphasizes technological interdependence in the spirit of Rosenberg (1979): "The social payoff of an innovation can rarely be identified in isolation. The growing productivity of industrial economies is the complex outcome of large numbers of interlocking, mutually reinforcing technologies, the individual components of which are of very limited economic consequence by themselves. The smallest relevant unit of observation is seldom a single innovation but, more typically, an interrelated clustering of innovations." (p. 28-29)

gregate productivity. Innovation is the arrival of a new technique. Such an innovation represents both a new way to produce a good and a new use for a different good.[3] Techniques that are not currently cost-effective still have value as the cost of the input may fall in the future. As new techniques are discovered, prices adjust, entrepreneurs substitute across suppliers, and cost savings diffuse through the network. How large those effects are hinge on the structure of the network.

The paper begins by studying the problem of a planner that, at each point in time, takes the network of techniques as given and organizes production to maximize the utility of a representative final consumer. The planner decides how much of an entrepreneur's production should be done with each of its techniques. The social cost of producing a good with a particular technique depends on both the technique's productivity and the social cost of the associated inputs. Since every technique has constant returns to scale, each entrepreneur generically produces using a single technique.

Solving for the optimal allocation requires characterizing the social costs of producing the various goods. For any realization of the network of techniques, there will be a distribution of social costs of production across goods. I show that the CDF of this distribution is, with probability one, the unique solution to a fixed point problem that depends on three primitives: the average number of techniques per entrepreneur, the distribution from which each technique's productivity is drawn, and the share of intermediate goods in production (relative to labor). That is, rather than characterizing the optimal allocation as a function of the realized network of techniques, the fixed point problem describes the distribution of social costs that is likely to arise, given the stochastic process governing the arrival of techniques. The CDF of this distribution is the main tool used to characterize other features of economic interest such as how customers are distributed across suppliers, the size distribution, and aggregate productivity.

When the share of intermediate goods in production is large, small differences in firms' marginal

[3] The idea that innovation is finding a new use for an existing good is related to recombinant growth of Weitzman (1998), in which innovation is finding a better way to combine inputs.

costs are amplified into large differences in size. With a larger input share, the price of the input accounts for more of the cost-effectiveness of a technique, so that charging a low price becomes more important in winning customers. Consequently, the lower-cost firms are selected as suppliers more frequently.

This amplification of cost differences matters for aggregate outcomes. The equilibrium input-output structure is a set of overlapping supply chains. Aggregate productivity depends on which firms are selected to produce at each step in those supply chains. With a larger intermediate goods share, aggregate productivity is higher because lower-cost firms are selected as suppliers more frequently.

The model predicts that entrepreneurs with a high cost of production are more likely to switch suppliers than those with a low cost, as a low-cost entrepreneur is less likely to discover a technique more cost-effective than the one she is using. It also predicts that entrepreneurs that switch suppliers grow more than those that keep the same supplier – in the model the only reason an entrepreneur would switch suppliers is if the new supplier delivered a lower cost of production.

While low-cost entrepreneurs are less likely to find new more cost-effective techniques, they play an important role in the propagation of cost-savings through the network. Since a low-cost entrepreneur is likely to have many customers, when its supplier's cost falls, it passes on the cost savings to many other entrepreneurs.

Because choices of suppliers hinge on input prices, how these prices are set matters for aggregate outcomes. I study equilibria of a particular market structure with monopolistic competition in sales of final goods and bilateral two-part pricing between all potential input-output pairs. Because techniques are specific to a particular input and output, there is market power in any vertical relationship. Two-part pricing - a price per unit and a fixed fee for positive purchases - allows for an arbitrary split of surplus through the fixed fee, while the price per unit determines the quantity

of inputs purchased. I consider the set of pairwise stable equilibria—contracting arrangements for which there are no profitable unilateral or pairwise deviations. In any such equilibrium, the price per unit in each two part contract is equal to the supplier's marginal cost, and the fixed fee is non-negative. Thus each pair avoids double marginalization, and production within each supply chain is efficient. There are many equilibria that decentralize the optimal allocation but differ in how profits are divided across firms.[4]

This paper makes a technical contribution in providing an approach to modeling the endogenous formation of a network. Each entrepreneur selects a supplier from among many potential suppliers, and these choices are mediated by prices. The endogeneity of network formation is advantageous in that it opens the door to the analysis of policy experiments that change incentives and thus the link formation process.

In particular, the model provides an explanation for a Pareto tail of the number of observed links in a network which differs from the mechanism of the influential preferential attachment model of Barabasi and Albert (1999).[5] Here, the Pareto tail stems from selection. Because entrepreneurs differ in marginal cost, they also differ in their ability to attract customers. While the distribution of *potential* customers follows a Poisson distribution, the distribution of *actual* customers does not. Some of those firms with many potential customers are able to offer lower prices and win over a larger fraction of those potential customers. Consequently, the distribution of links is driven by a feature of the environment – the share of intermediate foods – rather than the stochastic process of the arrival of links.

[4]There is also always at least one equilibrium in which no production occurs, although that equilibrium is fragile with respect to a perturbation of the environment.

[5]The preferential attachment model was designed because network models with uniform arrival of links exhibit thin tails, whereas many real world networks feature Pareto tails. In the preferential attachment model, there is an initial network and, over time, new links are formed and new nodes are born. The probability that a new link involves a particular node is increasing in the number of links that node already has. Variants of this model have been used to explain the cross-sectional distribution of suppliers (Atalay et al. (2011)) and the network structure of international shipments of goods (Chaney (2011)).

The structure of the model is related in some respects to the work of Kortum (1997), Eaton and Kortum (2002), Alvarez et al. (2008), and Lucas (2009). In particular, the model nests a simple version of Kortum (1997) when the share of intermediate goods in production is zero—a special case in which the network structure plays no role.

The paper proceeds as follows: Section 1 describe the economic environment, setting up and solving a planner's problem. Section 2 defines an equilibrium for a particular market structure and shows that there are equilibria that decentralize the planner's problem. Section 3 describes how customers are distributed across suppliers, derives the cross-sectional distribution of employment, and describes how these relate to aggregate output. Section 4 discusses dynamics, focusing on the duration of vertical relationships, the value of new techniques, and the diffusion of cost-savings through the network. Section 5 concludes.

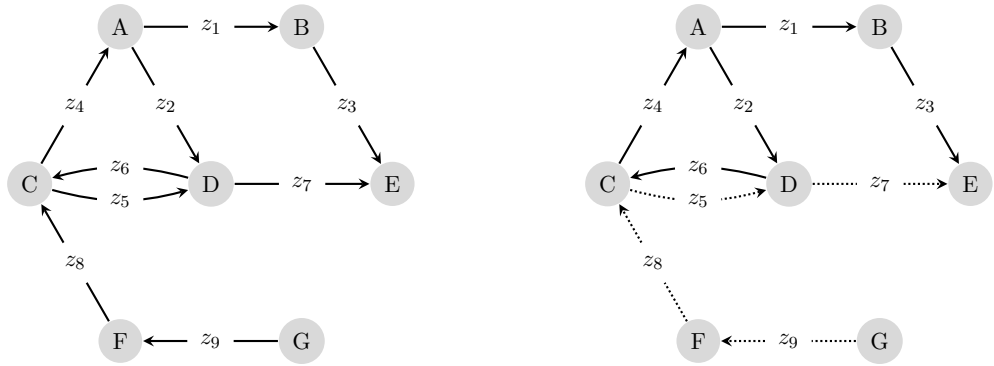
1 The Baseline Model

There is a unit mass of infinitely-lived firms, each associated with producing a particular good.[6] Each good is used for final consumption and potentially as an intermediate input by other firms. A representative consumer supplies L units of labor inelastically and has Dixit-Stiglitz preferences over all of the goods with elasticity of substitution ε.

Over time, each firm accumulates production techniques. For a firm, a technique is a method of producing its good using labor and some other firm's good as an intermediate input. Each technique is a production function fully described by the triple $\phi = \{\text{buyer, supplier}, \text{productivity}\}$. These are: (i) the good that is produced, $b(\phi)$; (ii) the good used as an input, $s(\phi)$; and (iii) a productivity parameter, $z(\phi)$:

$$y_b = \frac{1}{\alpha^\alpha (1-\alpha)^{1-\alpha}} z x_s^\alpha l^{1-\alpha}$$

[6] The words entrepreneur and firm will be used interchangeably.



(a) The Set of Techniques

(b) Techniques that are Used

Figure 1: A Graphical Representation of the Input-Output Structure

Figure 1a gives an example of a set of techniques at a point in time. Each node is a firm and edges correspond to techniques. The direction indicates which firm produces the output and which provides the input. Each edge has a weight associated with the productivity of the technique. In Figure 1b, solid arrows are techniques that are currently used, while dashed arrows represent unused techniques.

where y_b is the quantity of output of good $b(\phi)$, x_s is the quantity of good $s(\phi)$ used as an input, and l is labor.[7] Given a menu of techniques and input prices, each firm produces using the technique that delivers the best combination of input cost and productivity.

Time is continuous and new techniques arrive to each firm at rate $\mu(t)$. When firm j discovers a new technique, ϕ, the identity of the supplier $s(\phi)$ is random and uniformly distributed across all firms in the economy.[8] A technique's productivity, z, is drawn from a fixed distribution with CDF H and remains constant over time. It is assumed that the support of H is bounded below by some $z_0 > 0$ and that the tail of the distribution is not too thick: there exists a $\beta > \varepsilon - 1$ such that $\lim_{z\to\infty} z^{\beta}[1 - H(z)] = 0$, a sufficient condition for aggregate output to be finite with probability one. Lastly, at $t = 0$, no firm had access to any techniques.

At a point in time, the state of the economy can be summarized by the set of available techniques, $\Phi(t)$. Figure 1 gives a visual representation of such a set as a weighted, directed graph. In the

[7] Notably, the productivity of any technique is specific to a particular input and a particular output. This differs from a large literature which typically models productivity as output-specific.

[8] To keep the notation manageable, I abstract from any ex-ante differences in goods' suitability for use as an intermediate input for any other type of good or for consumption. The model can easily accommodate such asymmetries by explicitly allowing for multiple types of firms with different characteristics.

figure, firms are represented by nodes. Each technique is represented by an edge or link connecting two nodes. Each edge has a direction that corresponds to the flow of goods for the technique, indicating which firm provides the intermediate input and which produces the output. In addition, each edge has a weight (a number) corresponding to the productivity of the technique.

Figure 1a shows the set of all available techniques. Note that several firms (C, D, and E) each have multiple upstream techniques. In equilibrium they will each select a single technique with which to produce. These selections of suppliers jointly determine the equilibrium supply chains that are used to produce each good. Figure 1b gives an example of such selections.

Note also that firm G has no techniques and hence cannot produce. Consequently, firm F is unable to produce, because, while it has a technique, it is unable to buy the inputs required to use it. A *viable supply chain* for firm j is an infinite sequence of techniques $\{\phi_k\}_{k=0}^{\infty}$ with the property that $j = b(\phi_0)$ and $s(\phi_k) = b(\phi_{k+1})$ for each k. It is feasible for a firm to produce only if it has at least one viable supply chain.[9]

The network determines two sets of techniques that are directly relevant for each firm. Let $U_j(t) = \{\phi \in \Phi(t) | b(\phi) = j\}$ be j's upstream techniques and let $D_j(t) = \{\phi \in \Phi(t) | s(\phi) = j\}$ be j's downstream techniques. These correspond to j's potential suppliers and potential customers.

Over time, the network evolves as new techniques are discovered. The arrival of a technique is an expansion of a firm's production possibilities. Firms have perfect recall and may costlessly switch suppliers at any time. While the productivity of a technique, z, is constant over time, the attractiveness of a technique varies with the cost of the associated inputs.

[9] In the example with a finite set of firms, any viable supply chain must cycle at some point. With a continuum of firms, a cycle is not necessary.

1.1 A Planner's Problem

This section focuses on a planner's problem in order to build intuition about the economic environment. Section 2 discusses a market structure with bilateral two-part pricing for each technique, and shows there are equilibria that decentralize the planner's solution.

Consider the problem of a planner that takes the evolution of the network of techniques as given. At each point in time the planner makes production decisions and allocates labor to maximize the instantaneous utility of the representative consumer. This is a sequence of static problems, so to increase readability, I will suppress time subscripts when unnecessary.

For each producer $j \in [0,1]$, the planner chooses the quantity to be produced for final consumption, y_j^0. In addition, for each of j's upstream techniques $\phi \in U_j$, the planner chooses a quantity of labor, $l(\phi)$; and a quantity of good $s(\phi)$ for j to use as an intermediate input, $x(\phi)$. It is feasible for each firm to produce using more than one technique. However, because each technique exhibits constant returns to scale, the use of a single technique will generically be optimal.

Formally, the planner faces the following problem:

$$\max_{\left\{y_j^0, \{x(\phi), l(\phi)\}_{\phi \in U_j}\right\}_{j \in [0,1]}} \left(\int_0^1 (y_j^0)^{\frac{\varepsilon-1}{\varepsilon}} dj \right)^{\frac{\varepsilon}{\varepsilon-1}}$$

subject to technological constraints:

$$y_j^0 + \sum_{\phi \in D_j} x(\phi) \leq \sum_{\phi \in U_j} \frac{1}{\alpha^\alpha (1-\alpha)^{1-\alpha}} z(\phi) x(\phi)^\alpha l(\phi)^{1-\alpha}, \qquad \forall j \in [0,1]$$

and the labor resource constraint:

$$\int_0^1 \left(\sum_{\phi \in U_j} l(\phi) \right) dj \leq L$$

The left-hand side of the technological constraint for good j consists of its uses – output for final consumption and for use as an input in other firms' production – and the right hand side is total production.

Let λ_j be the multiplier on the constraint for good j, which can be interpreted as the marginal cost to the planner of good j. Let w be the multiplier on the labor clearing constraint, the marginal cost to the planner of labor. For each technique $\phi \in \Phi$, the first order conditions with respect to $x(\phi)$ and $l(\phi)$ imply

$$\begin{aligned} \lambda_{b(\phi)} \alpha \frac{y(\phi)}{x(\phi)} &\leq \lambda_{s(\phi)}, \text{ with equality if } x(\phi) > 0 \\ \lambda_{b(\phi)} (1-\alpha) \frac{y(\phi)}{l(\phi)} &\leq w, \text{ with equality if } l(\phi) > 0 \end{aligned}$$

Together with the production function associated with ϕ, these imply that $\lambda_{b(\phi)} \leq \frac{1}{z(\phi)} \lambda_{s(\phi)}^{\alpha} w^{1-\alpha}$, with equality if $y(\phi) > 0$. Collecting these conditions for each of j's upstream techniques gives

$$\lambda_j = \min_{\phi \in U_j} \frac{1}{z(\phi)} \lambda_{s(\phi)}^{\alpha} w^{1-\alpha} \tag{1}$$

It will be convenient to define $q_j \equiv \frac{w}{\lambda_j}$ as a measure of the *efficiency* with which the planner can produce good j in units of labor. Equation (1) can be written as

$$q_j = \max_{\phi \in U_j} z(\phi) q_{s(\phi)}^{\alpha} \tag{2}$$

with $q_j = 0$ if U_j is empty. The efficiency delivered by each of j's upstream techniques depends on the productivity of the technique and the efficiency with which the associated input can be produced. The efficiency with which the planner can produce j is then the maximum delivered by any of the available techniques.

Lemma 1 shows how aggregate output depends on firms' efficiencies. Let $Y^0 \equiv \left(\int_0^1 \left(y_j^0\right)^{\frac{\varepsilon-1}{\varepsilon}} dj\right)^{\frac{\varepsilon}{\varepsilon-1}}$ denote the aggregate of final output consumed by the representative consumer and let $Q \equiv \left(\int_0^1 q_j^{\varepsilon-1} dj\right)^{\frac{1}{\varepsilon-1}}$. Lemma 1 shows that Q measures aggregate productivity.

Lemma 1. *In the optimal allocation, $Y^0 = QL$.*

Proof. In characterizing the allocation of resources, it will be useful to organize production into the supply chains. Let $\mathfrak{C}_j^\infty$ be the set of viable supply chains to produce j. For any chain $\mathfrak{c} \in \mathfrak{C}_j^\infty$, we consider all of the resources used throughout the supply chain to produce final output of good j. Let $y_j^0(\mathfrak{c})$ be the total amount of final output of j produced using supply chain $\mathfrak{c}$.[10] Let $l_{j,0}(\mathfrak{c})$ and $x_{j,0}(\mathfrak{c})$ be the labor and inputs used by j to produce $y_j^0(\mathfrak{c})$. Similarly, let $l_{j,1}(\mathfrak{c})$ and $x_{j,1}(\mathfrak{c})$ be the labor and intermediate inputs used by j's supplier to produce $x_{j,0}(\mathfrak{c})$. In this manner, at each step in the supply chain $\mathfrak{c}$, let $l_{j,k+1}(\mathfrak{c})$ and $x_{j,k+1}(\mathfrak{c})$ be the labor and intermediates used to produce $x_{j,k}(\mathfrak{c})$. Similarly, let $\lambda_{j,k}(\mathfrak{c})$ be the marginal social cost of the kth-to-last intermediate input and let $z_{j,k}(\mathfrak{c})$ be the productivity of the technique used at the kth-to-last step.

If $w > 0$, an optimal allocation of labor across stages of production requires $l_{j,k+1}(\mathfrak{c}) = \alpha l_{j,k}(\mathfrak{c})$: the labor used at each stage is a constant fraction of the labor used in the subsequent stage. To see this, note that the optimal choices of inputs for the kth and $k+1$th suppliers in the chain $\mathfrak{c}$ imply $\frac{\lambda_{j,k}(\mathfrak{c})x_{j,k}(\mathfrak{c})}{\alpha} = \frac{wl_{j,k}(\mathfrak{c})}{1-\alpha}$ and $\frac{wl_{j,k+1}(\mathfrak{c})}{1-\alpha} = \lambda_{j,k}(\mathfrak{c})x_{j,k}(\mathfrak{c})$. Total labor used in chain $\mathfrak{c}$ across all stages in production of good j for final consumption, $\bar{l}_j(\mathfrak{c})$, is then:

$$\bar{l}_j(\mathfrak{c}) \equiv \sum_{k=0}^{\infty} l_{j,k}(\mathfrak{c}) = \sum_{k=0}^{\infty} \alpha^k l_{j,0}(\mathfrak{c}) = \frac{l_{j,0}(\mathfrak{c})}{1-\alpha}$$

The optimal choice of $l_{j,0}(\mathfrak{c})$ implies that $\lambda_j y_j^0(\mathfrak{c}) = \frac{wl_{j,0}(\mathfrak{c})}{1-\alpha}$, or $\lambda_j y_j^0(\mathfrak{c}) = w\bar{l}_j(\mathfrak{c})$. Note that $y_j^0(\mathfrak{c}) > 0$ implies that $x_{j,k}(\mathfrak{c}) > 0$ and $l_{j,k}(\mathfrak{c}) > 0$ for all k, so that $q_{b(\phi)} = z(\phi)q_{s(\phi)}^\alpha$ for each ϕ in the supply

[10] This is distinct from the total quantity of j produced using supply chain $\mathfrak{c}$, as that might include production of intermediates for use by other firms. Because each technique has constant returns to scale, this is well defined.

chain $\mathfrak{c}$. Thus for each $\mathfrak{c} \in \mathfrak{C}_j^\infty$, $q_j \geq \prod_{k=0}^\infty z_{j,k}(\mathfrak{c})^{\alpha^k}$ with equality if $y_j^0(\mathfrak{c}) > 0$.

Let $\bar{l}_j \equiv \sum_{\mathfrak{c}\in\mathfrak{C}_j} \bar{l}_j(\mathfrak{c})$ be the total quantity of labor used to produce good j for consumption. Summing across supply chains yields

$$y_j^0 = \sum_{\mathfrak{c}\in\mathfrak{C}_j^\infty} y_j^0(\mathfrak{c}) = \frac{w}{\lambda_j} \sum_{\mathfrak{c}\in\mathfrak{C}_j^\infty} \bar{l}_j(\mathfrak{c}) = q_j \bar{l}_j \tag{3}$$

With equation (3), we can derive an expression for aggregate consumption. The optimal choice of output for final consumption of j implies $\left(y_j^0/Y^0\right)^{-1/\varepsilon} = \lambda_j$. Together, these imply $\int_0^1 \lambda_j^{1-\varepsilon} dj = 1$, so that $y_j^0/Y^0 = (q_j/Q)^\varepsilon$. With this, the labor resource constraint can then be used with equation (3) to solve for aggregate output:

$$L = \int_0^1 \bar{l}_j dj = \int_0^1 \frac{y_j^0}{q_j} dj = \int_0^1 Y^0 Q^{-\varepsilon} q_j^{\varepsilon-1} dj = Y^0/Q$$

■

Lemma 1 implies that to solve for aggregate output, one must characterize firms' efficiencies. Given the network of techniques, Φ, this can be done using equation (2). One approach is to look for a vector of efficiencies $\{q_j\}_{j\in[0,1]}$ that is a fixed point of equation (2).[11] However, with a continuum of firms, this is neither computationally feasible nor would it be particularly illuminating. Section 1.2 describes an alternative approach which uses equation (2) along with the probabilistic structure of the model to sharpen the characterization of the optimal allocation.

[11] Taking logs of both sides gives an operator T, where the jth element of $T\left(\{\log q_j\}_{j\in[0,1]}\right)$ is $\max_{\phi\in U_j}\{\log z(\phi) + \alpha \log q_{s(\phi)}\}$. T satisfies monotonicity and discounting. If the support of z were bounded above, the operator would be a contraction on the appropriately bounded function space. This approach to solving the model is useful for simulations with a finite set of firms.

1.2 Using the Probabilistic Structure

The set of techniques that exist at a point in time, $\Phi(t)$, follows a stochastic process: for each firm, new techniques arrive at rate $\mu(t)$, and when a technique arrives, both its productivity and the identity of the associated supplier are randomly drawn. This section uses the law of large numbers to solve for the *distribution* of efficiencies that is likely to arise given this probabilistic structure. While any individual firm's efficiency varies across realizations of the economy, the cross-sectional distribution of efficiencies does not. In particular, this section shows that the CDF of this cross-sectional distribution at any point in time is the unique solution to a fixed point problem.

Let $F_t(q)$ be the fraction of firms with efficiency no greater than q at time t given the decisions of the planner. This function is endogenous and will need to be solved for. Our strategy exploits the fact that, at time t, F_t describes the distribution the efficiency of a firm and that of each of its potential suppliers.

What is the probability that a firm has efficiency no greater than q at time t? This depends on the number of techniques the firm has discovered and the efficiency each of those techniques delivers. Given the history of arrival rates of new techniques, $\{\mu(\tau)\}_{\tau\in[0,t]}$, the number of upstream technique each firm has at t, $|U_j(t)|$, follows a Poisson distribution. This distribution is fully described by its mean, M_t, which satisfies:

$$M_t = \int_0^t \mu(\tau)d\tau \tag{4}$$

Since M_t is the average number of techniques per firm, I will refer to it as the *density* of techniques in the network.

Two elements determine how cost-effective a technique is at t: (i) the productivity parameter, $z(\phi)$, drawn from an exogenous distribution H, and (ii) the efficiency of the supplier, $q_{s(\phi)}$. Since

the identity of the supplier is drawn randomly, the probability that the supplier's efficiency is no greater than $\tilde{q}$ is $F_t(\tilde{q})$. In particular, the efficiency of the supplier at time t is independent of when the technique arrived.

Let $G_t(q)$ be the probability that the efficiency delivered by a single random technique at t is no greater than q. Recalling that the efficiency delivered by the technique is $z(\phi)q^{\alpha}_{s(\phi)}$, this is:

$$G_t(q) = \int_0^{\infty} F_t\left((q/z)^{1/\alpha}\right) dH(z) \tag{5}$$

To interpret this, note that for each z, $F_t\left((q/z)^{1/\alpha}\right)$ is the portion of potential suppliers that, in combination with that z, leaves the firm with efficiency no greater than q.

Now, the probability that, given all of its techniques, a firm has efficiency no greater than q at t is:

$$\begin{aligned}\Pr(q_j \leq q) &= \sum_{n=0}^{\infty} \Pr(n \text{ techniques}) \Pr(\text{All } n \text{ techniques are} \leq q) = \sum_{n=0}^{\infty} \frac{M_t^n e^{-M_t}}{n!} G_t(q)^n \\ &= e^{-M_t[1-G_t(q)]}\end{aligned}$$

To interpret this last expression, note that if $M_t[1 - G_t(q)]$ is a parameter of a Poisson distribution (the arrival rate of techniques that would provide efficiency better than q), then $e^{-M_t[1-G_t(q)]}$ is the probability of no such techniques.

The law of large numbers implies that with probability one, $\Pr(q_j \leq q) = F_t(q)$.[12] Using the

[12]The proof of Proposition 1 uses such a law of large numbers for a continuum of random variables described by Uhlig (1996). To use this, one must verify that firms' efficiencies are pairwise uncorrelated. That is not immediately obvious in this context, as it is possible that two firms' supply chains overlap or that one is in the other's supply chain. However, by assumption the network is sufficiently sparse that with high probability the supply chains will not overlap: there is a continuum of firms but only a countable set of those are in any of a given firm's potential supply chains. Therefore, for any two firms, the probability that their supply chains overlap is zero. The law of large numbers then implies that equation (6) holds with probability one.

expression for $G_t(q)$ from equation (5) gives a fixed point problem for the CDF of efficiency $F_t(q)$:

$$F_t(q) = e^{-M_t \int_0^\infty \left[1 - F_t\left((q/z)^{1/\alpha}\right)\right] dH(z)} \tag{6}$$

This recursive equation is the key to characterizing the planner's solution.

Consider the space $\bar{\mathcal{F}}$ of non-decreasing functions $f : \mathbb{R}^+ \mapsto [0, 1]$, and consider the operator T on this space defined as

$$Tf(q) \equiv e^{-M_t \int_0^\infty \left[1 - f\left((q/z)^{1/\alpha}\right)\right] dH(z)}$$

Appendix A constructively defines a subset $\mathcal{F} \subset \bar{\mathcal{F}}$ that depends on M_t and H. Proposition 1 shows that the optimal allocation is the unique fixed point of T on $\mathcal{F}$. The proof, contained in Appendix A, uses the Tarski fixed point theorem to show existence, which also provides a numerical algorithm to solve for it.

Proposition 1. *There exists a unique fixed point of T on $\mathcal{F}$, F_t^{sp}. With probability one, F_t^{sp} is the CDF of the cross-sectional distribution of efficiencies in the solution to the planner's problem and aggregate productivity is $Q_t = \left(\int_0^\infty q^{\varepsilon-1} dF_t^{sp}(q)\right)^{\frac{1}{\varepsilon-1}}$.*

The qualitative features of the economy depend on whether the average number of techniques, M_t, is greater or less than one. If $M_t \leq 1$, there are so few techniques available that the probability that any individual firm has access to a viable supply chain is zero.[13] This paper focuses on the more interesting case in which the average number of techniques exceeds one.

When M_t is larger than the critical value of 1, there are at least three fixed points of the operator T on $\bar{\mathcal{F}}$, only one of which corresponds to the planners solution.[14] Because of the multiplicity of

[13] As shown in Appendix D, the probability that a firm has no viable chains is the smallest root ρ of $\rho = e^{-M_t(1-\rho)}$. For $M_t \leq 1$, the smallest root is one, while for $M_t > 1$, it is strictly less than one. This uses a standard result from the theory of branching processes. In this case T is a contraction, with the unique solution $f = 1$. A phase transition when the average number of links per node crosses 1 is a typical property of random graphs, a result associated with the Erdos-Renyi Theorem. Kelly (1997) and Kelly (2005) give such a phase transition an economic interpretation.

[14] There are two solutions in which $f(q)$ is constant for all q which stem from the fact that equation (6) is formulated

fixed points, it is important to restrict the fixed point problem to the function space $\mathcal{F}$, a space for which there is a unique solution which corresponds to the planner's allocation.

We therefore have that the cross-sectional distribution of efficiencies at a point in time depends on three primitives: the distribution from which technique-specific productivities are drawn, H; the average number of techniques per firm, M_t; and the share of intermediate goods.

Section 3 shows how the cross-sectional distribution of efficiencies can be used to characterize other features of economic interest.

2 Market Structure

There is bilateral market power in any input-output relationship. Because choices of suppliers hinge on input prices, how these prices are set matters for aggregate outcomes.

We will consider a market structure in which there is monopolistic competition in sales for consumption but bilateral two-part pricing for intermediate goods. We will characterize the set of contracting arrangements that are *pairwise stable*—arrangements for which the are no profitable unilateral or mutually beneficial pairwise deviations.[15]

recursively. The first is $f(q) = 1$ for all q, which corresponds to zero efficiency (infinite marginal social cost) for all goods; if the marginal social cost of every input is infinite, then the marginal social cost of each output is infinite as well. The corresponding allocation is feasible but dominated by another feasible allocation and is therefore not the solution to the planner's problem. A second, $f(q) = \rho \in (0, 1)$, implies infinite efficiency for firms with viable supply chains and zero efficiency for those without; if inputs have zero marginal social cost, output has zero social cost. This does not correspond to a feasible allocation and is therefore not a solution to the planner's problem.

Each of these correspond to a fixed point of equation (2). Multiple solutions to first order necessary conditions of the planner's problem is actually a typical feature of models in which a portion of output is used simultaneously as an input, such as a standard growth model with roundabout production. If the same good enters a technological constraint as both an input and an output, the Lagrange multiplier on that good will be on both sides of a first order condition. Consequently, the first order condition will be satisfied if the Lagrange multiplier takes the value of zero or infinity. One can usually sidestep this issue by finding an alternative way to describe the production technology, i.e., solving for final output as a function of primary inputs. Much of the work in the proof of Proposition 1 is in finding and characterizing such an alternative description of production possibilities.

[15] As an equilibrium concept, pairwise stability is used frequently in the networks literature (Jackson (2008) provides an excellent survey). For many applications in this literature, payoffs are a function of the particular links that are formed, and the idea is to find a network of links for which no pair of nodes wish to change whether or not they are connected (see Jackson (2003)).

2.1 Pairwise Stable Equilibrium

This section defines a pairwise stable equilibrium. This has two stages. First an *arrangement* determines which techniques are used and pricing for each of those techniques. Second, each firm sets a price for sales to the final consumer and selects its input mix.

An arrangement is a subset of techniques to be used in production and, for each of those techniques, the terms at which the buyer can purchase the associated input.

Definition 1. *An* ***arrangement*** *consists of: (i) A subset of techniques,* $\hat{\Phi} \subseteq \Phi$*, such that if* $\phi', \phi \in \hat{\Phi}$ *then* $b(\phi) \neq b(\phi')$ *(ii) For each* $\phi \in \hat{\Phi}$*, a price per unit and fixed fee,* $\{p(\phi), \tau(\phi)\}_{\phi \in \hat{\Phi}}$

The first condition implies that $\hat{\Phi}$ contains at most one technique for each firm to use. The arrangement determines several objects that will be useful in characterizing an equilibrium. First, let $\hat{J}$ be the set of firms with upstream techniques in the arrangement, $\hat{J} = \{b(\phi)\}_{\phi \in \hat{\Phi}}$. Second, for each firm $j \in \hat{J}$, let ϕ_j be technique used. Lastly, let $\hat{D}_j$ be the subset of j's downstream techniques that are used, $\hat{D}_j \equiv \{\phi \in \hat{\Phi} | s(\phi) = j\}$.

In the second stage, firms make production decisions taking as given the arrangement, the wage, and the choices of others. Each $j \in \hat{J}$ chooses a price for final consumption, p_j^0; a quantity of labor, l_j; and a quantity of good $s(\phi_j)$, $x(\phi_j)$.

Deviations:

An arrangement is said to be pairwise stable if there are no profitable unilateral or pairwise deviations (payoffs will be defined below). These deviations are defined as follows:

Definition 2. *For the arrangement* $\left\{\hat{\Phi}, \{p(\phi), \tau(\phi)\}_{\phi \in \hat{\Phi}}\right\}$*:*

- *A* ***unilateral deviation*** *for firm* $j \in [0, 1]$ *is an alternative arrangement* $\left\{\tilde{\Phi}, \{p(\phi), \tau(\phi)\}_{\phi \in \tilde{\Phi}}\right\}$ *where* $\tilde{\Phi} = \hat{\Phi} \setminus \tilde{\Phi}_j$ *and* $\tilde{\Phi}_j \subseteq \left((U_j \cup D_j) \cap \hat{\Phi}\right)$ *is a subset* j*'s upstream and downstream techniques.*

- *A **pairwise deviation** for the technique $\phi_0 \in \Phi$ is an alternative arrangement* $\left\{\tilde{\Phi}, \{\tilde{p}(\phi), \tilde{\tau}(\phi)\}_{\phi \in \tilde{\Phi}}\right\}$ *with* $\tilde{\Phi} = \{\phi_0\} \cup \{\phi \in \hat{\Phi} | b(\phi) \neq b(\phi_0)\}$ *and* $\{p(\phi), \tau(\phi)\} = \{\tilde{p}(\phi), \tilde{\tau}(\phi)\}$ *for each* $\phi \in \left(\tilde{\Phi} \setminus \{\phi_0\}\right)$.

For firm $j \in [0, 1]$, a unilateral deviation is the removal from the arrangement of some subset of j's upstream and downstream techniques. For example, a firm may choose not to supply goods to any of her customers and/or not to purchase inputs from her designated supplier.

A pairwise deviation for a technique $\phi \in \Phi$ is the addition to the arrangement of a new two part tariff for the technique, along with the removal from the arrangement of any other upstream technique for the buyer. There are two types of pairwise deviations. First, a firm could agree to terms with a new supplier on a technique that is not in the arrangement. Second, a buyer and supplier could agree to new terms for a technique already in the arrangement.

Payoffs:

Given an arrangement, define the payoff for any $j \in \hat{J}$ to be

$$\pi_j \equiv \max_{p_j^0, y_j^0, l_j, x(\phi_j)} p_j^0 y_j^0 + \sum_{\phi \in \hat{D}_j} [p(\phi)x(\phi) + \tau(\phi)] - [p(\phi_j)x(\phi_j) + \tau(\phi_j)] - wl_j$$

subject to the demand curve for final output and the technological constraint:

$$y_j^0 + \sum_{\phi \in \hat{D}_j} x(\phi) \leq \frac{1}{\alpha^\alpha (1-\alpha)^{1-\alpha}} z(\phi_j) x(\phi_j)^\alpha l_j^{1-\alpha}$$

Thus j takes as given the arrangement, the wage, the prices for final sales set by each other firm, and the quantity demanded by each of its customers, $\{x(\phi)\}_{\phi \in \hat{D}_j}$, and makes choices to maximize the sum of revenue from sales for consumption and for intermediate use net of the cost of the intermediate input and labor.

Payoffs must be well-defined for any arrangement. In particular, we must account for an ar-

rangement in which there is a $j \in \hat{J}$ that is unable to produce, because, for example, its supplier $i = s(\phi_j)$ has no way to produce ($i \notin \hat{J}$). Such an arrangement could arise following a unilateral deviation in which i's supplier removes ϕ_i from the arrangement.

To this end, I adopt the following convention (which will only be relevant off the equilibrium path): If the arrangement dictates that $i \in [0, 1]$ produce goods for $j \in \hat{J}$, but i is unable to produce (either because $i \notin \hat{J}$ or because i's supplier $s(\phi_i)$ is unable to produce), then i must compensate j for lost profit, so that j's payoff remains π_j.

Thus while payoffs are defined for all firms, only the production decisions of firms with viable supply chains will be relevant for the allocations of resources. Given an arrangement, define J to be the set of firms that are able to produce – those for which $\hat{\Phi}$ contains a viable supply chain. Thus π_j is the payoff for each $j \in \hat{J}$, but the choices $\{p_j^0, y_j^0, l_j, x(\phi_j)\}$ affect the allocation of resources only if $j \in J$.

Pairwise Stability:

Finally, we define a pairwise stable equilibrium.[16]

Definition 3. *A **pairwise stable equilibrium** is an arrangement* $\left\{\hat{\Phi}, \{p(\phi), \tau(\phi)\}_{\phi \in \hat{\Phi}}\right\}$*, firms' choices,* $\left\{p_j^0, y_j^0, x(\phi_j), l_j\right\}_{j \in J}$*, and a wage* w *such that (i) Given the wage and total profit, the representative consumer maximizes utility; (ii) For each* $j \in \hat{J}$*,* $\left\{p_j^0, y_j^0, x(\phi_j), l_j\right\}$ *maximize firm* j*'s profit given the arrangement, the wage, and the final demand; (iii) Labor and final goods markets clear; (iv) There are no unilateral deviations that would increase a firm's payoff or pairwise deviations that would increase one firm's payoff without lowering the other's.*

[16] In closely related environments, Ostrovsky (2008) and Hatfield et al. (2012) take an alternative approach to modeling contracting arrangements. Both focus on a space of contracts that effectively dictate a price and quantity to be traded between each pair of firms, and use stronger notions of stability—that arrangements are stable with respect to coalitional deviations. Two see how that approach is related to the one used here, consider the following example. Suppose that according to the arrangement, j is purchasing from i, and wants to deviate and purchase from $\tilde{i}$ instead. In that deviation, for $\tilde{i}$ to produce the goods for j, she would would want to use more inputs from her supplier, $s(\phi_{\tilde{i}})$. With two-part pricing, the terms at which $\tilde{i}$ purchases the additional goods from her supplier are already part of the original arrangement; she just need to place a larger order. If, instead, contracts dictated a price and quantity between each pair, this would require a new contract between $\tilde{i}$ and her supplier.

2.2 Equilibrium Allocations

This section describes how the arrangement determines the allocation of resources across firms and aggregate output. Without loss of generality, we restrict attention to equilibria for which the arrangement contains a viable supply chain for each $j \in \hat{J}$.

Given an arrangement, each firm takes factor prices as given and optimizes over inputs, so that marginal cost for firm $j \in J$ is $\lambda_j \equiv \frac{1}{z(\phi_j)} p(\phi_j)^{\alpha} w^{1-\alpha}$. If $j \notin J$, $\lambda_j \equiv \infty$. Define $q_j \equiv \frac{w}{\lambda_j}$ to be firm j's efficiency. Proposition 2 describes implications of pairwise stability. Appendix C shows:

Proposition 2. *In any pairwise stable equilibrium:*

1. *For each technique $\phi \in \hat{\Phi}$, price is equal to marginal cost, $p(\phi) = \lambda_{s(\phi)}$, and the fixed fee is nonnegative, $\tau(\phi) \geq 0$;*

2. *For each $j \in J$, $p_j^0 = \frac{\varepsilon}{\varepsilon - 1} \lambda_j$;*

3. *For each $j \in [0, 1]$, $q_j = \max_{\phi \in U_j} z(\phi) q_{s(\phi)}^{\alpha}$ with $q_j = 0$ if U_j is empty.*

The first condition says that for any technique used in equilibrium, the price per unit of the input equals the supplier's marginal cost. This means that there is no double marginalization in any equilibrium supply chain, so that production within each supply chain is efficient. Given that price per unit is equal to marginal cost, a non-negative fixed fee implies that suppliers at worst break even in sales of intermediate goods. The second condition says that each firm charges the usual markup in final sales given the constant elasticity of final demand.[17]

These conditions are the same as the necessary conditions to the planner's problem. In fact, there are many pairwise stable equilibria that decentralize the planner's solution. In each of these,

[17] This is not immediate because it is possible for a firm to be in its own supply chain (if the supply chain is a cycle). In that case, selling more final goods would also raise sales of intermediate goods. However, since the price on sales of intermediate goods is equal to the firm's marginal cost, selling more intermediate goods has no impact on the firm's payoff.

marginal costs are the same as in the planner's solution, but the various equilibria differ in the fixed part of the tariffs. In each of these equilibria, the markup in sales to the final consumer generates a profit for each good. This profit is then divided up across the supply chain to produce that good according to the fixed fees. As one might expect, conditioning on the techniques that are used and prices per unit, the fixed fees play no role in determining the allocation of real resources. These are lump some transfers from one firm to another, and since all firms are owned by the representative consumer, total profit is independent of these transfers.[18] The markups on final sales distort the consumption–leisure margin, but since labor is supplied inelastically there is no impact on the allocation of goods and labor. If labor supply were elastic, all of these equilibria would generate the *same aggregate production function* as the optimal allocation, but all production would scale in proportion to aggregate labor.

There are additional equilibria that do not decentralize the planner's problem. For example, there is always an equilibrium in which $\hat{\Phi}$ is empty so that no firms produce (this corresponds to one of constant fixed points of equation (6), as discussed in footnote (14)).

The remainder of the paper restricts attention to the set of equilibria that decentralize the planner's solution. Such a focus can be justified by appealing to an equilibrium refinement. Consider an alternative environment in which each firm has the ability to produce using only labor, with the production function $y = \underline{q}L$, with $\underline{q}$ common to all firms. Appendix C shows that such an environment allows for a welfare theorem: every pairwise stable equilibrium decentralizes the planner's solution. This result holds even when $\underline{q}$ is arbitrarily close to zero.[19]

[18] If firms made entry/exit decisions or chose the intensity with which to search for new techniques, the ex-post distribution of profit across firms would play a more central role.

[19] Why does adding a labor only option reduce the set of equilibria? The trivial equilibrium is circular in that if all of a firm's potential suppliers cannot produce, the firm cannot produce either. The labor only option breaks this circularity.

3 Cross-Sectional Implications

This section studies how the choices of which technique to use shape the cross-sectional features of the economy, such as how input-output links are distributed across firms, the size distribution, and, most importantly, aggregate output.

To illustrate these, it will be useful to focus on a particular functional-form assumption that proves to be analytically tractable. This will allow a closed-form expression for the distribution of efficiencies and for aggregate output, and provides a more transparent connection between the model parameters and economic outcomes.

Assume that each technique's productivity is drawn from a Pareto distribution, $H(z) = 1 - \left(\frac{z}{z_0}\right)^{-\zeta}$, with the restriction that $\zeta > \varepsilon - 1$. The regularly varying tail of the Pareto distribution is convenient, but the discontinuity at z_0 is not. We will therefore focus on a particular limiting economy. Towards this, rescale the arrival rates of techniques so that the average number of techniques per firm is $M_t = m_t z_0^{-\zeta}$. One can compare economies with different values of z_0 (holding $\{m_t\}$ fixed).[20] In an economy with a lower z_0, (i) each technique has stochastically lower productivity, and (ii) new techniques arrive more frequently. In fact, the parameterization is such that varying z_0 has no impact on the average number of existing techniques with a given level of productivity $\hat{z}$, $M_t H'(\hat{z})$, as the two effects offset exactly. The only difference with a lower z_0 is that firms discover additional relatively unproductive techniques.

We then look at the limit of a sequence of economies as $z_0 \to 0$. This adds many relatively unproductive techniques (low z) to the economy without changing the measure of productive techniques (high z). In the limit, the measure of firms without any techniques shrinks to zero. Since this section focuses on cross-sectional implications, time subscripts will be suppressed.

[20] At a more primitive level, we rescale the arrival rates so that $\mu(t) = \bar{\mu}(t) z_0^{-\zeta}$. Letting $m_t = \int_0^t \bar{\mu}(\tau) d\tau$, we compare economies with different values of z_0 holding $\{\bar{\mu}_t\}$ fixed.

In this special case, every solution F to equation (6) is the CDF of a Frechet distribution. To see this, note that one can use the change of variables $x = (q/z)^{1/\alpha}$ to write equation (5) as

$$1 - G(q) = \int_0^{(q/z_0)^{1/\alpha}} H'\left(q/x^\alpha\right)\left(1 - F(x)\right) q\alpha x^{-\alpha-1} dx$$

Using the functional forms $H'(z) = \zeta z_0^\zeta z^{-\zeta-1}$ and $M = m z_0^{-\zeta}$, one can then write

$$\begin{aligned} M\left[1 - G(q)\right] &= m z_0^{-\zeta} \int_0^{(q/z_0)^{1/\alpha}} \zeta z_0^\zeta \left(q/x^\alpha\right)^{-\zeta-1}\left(1 - F(x)\right) q\alpha x^{-\alpha-1} dx \\ &= q^{-\zeta} m \int_0^{(q/z_0)^{1/\alpha}} \alpha\zeta x^{\alpha\zeta-1}\left(1 - F(x)\right) dx \end{aligned}$$

For any q, as $z_0 \to 0$, this expression approaches $q^{-\zeta}$ multiplied by a constant. Label this constant θ, so that equation (6) can be written as $F(q) = e^{-\theta q^{-\zeta}}$, the CDF of a Frechet random variable. Note that the exponent ζ is the same as that of the Pareto distribution H. This means that the distribution of efficiencies q inherits the tail behavior of the productivity draws, z.

I next solve for θ, which was defined to satisfy $\theta = m \int_0^\infty \alpha\zeta x^{\alpha\zeta-1}\left(1 - F(x)\right) dx$. Substituting in the expression for F and manipulating gives $\theta = \Gamma(1-\alpha) m \theta^\alpha$, or more simply

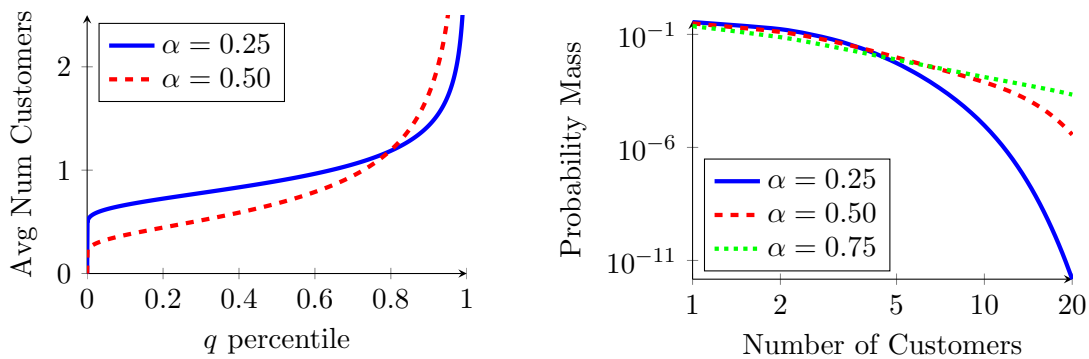
$$\theta = \left[\Gamma(1-\alpha) m\right]^{\frac{1}{1-\alpha}} \tag{7}$$

where $\Gamma(\cdot)$ is the gamma function.[21]

3.1 Distribution of Customers

To this point we have described how a firm's efficiency depends on the arrival of new upstream techniques and the availability of efficient supply chains. This section looks at techniques from the

[21] $\theta = \Gamma(1-\alpha) m \theta^\alpha$ has three non-negative roots: the one given in equation (7), zero, and infinity. The latter two correspond to the two constant fixed points of equation (6), as discussed in footnote (14).



(a) Conditional Means

(b) Unconditional Distribution

Figure 2: Distribution of Customers

Figure 2a shows the mean number of actual customers for each percentile in the efficiency distribution. Figure 2b gives the mass of firms with n customers on a log-log plot.

opposite perspective: how they shape demand for a firm's good as an input. Together, these determine a firm's contribution to aggregate productivity. Proposition 3 characterizes how customers are distributed across suppliers.

Proposition 3. *Among firms with efficiency q, the number of actual customers follows a Poisson distribution with mean*

$$M \int_0^\infty \frac{1}{q^\alpha} H'\left(\frac{\tilde{q}}{q^\alpha}\right) \frac{F(\tilde{q}) - e^{-M}}{G(\tilde{q})(1 - e^{-M})} d\tilde{q} \tag{8}$$

With the functional form assumptions, this simplifies to $\frac{m}{\theta} q^{\alpha\zeta}$.

To interpret equation (8): for a firm with efficiency q, $\frac{1}{q^\alpha} H'\left(\frac{\tilde{q}}{q^\alpha}\right)$ is the likelihood that a single downstream technique delivers efficiency $\tilde{q}$, $\frac{F(\tilde{q})-e^{-M}}{G(\tilde{q})(1-e^{-M})}$ is the probability that a buyer has no other techniques that deliver efficiency better than $\tilde{q}$, and M is the average number of downstream techniques.

Figure 2a shows the average number of customers at each percentile in the efficiency distribution for different values of α. Each curve in Figure 2a is increasing, indicating high-efficiency firms attract more customers: the distribution of customers among high-efficiency suppliers first order stochastically dominates the distribution among low-efficiency suppliers.

When the share of intermediate goods is larger, differences in firms' efficiency are amplified. Recall that the efficiency delivered by a single technique is $z(\phi)q^{\alpha}_{s(\phi)}$, where z is the productivity embedded in the technique and q is the efficiency of supplier. α determines the relative impact of each of these factors on the cost-effectiveness of a technique and, consequently, on the selection of supplier. When α is higher, the cost of the inputs matters more, which means the most efficient producers are selected as suppliers more frequently. The shift is evident in Figure 2a: with higher α, the high-efficiency firms capture a larger share of customers.[22]

We next look at the unconditional distribution over the number of customers among all firms. To find the mass of suppliers with n customers, we simply integrate over suppliers of each efficiency. The resulting formula is given in Proposition 4, which also describes the tail of the distribution:

Proposition 4. *Let p_n be the mass of firms with n customers. With the functional form assumptions,*

$$p_n = \int_0^{\infty} \frac{\left(\frac{w^{-\alpha}}{\Gamma(1-\alpha)}\right)^n e^{-\left(\frac{w^{-\alpha}}{\Gamma(1-\alpha)}\right)}}{n!} e^{-w} dw$$

The counter cumulative distribution has a tail index of $1/\alpha$.

Figure 2b shows the unconditional distribution of customers for different values of α. With a larger intermediates share, the distribution has a thicker tail, as the high-efficiency firms are more likely to attract a disproportionate number of customers. In other words, when the intermediate goods share is large, the equilibrium network features more star suppliers – firms with many downstream customers.

Both the conditional means and unconditional distribution of number of customers depend on a single parameter, α.[23] Notably absent from in the distribution of number of customers is

[22] There is a single crossing property: curves for different values of α in Figure 2a cross exactly once.

[23] Proposition 4 shows this for the unconditional distribution. To see this for the conditional distributions, let p be a percentile in the distribution, so that $p = F(q) = e^{-\theta q^{-\zeta}}$. Then the average number of customers among plants at the p percentile of the efficiency distribution is $\frac{m}{\theta} q^{\alpha\zeta} = \frac{\left[\theta q^{-\zeta}\right]^{-\alpha}}{\Gamma(1-\alpha)} = \frac{\left[\log(1/p)\right]^{-\alpha}}{\Gamma(1-\alpha)}$.

the parameter ζ which governs variation in firms' efficiencies. In particular, if ζ were large there would be little variation in firms' marginal costs, yet the distribution of customers across firms will have a Pareto tail with tail index $1/\alpha$, and lower cost firms will have a disproportionate share of customers.[24]

Going back to Rosen (1981), the superstar literature studies how small differences in talent can lead to large differences in compensation. The key factor is the relationship between differences in talent and how much customers are willing to trade off differences in talent for a lower price.

While the analogy is not perfect, similar forces are at play here. Aside from the share of intermediate goods, the distribution of customers depends on the relationship between differences in cost of inputs and customers' willingness to switch to lower cost suppliers. The latter depends on variation in technique-specific productivity draws. If all techniques had the same productivity, each firm would select the technique with the least expensive inputs, and the distribution of customers would be heavily skewed toward the most efficient firms. In contrast, if all inputs had the same cost, each firm would choose the technique with the best fit (highest z), and the distribution of customers would be Poisson. Thus what matters is the level of variation in productivity draws relative to variation in suppliers' marginal costs.

The parameter ζ drives variation in both. When ζ is larger so that there is less variation in productivity draws, there is also less variation in suppliers' cost; those costs depend on the suppliers' upstream techniques, which are themselves determined by productivity draws. With the functional forms, these two channels offset exactly in determining the distribution of customers.[25]

Also absent is the density of the network, m. While a higher m raises the number of potential

[24] Foster et al. (2008) provide evidence that dispersion in size is significantly larger than dispersion in marginal costs. Using data from US Census of Manufactures, they focus on industries with homogenous goods so that comparisons of units and marginal cost are meaningful. They report that the standard deviation of log physical output is four times larger than the standard deviation of log physical TFP. Because they use common materials price deflators, their measure of TFPQ is the appropriate analog to a firm's efficiency.

[25] As mentioned above, the distribution of efficiencies inherits the tail of the distribution of productivity draws.

customers for any firm, it also raises the number of alternatives available to each of those potential customers. Because of the regularly varying tail of productivity draws, these offset each other. In particular, the functional form assumptions allow for a separation of the level of the economy (which depends on m) from cross-sectional patterns.

3.2 The Distribution of Employment

The results in Section 3.1 can be used to characterize the distribution of employment.[26] An individual's employment depends on how much it produces for the final consumer and how much it produces for other firms. The latter depends on both the number of actual customers and the size of each of those customers (which depends on how many customers they have, etc.). However, because an individual firm's employment depends on the size of each of its customers, the distribution can be characterized recursively.

Because a firm's employment is the sum of several components—the labor used to produce inputs for each of its customers and for the final consumer—rather than working with the CDF of the size distribution, $\mathcal{L}$, it will be easier to work with its Laplace-Stieltjes transform, $\hat{\mathcal{L}}(s) \equiv \int_0^\infty e^{-sl} d\mathcal{L}(l)$. Proposition 5 characterizes this transform and its implications for the size distribution. The proof is given in Appendix B.1.

Proposition 5. *Let $\hat{\mathcal{L}}(s)$ be the Laplace-Stieltjes transform of the distribution of employment. With the functional form assumptions, this satisfies:*

$$\hat{\mathcal{L}}(s) = \int_0^\infty \exp\left\{ -s \frac{t^{-\frac{\varepsilon-1}{\zeta}}}{\Gamma\left(1 - \frac{\varepsilon-1}{\zeta}\right)} (1-\alpha) L - \frac{t^{-\alpha}}{\Gamma(1-\alpha)} \left[1 - \hat{\mathcal{L}}(\alpha s)\right] - t \right\} dt \tag{9}$$

[26] The size distribution of firms/establishments has been a central object of interest, studied by, among others, Hopenhayn (1992), Klette and Kortum (2004), Luttmer (2007), Rossi-Hansberg and Wright (2007), and Arkolakis (2011)

Suppose $\rho \equiv \min\left\{\frac{1}{\alpha}, \frac{1}{(\varepsilon-1)/\zeta}\right\}$ *is not an integer. Then*

$$1 - \mathcal{L}(l) \sim K l^{-\rho}$$

where $K \equiv \frac{1}{1-\alpha^{\rho}}\left(\frac{(1-\alpha)\mathbb{I}_{\rho=\frac{\varepsilon-1}{\zeta}}+\alpha\mathbb{I}_{\rho=\alpha}}{\Gamma(1-\rho)}\right)^{\rho}$

The first part of Proposition 5 gives several implications. First, as L changes, the distribution of labor scales uniformly. Second the shape of distribution depends on only two parameters, α and $\frac{\varepsilon-1}{\zeta}$. The share of intermediates matters for the same reason as before; it determines the distribution of customers. $\frac{\varepsilon-1}{\zeta}$ is a composite of two parameters, the elasticity of substitution in final consumption, and ζ, the tail index of both the productivity distribution (H) and the efficiency distribution F. In combination, these parameters determine the tail index of the distribution of final consumption. When ζ is small, the efficiency distribution has a thicker tail, inducing a thicker tail in the distribution of final consumption. When ε is high, consumers are more willing to substitute toward low cost goods, also thickening the tail of final consumption.

The second part of Proposition 5 describes the determinants of the upper tail of the size distribution. The slope of the upper tail is governed by either labor used to make intermediate inputs or by labor used to make final consumption. Across firms, the upper tail of labor used to make intermediate inputs for other firms has a Pareto tail with exponent $1/\alpha$, while labor used to make consumption has a Pareto tail of $\zeta/(\varepsilon-1)$. The proposition says that one of these features will dominate and determine the upper tail of the overall size distribution.

Figure 3 shows the how size distribution of the model compare with the size distribution among manufacturing establishments in the US.[27] The data is an extract from the Statistics of US Businesses made available by Rossi-Hansberg and Wright (2007), and corresponds to the distribution

[27] This paper has no theory of the firm. Because firms are more likely to own several steps in a supply chain or produce multiple goods, manufacturing establishments provide a better empirical analogue to individual producers in the model.

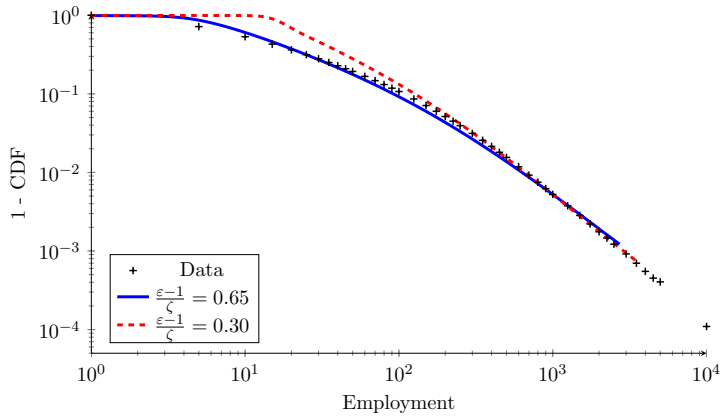


Figure 3: The Distribution of Employment

This figure plots the right CDF of the size distribution a log-log plot to show the shape of the upper tail. The points labeled Data show the fraction of manufacturing establishments in 1990 with employment greater than the specified cutoff, from the SUSB. In each curve, $\alpha = 0.65$ and L is chosen so that all three series intersect when employment is 1000.

of employment across manufacturing establishments in 1990.

In Figure 3, the model is parameterized so that the share of intermediate goods is $\alpha = 0.65$, roughly the cost share of intermediate inputs in the US manufacturing sector. When labor used to produce intermediates dominates the tail of the distribution ($\alpha > \frac{\varepsilon-1}{\zeta}$), the model fits the upper tail of the establishment size distribution well regardless of the value of the composite parameter $\frac{\varepsilon-1}{\zeta}$. The fit toward the bottom of the distribution varies, but since the model has no entry and exit, the model should not be expected to fit that feature well.

3.3 Aggregate Output

Lemma 1 showed that aggregate output in the economy is $Y^0 = QL$. With the functional forms, aggregate productivity is:

$$Q = \left(\int_0^\infty q^{\varepsilon-1} dF(q)\right)^{\frac{1}{\varepsilon-1}} = \theta^{1/\zeta}\Gamma\left(1 - \frac{\varepsilon-1}{\zeta}\right)^{\frac{1}{\varepsilon-1}}$$

Using the expression for θ from equation (7) gives an expression for final consumption:

$$Y^0 = \left[\Gamma\left(1 - \frac{\varepsilon - 1}{\zeta}\right)^{\frac{\zeta}{\varepsilon-1}} \Gamma(1-\alpha)^{\frac{1}{1-\alpha}} m^{\frac{1}{1-\alpha}}\right]^{\frac{1}{\zeta}} L \tag{10}$$

There are several immediate implications. First, aggregate output is increasing in the density of the network, m. In a more densely populated network, firms on average have a larger set of supply chains to choose from and, hence, are more likely to have found an efficient one. The exponent $\frac{1}{\zeta}$ measures the direct contribution of additional techniques output elasticity of output. When ζ is small, there is a thicker tail of productivity draws, so the impact of any single technique is likely to be larger. The term $\Gamma\left(1 - \frac{\varepsilon-1}{\zeta}\right)^{\frac{\zeta}{\varepsilon-1}}$ reflects the representative consumer's ability to consume more of less expensive goods. ε, the elasticity of substitution across varieties, measures how much the final consumer is willing to substitute toward lower cost goods, while ζ, the tail index of distribution of efficiencies, indicates how much cheaper these lower cost goods are.[28]

Second, the share of intermediate goods in production α plays two roles. As in other models with roundabout production, it determines the extent to which lower input prices feed back into lower cost of production. This is the exponent $\frac{1}{1-\alpha}$ that appears in several places.

A separate, more interesting, role is that α determines the composition of the firms supplying intermediate inputs. Recall from Section 3.1 that the intermediate goods share determines the likelihood that the lower-cost producers are *selected* as suppliers. In other words, when α is closer to one, the lowest cost producers are more likely to become star suppliers and be more relevant for aggregate production. Mathematically, this shows up in the term $\Gamma(1-\alpha)$.[29]

In sum, when α is high, each supplier is able to pass through cost savings to its customers at a

[28] These terms are familiar from other studies with Frechet distributed productivities. In the special case in which the share of intermediate goods goes to zero, the input-output relationships play no role and the expression for aggregate output is the same as Kortum (1997).

[29] Note that $\Gamma(x)$ is decreasing on $(0,1)$, with $\lim_{x\searrow 0}\Gamma(x) = \infty$ and $\Gamma(1) = 1$.

higher rate *and* the most efficient firms are selected to be suppliers more frequently, so that these suppliers' low costs can be passed through to a larger share of customers.

4 The Sources of Growth

Over time, some firms discover new techniques while others substitute across techniques in response to changing input prices. For aggregate quantities, all of these changes can be summarized by the evolution of the density of the network, m: $\frac{\dot{Y}}{Y} = \frac{1}{1-\alpha}\frac{1}{\zeta}\frac{\dot{m}}{m}$.

These simple aggregate dynamics hide constant reallocation and reorganization of production. This section discusses the nature of these changes: which firms switch to new techniques, how the cost savings from new techniques diffuse through the network, and how this depends on the existing network structure.

4.1 Dynamics of Efficiency

As an intermediate step, it will be useful to characterize the evolution of firm level efficiency. Following an individual firm poses a challenge as the relevant state of a firm is its position in the network: changes in a firm's efficiency depend on changes in any of its many potential supply chains. Rather than following individual firms, it will be useful to focus on the joint distribution of efficiency at two points in time. Just as the distribution of efficiency was the main tool in characterizing cross-sectional properties of the economy, this joint distribution will be the main tool in characterizing its dynamics.

For $\hat{t} > t$, let $F_{t,\hat{t}}(q, \hat{q})$ be the fraction of firms with efficiency no greater than q at time t and no greater than $\hat{q}$ at time $\hat{t}$. This is the product of the probabilities of two independent events: (i) that all of the firm's upstream techniques discovered before t deliver efficiency no greater than q at t and no greater than $\hat{q}$ at $\hat{t}$; and (ii) that all of the techniques discovered between t and $\hat{t}$ deliver

efficiency no greater than $\hat{q}$ at $\hat{t}$. Each can be characterized recursively with the same method used to characterize the cross-sectional distribution, equation (6).[30]

Proposition 6. *The fraction of firms with efficiency no greater than q at t and no greater than $\hat{q}$ at $\hat{t}$ satisfies*

$$F_{t,\hat{t}}(q,\hat{q}) = e^{-M_t \int_0^\infty \left[1-F_{t,\hat{t}}\left(\left(\frac{q}{z}\right)^{\frac{1}{\alpha}},\left(\frac{\hat{q}}{z}\right)^{\frac{1}{\alpha}}\right)\right] dH(z)} e^{-\left(M_{\hat{t}}-M_t\right)\int_0^\infty \left[1-F_{\hat{t}}\left(\frac{\hat{q}}{z}\right)^{\frac{1}{\alpha}}\right] dH(z)}$$

Proof. Let $G_{t,\hat{t}}(q,\hat{q})$ be the probability that a single technique that exists at t delivers efficiency greater than q at t or greater than $\hat{q}$ at $\hat{t}$. This can be written as

$$G_{t,\hat{t}}(q,\hat{q}) = \int_{z_0}^{\infty} F_{t,\hat{t}}\left(\left(\frac{q}{z}\right)^{\frac{1}{\alpha}}, \left(\frac{\hat{q}}{z}\right)^{\frac{1}{\alpha}}\right) dH(z)$$

To interpret this, $F_{t,\hat{t}}\left(\left(\frac{q}{z}\right)^{\frac{1}{\alpha}}, \left(\frac{\hat{q}}{z}\right)^{\frac{1}{\alpha}}\right)$ is the fraction of suppliers with efficiency low enough at t and $\hat{t}$ so that, in combination with efficiency z, the technique delivers efficiency no greater than q at t and no greater than $\hat{q}$ at $\hat{t}$. The number of techniques discovered by each firm before t follows a Poisson distribution with mean M_t, so the probability that all such techniques deliver efficiency no greater than q at t and no greater than $\hat{q}$ at $\hat{t}$ is is $e^{-M_t\left[1-G_{t,\hat{t}}(q,\hat{q})\right]}$.

Second, a firm discovers techniques between t and $\hat{t}$. The probability that a single one of these techniques delivers efficiency no greater than $\hat{q}$ at $\hat{t}$ is $G_{\hat{t}}(\hat{q}) = \int_{z_0}^{\infty} F_{\hat{t}}\left((\hat{q}/z)^{1/\alpha}\right) dH(z)$. The number of techniques discovered between t and $\hat{t}$ follows a Poisson distribution with mean $M_{\hat{t}} - M_t$, so the probability that the firm has discovered no such techniques is $e^{-(M_{\hat{t}}-M_t)\left[1-G_{\hat{t}}(\hat{q})\right]}$.

$F_{t,\hat{t}}(q,\hat{q})$ is the product of these two terms. ■

[30] This method is easily extended to characterize the joint distribution at any finite number of points in time.

4.2 Persistence of Vertical Relationships

Between t and $\hat{t}$, a firm could keep using the same technique, switch to a technique it was not using, or switch to a newly discovered technique. This section studies how frequently these switches happen, which firms are most likely to switch, and how these are shaped by features of the environment.

Among firms with efficiency q at t, define the following three conditional distributions: Let $R^{same}_{t,\hat{t}}(\hat{q}|q)$ be the probability that the technique used in equilibrium at t delivers efficiency no greater than $\hat{q}$ at $\hat{t}$; let $R^{unused}_{t,\hat{t}}(\hat{q}|q)$ be the probability that among techniques that were available but not used at t, none deliver efficiency greater than $\hat{q}$ at $\hat{t}$; and let $R^{new}_{t,\hat{t}}(\hat{q})$ be the probability that none of the techniques discovered between t and $\hat{t}$ deliver efficiency greater than $\hat{q}$ at $\hat{t}$. These comprise all of the possibilities, so that $F_{t,\hat{t}}(q,\hat{q}) = \int_0^q R^{same}_{t,\hat{t}}(\hat{q}|q')R^{unused}_{t,\hat{t}}(\hat{q}|q')R^{new}_{t,\hat{t}}(\hat{q})dF_t(q')$. Appendix F.2 derives expressions for each of these conditional CDFs.

These conditional distributions can be used to characterize the probability that various firms switch to new techniques. Among firms with efficiency q at t, the hazard of switching to a newly discovered technique is

$$\lim_{\hat{t}\to t} \frac{1}{\hat{t}-t} \int_q^\infty R^{same}_{t,\hat{t}}(\hat{q}|q)R^{unused}_{t,\hat{t}}(\hat{q}|q)dR^{new}_{t,\hat{t}}(\hat{q})$$

To interpret this, $dR^{new}_{t,\hat{t}}(\hat{q})$ is the likelihood that the best newly discovered technique delivers efficiency $\hat{q}$, while $R^{same}_{t,\hat{t}}(\hat{q}|q)R^{unused}_{t,\hat{t}}(\hat{q}|q)$ is the probability that none of the firm's old techniques deliver efficiency greater than $\hat{q}$ at $\hat{t}$. The integral therefore gives the probability that a firm with efficiency q at t uses a newly discovered technique at $\hat{t}$. Similarly, the hazard of switching to an

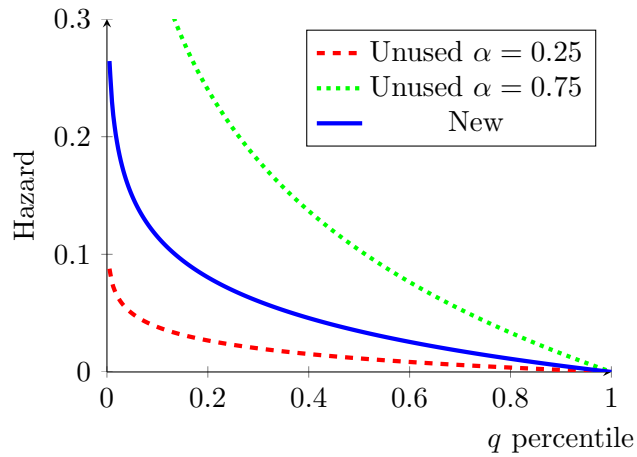


Figure 4: Hazard of Switch

For each percentile in the efficiency distribution, Figure 4 shows the instantaneous hazards of switching to a new technique and to an available but unused technique.

available but unused supplier is

$$\lim_{\hat{t}\to t} \frac{1}{\hat{t}-t} \int_q^\infty R^{same}_{t,\hat{t}}(\hat{q}|q) R^{new}_{t,\hat{t}}(\hat{q}) dR^{unused}_{t,\hat{t}}(\hat{q}|q)$$

With the functional form assumptions, each of these depends only on α and the arrival rate of new techniques (relative to the existing stock).

Figure 4 shows these hazards for various percentiles in the productivity distribution. Both hazards are downward sloping, so the higher efficiency firms are less likely to switch to a new supplier. For newly discovered techniques this is intuitive: the higher a firm's initial efficiency, the lower the probability that a new technique delivers efficiency high enough to induce the firm to switch. For unused techniques, the logic is similar though less obvious, as high efficiency firms typically have more unused techniques than low efficiency firms.

Since larger firms tend to have higher efficiencies, we have the following two testable predictions. First, larger and higher efficiency firms firms will have longer relationships with their suppliers. Second, a firm is likely to have longer relationships with customers that have higher efficiency.

Each of these matches the findings of Carvalho and Gofman (2012), who use a proprietary database on supplier-customer relationships based on information from SEC filings, press releases, websites, and other sources. They find that both higher productivity of the customer and of the supplier predict longer vertical relationships. They also find that among a supplier's customers, the hazard of the relationship ending is lower when the customer is more productive.

When the share of intermediate goods in production is larger, the total hazard of switching is higher. The hazard of switching to a newly discovered technique is actually independent of α; it only depends on the growth rate of the total number of techniques.[31] In contrast, the hazard of switching to an available but unused technique is increasing in α. The efficiency delivered by an existing technique improves only when the supplier lowers its price. When the intermediate goods share is small, any such changes have a small impact on the efficiency of the technique. This is most clear when $\alpha = 0$, an extreme case in which no firm ever switches to an unused technique. Putting these together, the overall hazard of switching is increasing in α.

4.3 Increases in Efficiency

There are three ways a firm's efficiency can increase between t and $\hat{t}$: it switches to a newly discovered technique with higher efficiency; it switches to a technique that is now more efficient because the associated input is cheaper; it keeps using the same supplier and the supplier lowers its price. This section asks how much of firms' increases in efficiency comes from each of these three channels.

Let $J_t(q)$ be the set of firms with efficiency q at t. Among such a set of firms, I will use the following formula to measure the average instantaneous increase in efficiency relative to the increase

[31] To see this among the whole population of firms: at $\hat{t}$, the efficiency delivered by each newly discovered technique follows the same distribution as the efficiency delivered by each technique discovered prior to t. Each of those techniques is equally likely to be the firm's best at $\hat{t}$, so the fraction of firms using techniques discovered after t is equal to the fraction of techniques discovered after t.

in aggregate productivity:

$$\lim_{\hat{t}\to t} \frac{\frac{1}{F_t'(q)} \int_{j\in J_t(q)} \left[\left(q_{j,\hat{t}}/q \right)^{\varepsilon-1} - 1 \right] dj}{\left(Q_{\hat{t}}/Q_t\right)^{\varepsilon-1} - 1} \tag{11}$$

This can be interpreted as the decline in the price index in sales to the final consumer for that subset of firms relative to the decline in the aggregate price index.[32] A convenient feature of this measure is that the cost reduction among all firms with efficiency q can be cleanly decomposed into each of the three channels. Let $J^{same}_{t,\hat{t}}(q)$ be the subset of $J_t(q)$ that use the same technique at t and $\hat{t}$. Similarly, let $J^{unused}_{t,\hat{t}}(q)$ and $J^{new}_{t,\hat{t}}(q)$ be the subsets of those firms that switch to available but unused techniques and to newly discovered techniques respectively. The numerator of equation (11) can be written as the sum of three separate integrals over each of these three subsets. Each of these can be computed in the same way. For example,

$$\int_{j\in J^{same}_{t,\hat{t}}(q)} \left[\left(\frac{q_{j,\hat{t}}}{q} \right)^{\varepsilon-1} - 1 \right] dj = \int_q^\infty \left[\left(\frac{\hat{q}}{q} \right)^{\varepsilon-1} - 1 \right] R^{unused}_{t,\hat{t}}(\hat{q}|q) R^{new}_{t,\hat{t}}(\hat{q}) dR^{same}_{t,\hat{t}}(\hat{q}|q)$$

Figure 5a shows the average instantaneous increase in efficiency among firms at each percentile of the efficiency distribution. This is decomposed into the increases coming from each of the three channels.

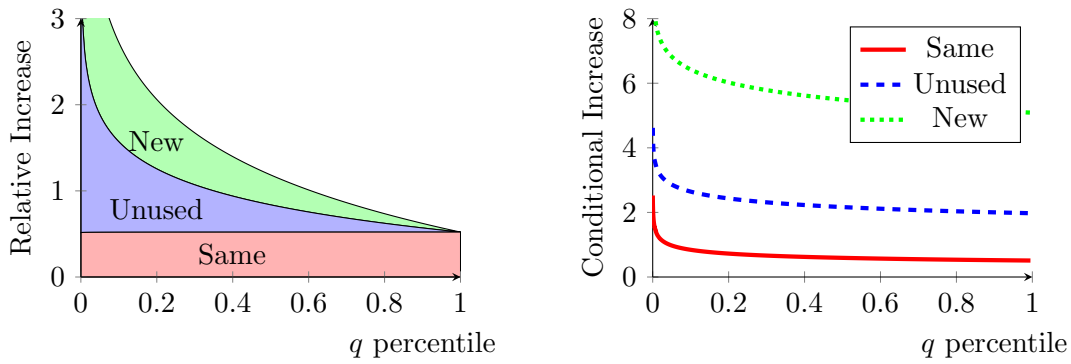
Overall, low efficiency firms typically experience larger increases in efficiency. Consistent with Figure 4 these firms derive most of their gains from switching to either newly discovered techniques or to unused suppliers. In contrast, high efficiency firms are unlikely to find new techniques that are worth switching to.

Notably low and high efficiency firms that keep the same supplier see, on average, the same

[32] To see this, the decline in the price index for $J_t(q)$ relative to the decline in the aggregate price index is

$$\lim_{\hat{t}\to t} \frac{\ln\left(\int_{j\in J_t(q)} q_{j,\hat{t}}^{\varepsilon-1} dj\right)^{\frac{1}{\varepsilon-1}} - \ln\left(\int_{j\in J_t(q)} q_{j,t}^{\varepsilon-1} dj\right)^{\frac{1}{\varepsilon-1}}}{\ln Q_{\hat{t}} - \ln Q_t} = \lim_{\hat{t}\to t} \frac{\left(\int_{j\in J_t(q)} q_{j,\hat{t}}^{\varepsilon-1} dj - \int_{j\in J_t(q)} q_{j,t}^{\varepsilon-1} dj\right) / \int_{j\in J} q_{j,t}^{\varepsilon-1} dj}{\left(Q_{\hat{t}}^{\varepsilon-1} - Q_t^{\varepsilon-1}\right) / Q_t^{\varepsilon-1}}$$

equation (11) follows because each $q_{j,t} = q$.



(a) Increase in Efficiency

(b) Conditional Increase in Efficiency

Figure 5: Increases in Efficiency

For each percentile in the efficiency distribution, Figure 5a shows average instantaneous increase in efficiency relative to the increase in aggregate productivity, decomposed into the three channels. For each percentile, Figure 5b shows the average increase in cost conditional on switching to a new supplier, unused supplier, or keeping the same supplier when the average number of techniques has grown by 10%. In each, $\frac{\varepsilon-1}{\zeta} = \frac{1}{2}$ and $\alpha = 0.75$.

increases in efficiency. This stems from the fact that, given the functional forms, there is no difference on average between suppliers of low and high efficiency firms.

Figure 5b shows how firms that switch suppliers fare relative to those that keep the same supplier when the average number of techniques grows by 10%. At all levels of efficiency, firms that switch to either new or unused techniques experience larger increases in efficiency than firms that keep the same supplier. The relationship is even stronger without conditioning on efficiency, as low efficiency firms experience larger increases and are more likely to switch.[33] Consequently, we have the additional testable prediction that switching suppliers is associated with increases in size and efficiency.

This also matches the findings of Carvalho and Gofman (2012), who find that turnover of suppliers is related to increases in sales, employment, and labor productivity.

[33] When $\hat{t} - t$ is small, the increase in efficiency conditional on switching suppliers is significantly larger than the increase among those that keep the same supplier. While the total instantaneous cost reductions from each of the three channels are of the same order of magnitude, the fraction of firms that switch to new suppliers is proportional to $\hat{t} - t$. Since there are very few switchers when $\hat{t} - t$ is small, the average increase in efficiency among the switchers is extremely large.

4.4 Contribution to Growth

Section 4.3 showed the sources of cost reductions for firms of various levels of efficiency. This section studies how those cost reductions diffuse through the network. To get at this, we define a firm's surplus to be its direct and indirect contribution to aggregate productivity. Specifically, the surplus created by firm j includes the consumption value of good j, plus the consumption value of j's contribution to raising other firms' efficiencies.

Let $S(q)$ be the average surplus created by firms with efficiency q. This can be characterized recursively:

$$S(q) = q^{\varepsilon-1} + M \int_{z_0}^{\infty} \int_{0}^{\infty} \left[S\left(\max\{\tilde{q}, zq^{\alpha}\}\right) - S(\tilde{q}) \right] d\tilde{F}(\tilde{q}) dH(z)$$

where $\tilde{F}(\tilde{q}) = \frac{F(\tilde{q})-e^{-M}}{G(\tilde{q})(1-e^{-M})}$ is the CDF of a firm's best alternative technique, derived in equation (13). The first term is the direct contribution to consumption. To interpret the second term, consider a single downstream technique that uses the firm's good as an input. Suppose that a technique had productivity z and was discovered by a buyer whose best other technique delivered efficiency $\tilde{q}$. With that technique, the buyer's efficiency would be $\max\{\tilde{q}, zq^{\alpha}\}$, so that the change in the surplus of the buyer would be $S\left(\max\{\tilde{q}, zq^{\alpha}\}\right) - S(\tilde{q})$. The expected surplus from a single downstream technique is can be found by integrating over the productivity of the downstream technique and the efficiency of the buyer's best alternative technique. Lastly, the expected surplus from all downstream techniques is the expected number of downstream techniques, M, multiplied by the expected surplus from each.

Appendix F.3 shows that with the functional forms we can derive a closed form expression for the average surplus among firms with efficiency q:

$$S(q) = q^{\varepsilon-1} + \frac{mq^{\zeta\alpha}}{\theta} \frac{1}{1-\alpha} \frac{\varepsilon-1}{\zeta} Q^{\varepsilon-1}$$

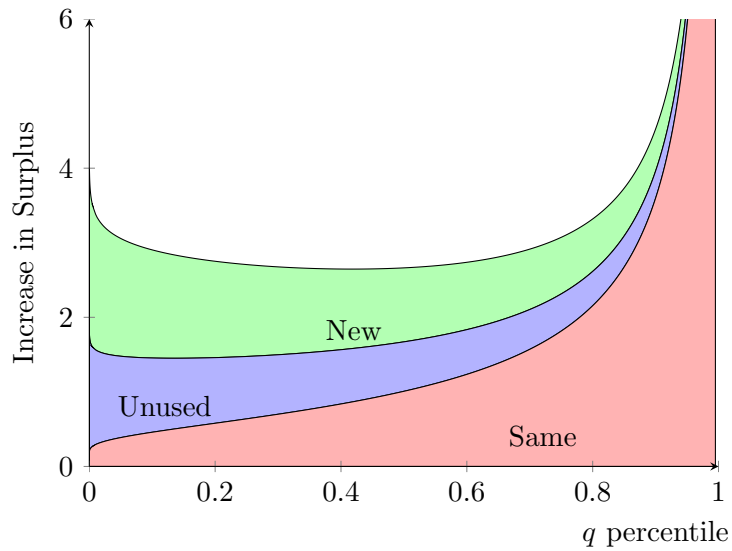


Figure 6: Change in Surplus

For each percentile in the efficiency distribution, Figure 6 shows average increase in surplus from increases in efficiency, decomposed into the three channels. In the figure, $\frac{\varepsilon-1}{\zeta} = \frac{1}{2}$ and $\alpha = 0.75$.

The first part is the firm's direct contribution to consumption, while the second part is the expected number of actual customers (given the firm's efficiency) multiplied by the contribution to the customers' surpluses. The term $\frac{\varepsilon-1}{\zeta}$ is average increase in a customer's direct contribution to consumption relative to that customer's next best option. The multiplier $\frac{1}{1-\alpha}$ accounts for the fact that increasing the customer's efficiency raises its customers' efficiencies, etc.

Figure 6 shows the average change in surplus from instantaneous increases in efficiency relative to the change in aggregate productivity.[34] In particular, the figure shows the contribution of each of the three channels to increasing surplus.

While the high efficiency firms have the smallest increases in efficiency, these cost reductions generate the most surplus as these firms pass on their cost savings to so many customers. Thus while these star suppliers do not switch to new techniques frequently, they play an especially large role in the diffusion of cost savings.

[34]Note that the cumulative increase in surplus shown in the graph integrates to a number larger than one. This is because there is double counting. For example, a new technique might lower the cost of several downstream firms. The graph shows the increase in surplus from each separately.

In contrast, lower efficiency firms have fewer customers to whom increases in efficiency can be passed on. However, the change in their surplus is likely to be large because the increases in efficiency are so large.

4.4.1 The Surplus of a New Technique

In the model, all growth in aggregate productivity comes either directly or indirectly from the arrival of new techniques. This section shows how the value of a new technique depends on the existing network structure.

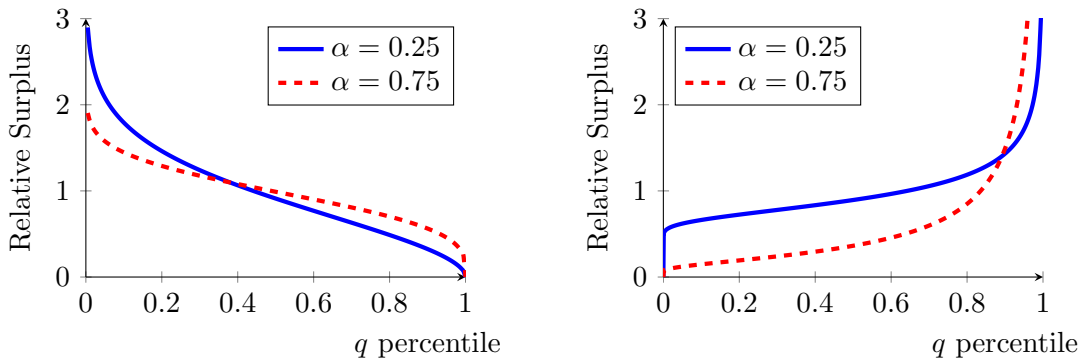
We first characterize the contribution of a new upstream technique. For a firm with productivity q_0, a new upstream technique that provides efficiency q_1 would generate additional surplus of $\max\{0, S(q_1) - S(q_0)\}$, so that the expected value of a single new technique is $\int_{q_0}^{\infty} [S(q_1) - S(q_0)]\, dG_t(q_1)$.

The surplus created by a new downstream technique depends on the productivity of the technique and the potential buyer's best alternative technique. For a firm with efficiency q, the surplus generated by an additional downstream technique is $\int_{z_0}^{\infty} \int_0^{qz^{\alpha}} [S(zq^{\alpha}) - S(\tilde{q})]\, dF_t(\tilde{q}) dH(z)$.

The surplus of an additional upstream technique is inversely related to efficiency. This is fairly intuitive; the higher a firm's initial efficiency, the less there is to gain from a new technique. In contrast, the surplus created by a new downstream technique is positively related to a firm's efficiency, as the new technique is more likely to be useful to potential buyer.

With the functional forms, there are closed form expressions for each. Appendix F.3.1 shows that relative to the average across all firms, the surplus generated by a new upstream technique for a firm with efficiency q is $(1-\alpha)\frac{\left(\theta q^{-\zeta}\right)^{1-\frac{\varepsilon-1}{\zeta}}}{\Gamma(2-\frac{\varepsilon-1}{\zeta})} + \alpha\frac{\left(\theta q^{-\zeta}\right)^{1-\alpha}}{\Gamma(2-\alpha)}$. Relative to the average across all firms, the surplus generated by a new downstream technique for a firm with efficiency q is $\frac{m}{\theta} q^{\zeta\alpha} = \frac{(\theta q^{-\zeta})^{-\alpha}}{\Gamma(1-\alpha)}$.

Figure 7 shows how the expected surplus of a new technique depends on efficiency of the firm that discovered it.



(a) Upstream

(b) Downstream

Figure 7: Value of a New Technique

For each percentile in the efficiency distribution, Figure 7a shows the expected surplus of a single upstream technique relative to the average expected surplus. Figure 7b shows the same for a single downstream technique. In each, $\frac{\varepsilon-1}{\zeta} = \frac{1}{2}$.

When the share of intermediate goods is larger, the difference in value of a new downstream technique between high and low efficiency firms is magnified. Since the price of the input matters more, there is a bigger difference in the efficiency the technique would deliver.

5 Conclusion

In this paper I have developed a theory of the formation of chains of production. An economy's productive capacity rests on technological relationships among firms. A firm creates value when it finds a supplier that can provide a useful input at a low price.

Despite the network structure, the model is analytically tractable. Expressions are given for aggregate productivity and for cross-sectional distributions of several firm-level characteristics. This allows for a transparent characterization of the channels through which deep parameters alter aggregate productivity and the organization of production.

In the model, there is a meaningful distinction between efficiency and demand. These jointly determine a firm's size and contribution to aggregate productivity. Aggregate productivity depends on how frequently the lower-cost firms are selected as suppliers. To be selected frequently, a firm

must have many potential customers who have techniques that use the firm's good as an input and those techniques must be cost-effective. When intermediate goods are more important in production, the ability to charge a low price becomes more important in winning customers, and the lower-cost firms are more likely to become star suppliers. This increases aggregate productivity and also increases the market concentration in sales of intermediate goods.

References

Acemoglu, Daron, Vasco M. Carvalho, Asuman E. Ozdaglar, and Alireza Tahbaz-Salehi, "The Network Origins of Aggregate Fluctuations," Technical Report January 2012.

Alvarez, Fernando E., Francisco J. Buera, and Jr. Robert E. Lucas, "Models of Idea Flows," NBER Working Papers 14135, National Bureau of Economic Research, Inc June 2008.

Arkolakis, Costas, "A Unified Theory of Firm Selection and Growth," NBER Working Papers 17553, National Bureau of Economic Research, Inc October 2011.

Atalay, Enghin, Ali Hortacsu, James Roberts, and Chad Syverson, "Network Structure of Production," *Proceedings of the National Academy of Sciences*, 03 2011, *108* (13), 5199–202.

Athreya, Krishna B. and Peter Ney, *Branching processes*, Springer-Verlag, Berlin, New York,, 1972.

Barabasi, Albert-Laszlo and Reka Albert, "Emergence of Scaling in Random Networks," *Science*, 10 1999, *286*, 509–512.

Bingham, NH and RA Doney, "Asymptotic properties of supercritical branching processes I: The Galton-Watson process," *Advances in Applied Probability*, 1974, pp. 711–731.

Caliendo, L. and F. Parro, "Estimates of the Trade and Welfare Effects of NAFTA," Technical Report 2011.

Carvalho, Vasco and Micahel Gofman, "Structure and Change in Production Networks: Evidence from US Firm Level Data," Technical Report 2012.

Chaney, Thomas, "The Network Structure of International Trade," NBER Working Papers 16753, National Bureau of Economic Research, Inc January 2011.

Dupor, Bill, "Aggregation and irrelevance in multi-sector models," *Journal of Monetary Economics*, April 1999, *43* (2), 391–409.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, September 2002, *70* (5), 1741–1779.

Foerster, Andrew T., Pierre-Daniel G. Sarte, and Mark W. Watson, "Sectoral versus Aggregate Shocks: A Structural Factor Analysis of Industrial Production," *Journal of Political Economy*, 2011, *119* (1), 1 – 38.

Foster, L., J. Haltiwanger, and C. Syverson, "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, 2008, *98* (1), 394–425.

Hatfield, J.W., S.D. Kominers, A. Nichifor, M. Ostrovsky, and A. Westkamp, "Stability and competitive equilibrium in trading networks," 2012.

Hopenhayn, Hugo A., "Entry, Exit, and firm Dynamics in Long Run Equilibrium," *Econometrica*, 1992, *60* (5), pp. 1127–1150.

Horvath, Michael, "Cyclicality and Sectoral Linkages: Aggregate Fluctuations from Independent Sectoral Shocks," *Review of Economic Dynamics*, October 1998, *1* (4), 781–808.

Jackson, Matthew O., "A Survey of Models of Network Formation: Stability and Efficiency," Game Theory and Information 0303011, EconWPA March 2003.

_ , *Social and Economic Networks*, Princeton, NJ, USA: Princeton University Press, 2008.

Jones, Charles I., "Intermediate Goods, Weak Links, and Superstars: A Theory of Economic Development," NBER Working Papers 13834, National Bureau of Economic Research, Inc March 2008.

Kelly, Morgan, "The Dynamics of Smithian Growth," *The Quarterly Journal of Economics*, August 1997, *112* (3), 939–64.

_ , "The Invention of Invention," Working Papers 200515, School Of Economics, University College Dublin September 2005.

Klette, T.J. and S. Kortum, "Innovating firms and aggregate innovation," *Journal of Political Economy*, 2004, *112* (5), 986–1018.

Kortum, Samuel S., "Research, Patenting, and Technological Change," *Econometrica*, November 1997, *65* (6), 1389–1420.

Long, John B and Charles I Plosser, "Real Business Cycles," *Journal of Political Economy*, February 1983, *91* (1), 39–69.

Lucas, Robert E., "Ideas and Growth," *Economica*, 02 2009, *76* (301), 1–19.

Luttmer, Erzo G. J., "Selection, Growth, and the Size Distribution of Firms," *The Quarterly Journal of Economics*, 08 2007, *122* (3), 1103–1144.

Ostrovsky, M., "Stability in supply chain networks," *The American Economic Review*, 2008, pp. 897–923.

Rosen, Sherwin, "The economics of superstars," *The American economic review*, 1981, pp. 845–858.

Rosenberg, Nathan, "Technological Interdependence in the American Economy," *Technology and Culture*, 1979, *20* (1), pp. 25–50.

Rossi-Hansberg, Esteban and Mark L. J. Wright, "Establishment Size Dynamics in the Aggregate Economy," *American Economic Review*, December 2007, *97* (5), 1639–1666.

Uhlig, Harald, "A Law of Large Numbers for Large Economies," *Economic Theory*, 1996, *8* (1), 41–50.

Weitzman, M.L., "Recombinant growth," *The Quarterly Journal of Economics*, 1998, *113* (2), 331–360.

Willmot, Gordon E., "Asymptotic Tail Behaviour of Poisson Mixtures with Applications," *Advances in Applied Probability*, 1990, *22* (1), pp. 147–159.

Appendix

A The Solution to the Planner's Problem

The strategy begins with defining a sequence of random variables $\{X_N\}_{N\in\mathbb{N}}$ with the property that the maximum feasible efficiency of a firm is given by the limit of this sequence, if such a limit exists. We then show that X_N converges to a random variable X^{sp} in $L^{\varepsilon-1}$. Next we show that the CDF of X^{sp} is the unique fixed point of T in $\mathcal{F}$, a subset of $\bar{\mathcal{F}}$ (and that such a fixed point exists). Letting F^{sp} be this fixed point, the law of large numbers implies the CDF of the cross-sectional distribution of efficiencies is F^{sp} and that aggregate productivity is $\|X^{sp}\|_{\varepsilon-1}$.

A.1 Existence of a Fixed Point of Equation (6)

We begin by defining three functions, $\bar{f}$, f^1, and $\underline{f}$, in $\bar{\mathcal{F}}$. To do so, we define several objects that will parameterize these functions. Let $\rho \in (0,1)$ be the smallest root of $\rho = e^{-M(1-\rho)}$. In the definition of $\bar{f}$, let $\beta > \varepsilon - 1$ be such that $\lim_{z\to\infty} z^{\beta}[1-H(z)] = 0$. Then there exists a $z_2 > 1$ such that $z > z_2$ implies $z^{\beta}[1-H(z)] < (1-\alpha)$. With this, define q_2 so that $q_2^{(1-\alpha)\beta} > M\left(z^{\beta}+1\right)$, which also implies $q_2^{1-\alpha} > z_2$. In the definition of $\underline{f}$, $q_0 = z_0^{\frac{1}{1-\alpha}}$.

$$
\begin{aligned}
\bar{f}(q) &\equiv \begin{cases} \rho, & q < q_2 \\ 1-(1-\rho)\left(\frac{q}{q_2}\right)^{-\beta}, & q \geq q_2 \end{cases} \\
f^1(q) &\equiv \begin{cases} \rho, & q < 1 \\ 1, & q \geq 1 \end{cases} \\
\underline{f}(q) &\equiv \begin{cases} \rho, & q < q_0 \\ 1, & q \geq q_0 \end{cases}
\end{aligned}
$$

On $\bar{\mathcal{F}}$, the set of right continuous, weakly increasing functions $f : \mathbb{R}^+ \to [0,1]$, consider the partial order given by the binary relation $\preceq$: $f_1 \preceq f_2 \Leftrightarrow f_1(q) \leq f_2(q), \forall q \geq 0$. Clearly $\bar{f} \preceq f^1 \preceq \underline{f}$. Let $\mathcal{F} \subset \bar{\mathcal{F}}$ be the

subset of set of nondecreasing functions $f : \mathbb{R}^+ \to [0,1]$ that satisfy $\bar{f} \preceq f \preceq \underline{f}$.

Lemma 2. *$T\underline{f} \preceq \underline{f}$ and $\bar{f} \preceq T\bar{f}$*

Proof. We first show $T\underline{f} \preceq \underline{f}$. For $q \geq q_0$, $T\underline{f}(q) \leq 1 = \underline{f}(q)$. For $q < q_0$:

$$T\underline{f}(q) = e^{-M\int_0^\infty \left[1-\underline{f}\left((q/z)^{1/\alpha}\right)\right]dH(z)} = e^{-M\int_{q/q_0^\alpha}^\infty [1-\rho]dH(z)} \leq e^{-M[1-\rho]\left(1-H\left(q_0^{1-\alpha}\right)\right)} = \rho = \underline{f}(q)$$

We proceed to $\bar{f}$. First, for $q < q_2$, we have $T\bar{f}(q) = e^{-M\int_0^\infty (1-\bar{f})dH(z)} \geq e^{-M(1-\rho)} = \rho = f(q)$. Next, as an intermediate step, we will show that for $q \geq q_2$:

$$\int_{z_0}^{q/q_2^\alpha} z^{\frac{\beta}{\alpha}} dH(z) + (q/q_2^\alpha)^{\frac{\beta}{\alpha}} \left[1 - H\left(q/q_2^\alpha\right)\right] < \left(z_2^\beta + 1\right)(q/q_2^\alpha)^{\frac{1-\alpha}{\alpha}\beta} \tag{12}$$

To see this note we can integrate by parts to get

$$\int_{z_2}^{q/q_2^\alpha} z^{\beta/\alpha} dH(z) = \left[1 - H\left(z_2\right)\right] z_2^{\beta/\alpha} - (q/q_2^\alpha)^{\beta/\alpha}\left[1 - H\left(q/q_2^\alpha\right)\right] + \int_{z_2}^{q/q_2^\alpha} \frac{\beta}{\alpha} z^{\beta/\alpha - 1}\left[1 - H(z)\right] dz$$

Rearranging this gives

$$H\left(z_2\right) z_2^{\beta/\alpha} + \int_{z_2}^{q/q_2^\alpha} z^{\beta/\alpha} dH(z) + (q/q_2^\alpha)^{\beta/\alpha}\left[1 - H\left(q/q_2^\alpha\right)\right] = z_2^{\beta/\alpha} + \frac{\beta}{\alpha}\int_{z_2}^{q/q_2^\alpha} z^{\beta/\alpha - 1}\left[1 - H(z)\right] dz$$

Since $q/q_2^\alpha > z_2$, equation (12) follows from this and three inequalities: (i) $\int_{z_0}^{z_2} z_2^{\beta/\alpha} dH(z) \leq H\left(z_2\right) z_2^{\beta/\alpha}$; (ii) $z_2^{\beta/\alpha} \leq z_2^\beta \left(q/q_2^\alpha\right)^{\beta/\alpha - \beta}$; and (iii) $\int_{z_2}^{q/q_2^\alpha} z^{\beta/\alpha - 1}\left[1 - H(z)\right] dz \leq \int_0^{q/q_2^\alpha} z^{\beta/\alpha - 1}\left[(1-\alpha)z^{-\beta}\right] dz$.

Next, beginning with $1 - T\bar{f}(q) \leq -\ln T\bar{f}(q)$, we have

$$\begin{aligned}
\frac{1-T\bar{f}(q)}{1-\rho} &\leq M\int_{z_0}^{\infty}\frac{1-f\left((q/z)^{1/\alpha}\right)}{1-\rho}dH(z) \\
&= M\int_{z_0}^{q/q_2^{\alpha}}\left(\frac{(q/z)^{1/\alpha}}{q_2}\right)^{-\beta}dH(z) + M\left[1-H\left(q/q_2^{\alpha}\right)\right] \\
&= \left(\frac{q}{q_2}\right)^{-\beta}\frac{M\left(z^{\beta}+1\right)}{q_2^{(1-\alpha)\beta}}\left\{\frac{\int_{z_0}^{q/q_2^{\alpha}} z^{\frac{\beta}{\alpha}}dH(z) + \left(q/q_2^{\alpha}\right)^{\frac{\beta}{\alpha}}\left[1-H\left(q/q_2^{\alpha}\right)\right]}{\left(z^{\beta}+1\right)\left(q/q_2^{\alpha}\right)^{\frac{1-\alpha}{\alpha}\beta}}\right\} \\
&\leq \left(\frac{q}{q_2}\right)^{-\beta} \\
&= \frac{1-\bar{f}(q)}{1-\rho}
\end{aligned}$$

This then gives, for $q \geq q_2$, $T\bar{f}(q) \geq \bar{f}(q)$. ■

Lemma 3. *There exist least and greatest fixed points of the operator T in $\mathcal{F}$, given by $\lim_{N\to\infty} T^N\bar{f}$ and $\lim_{N\to\infty} T^N\underline{f}$ respectively.*

Proof. The operator T is order preserving, and $\mathcal{F}$ is a complete lattice. By the Tarski fixed point theorem, the set of fixed points of T in $\mathcal{F}$ is also a complete lattice, and hence has a least and a greatest fixed point given by $\lim_{N\to\infty} T^N\bar{f}$ and $\lim_{N\to\infty} T^N\underline{f}$ respectively. ■

A.2 Existence of a Limit

We begin with some notation. Given the set of all techniques, Φ, we define several objects. A chain is a sequence of techniques (finite or infinite) $\phi_0\phi_1\phi_2...$ with the property $s(\phi_{k+1}) = b(\phi_k)$.

Let $C(j)$ be the set of chains of techniques with the additional property that $b(\phi_0) = j$. These are the distinct chains of techniques that end with j. For example, if firm j has access to a single technique ϕ_0 with supplier j_1, and j_1 has access to two techniques ϕ_1 and ϕ_1', then $C(j)$ contains the three distinct chains ϕ_0, $\phi_0\phi_1$, and $\phi_0\phi_1'$.

Let $C_N(j) \subseteq C(j)$ be the set of chains of length N that end with j. In the example, $C_1(j) = \{\phi_0\}$ and $C_2(j) = \{\phi_0\phi_1, \phi_0\phi_1'\}$.

Lastly, let $C^{\infty}(j) \subseteq C(j)$ be the set of infinite chains that end with j. These are all of the viable supply

chains to produce good j. $C_N^\infty(j) \subseteq C_N(j)$ is the set of distinct chains of length N that form the beginning of an infinite supply chain. For example, $\phi_0\phi_1\phi_2... \in C^\infty(\mathrm{j})$, then $\phi_0\phi_1 \in C_2^\infty(j)$.

We first show that the planer's problem is well defined. For each chain $c \in C^\infty(j)$ and each $n \geq 0$, let $z_n(c)$ be the productivity of the nth technique in chain c. In other words, for a chain c of techniques $\phi_0\phi_1\phi_2...$ with ϕ_0 furthest downstream, $z_n(c) = z(\phi_n)$. Finally let $\mathfrak{q}(c) \equiv \prod_{n=0}^\infty z_n(c)^{\alpha^n}$.

We want to define q_j to be the efficiency provided by the most cost effective supply chain available to produce j, or more formally that $q_j = \sup_{c\in C^\infty(j)} \mathfrak{q}(c)$. To do this, we first argue that for each c, $\mathfrak{q}(c)$ is well defined. The concern is that for some c the sequence $\left\{\prod_{n=0}^N z_n(c)^{\alpha^n}\right\}_{N=0}^\infty$ might not converge. Lemma 4 shows that it does.

Define $\mathfrak{q}_N(c) \equiv \prod_{n=0}^{N-1} z_n(c)^{\alpha^n}$ for $N \geq 1$ and $\mathfrak{q}_0(c) \equiv 1$.

Lemma 4. *Assume $z_0 > 0$. Then for each $c \in C^\infty(j)$, $\lim_{N\to\infty} \mathfrak{q}_N(c)$ exists.*

Proof. For each n, $\log z_n(c)$ can be decomposed into $\log z_n^+(c) - \log z_n^-(c)$, where $\log z_n^+(c) = \max\{\log z_n(c), 0\}$ and $\log z_n^-(c) = \max\{-\log z_n(c), 0\}$, so that $\log \mathfrak{q}_N(c) = \sum_{n=0}^{N-1} \alpha^n \log z_n^+(c) - \sum_{n=0}^{N-1} \alpha^n \log z_n^-(c)$. $\sum_{n=0}^{N-1} \alpha^n \log z_n^+(c)$ is a monotone sequence so it converges to a (possibly infinite) limit. $\sum_{n=0}^{N-1} \alpha^n \log z_n^-(c)$ is a monotone sequence bounded by $\frac{\log(1/z_0)}{1-\alpha}$ so it converges to a limit in the range $\left[0, \frac{\log(1/z_0)}{1-\alpha}\right]$. Thus $\mathfrak{q}_N(c)$ converges to a (possibly infinite) limit. ■

If $C^\infty(j)$ is non-empty, we define the random variable $X_N(j) = \max_{c\in C^\infty(j)} \mathfrak{q}_N(c)$. Roughly, the remainder of this subsection shows that $X_N(j)$ converges to q_j. Since $q_j = \sup_{c\in C^\infty(j)} \lim_{N\to\infty} \mathfrak{q}_N(c)$, we are essentially proving that the limit can be passed through the sup. A useful property of $X_N(j)$ is that its CDF is $T^N f^1$ (the logic is similar to the derivation of equation (6)).

It will also be useful to construct random variables $\bar{Y}_N(j)$ and $\underline{Y}_N(j)$ whose CDFs are $T^N \bar{f}$, and $T^N \underline{f}$ respectively. Toward this, given a realization of Φ, let $\{\tilde{q}(c)\}_{\forall c\in\bigcup_{N=0}^\infty C_N^\infty(j), \forall j\in J}$ be IID random variables, each with CDF $\frac{\bar{f}-\rho}{1-\rho}$.[35] With this we define $\bar{\mathfrak{q}}_N(c) \equiv \mathfrak{q}_N(c)\tilde{q}(c)^{\alpha^N}$ and $\underline{\mathfrak{q}}_N(c) \equiv \mathfrak{q}_N(c)q_0^{\alpha^N}$ (q_0 is the same as a random variable with CDF $\frac{\underline{f}-\rho}{1-\rho}$). Lastly, for $N \geq 1$, let $\bar{Y}_N(j) \equiv \max_{c\in C_N^\infty(j)} \bar{\mathfrak{q}}_N(c)$ and $\underline{Y}_N(j) \equiv \max_{c\in C_N^\infty(j)} \underline{\mathfrak{q}}_N(c)$. Also let $X_0(j) = 1$, $\underline{Y}_0(j) = q_0$, and $\bar{Y}_0(j)$ to have CDF $\frac{\bar{f}-\rho}{1-\rho}$.

[35] $\tilde{q}(c)$ is designed to be the analog of the efficiency of the firm furthest upstream in c, conditioning on the fact that c is viable, if the cross sectional distribution of CDF efficiencies were $\bar{f}$.

If $C^{\infty}(j)$ is empty, then $X_N(j) \equiv 0$, $\bar{Y}_N(j) \equiv 0$, $\underline{Y}_N(j) \equiv 0$ for all $N \geq 0$.

To improve readability, the subscript j will be suppressed when not necessary.

Lemma 5. *$\{X_N\}_{N\in\mathbb{N}}$, $\{\bar{Y}_N\}_{N\in\mathbb{N}}$, and $\{\underline{Y}_N\}_{N\in\mathbb{N}}$ are uniformly integrable in $L^{\varepsilon-1}$.*

Proof. First, recall that $\bar{Y}_0$ is defined so that its CDF is $\bar{f}$. Since the T is order preserving, the relations $T^N f^1 \succeq T^N \bar{f}$ and $T^N \bar{f} \succeq T^{N-1}\bar{f}$ imply that $T^N f^1 \succeq \bar{f}$. As a consequence, $\bar{Y}_0$ first-order stochastically dominates each X_N and $\bar{Y}_N$, and, by the identical argument, $\underline{Y}_N$. Therefore $\mathbb{E}\left|\bar{Y}_0\right|^{\varepsilon-1} = \frac{q_2^{\varepsilon-1}}{1-\frac{\varepsilon-1}{\beta}} < \infty$ serves as a uniform bound on each $\mathbb{E}\left|X_N\right|^{\varepsilon-1}$, $\mathbb{E}\left|\bar{Y}_N\right|^{\varepsilon-1}$, and $\mathbb{E}\left|\underline{Y}_N\right|^{\varepsilon-1}$. ■

Lemma 6. *There exists a random variable X^{sp} such that X_N converges to X^{sp} almost surely and in $L^{\varepsilon-1}$.*

Proof. Let $P_N \equiv \frac{X_N}{\prod_{n=0}^{N}\mu_n}$ where $\mu_n \equiv \int_{z_0}^{\infty} w^{\alpha^n} \frac{M(1-\rho)H'(w)e^{-M(1-\rho)[1-H(w)]}}{1-e^{-M(1-\rho)}}dw$. We first show that $\{P_N\}$ is a submartingale with respect to $\{C_N^{\infty}\}$.

If C_1^{∞} is empty then $\mathbb{E}\left[P_N|C_{N-1}^{\infty}\right] = P_{N-1} = 0$. Otherwise, define a set $\mathcal{D}_N$ as follows: Let $c_N^* \in \arg\max_{c\in C_N^{\infty}} \mathfrak{q}_N(c)$ so that $X_N = \mathfrak{q}_N\left(c_N^*\right)$. Let $\mathcal{D}_N \subseteq C_N^{\infty}$ be the set of chains in C_N^{∞} for which the first $N-1$ links are c_{N-1}^*. In other words, all chains in $\mathcal{D}_N$ are of the form $c_{N-1}^*\phi$ for some ϕ.

Define the random variable $D_N = \max_{c\in\mathcal{D}_N} \mathfrak{q}_N(c)$. Since $\mathcal{D}_N \subseteq C_N^{\infty}$, it must be that $X_N \geq D_N$. We now show that $E\left[D_N|C_{N-1}^{\infty}\right] \geq \mu_N X_{N-1}$:

The probability that $|\mathcal{D}_N| = k$ is $\frac{e^{-M[1-\rho]}[M(1-\rho)]^k}{\left[1-e^{-M[1-\rho]}\right]k!}$ for $k \geq 1$. To see this, note that for any node, the number of techniques is poisson with mean M. Each of those has probability $1-\rho$ of being viable (having a chain that continues infinitely), and we are conditioning on at least one viable technique.

Each of those techniques has a productivity drawn from H. For any ϕ such that $c_{N-1}^*\phi \in \mathcal{D}_N$, we have that

$$\Pr\left(\mathfrak{q}_N\left(c_N^*\phi\right) < x|C_{N-1}^{\infty}\right) = \Pr\left(z(\phi)^{\alpha^n} < x/X_{N-1}\right) = H\left(\left(x/X_{N-1}\right)^{\alpha^{-N}}\right)$$

Given X_{N-1}, if $\mathcal{D}_N$ consists of k chains, the probability that $D_N < x$ is

$$\Pr\left(D_N < x|C_{N-1}^{\infty}, |\mathcal{D}_N| = k\right) = H\left(\left(x/X_{N-1}\right)^{\alpha^{-N}}\right)^k$$

With this, the CDF of D_N, given C_{N-1}^{∞}, is

$$\begin{aligned}
\Pr\left(D_N < x | C_{N-1}^{\infty}\right) &= \sum_{k=1}^{\infty} \Pr\left(D_N < x | X_{N-1}, |\mathcal{D}_N| = k\right) \Pr\left(|\mathcal{D}_N| = k\right) \\
&= \sum_{k=1}^{\infty} H\left(\left(x/X_{N-1}\right)^{\alpha^{-N}}\right)^k \frac{e^{-M[1-\rho]}\left[M(1-\rho)\right]^k}{\left[1-e^{-M[1-\rho]}\right]k!} \\
&= \frac{e^{M[1-\rho]H\left(\left(x/X_{N-1}\right)^{\alpha^{-N}}\right)} - 1}{e^{M[1-\rho]} - 1}
\end{aligned}$$

We can now compute the conditional expectation of D_N (using the change of variables $w = \left(x/X_{N-1}\right)^{\alpha^{-N}}$):

$$\mathbb{E}\left[D_N | C_{N-1}^{\infty}\right] = X_{N-1} \int_0^{\infty} w^{\alpha^N} M(1-\rho) H'(w) \frac{e^{M[1-\rho]H(w)}}{e^{M[1-\rho]} - 1} dw = \mu_N X_{N-1}$$

Putting this together, we have

$$\mathbb{E}\left[P_N | C_{N-1}^{\infty}\right] = \frac{1}{\prod_{n=0}^{N} \mu_n} \mathbb{E}\left[X_N | C_{N-1}^{\infty}\right] \geq \frac{1}{\prod_{n=0}^{N} \mu_n} \mathbb{E}\left[D_N | C_{N-1}^{\infty}\right] = \frac{1}{\prod_{n=0}^{N} \mu_n} \mu_N X_{N-1} = P_{N-1}$$

We next need to show that $\{P_N\}$ is uniformly integrable, i.e., that $\sup_N \mathbb{E}\left[P_N\right] < \infty$. Since $\sup_N \mathbb{E}\left[X_N\right] < \infty$, it suffices to show a uniform lower bound on $\left\{\prod_{n=0}^{N} \mu_n\right\}$. Since each $\mu_n \geq z_0^{\alpha^n}$ and $z_0 < 1$, we have that $\prod_{n=0}^{N} \mu_n \geq \prod_{n=0}^{N} z_0^{\alpha^n} \geq \prod_{n=0}^{\infty} z_0^{\alpha^n} = z_0^{\frac{1}{1-\alpha}}$.

We have therefore established that $\{P_N\}_{N \in \mathbb{N}}$ is a uniformly integrable (in L^1) submartingale, so by the martingale convergence theorem, there exists an P such that P_N converges to P almost surely. By the continuous mapping theorem, there exists an X^{sp} such that X_N converges to X^{sp} almost surely. Since each $X_N^{\varepsilon-1}$ is dominated by the integrable random variable $\bar{Y}_0^{\varepsilon-1}$, by dominated convergence we have that X_N converges to X^{sp} in $L^{\varepsilon-1}$. ■

Lemma 7. *If C^{∞} is nonempty then with probability one, $X^{sp} = \sup_{c \in C^{\infty}} \mathfrak{q}(c)$*

Proof. We first show that $X^{sp} \geq \sup_{c \in C^{\infty}} \mathfrak{q}(c)$ with probability one. Consider any realization of techniques, Φ. For any $\nu > 0$, there exists a $c^* \in C^{\infty}$ such that $\mathfrak{q}(c^*) > \sup_{c \in C^{\infty}} \mathfrak{q}(c) - \nu$. There also exists an N_1 such that $N > N_1$ implies $\mathfrak{q}_N(c^*) > \mathfrak{q}(c^*) - \nu$. Lastly, with probability one there exists an N_2 such that $N > N_2$

implies $X_N < X^{sp} + \nu$. We then have for $N > \max\{N_1, N_2\}$ that

$$X^{sp} > X_N - \nu = \max_{c \in C_N^\infty} \mathfrak{q}_N(c) - \nu \geq \mathfrak{q}_N(c^*) - \nu > \mathfrak{q}(c^*) - 2\nu > \sup_{c \in C^\infty} \mathfrak{q}(c) - 3\nu, \qquad w.p.1$$

This is true for any $\nu > 0$, so $X^{sp} \geq \sup_{c \in C^\infty} \mathfrak{q}(c)$. We next show the opposite inequality. For any N, we have

$$\sup_{c \in C^\infty} \mathfrak{q}(c) \geq \sup_{c \in C^\infty} \mathfrak{q}_N(c) z_0^{\frac{\alpha^N}{1-\alpha}} = X_N z_0^{\frac{\alpha^N}{1-\alpha}}$$

Since this is true for any N and $\lim_{N \to \infty} z_0^{\frac{\alpha^N}{1-\alpha}} = 1$, we can take the limit to get $\sup_{c \in C^\infty} \mathfrak{q}(c) \geq X^{sp}$ with probability one. ■

A.3 Characterization of the Limit

We will show below that $\log \bar{Y}_N - \log \underline{Y}_N$ converges to 0 in probability. Since $X_N \in [\underline{Y}_N, \bar{Y}_N]$, it must be that both $\bar{Y}_N$ and $\underline{Y}_N$ converge to X^{sp} in probability. Convergence in probability implies convergence in distribution, which gives two implications. First, $T^N \bar{f}$ and $T^N \underline{f}$ converge to the same limiting function. Since these are the least and greatest fixed points of T in $\mathcal{F}$, this limiting function, F^{sp}, is the unique fixed point of T in $\mathcal{F}$. Second, since $T^N \bar{f} \preceq T^N f^1 \preceq T^N \underline{f}$, F^{sp} is the CDF of X^{sp}.

We first show that $\log \bar{Y}_N - \log \underline{Y}_N$ converges to zero in probability.

Lemma 8. *If C^∞ is nonempty, then for any $\eta > 1$, $\lim_{N \to \infty} \Pr\left(\bar{Y}_N / \underline{Y}_N > \eta\right) = 0$.*

Proof. Conditional on the set of techniques Φ, we have that for each chain c in C_N^∞ that $\bar{\mathfrak{q}}_N(c)/\underline{\mathfrak{q}}_N(c) = \left(\frac{\tilde{q}(c)}{q_0}\right)^{\alpha^N}$. We therefore have:

$$\Pr\left(\bar{\mathfrak{q}}_N(c)/\underline{\mathfrak{q}}_N(c) \leq \eta | \Phi\right) = \Pr\left(\left(\frac{\tilde{q}(c)}{q_0}\right)^{\alpha^N} \leq \eta | \Phi\right) = \frac{\bar{f}\left(\eta^{\alpha^{-N}} q_0\right) - \rho}{1 - \rho}$$

If $|C_N^\infty|$ is the number of distinct chains of length N that are viable, the probability that every chain $c \in C_N^\infty$

satisfies $\bar{\mathfrak{q}}_N(c)/\underline{\mathfrak{q}}_N(c) \leq \eta$ is $\left(\frac{\bar{f}\left(\eta^{\alpha^{-N}} q_0\right)-\rho}{1-\rho}\right)^{|C_N^\infty|}$ so that

$$\Pr\left(\bar{Y}_N/\underline{Y}_N \leq \eta|\Phi\right) = \left(\frac{\bar{f}\left(\eta^{\alpha^{-N}} q_0\right)-\rho}{1-\rho}\right)^{|C_N^\infty|} \geq \left(\frac{\bar{f}\left(\eta^{\alpha^{-N}} q_0\right)-\rho}{1-\rho}\right)^{|C_N|}$$

A standard result from the theory of branching processes (see Appendix D for a derivation) is that for any x, $\mathbb{E}\left[x^{|C_N|}\right] = \varphi^{(N)}(x)$ where $\varphi^{(N)}$ is the N-fold composition of $\varphi(x) \equiv e^{-M(1-x)}$, the probability generating function for $|C_1|$, and expectations are taken over realizations of Φ. This implies

$$\begin{aligned}
\Pr\left(\bar{Y}_N/\underline{Y}_N \leq \eta\right) &= \mathbb{E}\left[\Pr\left(\bar{Y}_N/\underline{Y}_N \leq \eta|\Phi\right)\right] \\
&\geq \mathbb{E}\left[\left(\frac{\bar{f}\left(\eta^{\alpha^{-N}} q_0\right)-\rho}{1-\rho}\right)^{|C_N|}\right] = \varphi^{(N)}\left(\frac{\bar{f}\left(\eta^{\alpha^{-N}} q_0\right)-\rho}{1-\rho}\right)
\end{aligned}$$

Put differently, $\lim_{N\to\infty} \Pr\left(\bar{Y}_N/\underline{Y}_N > \eta\right) \leq \lim_{N\to\infty} 1 - \varphi^{(N)}\left(\frac{\bar{f}\left(\eta^{\alpha^{-N}} q_0\right)-\rho}{1-\rho}\right)$. We complete the proof by showing $\lim_{N\to\infty} 1 - \varphi^{(N)}\left(\frac{\bar{f}\left(\eta^{\alpha^{-N}} q_0\right)-\rho}{1-\rho}\right) = 0$.

To do this, we first show that for $x \in [0,1]$, $\frac{d}{dx}\varphi^{(N)}(x) \leq M^N$. To see this, note that φ is convex and $\varphi'(1) = M$, so that $\varphi'(x) \leq M$ for $x \leq 1$. In addition, if $x \in [0,1]$ then $\varphi(x) \in (0,1]$, which implies $\varphi^{(N)}(x) \in (0,1]$ for each N. We then have

$$\frac{d}{dx}\varphi^{(N)}(x) = \prod_{n=1}^{N} \varphi'\left(\varphi^{(n-1)}(x)\right) \leq M^N$$

With this, for any x, we can bound $\varphi^{(N)}(x)$ by

$$\varphi^{(N)}(x) = \varphi^{(N)}(1) - \int_x^1 \varphi^{(N)\prime}(w)dw \geq 1 - M^N \int_x^1 dw \geq 1 - M^N\left[1-x\right]$$

Lastly, $\lim_{N\to\infty} M^N\left[1 - \frac{\bar{f}\left(\eta^{\alpha^{-N}} q_0\right)-\rho}{1-\rho}\right] = \lim_{N\to\infty} M^N q_2^\beta\left(\eta^{-\beta\alpha^{-N}} q_0^{-\beta}\right) = 0$. ■

We now come to the main result.

Proposition (1). *There is a unique fixed point of T on $\mathcal{F}$, F^{sp}. F^{sp} is CDF of X^{sp}. Aggregate productivity*

is $Q = \left(\int_0^\infty q^{\varepsilon-1} dF^{sp}(q)\right)^{\frac{1}{\varepsilon-1}}$ *with probability one.*

Proof. If C^∞ is nonempty, the combination of $\log \bar{Y}_N - \log \underline{Y}_N \xrightarrow{p} 0$, $\bar{Y}_N \geq X_N \geq \underline{Y}_N$, and $X_N \xrightarrow{p} X^{sp}$ implies that $\bar{Y}_N \xrightarrow{p} X^{sp}$ and $\underline{Y}_N \xrightarrow{p} X^{sp}$. If C^∞ is empty, then $X^{sp} = 0$, so that $\bar{Y}_N \xrightarrow{p} X^{sp}$ and $\underline{Y}_N \xrightarrow{p} X^{sp}$. Together, these imply that $\bar{Y}_N \xrightarrow{p} X^{sp}$ and $\underline{Y}_N \xrightarrow{p} X^{sp}$ unconditionally.

We first show that there is a unique fixed point, which is also the CDF of X^{sp}. The CDFs of $\bar{Y}_N$ and $\underline{Y}_N$ are $T^N \bar{f}$ and $T^N \underline{f}$ respectively. The least and greatest fixed points of T in $\mathcal{F}$ are $\lim_{N\to\infty} T^N \bar{f}$ and $\lim_{N\to\infty} T^N \underline{f}$ respectively. Convergence in probability implies convergence in distribution, so the least and greatest fixed point are the same, and that the fixed point is the CDF of X^{sp}. Call this fixed point F^{sp}.

Since $\{\bar{Y}_N\}$ and $\{\underline{Y}_N\}$ are uniformly integrable in $L^{\varepsilon-1}$, we have by Vitali's convergence theorem that $\bar{Y}^N \to X^{sp}$ in $L^{\varepsilon-1}$ and $\underline{Y}^N \to X^{sp}$ in $L^{\varepsilon-1}$.

Putting all of these pieces together, we have that the CDF of q_j is F^{sp}. We next show that aggregate productivity is the $Q = \left(\int_0^\infty q^{\varepsilon-1} dF^{sp}(q)\right)^{\frac{1}{\varepsilon-1}}$. For this we simply apply the law of large numbers for a continuum economy of Uhlig (1996). To do this, we must verify that the efficiencies are pairwise uncorrelated. This is trivial: consider two firms, j and i. Since the set of firms in any of j's supply chains is countable, the probability that i and j have overlapping supply chains is zero. The theorem in Uhlig (1996) also requires that the variable in question has a finite variance, and if it does, the then the L^2 integral exists. Here we are interested in the $L^{\varepsilon-1}$ norm, so we require that X^{sp} is $L^{\varepsilon-1}$ integrable. Therefore we have that $Q = \left(\int_0^\infty q^{\varepsilon-1} dF^{sp}(q)\right)^{\frac{1}{\varepsilon-1}}$ with probability one. ■

B Distribution of Customers

Proof of Proposition 3. Consider a single draw of a technique that uses firm i's good as an input. Given the firm's efficiency, q_i, we can compute the probability that the technique is the potential buyer's best available technique. To do this, we first ask: for a potential customer that has drawn a technique that uses i, what is the probability that it has no other techniques that deliver efficiency better than q?

Given the distribution of techniques, a firm will have $n-1$ other techniques with probability $\frac{e^{-M}M^n}{n!(1-e^{-M})}$. The CDF of efficiency delivered by each of these techniques is $G(q)$. Therefore the probability that the

potential buyer has no other technique that delivers better than q, conditioning on having at least one technique, is:

$$\frac{\sum_{n=1}^{\infty}\frac{e^{-M}M^n}{n!}G(q)^{n-1}}{1-e^{-M}} = \frac{1}{G(q)(1-e^{-M})}\left[\sum_{n=0}^{\infty}\frac{e^{-M}M^n}{n!}G(q)^n - e^{-M}\right] = \frac{F(q)-e^{-M}}{G(q)(1-e^{-M})} \tag{13}$$

Among techniques that use i as a supplier, the fraction that deliver efficiency no greater than q is $H\left(\frac{q}{q_i^\alpha}\right)$, with density $\frac{1}{q_i^\alpha}H'\left(\frac{q}{q_i^\alpha}\right)$. The probability that a particular technique is the potential buyer's best technique is thus:

$$\Pr\left(\phi \text{ is used}|q_i\right) \quad = \quad \int_0^\infty \frac{1}{q_i^\alpha}H'\left(\frac{\tilde{q}}{q_i^\alpha}\right)\frac{F(\tilde{q})-e^{-M}}{G(\tilde{q})(1-e^{-M})}d\tilde{q}$$

How many potential customers does a given supplier have? In other words, how many existing techniques would use the firm's good as an input? Across all firms, potential customers follow a Poisson law with mean M. Each one of those techniques has an equal chance of being the potential buyer's best technique. Thus, among suppliers with efficiency q_i the distribution of *actual* customers also follows a Poisson law, with mean

$$M\int_0^\infty \frac{1}{q_i^\alpha}H'\left(\frac{\tilde{q}}{q_i^\alpha}\right)\frac{F(\tilde{q})-e^{-M}}{G(\tilde{q})(1-e^{-M})}d\tilde{q}$$

Using the functional form for H and taking the limit as $z_0 \to 0$ this can be simplified considerably. Since $\lim_{z_0\to 0} e^{-mz_0^{-\zeta}} = 0$ and $\lim_{z_0\to 0} G(q) = 1$, we have

$$\lim_{z_0\to 0} Mz_0^\zeta q_i^{\alpha\zeta}\frac{1}{\theta}\left[1-e^{-\theta(z_0q_i^\alpha)^{-\zeta}}\right] = \frac{m}{\theta}q_i^{\alpha\zeta}$$

So among firms with efficiency q_i, the distribution over the number of customers is a Poisson distribution with mean $\frac{m}{\theta}q_i^{\alpha\zeta}$. ■

Proof of Proposition 4. For a firm with efficiency q, the number of customers follows a Poisson distribution with mean $\frac{m}{\theta}q^{\alpha\zeta}$. Integrating across efficiencies, the unconditional probability that a firm has

n customers is

$$p_n = \int_0^\infty \frac{\left(\frac{m}{\theta} q^{\alpha\zeta}\right)^n e^{-\frac{m}{\theta} q^{\alpha\zeta}}}{n!} dF(q) = \int_0^\infty \frac{\left(\frac{w^{-\alpha}}{\Gamma(1-\alpha)}\right)^n e^{-\left(\frac{w^{-\alpha}}{\Gamma(1-\alpha)}\right)}}{n!} e^{-w} dw$$

The second equality uses $\theta = \Gamma(1-\alpha) m \theta^\alpha$. This can be written as

$$p_n = \int_0^\infty \frac{u^n e^{-u}}{n!} \frac{e^{-[\Gamma(1-\alpha)u]^{-1/\alpha}}}{\Gamma(1-\alpha)^{\frac{1}{\alpha}} \alpha} u^{-\frac{1}{\alpha}-1} du$$

Since $\frac{e^{-[\Gamma(1-\alpha)u]^{-1/\alpha}}}{\Gamma(1-\alpha)^{\frac{1}{\alpha}} \alpha}$ is slowly varying as $u \to \infty$, theorem 2.1 of Willmot (1990) gives

$$p_n \sim \frac{e^{-[\Gamma(1-\alpha)n]^{-1/\alpha}}}{\Gamma(1-\alpha)^{\frac{1}{\alpha}} \alpha} n^{-\frac{1}{\alpha}-1}, \qquad n \to \infty$$

Therefore the counter-cumulative distribution has a tail index of $1/\alpha$. ■

B.1 The Distribution of Employment

Let $\mathcal{L}$ be the CDF of the size distribution, so that $\mathcal{L}(l) = \Pr(l_j \leq l)$. Similarly we can define the CDF of the conditional size distribution: among firms with efficiency q, let $\mathcal{L}(l|q)$ be the fraction with size no greater than l. Let $\hat{\mathcal{L}}(s) \equiv \int_0^\infty e^{-sl} d\mathcal{L}(l)$ and $\hat{\mathcal{L}}(s|q) \equiv \int_0^\infty e^{-sl} d\mathcal{L}(l|q)$ be their respective Laplace-Stieltjes tranforms. These are related in that $\mathcal{L}(l) = \int_0^\infty \mathcal{L}(l|q) dF(q)$ and $\hat{\mathcal{L}}(l) = \int_0^\infty \hat{\mathcal{L}}(l|q) dF(q)$.

We first derive a the relationship between the conditional size distributions among firms with different efficiencies. Recall that $\tilde{F}(q) \equiv \frac{F(q)-e^{-M}}{G(q)(1-e^{-M})}$ describes the CDF of a potential buyer's best alternative technique.

Proposition 7. *The transforms* $\{\hat{\mathcal{L}}(\cdot|q)\}$ *satisfy*

$$\hat{\mathcal{L}}(s|q) = e^{-s(1-\alpha)(q/Q)^{\varepsilon-1} L} e^{-M \int_0^\infty \tilde{F}(zq^\alpha)\left[1-\hat{\mathcal{L}}(\alpha s|zq^\alpha)\right] dH(z)}$$

Proof. Total labor used by a firm is the sum of labor used to make output for consumption and for use as an intermediate input. We use the fact that the Laplace-Stieltjes transform of a sum of random variables is

the product of the transforms of each.

A firm with efficiency q uses $(1-\alpha)(q/Q)^{\varepsilon-1}L$ units of labor in making goods for final consumption. The transform of this is $e^{-s(1-\alpha)(q/Q)^{\varepsilon-1}L}$.

We next consider labor used to make intermediate inputs. Recall from the proof of Lemma 1 the convenient fact that if firm j uses l_j units of labor, j's supplier will use αl_j units of labor to make the inputs for j. Thus if the transform of labor used by a buyer with efficiency q_b is $\hat{\mathcal{L}}(s|q_b)$, then the transform of labor used by its supplier to make intermediates is

$$\int_0^\infty \frac{1}{\alpha}\Pr\left(l_j=\frac{l}{\alpha}\right)e^{-sl}dl=\int_0^\infty \Pr\left(l_j=\frac{l}{\alpha}\right)e^{-(\alpha s)\frac{l}{\alpha}}d\left(\frac{l}{\alpha}\right)=\hat{\mathcal{L}}(\alpha s|q_b)$$

For a firm with efficiency q, consider a single downstream technique of with productivity z, so that the technique delivers efficiency to the buyer of zq^α. With probability $\tilde{F}(zq^\alpha)$ it is the buyer's best technique, in which case the transform of labor used to create intermediates for that customer is $\hat{\mathcal{L}}(\alpha s|zq^\alpha)$. With probability $1-\tilde{F}(zq^\alpha)$ the potential buyer uses an alternative supplier, in which case the transform of labor used to create intermediates for that customer is simply 1. Putting these together and integrating over possible realizations of productivity, the transform of labor use to make intermediates for a single *potential* customer is

$$\int_0^\infty \left\{\left[1-\tilde{F}(zq^\alpha)\right]+\tilde{F}(zq^\alpha)\hat{\mathcal{L}}(\alpha s|zq^\alpha)\right\}dH(z)=1-\int_0^\infty \tilde{F}(zq^\alpha)\left[1-\hat{\mathcal{L}}(\alpha s|zq^\alpha)\right]dH(z)$$

Each firm has n potential customers with probability $\frac{M^n e^{-M}}{n!}$, so the transform over labor used to create all intermediate goods (summing across all potential customers) is

$$\sum_{n=0}^\infty \frac{M^n e^{-M}}{n!}\left(1-\int_0^\infty \tilde{F}(zq^\alpha)\left[1-\hat{\mathcal{L}}(\alpha s|zq^\alpha)\right]dH(z)\right)^n=e^{-M\int_0^\infty \tilde{F}(zq^\alpha)\left[1-\hat{\mathcal{L}}(\alpha s|zq^\alpha)\right]dH(z)}$$

$\hat{\mathcal{L}}(s|q)$ is simply the product of the transforms of labor used for to make final consumption and labor used to make intermediate inputs. ■

Using the functional forms, we can get at the overall size distribution without the intermediate step of

solving for the conditional size distributions.

Proposition 8. *Define $v \equiv \frac{\varepsilon-1}{\zeta}$. With the functional form assumptions, the transform $\hat{\mathcal{L}}(\cdot)$ satisfies*

$$\hat{\mathcal{L}}(s) = e^{-s(1-\alpha)\frac{t^{-v}}{\Gamma(1-v)}L} e^{-\frac{t^{-\alpha}}{\Gamma(1-\alpha)}\left[1-\hat{\mathcal{L}}(\alpha s)\right]} e^{-t} dt$$

Proof. First, using the functional forms, the term $M\int_0^\infty \tilde{F}(zq^\alpha)\left[1-\hat{\mathcal{L}}(\alpha s|zq^\alpha)\right]dH(z)$ can be written as $mz_0^{-\zeta}\int_{z_0}^\infty \tilde{F}(zq^\alpha)\left[1-\hat{\mathcal{L}}(\alpha s|zq^\alpha)\right]\zeta z_0^\zeta z^{-\zeta-1}dz$. Since $\tilde{F}(zq^\alpha) \to e^{\theta(zq^\alpha)^{-\zeta}}$, this becomes (using the change of variables $w = zq^\alpha$):

$$\frac{mq^{\alpha\zeta}}{\theta}\int_0^\infty e^{\theta w^{-\zeta}}\left[1-\hat{\mathcal{L}}(\alpha s|w)\right]\zeta\theta w^{-\zeta-1}dw = \frac{mq^{\alpha\zeta}}{\theta}\left[1-\hat{\mathcal{L}}(\alpha s)\right]$$

where the last step follows because $e^{\theta w^{-\zeta}}\zeta\theta w^{-\zeta-1}dw = dF(w)$. We use this express $\hat{\mathcal{L}}(\cdot)$:

$$\hat{\mathcal{L}}(s) = \int_0^\infty \hat{\mathcal{L}}(s|q)dF(q) \to \int_0^\infty e^{-s(1-\alpha)(q/Q)^{\varepsilon-1}L} e^{-\frac{mq^{\alpha\zeta}}{\theta}\left[1-\hat{\mathcal{L}}(\alpha s)\right]}\zeta\theta q^{-\zeta-1}e^{-\theta q^{-\zeta}}dq$$

The conclusion follows from the substitutions $Q^{\varepsilon-1} = \Gamma(1-v)\theta^v$ and $m = \frac{\theta^{1-\alpha}}{\Gamma(1-\alpha)}$, and the change of variables $t = \theta q^{-\zeta}$. ■

B.2 Tail Behavior

Let $\rho = \min\left\{\alpha^{-1}, v^{-1}\right\}$ and let N be the greatest integer that is strictly less then ρ. For any integer n, let $\mu_n \equiv \int_0^\infty l^n d\mathcal{L}(l) = (-1)^n\hat{\mathcal{L}}^{(n)}(0)$ be the nth moment of the size distribution. The strategy is to show that $\mu_N - (-1)^N\hat{\mathcal{L}}^{(N)}(s)$ is regularly varying with index $\rho - N$ as $s \searrow 0$. Using the Tauberian theorem of Bingham and Doney (1974), this will imply that $1-\mathcal{L}(l)$ is regularly varying with index $-\rho$ as $l \to \infty$. The theorem gives this implication only when $0 < \rho - N < 1$, so we restrict attention to that case.

Define $\varphi(s;t) \equiv s\frac{(1-\alpha)}{\Gamma(1-v)}t^{-v} + \frac{1-\hat{\mathcal{L}}(\alpha s)}{\Gamma(1-\alpha)}t^{-\alpha}$ so that $\hat{\mathcal{L}}(s) = \int_0^\infty e^{-\varphi(s;t)}e^{-t}dt$. Since we will be interested in $\hat{\mathcal{L}}^{(n)}$, it will be useful to derive an expression for $\frac{d^n}{ds^n}\left[e^{-\varphi(s;t)}\right]$. By Faa di Bruno's formula (a generalization

of the chain rule to higher derivatives), we have that

$$\frac{d^n}{ds^n}\left[e^{-\varphi(s;t)}\right] = e^{-\varphi(s;t)} \sum_{\iota \in I_n} \frac{n!}{\iota_1!(1!)^{\iota_1}...\iota_n!(n!)^{\iota_n}} \prod_{j=1}^{n} \left[-\varphi^{(j)}(s;t)\right]^{\iota_j}$$

where I_n is the set of all n-tuples of non-negative integers $\iota = (\iota_1, ..., \iota_n)$ such that $1\iota_1 + ... + n\iota_n = n$.

Since $\varphi^{(1)}(s;t) = \frac{(1-\alpha)}{\Gamma(1-v)}t^{-v} + \frac{-\alpha\hat{\mathcal{L}}^{(1)}(\alpha s)}{\Gamma(1-\alpha)}t^{-\alpha}$ and $\varphi^{(j)}(s;t) = \frac{-\alpha^j\hat{\mathcal{L}}^{(j)}(\alpha s)}{\Gamma(1-\alpha)}t^{-\alpha}$ for $j \geq 2$, we have for each $\iota \in I_n$ that

$$\begin{aligned}
\prod_{j=1}^{n}\left[-\varphi^{(j)}(s;t)\right]^{\iota_j} &= \sum_{k=0}^{\iota_1}\binom{\iota_1}{k}\left[\frac{-(1-\alpha)}{\Gamma(1-v)}t^{-v}\right]^k\left[\frac{\alpha\hat{\mathcal{L}}^{(1)}(\alpha s)}{\Gamma(1-\alpha)}t^{-\alpha}\right]^{\iota_1-k}\prod_{j=2}^{n}\left[\frac{\alpha^j\hat{\mathcal{L}}^{(j)}(\alpha s)}{\Gamma(1-\alpha)}t^{-\alpha}\right]^{\iota_j} \\
&= \sum_{k=0}^{\iota_1}\binom{\iota_1}{k}\left[\frac{-(1-\alpha)}{\Gamma(1-v)}\frac{\Gamma(1-\alpha)}{\alpha\hat{\mathcal{L}}^{(1)}(\alpha s)}\right]^k\prod_{j=1}^{n}\left[\frac{\alpha^j\hat{\mathcal{L}}^{(j)}(\alpha s)}{\Gamma(1-\alpha)}\right]^{\iota_j} t^{-\left[\alpha\left(\sum_{j=1}^{n}\iota_j-k\right)+vk\right]}
\end{aligned}$$

Thus we can write

$$\frac{d^n}{ds^n}\left[e^{-\varphi(s;t)}\right] = e^{-\varphi(s;t)}\sum_{i\in I_n}\sum_{k=0}^{\iota_1} B_{n,\iota,k}(s)t^{-\beta(\iota,k)} \tag{14}$$

where

$$\begin{aligned}
B_{n,\iota,k}(s) &\equiv \frac{n!}{\iota_1!(1!)^{\iota_1}...\iota_n!(n!)^{\iota_n}}\binom{\iota_1}{k}\left[\frac{-(1-\alpha)}{\Gamma(1-v)}\frac{\Gamma(1-\alpha)}{\alpha\hat{\mathcal{L}}^{(1)}(\alpha s)}\right]^k\prod_{j=1}^{n}\left[\frac{\alpha^j\hat{\mathcal{L}}^{(j)}(\alpha s)}{\Gamma(1-\alpha)}\right]^{\iota_j} \\
\beta(\iota,k) &\equiv \alpha\left(\sum_{j=1}^{n}\iota_j - k\right) + vk
\end{aligned}$$

Note that if $n < \rho$ then each $\beta(\iota,k) \in (0,1)$. We first show that the nth derivative can be taken inside the integral

Lemma 9. *For any integer $n < \rho$, $\hat{\mathcal{L}}^{(n)}(s) = \int_0^\infty \frac{d^n}{ds^n}\left[e^{-\varphi(s;t)}\right]e^{-t}dt$.*

Proof. We first show that for each $\iota \in I_{n-1}$,

$$\frac{d}{ds}\int_0^\infty e^{-\varphi(s;t)}t^{-\beta(\iota,k)}e^{-t}dt = \int_0^\infty \frac{d}{ds}\left[e^{-\varphi(s;t)}\right]t^{-\beta(\iota,k)}e^{-t}dt$$

Since $\frac{d}{ds}\left[e^{-\varphi(s;t)}\right] = \left(\frac{(1-\alpha)}{\Gamma(1-v)}t^{-v} + \frac{-\alpha\hat{\mathcal{L}}^{(1)}(\alpha s)}{\Gamma(1-\alpha)}t^{-\alpha}\right)e^{-\varphi(s;t)}$, the integrand on RHS is dominated by $\left(\frac{(1-\alpha)}{\Gamma(1-v)}t^{-v} + \frac{-\alpha\hat{\mathcal{L}}^{(1)}(\alpha s)}{\Gamma(1-\alpha)}t^{-\alpha}\right.$

which is integrable for each s because $\alpha + \beta(\iota, k) < 1$ and $v + \beta(\iota, k) < 1$.

With this, we proceed by induction. Trivially, the conclusion holds for $n = 0$. Now for $n < \rho$, assume that $\hat{\mathcal{L}}^{(n-1)}(s) = \int_0^\infty \frac{d^{n-1}}{ds^{n-1}} \left[e^{-\varphi(s;t)}\right] e^{-t} dt$. Equation (14) implies

$$\hat{\mathcal{L}}^{(n-1)}(s) = \sum_{\iota \in I_{n-1}} \sum_{k=0}^{\iota_1} B_{n-1,\iota,k}(s) \int_0^\infty e^{-\varphi(s;t)} t^{-\beta(\iota,k)} e^{-t} dt$$

Differentiating each side gives

$$\begin{aligned}
\hat{\mathcal{L}}^{(n)}(s) &= \sum_{\iota \in I_{n-1}} \sum_{k=0}^{\iota_1} \frac{dB_{n-1,\iota,k}(s)}{ds} \int_0^\infty e^{-\varphi(s;t)} t^{-\beta(\iota,k)} e^{-t} dt + B_{n-1,\iota,k}(s) \frac{d}{ds} \int_0^\infty e^{-\varphi(s;t)} t^{-\beta(\iota,k)} e^{-t} dt \\
&= \int_0^\infty \sum_{\iota \in I_{n-1}} \sum_{k=0}^{\iota_1} \left[\frac{dB_{n-1,\iota,k}(s)}{ds} e^{-\varphi(s;t)} + B_{n-1,\iota,k}(s) \frac{de^{-\varphi(s;t)}}{ds} \right] t^{-\beta(\iota,k)} e^{-t} dt \\
&= \int_0^\infty \frac{d^n}{ds^n} \left[e^{-\varphi(s;t)} \right] e^{-t} dt
\end{aligned}$$

■

Lemma 10. *For any integer $n < \rho$, $\mu_n < \infty$.*

Proof. Again, we proceed by induction. $\mu_0 = 1$. Now assume that $\mu_0, ..., \mu_{n-1} < \infty$. We begin with the expression for $\hat{\mathcal{L}}^{(n)}(s)$:

$$\hat{\mathcal{L}}^{(n)}(s) = \sum_{\iota \in I_n} \sum_{k=0}^{\iota_1} B_{n,\iota,k}(s) \int_0^\infty e^{-\varphi(s;t)} t^{-\beta(\iota,k)} e^{-t} dt \tag{15}$$

Define $\tilde{I}_n \equiv I_n \setminus (0, ..., 0, 1)$. Then pulling out from the sum the term for $\iota = (0, ..., 0, 1)$, we have

$$\hat{\mathcal{L}}^{(n)}(s) = \frac{\alpha^n \hat{\mathcal{L}}^{(n)}(\alpha s)}{\Gamma(1-\alpha)} \int_0^\infty e^{-\varphi(s;t)} t^{-\alpha} e^{-t} dt + \sum_{\iota \in \tilde{I}_n} \sum_{k=0}^{\iota_1} B_{n,\iota,k}(s) \int_0^\infty e^{-\varphi(s;t)} t^{-\beta(\iota,k)} e^{-t} dt$$

We can take the limit as $s \searrow 0$ of each side. Since $e^{-\varphi(s;t)}$ is dominated by 1, the limit can be taken inside of each integral, and since $\lim_{s \searrow 0} \varphi(s;t) = 0$, we have

$$\hat{\mathcal{L}}^{(n)}(0) = \alpha^n \hat{\mathcal{L}}^{(n)}(0) + \sum_{\iota \in \tilde{I}_n} \sum_{k=0}^{\iota_1} B_{n,\iota,k}(0) \Gamma \left\{1 - \beta(\iota, k)\right\}$$

For each $\iota \in \tilde{I}_n$, and for each k, $B_{n,\iota,k}(0)$ is proportional to a product of derivatives of $\hat{\mathcal{L}}$, and each of

those derivatives is of order less than n. Since all of these are finite, $\hat{\mathcal{L}}^{(n)}(0) < \infty$. ■

With this we show that $\mu_N - (-1)^N \hat{\mathcal{L}}^{(N)}(s)$ is regularly varying as $s \searrow 0$. First, we have

$$\begin{aligned}
\mu_N - (-1)^N \hat{\mathcal{L}}^{(N)}(s) &= (-1)^N \left[\hat{\mathcal{L}}^{(N)}(0) - \hat{\mathcal{L}}^{(N)}(s)\right] \\
&= (-1)^N \sum_{\iota \in I_n} \sum_{k=0}^{\iota_1} B_{n,\iota,k}(0) \int_0^\infty t^{-\beta(\iota,k)} e^{-t} dt - B_{n,\iota,k}(s) \int_0^\infty e^{-\varphi(s;t)} t^{-\beta(\iota,k)} e^{-t} dt
\end{aligned}$$

We can decompose the object of interest into three terms:

$$\frac{\mu_N - (-1)^N \hat{\mathcal{L}}^{(N)}(s)}{s^{\rho - N}} = A_1(s) + A_2(s) + A_3(s)$$

where A_1, A_2 and A_3 are defined as

$$\begin{aligned}
A_1(s) &\equiv s^{N-\rho}(-1)^N \sum_{\iota \in I_N} \sum_{k=0}^{\iota_1} [B_{N,\iota,k}(0) - B_{N,\iota,k}(s)] \int_0^\infty t^{-\beta(\iota,k)} e^{-t} dt \\
A_2(s) &\equiv s^{N-\rho}(-1)^N \sum_{\iota \in I_N} \sum_{k=0}^{\iota_1} B_{N,\iota,k}(s) \int_0^\infty t^{-\beta(\iota,k)} \left[1 - e^{-\varphi(s;t)}\right] dt \\
A_3(s) &\equiv s^{N-\rho}(-1)^{N+1} \sum_{\iota \in I_N} \sum_{k=0}^{\iota_1} B_{N,\iota,k}(s) \int_0^\infty t^{-\beta(\iota,k)} \left[1 - e^{-\varphi(s;t)}\right] \left[1 - e^{-t}\right] dt
\end{aligned}$$

This particular decomposition is useful because it will allow for the use of the Monotone Convergence Theorem in characterizing the limiting behavior of A_2 and A_3.

Lemma 11. *If $\rho \notin \mathbb{N}$, $\lim_{s \searrow 0} A_1(s) = \alpha^\rho \lim_{s \to 0} \frac{\mu_N - (-1)^N \hat{\mathcal{L}}^{(N)}(s)}{s^{\rho - N}}$.*

Proof. As above, we can separate the term for $\iota = (0, ..., 0, 1)$ from $\tilde{I}_N$, to write

$$\begin{aligned}
\lim_{s \searrow 0} A_1(s) &= \lim_{s \to 0} s^{N-\rho}(-1)^N \sum_{\iota \in I_N} \sum_{k=0}^{\iota_1} [B_{N,\iota,k}(0) - B_{N,\iota,k}(s)] \Gamma(1 - \beta(\iota,k)) \\
&= \lim_{s \searrow 0} (-1)^N \frac{\left[\frac{\alpha^N \hat{\mathcal{L}}^{(N)}(0)}{\Gamma(1-\alpha)} - \frac{\alpha^n \hat{\mathcal{L}}^{(n)}(\alpha s)}{\Gamma(1-\alpha)}\right]}{s^{\rho - N}} \Gamma(1 - \alpha) \\
&\quad + \lim_{s \searrow 0} (-1)^N \sum_{\iota \in \tilde{I}_N} \sum_{k=0}^{\iota_1} \frac{[B_{N,\iota,k}(0) - B_{N,\iota,k}(s)]}{s^{\rho - N}} \Gamma(1 - \beta(\iota,k))
\end{aligned}$$

The first term is simply $\alpha^\rho \lim_{s \searrow 0} \frac{\mu_N - (-1)^N \hat{\mathcal{L}}^{(N)}(\alpha s)}{(\alpha s)^{\rho - N}}$. The second term equals zero: After using L'Hospital's

rule, the numerator becomes a mulitnomial of derivatives of $\hat{\mathcal{L}}$ or order no greater than N, (so all of these are finite) while the denominator becomes $(\rho - N)s^{\rho-N-1}$ which goes to infinity. ■

To characterize the limiting behavior of A_2 and A_3, it will be useful to define

$$\begin{aligned} \kappa &\equiv \frac{1}{\Gamma(1-\rho)}\left((1-\alpha)\mathbb{I}_{v\geq\alpha} + \alpha\mathbb{I}_{\alpha\geq v}\right) \\ \tilde{\varphi}(s;w) &\equiv \varphi\left(s; \left[1-\hat{\mathcal{L}}(\alpha s)\right]^{\rho}\kappa^{\rho}\alpha^{-\rho}w^{-\rho}\right) \end{aligned}$$

$\tilde{\varphi}$ is defined this way so that $\lim_{s\to 0}\tilde{\varphi}(s;w) = w$ (this can be easily verified for each of the three cases: $\alpha > v$, $\alpha = v$, and $\alpha < v$).

Lemma 12. *$\tilde{\varphi}(s;w)$ is non-decreasing in s in the neighborhood of zero.*

Proof. Using the definitions of $\tilde{\varphi}$ and φ, we have

$$\tilde{\varphi}(s;w) = s^{1-\rho v}(1-\alpha)\frac{\kappa^{-\rho v}\left[\frac{1-\hat{\mathcal{L}}(\alpha s)}{\alpha s}\right]^{-\rho v}}{\Gamma(1-v)}w^{\rho v} + \frac{(\kappa/\alpha)^{-\rho\alpha}}{\Gamma(1-\alpha)}\left[1-\hat{\mathcal{L}}(\alpha s)\right]^{1-\alpha\rho}w^{\alpha\rho}$$

We first show that $\frac{1-\hat{\mathcal{L}}(s)}{s}$ is non-increasing in a neighborhood of zero. Since $\mu_2 > 0$,

$$\lim_{s\searrow 0}\frac{d}{ds}\left(\frac{1-\hat{\mathcal{L}}(s)}{s}\right) = \lim_{s\searrow 0}\frac{-s\hat{\mathcal{L}}^{(1)}(s) - \left[1-\hat{\mathcal{L}}(s)\right]}{s^2} = -\hat{\mathcal{L}}^{(2)}(0) < 0 \tag{16}$$

Next, since $1-\hat{\mathcal{L}}(\alpha s)$ is non-decreasing in s, and since both $1-\rho v$ and $1-\rho\alpha$ are non-negative, $\tilde{\varphi}(s;w)$ is nondecreasing in the neighborhood of zero. ■

Lemma 13. $\lim_{s\searrow 0} A_2(s) = \kappa^{\rho}\frac{\rho}{\rho-N}\Gamma(1-\rho+N)$.

Proof. Using the change of variables $w = \left[1-\hat{\mathcal{L}}(\alpha s)\right]\kappa\alpha^{-1}t^{-1/\rho}$, we have

$$A_2(s) = (-1)^N\sum_{\iota\in I_N}\sum_{k=0}^{\iota_1}s^{N-\rho\beta(\iota,k)}B_{N,\iota,k}(s)\left\{\frac{\left[1-\hat{\mathcal{L}}(\alpha s)\right]}{\alpha s}\kappa\right\}^{\rho-\rho\beta(\iota,k)}\int_0^{\infty}\left[1-e^{-\tilde{\varphi}(s;w)}\right]\rho w^{-\rho[1-\beta(\iota,k)]-1}dw$$

We next take the limit of each side as s goes to zero. Lemma 12 and the monotone convergence theorem

imply that the limit can be brought inside the integral to yield

$$\lim_{s\searrow 0} A_2(s) = (-1)^N \sum_{\iota\in I_N}\sum_{k=0}^{\iota_1}\left(\lim_{s\searrow 0} s^{N-\rho\beta(\iota,k)}\right) B_{N,\iota,k}(0)\kappa^{\rho-\rho\beta(\iota,k)}\int_0^\infty \left[1-e^{-w}\right]\rho w^{-\rho[1-\beta(\iota,k)]-1}dw$$

Noting that $N \geq \rho\beta(\iota,k)$, the term $\lim_{s\to 0} s^{N-\rho\beta(\iota,k)}$ is zero unless $N=\rho\beta(\iota,k)$. Thus $\lim_{s\searrow 0} A_2(s)$ can be written as

$$\lim_{s\searrow 0} A_2(s) = (-1)^N\kappa^{\rho-N}\int_0^\infty \left[1-e^{-w}\right]\rho w^{-(\rho-N)-1}dw \sum_{\iota\in I_N}\sum_{k=0}^{\iota_1} B_{N,\iota,k}(0)\mathbb{I}_{N=\rho\beta(\iota,k)}$$

The integral is $\int_0^\infty \left[1-e^{-w}\right]\rho w^{-(\rho-N)-1}dw = \frac{\rho}{\rho-N}\Gamma(1-\rho+N)$. To finish the proof, we show that

$$(-1)^N \sum_{\iota\in I_N}\sum_{k=0}^{\iota_1} B_{N,\iota,k}(0)\mathbb{I}_{N=\rho\beta(\iota,k)} = \kappa^N \tag{17}$$

To see this note first that $N=\rho\beta(\iota,k)$ requires $\iota=(N,0,...,0)$. If $\alpha>v$, $N=\rho\beta(\iota,k)$ also requires $k=0$, whereas if $\alpha<v$, it requires $k=N$. If $\alpha=v$, $N=\rho\beta(\iota,k)$ for each $k\in\{0,...,N\}$. For each of these three cases, one can compute each non-zero term in the sum and verify equation (17).

■

Lemma 14. $\lim_{s\searrow 0} A_3(s) = 0$

Proof. The strategy is the same as in the previous lemma. Using the same change of variables $w = \left[1-\hat{\mathcal{L}}(\alpha s)\right]\kappa\alpha^{-1}t^{-1/\rho}$, we have

$$\begin{aligned} A_3(s) &= (-1)^N \sum_{\iota\in I_N}\sum_{k=0}^{\iota_1} s^{N-\rho\beta(\iota,k)} B_{N,\iota,k}(s)\left\{\frac{\left[1-\hat{\mathcal{L}}(\alpha s)\right]}{\alpha s}\kappa\right\}^{\rho-\rho\beta(\iota,k)} \\ &\quad\times\int_0^\infty \left[1-e^{-\tilde{\varphi}(s;w)}\right]\left[1-e^{-\left[1-\hat{\mathcal{L}}(\alpha s)\right]^\rho(\kappa/\alpha)^\rho w^{-\rho}}\right]\rho w^{-\rho[1-\beta(\iota,k)]-1}dw \end{aligned}$$

We can take a limit of each side. Since both $\tilde{\varphi}(s;w)$ and $\left[1-\hat{\mathcal{L}}(\alpha s)\right]^\rho(\kappa/\alpha)^\rho w^{-\rho}$ are non-decreasing in s in the neighborhood of $s=0$, we can use the monotone convergence theorem to bring the limit inside the

integral. Thus we have

$$\begin{aligned}\lim_{s\searrow 0} A_3(s) &= (-1)^N \sum_{\iota\in I_N}\sum_{k=0}^{\iota_1}\left(\lim_{s\searrow 0} s^{N-\rho\beta(\iota,k)}\right) B_{N,\iota,k}(0)\kappa^{\rho-\rho\beta(\iota,k)} \\ &\quad\times \int_0^\infty \lim_{s\searrow 0}\left[1-e^{-\tilde{\varphi}(s;w)}\right]\left[1-e^{-\left[1-\hat{\mathcal{L}}(\alpha s)\right]^\rho(\kappa/\alpha)^\rho w^{-\rho}}\right]\rho w^{-\rho[1-\beta(\iota,k)]-1}dw\end{aligned}$$

Since the limit of the integrand is zero for each integral, we have $\lim_{s\searrow 0} A_3(s) = 0$. ■

We finally come to the main result.

Proposition 9. *If $\rho \notin \mathbb{N}$, $1-\mathcal{L}(l) \sim \frac{\kappa^\rho}{1-\alpha^\rho} l^{-\rho}$*

Proof. The previous lemmas imply that

$$\lim_{s\searrow 0}\frac{\mu_N-(-1)^N\hat{\mathcal{L}}^{(N)}(s)}{s^{\rho-N}} = \alpha^\rho \lim_{s\searrow 0}\frac{\mu_N-(-1)^N\hat{\mathcal{L}}^{(N)}(s)}{s^{\rho-N}} + \kappa^\rho\frac{\rho}{\rho-N}\Gamma(1-\rho+N)+0$$

or

$$\lim_{s\searrow 0}\frac{\mu_N-(-1)^N\hat{\mathcal{L}}^{(N)}(s)}{s^{\rho-N}} = \frac{\kappa^\rho}{1-\alpha^\rho}\frac{\rho}{\rho-N}\Gamma(1-\rho+N)$$

By Theorem A in Bingham and Doney (1974) we therefore have that

$$\lim_{l\to\infty}\frac{1-\mathcal{L}(l)}{l^{-\rho}} = \frac{\kappa^\rho}{1-\alpha^\rho}\left[(-1)^N\frac{\Gamma(\rho-N)}{\Gamma(\rho)}\frac{\Gamma(1-\rho+N)}{\Gamma(1-\rho)}\right]$$

Since $\Gamma(\rho) = \Gamma(\rho-N)\prod_{k=1}^N(\rho-k)$ and $(-1)^N\Gamma(1-\rho+N) = \Gamma(1-\rho)\prod_{k=1}^N(-1)(1-\rho+N-k) = \Gamma(1-\rho)\prod_{k=1}^N(\rho-k)$, the term in brackets equals unity, completing the proof. ■

C Pairwise Stable Equilibria

We are careful here to use notation that allows for the possibility that a firm uses a production chain with a cycle: at least some of the input from j is used in as an intermediate good in the supply chain used by j. In such a case, if a firm decides to lower its price of final output, it sells more final output but also raises the demand for its good as an intermediate input. Given an arrangement, for each $\phi \in \hat{D}_j$, let $x(\phi, y_j^0)$

be the quantity of good j sold to firm $b(\phi)$ given y_j^0, holding constant the final output of other firms. x is non-decreasing and weakly concave in its second argument. Given an arrangement, let $\delta(p^0) \equiv Y^0 \left(p^0/P^0\right)^{-\varepsilon}$ be final demand given price p^0. Firm j's profit is:

$$\pi_j = \max_{p_j^0, y_j^0, l_j, x} p_j^0 y_j^0 + \sum_{\phi \in \hat{D}_j} \left[p(\phi)x(\phi, y_j^0) + \tau(\phi)\right] - wl_j - p(\phi_j)x - \tau(\phi_j)$$

subject to final demand, $y_j^0 \leq \delta(p_j^0)$, and

$$y_j^0 + \sum_{\phi \in \hat{D}_j} x(\phi, y_j^0) \quad \leq \quad \frac{1}{\alpha^\alpha (1-\alpha)^{1-\alpha}} z(\phi_j) x^\alpha l_j^{1-\alpha}$$

The first order conditions with respect to inputs imply cost minimization so that firm j's marginal cost is $\lambda_j \equiv \frac{1}{z(\phi_j)} p(\phi_j)^\alpha w^{1-\alpha}$. The price of final output p_j^0 then satisfies:

$$p_j^0 = \arg\max_p (p - \lambda_j)\delta(p_j^0) + \sum_{\phi \in \hat{D}_j} (p(\phi) - \lambda_j) x\left(\phi, \delta(p_j^0)\right)$$

Note that if good j is in the supply chain to produce good j (there is a cycle), firm j internalizes how changing final output would affect its sales of j as an intermediate.

Proposition (2a). *In any pairwise stable equilibrium, for any technique $\phi \in \hat{\Phi}$, then $p(\phi) = \lambda_{s(\phi)}$.*

Proof. Assume there is a $j \in \hat{J}$ such that $p(\phi_j) \neq \lambda_{s(\phi)}$, and let $i = s(\phi)$. Consider the deviation

$$\begin{aligned} \tilde{p}(\phi_j) &= \lambda_i \\ \tilde{\tau}(\phi_j) &= \tau(\phi_j) + \left[p(\phi_j) - \lambda_i\right] x(\phi_j) + K \end{aligned}$$

where $K \equiv \frac{1}{2}\left\{\left(\tilde{p}_j^0 - \tilde{\lambda}_j\right)\tilde{y}_j^0 - \left(p_j^0 - \tilde{\lambda}_j\right) y_j^0 + \sum_{\phi \in \hat{D}_j} \left(p(\phi) - \tilde{\lambda}_j\right)\left[x(\phi, \tilde{y}_j^0) - x(\phi, y_j^0)\right]\right\}$ and $\tilde{\lambda}_j = \frac{1}{z(\phi_j)} \lambda_i^\alpha w^{1-\alpha}$ is j's marginal cost given the deviation. K is strictly positive because $\hat{p}_j^0$ and $\hat{y}_j^0$ are optimal when j's marginal cost is $\tilde{\lambda}_j$. The change in i's profit from the deviation is $K > 0$. The change in j's profit from the deviation

is

$$\tilde{\pi}_j - \pi_j = \tilde{p}_j^0 \tilde{y}_j^0 - p_j^0 y_j^0 + \sum_{\phi \in \hat{D}_j} p(\phi) \left[x(\phi, \tilde{y}_j^0) - x(\phi, y_j^0) \right] - \left[\tilde{\lambda}_j \tilde{y}_j - \lambda_j y_j \right] - \left[\tilde{\tau}(\phi_j) - \tau(\phi_j) \right]$$

Using expressions for $\tilde{\tau}(\phi_j)$ and $\tilde{p}(\phi_j)$ along with $p(\phi_j)x(\phi_j) = \alpha \lambda_j y_j$ and $\frac{\lambda_j}{\tilde{\lambda}_j} = \left(\frac{p(\phi_j)}{\lambda_i} \right)^\alpha$ we have

$$\begin{aligned} \tilde{\pi}_j - \pi_j &= \left(\tilde{p}_j^0 - \tilde{\lambda}_j \right) \tilde{y}_j^0 - \left(p_j^0 - \tilde{\lambda}_j \right) y_j^0 + \sum_{\phi \in \hat{D}_j} \left(p(\phi) - \tilde{\lambda}_j \right) \left[x(\phi, \hat{y}_j^0) - x(\phi, y_j^0) \right] - K \\ &\quad + \left[\left((1-\alpha) + \frac{\lambda_i}{p(\phi_j)} \alpha \right) \left(\frac{p(\phi_j)}{\lambda_i} \right)^\alpha - 1 \right] \tilde{\lambda}_j y_j \end{aligned}$$

From the definition of K, the first line of the RHS is strictly positive. The second line is nonnegative because Jensen's inequality implies $[(1-\alpha) + r\alpha]\, r^{-\alpha} \geq 1$ for $\alpha \in [0,1]$ (r^t is a convex function of t, so $(1-\alpha)r^{-\alpha} + \alpha r^{1-\alpha} \geq r^{(1-\alpha)(-\alpha)+\alpha(1-\alpha)} = 1$). Therefore the deviation is mutually beneficial. ■

Proposition (2b). *In any pairwise stable equilibrium, for any technique $\phi \in \hat{\Phi}$, $\tau(\phi) \geq 0$.*

Proof. Assume a pairwise stable equilibrium in there is a $\phi \in \hat{\Phi}$ such that $\tau(\phi) < 0$. Consider the unilateral deviation in which the suppler, $s(\phi)$ removes the technique ϕ from the arrangement. After the deviation, every firm's input price remains the same except the buyer, $b(\phi)$. The quantity produced by $s(\phi)$ changes, as it will no longer produce inputs for $b(\phi)$. However, since price equals marginal cost, this change in quantity produced has no effect on $s(\phi)$'s profit. Thus $\tilde{\pi}_{s(\phi)} = \pi_{s(\phi)} - \tau(\phi) > \pi_{s(\phi)}$, confirming the unilateral deviation.

■

Proposition (2c). *In any pairwise stable equilibrium, for each $j \in [0,1]$, $\lambda_j = \min_{\phi \in U_j} \frac{1}{z(\phi)} \lambda_{s(\phi)}^\alpha w^{1-\alpha}$.*

Proof. Toward a contradiction, suppose there is a j such $\lambda_j > \min_{\phi \in U_j} \frac{1}{z(\phi)} \lambda_{s(\phi)}^\alpha w^{1-\alpha}$, and let $\tilde{\phi} = \arg\min_{\phi \in U_j} \frac{1}{z(\phi)} \lambda_{s(\phi)}^\alpha w^{1-\alpha}$. Consider the deviation for $\tilde{\phi}$ of

$$\begin{aligned} \tilde{p}(\tilde{\phi}) &= \lambda_{s(\tilde{\phi})} \\ \tilde{\tau}(\tilde{\phi}) &= \frac{1}{2} \left[\max_p \left(p - \tilde{\lambda}_j \right) \delta(p) - \max_p \left(p - \lambda_j \right) \delta(p) \right] > 0 \end{aligned}$$

where $\tilde{\lambda}_j \equiv \frac{1}{z(\tilde{\phi})} \lambda_{s(\tilde{\phi})}^\alpha w^{1-\alpha}$. In the deviation, j's marginal cost falls, but every other firm's marginal cost

is unchanged. Therefore j lowers its price of final output, but no other firm does, so the quantity of final output of j rises.

Firm j finds the deviation profitable, $\tilde{\pi}_j > \pi_j$, because marginal cost has fallen: (i) j now makes a positive profit on each unit of output sold as intermediate output (if j is in a cycle, then profit from intermediates would rise even more); (ii) The change in profit from final sales is greater than the fee $\tilde{\tau}(\tilde{\phi})$; and (iii) j no longer has to pay $\tau(\phi_j)$ (which is nonnegative). Firm $s(\tilde{\phi})$ also finds the deviation profitable, $\hat{\pi}_{s(\tilde{\phi})} > \pi_{s(\tilde{\phi})}$, because it now collects $\hat{\tau}(s(\tilde{\phi}))$. Since price equals marginal cost, changes in quantity produced as intermediates have no impact on its profit. Therefore the deviation is mutually beneficial. ■

C.1 Labor Only Option

We now show that if each firm has access to a production technology uses only labor, $y_j = \underline{q}L$ with common productivity $\underline{q}$, then in every pairwise stable equilibrium aggregate productivity Q is the same as in the planner's problem.

Let q_j^{pw} be the efficiency of firm j in a pairwise stable equilibrium, and let $F^{pw}(q)$ be the fraction of firms in that equilibrium with efficiency no greater than q. We can compare this to the CDF of efficiencies that solve the planner's problem, F^{sp}.

Proposition (??). *In any pairwise stable equilibrium, $F^{pw} = F^{sp}$ and $Q^{pw} = Q^{sp}$.*

Sketch of Proof. In both the planner's problem and in any pairwise stable equilibrium, each firm chooses the technique that delivers the most cost effective combination:

$$q_j = \max\left\{\underline{q}, \max_{\phi \in U_j}\left\{z(\phi)q_{s(\phi)}^{\alpha}\right\}\right\}$$

For each firm j, we are interested in the efficiency in the social planner's problem, q_j^{sp}, and efficiency in a

pairwise equilibrium. These satisfy the following three equations:

$$
\begin{aligned}
q_j^{sp} &= \max\left\{\sup_{c\in C^{\infty}(j)} \mathfrak{q}(c), \sup_{N\in\mathbb{N}} \max_{c\in C_N(j)} \mathfrak{q}_N(c)\underline{q}^{\alpha^N}\right\}, \qquad \text{a.e. } j\in[0,1] \\
q_j^{pw} &\geq \sup_{N\in\mathbb{N}} \max_{c\in C_N(j)} \mathfrak{q}_N(c)\underline{q}^{\alpha^N} \qquad \forall j\in[0,1] \\
q_j^{pw} &\leq q_j^{sp}, \qquad \text{a.e. } j\in[0,1]
\end{aligned}
$$

These can be interpreted as follows: First, the chain with the maximum feasible efficiency available to the planner is either an infinite chain or a finite chain that ends with the furthest firm upstream using the outside option. Second, in any pairwise stable equilibrium, the efficiency of a firm must be at least that of any finite chain that ends with a firm using its outside option $\underline{q}$. If otherwise, at least one of the firms in that chain must be using a suboptimal technique. Third, efficiency in a pairwise stable equilibrium must be feasible. The first and third statements are satisfied for almost every firm because the planner may deviate from the maximum feasible efficiency for a set of firms with measure zero.

The remainder of the proof follows broadly along the same lines as Appendix A with some minor modifications. As a result I will only provide a sketch, pointing out the differences. We can construct an operator in the same manner as before:

$$
Tf(q) \equiv \begin{cases} 0, & q<\underline{q} \\ e^{-M\left[1-\int_0^{\infty} f\left(\left(\frac{q}{z}\right)^{\frac{1}{\alpha}}\right)dH(z)\right]}, & q\geq\underline{q} \end{cases}
$$

We can also define three functions.

$$
\begin{aligned}
\underline{f}(q) &\equiv \begin{cases} 0, & q<\underline{q} \\ 1, & q\geq\underline{q} \end{cases} \\
f^1(q) &\equiv \begin{cases} 0, & q<1 \\ 1, & q\geq 1 \end{cases} \\
\bar{f}(q) &\equiv \begin{cases} 0, & q<q_2 \\ 1-\left(\frac{q}{q_2}\right)^{-\beta}, & q\geq q_2 \end{cases}
\end{aligned}
$$

With the additional restriction that $q_2 > \underline{q}$.

Let $\mathcal{F}$ be the subset of $\bar{\mathcal{F}}$ defined by $\{f | \bar{f} \preceq f \preceq \underline{f}\}$. In the same manner as before, we can show that $T\bar{f} \succeq \bar{f}$ and $T\underline{f} \preceq \underline{f}$. This means that there is a least and greatest fixed point of T in $\mathcal{F}$, given by $\lim_{N\to\infty} T^N \bar{f}$ and $\lim_{N\to\infty} T^N \underline{f}$ respectively.

Define $\underline{Y}_N \equiv \max_{\hat{N}\leq N} \max_{c\in C_{\hat{N}}} \prod_{n=0}^{\hat{N}-1} \mathfrak{q}(c)\underline{q}^{\alpha^{\hat{N}}}$. With this we can define variables $X_N \equiv \max\left\{\underline{Y}_N, \max_{c\in C_N} \prod_{n=0}^{N-1} \mathfrak{q}(c)\right\}$ and $\bar{Y}_N \equiv \max\left\{\underline{Y}_N, \max_{c\in C_N} \prod_{n=0}^{N-1} \mathfrak{q}(c)\tilde{q}(c)^{\alpha^N}\right\}$, where $\tilde{q}(c)$ is a random variable drawn independently for each c in C_N for each realization of Φ, each with CDF $\bar{f}$. The CDFs of X_N, $\bar{Y}_N$, and $\underline{Y}_N$ are $T^N f^1$, $T^N \bar{f}$, and $T^N \underline{f}$ respectively.

Let $X^{sp} \equiv \lim_{N\to\infty} X_N$ and $X^{pw} \equiv \lim_{N\to\infty} \underline{Y}_N$ In the same manner as before we can show that these limits exists and that with probability 1 we have both $X^{sp} = q^{sp}$ and $X^{pw} \leq q^{pw} \leq X^{sp}$. Lastly, we can show that F^{sp} is the unique fixed point of T on $\mathcal{F}$, and is the CDF of both X^{sp} and X^{pw}. Applying the law of large numbers gives $Q^{sp} = Q^{pw}$. ■

D The Number of Supply Chains

For completeness, we show the derivation of several results from the theory of branching processes that are used in this paper (see for example Athreya and Ney (1972)).

Let $B_N(j) \equiv |C_N(j)|$ be the number of distinct chains of length N. Let $p(k)$ be the probability that a firm has exactly k techniques, in this case equal to $\frac{e^{-M}M^k}{k!}$, and let $P_N(l,k)$ be the probability that, in total, l different firms have among them k supply chains of length N (i.e. $\sum_{i=1}^{l} B_N(j_i) = k$). Note that $P_N(1,k)$ is the probability a firm has exactly k supply chains of length N (i.e., the probability that $B_N = k$).

Define $\varphi(x) = \sum_{k=0}^{\infty} p(k)x^k$ to be the probability generating function for the random variable B_1. In this case $\varphi(x) = e^{-M(1-x)}$. Also, for each N, let $\varphi_N(\cdot)$ be the probability generating function associated with B_N. If $\varphi^{(N)}$ is the N-fold composition of φ then we have the convenient result:

Lemma 15. $\varphi_N(x) = \varphi^{(N)}(x)$

Proof. We proceed by induction. By definition, the statement is true for $N = 1$. Noting that $\sum_{k=0}^{\infty} P_1(l,k)x^k = \varphi(x)^l$, we have

$$\begin{aligned}
\varphi_{N+1}(x) &= \sum_{l=0}^{\infty} P_{N+1}(1,l)x^l = \sum_{l=0}^{\infty}\sum_{k=0}^{\infty} P_N(1,k)P_1(k,l)x^l = \sum_{k=0}^{\infty} P_N(1,k)\sum_{l=0}^{\infty} P_1(k,l)x^l \\
&= \sum_{k=0}^{\infty} P_N(1,k)\varphi(x)^k = \varphi_N(\varphi(x))
\end{aligned}$$

■

We immediately have the following:

Claim 1. *For any x, $\mathbb{E}\left[x^{B_N}\right] = \varphi^{(N)}(x)$.*

Proof. $\mathbb{E}\left[x^{B_N}\right] = \sum_{k=0}^{\infty} P_N(1,k)x^k = \varphi_N(x) = \varphi^{(N)}(x)$ ■

We next study the probability that a firm has no chains that continue indefinitely.

Claim 2. *The probability that a single firm has no chains that continue indefinitely is the smallest root, ρ, of $y = \varphi(y)$.*

Proof. The probability that a firm has no chains greater than length N is $P_N(1,0)$, or $\varphi_N(0)$. Then the probability that a single firm has no chains that continue indefinitely is $\lim_{N\to\infty}\varphi^{(N)}(0)$. Next note that that φ is increasing and convex, $\varphi(1)=1$, and $\varphi(0)\geq 0$. This implies that in the range $[0,1]$ the equation $\varphi(y)=y$ has a either a unique root at $y=1$ or two roots, $y=1$ and a second in $(0,1)$.

Let ρ be the smallest root. Noting that for $y\in[0,\rho)$, $y<\varphi(y)<\rho$, while for $y\in(\rho,1)$ (if such y exist), $\rho<\varphi(y)<y$. Together these imply that if $y\in[0,1)$, the sequence $\left\{\varphi^{(N)}(y)\right\}$ is monotone and bounded, and therefore has a limit. We have $\varphi^{(N+1)}(0)=\varphi(\varphi^{(N)}(0))$. Taking limits of both sides (and noting that φ is continuous) gives $\lim_{N\to\infty}\varphi^{(N+1)}(0)=\varphi(\lim_{N\to\infty}\varphi^{(N)}(0))$. Therefore the limit is a root of $y=\varphi(y)$, and therefore must be ρ. In other words, $\lim_{N\to\infty}\varphi^{(N)}(0)=\rho$. ■

Claim 3. *If $M\leq 1$ then with probability 1 the firm has no chains that continue indefinitely. If $M>1$ then there is a strictly positive probability the firm has a chain that continues indefinitely.*

Proof. In this case, we have $\varphi(x)=e^{-M(1-x)}$. If $M\leq 1$ then the smallest root of $y=\varphi(y)$ is $y=1$. If $M>1$ the smallest root is strictly less than 1. ■

E Multiple Inputs

In the preceding analysis, each firm used a single input in production. While this is assumption simplifies the analysis (and visualization) of the model, it is obviously at odds with production in the real world. This section extends the analysis so that firms use multiple inputs. There are several ways to generalize a technique. Here, the focus is on a version I consider both natural and tractable.

To produce, a firm needs to accomplish a fixed set of tasks. Over time, firms discover new techniques to accomplishing each task, and each of these techniques is a sub-production function that requires a particular input. In this way, a firm can improve production by finding techniques that are more cost effective in accomplishing individual tasks.[36]

The central assumption is that each firm uses a production function that combines N tasks into a unit of output. To accomplish each task, the firm uses a technique. For the technique used for the nth task, let z_n

[36] An alternative modeling strategy is that a new technique requires an entirely new input bundle. In that model, a firm would change suppliers all at once, rather than one at a time.

be the productivity and q_n be the efficiency of the supplier associated with that task. Then the production is such that the efficiency of the firm is

$$q = \prod_{n=1}^{N} \left(z_n q_n^{\alpha}\right)^{\frac{1}{N}}$$

The production function that justifies this can be written in two ways that are equivalent:

$$\begin{aligned} y &= \prod_{n=1}^{N} \left(\frac{z_n x_n^{\alpha} L_n^{1-\alpha}}{\left(\frac{1-\alpha}{N}\right)^{1-\alpha} \left(\frac{\alpha}{N}\right)^{\alpha}} \right)^{\frac{1}{N}} \\ y &= \frac{1}{(1-\alpha)^{1-\alpha}} \prod_{n=1}^{N} \left(\frac{\alpha}{N}\right)^{-\frac{\alpha}{N}} \prod_{n=1}^{N} z_n^{\frac{1}{N}} \left(\prod_{n=1}^{N} x_n^{\frac{1}{N}} \right)^{\alpha} L^{1-\alpha} \end{aligned}$$

The difference between the two is that in the first version labor is explicitly allocated separately to each task. Note that, as before, α is the share of intermediate inputs.

Let $F(q)$ be the fraction of firms with efficiency no greater than q. We will use the same method as before to define a fixed point problem for F.

For a firm, let v_n be the efficiency of a firm in accomplishing the nth task. Also, the number of techniques a firm has to accomplish a given task is follows a Poisson distribution with mean M (which is the same for each task). Let $K(v)$ be probability that $v_n < v$.[37] Then by the same analysis as before, we can write

$$K(v) = e^{-M \int_0^{\infty} \left[1 - F\left(\left(\frac{v}{z}\right)^{1/\alpha}\right)\right] dH(z)}$$

Since $q = \prod_{n=1}^{N} v_n^{\frac{1}{N}}$, we can write $F(q)$ as

$$F(q) = \int_0^{\infty} ... \int_0^{\infty} K\left(\frac{q^N}{\prod_{n=2}^{N} v_n}\right) dK\left(v_2\right) dK\left(v_3\right) ... dK\left(v_N\right)$$

We now impose the parametric assumptions $H(z) = 1 - (z/z_0)^{-\zeta}$ and $M = m z_0^{-\zeta}$. As shown in Appendix E.1, in the limiting economy as $z_0 \to 0$, the fraction of tasks with techniques delivering efficiency

[37] The analysis can easily be extended so that M varied by task, which would imply that the CDFs of each v_n would be different (i.e. $K_n(\cdot)$).

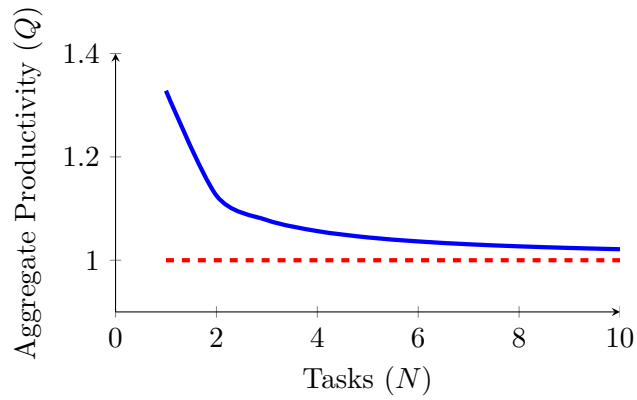


Figure 8: Aggregate Productivity by Number of Inputs

Figure 8 shows aggregate productivity Q the as the number of inputs changes. The parameters are $\alpha = 1/2$, $\zeta = 4$, and $\epsilon = 3$, with λ set so that the limiting aggregate productivity when the number of inputs grows large is 1. The dashed line is the asymptotic aggregate productivity as the number of inputs grows large.

no greater than v converges to a Frechet distribution

$$\begin{aligned} K(v) &= e^{-\theta v^{-\zeta}} \\ \theta &= m\theta^{\alpha}\Gamma\left(1-\frac{\alpha}{N}\right)^{N} \end{aligned}$$

The expression for F is messier, but we can derive an expression for aggregate productivity $Q = \left(\int_0^\infty q^{\varepsilon-1}dF(q)\right)^{\frac{1}{\varepsilon-1}}$, also in Appendix E.1:

$$Q = \Gamma\left(1-\frac{(\varepsilon-1)/\zeta}{N}\right)^{\frac{N}{\varepsilon-1}} m^{\frac{1}{(1-\alpha)\zeta}}\Gamma\left(1-\frac{\alpha}{N}\right)^{N\frac{1}{(1-\alpha)\zeta}}$$

This generalizes the formula for aggregate productivity with a single input given by equation (10). Since $\Gamma\left(1-\frac{\alpha}{N}\right)^{N}$ and $\Gamma\left(1-\frac{(\varepsilon-1)/\zeta}{N}\right)^{\frac{N}{\varepsilon-1}}$ are decreasing in N, aggregate productivity is declining in the number of inputs, as shown in Figure 8.

Aggregate productivity declines with the number of inputs because the tail of superstar suppliers shrinks. This happens because of two mutually reinforcing mechanisms. First, with a large number of tasks, there is less cross-sectional variation in efficiency. With a single input, a firm would end up with a high efficiency with one great technique. With more inputs, a firm that gets lucky enough to draw a great technique for one of its tasks may only have low efficiency techniques for its other tasks. The more inputs there are, the harder it is for a firm to have a high overall efficiency, as the the law of large numbers (across tasks) begins

to kick in.

On top of this, with less variation in the prices of potential suppliers, there is less variation in cost-effectiveness of individual techniques that are drawn. The efficiency provided by a technique is zq^{α}, and less dispersion in q implies that relatively high efficiency techniques are drawn less frequently.

A stark feature of the model that drives this result is that all tasks enter symmetrically into production. If instead production were dominated by relatively few core tasks, there would be more variation in efficiency: to have a high efficiency, a firm would only need a good draw for one of its core tasks.

Alternatively, if tasks are substitutes then the high efficiency tasks will have a large cost share. In the limit as the elasticity of substitution goes to infinity, the N input model would converge to the single input model.

E.1 Multiple Inputs

The fraction of tasks that deliver efficiency no greater than v is $K(v)$:

$$K(v) = e^{-M \int_0^{\infty} \left[1 - F\left(\left(\frac{v}{z}\right)^{1/\alpha}\right)\right] dH(z)}$$

Since $q = \prod_{n=1}^{N} v_n^{\frac{1}{N}}$, we the cross sectional distribution of efficiency $F(q)$ is

$$F(q) = \int_0^{\infty} ... \int_0^{\infty} K\left(\frac{q^N}{\prod_{n=2}^{N} v_n}\right) dK\left(v_2\right) dK\left(v_3\right) ... dK\left(v_N\right)$$

Imposing the parametric assumptions $H(z) = 1 - (z/z_0)^{-\zeta}$ and $M = mz_0^{-\zeta}$ gives

$$K(v) \quad = \quad e^{-m \int_{z_0}^{\infty} \left[1 - F\left(\left(\frac{v}{z}\right)^{1/\alpha}\right)\right] \zeta z^{-\zeta - 1} dz} = e^{-mv^{-\zeta} \int_0^{(v/z_0)^{1/\alpha}} [1 - F(r)] \alpha \zeta r^{\alpha\zeta - 1} dr}$$

Taking a limit as $z_0 \to 0$ gives

$$K(v) = e^{-\theta v^{-\zeta}}$$

where $\theta = m\int_0^\infty [1-F(r)]\,\alpha\zeta r^{\alpha\zeta-1}dr$. Next, we solve for θ. To do this we can rewrite $F(q)$ (using the substitution $w_n = \theta v_n^{-\zeta}$) as:

$$\begin{aligned} F(q) &= \int_0^\infty ... \int_0^\infty e^{-\theta\left(q^N/\prod_{n=2}^N v_n\right)^{-\zeta}} \theta\zeta v_2^{-\zeta-1} e^{-\theta v_2^{-\zeta}} dv_2 ... \theta\zeta v_N^{-\zeta-1} e^{-\theta v_N^{-\zeta}} dv_N \\ &= \int_0^\infty ... \int_0^\infty e^{-\left(\theta q^{-\zeta}\right)^N/\prod_{n=2}^N w_n} e^{-w_2} dw_2 ... e^{-w_N} dw_N \end{aligned}$$

so that the density is

$$F'(q) = \frac{1}{q} N\zeta\left(\theta q^{-\zeta}\right)^N \int_0^\infty ... \int_0^\infty e^{-\left(\theta q^{-\zeta}\right)^N/\prod_{n=2}^N w_n} e^{-w_2}\frac{dw_2}{w_2} ... e^{-w_N}\frac{dw_N}{w_N}$$

We can take the expression for θ, integrate by parts, and substitute in for F' to get

$$\begin{aligned} \theta &= m\int_0^\infty [1-F(r)]\,\alpha\zeta r^{\alpha\zeta-1}dr \\ &= m\int_0^\infty F'(r) r^{\alpha\zeta} dr \\ &= m\int_0^\infty \frac{1}{r} N\zeta\left(\theta r^{-\zeta}\right)^N \int_0^\infty ... \int_0^\infty e^{-\left(\theta r^{-\zeta}\right)^N/\prod_{n=2}^N w_n} e^{-w_2}\frac{dw_2}{w_2} ... e^{-w_N}\frac{dw_N}{w_N} r^{\alpha\zeta} dr \end{aligned}$$

Making the substitution $u = \frac{\left(\theta r^{-\zeta}\right)^N}{\prod_{n=2}^N w_n}$, this becomes

$$\begin{aligned} \theta &= m\theta^\alpha \int_0^\infty \int_0^\infty ... \int_0^\infty e^{-u} e^{-w_2}\frac{dw_2}{w_2} ... e^{-w_N}\frac{dw_N}{w_N} u^{1-\frac{\alpha}{N}} \left(\prod_{n=2}^N w_n\right)^{1-\frac{\alpha}{N}} \frac{du}{u} \\ &= m\theta^\alpha \Gamma\left(1-\frac{\alpha}{N}\right)^N \end{aligned}$$

Aggregate productivity Q can be computed as follows:

$$\begin{aligned} Q^{\varepsilon-1} &= \int_0^\infty q^{\varepsilon-1} dF(q) \\ &= \int_0^\infty q^{\varepsilon-1}\frac{1}{q} N\zeta\left(\theta q^{-\zeta}\right)^N \int_0^\infty ... \int_0^\infty e^{-\left(\theta q^{-\zeta}\right)^N/\prod_{n=2}^N w_n} e^{-w_2}\frac{dw_2}{w_2} ... e^{-w_N}\frac{dw_N}{w_N} dq \end{aligned}$$

Making the substitution $u = \frac{\left(\theta q^{-\zeta}\right)^N}{\prod_{n=2}^N w_n}$ and this becomes

$$\begin{aligned} Q^{\varepsilon-1} &= \theta^{\frac{\varepsilon-1}{\zeta}} \int_0^\infty ... \int_0^\infty \left(\prod_{n=2}^N w_n\right)^{1-\frac{\varepsilon-1}{\zeta}/N} u^{1-\frac{\varepsilon-1}{\zeta}/N} e^{-u} e^{-w_2} \frac{dw_2}{w_2} ... e^{-w_N} \frac{dw_N}{w_N} \frac{du}{u} \\ &= \theta^{\frac{\varepsilon-1}{\zeta}} \Gamma\left(1 - \frac{(\varepsilon-1)/\zeta}{N}\right)^N \end{aligned}$$

so that aggregate productivity is

$$Q = \Gamma\left(1 - \frac{(\varepsilon-1)/\zeta}{N}\right)^{\frac{N}{\varepsilon-1}} m^{\frac{1}{(1-\alpha)\zeta}} \Gamma\left(1 - \frac{\alpha}{N}\right)^{N \frac{1}{(1-\alpha)\zeta}}$$

F Dynamics

The joint distribution over efficiencies at t and $\hat{t}$ is

$$F_{t,\hat{t}}(q,\hat{q}) = e^{-M_t \int_0^\infty \left[1 - F_{t,\hat{t}}\left(\left(\frac{q}{z}\right)^{\frac{1}{\alpha}}, \left(\frac{\hat{q}}{z}\right)^{\frac{1}{\alpha}}\right)\right] dH(z)} e^{-(M_{\hat{t}} - M_t) \int_0^\infty \left[1 - F_{\hat{t}}\left(\frac{\hat{q}}{z}\right)^{\frac{1}{\alpha}}\right] dH(z)}$$

In this section, the subscripts t and $\hat{t}$ will be omitted except when necessary for clarity.

Claim 4. *With the functional form, $F_{t,\hat{t}}(q,\hat{q}) = F_t(q)^{\tau(\hat{q}/q)}$ where τ satisfies*

$$\tau(x) = \tau\left(x^{1/\alpha}\right)^\alpha + \left(\frac{\theta_{\hat{t}}}{\theta_t} - \left(\frac{\theta_{\hat{t}}}{\theta_t}\right)^\alpha\right) x^{-\zeta} \tag{18}$$

Proof. With the functional forms, we have for $k > 0$

$$\begin{aligned} F_{t,\hat{t}}(kq, k\hat{q}) &= e^{-m_t \int_0^\infty \left[1 - F_{t,\hat{t}}\left(\left(\frac{kq}{z}\right)^{\frac{1}{\alpha}}, \left(\frac{k\hat{q}}{z}\right)^{\frac{1}{\alpha}}\right)\right] \zeta z^{-\zeta-1} dz} e^{-(m_{\hat{t}} - m_t) \int_0^\infty \left[1 - F_{\hat{t}}\left(\frac{k\hat{q}}{z}\right)^{\frac{1}{\alpha}}\right] \zeta z^{-\zeta-1} dz} \\ &= e^{-m_t k^{-\zeta} \int_0^\infty \left[1 - F_{t,\hat{t}}\left(\left(\frac{q}{w}\right)^{\frac{1}{\alpha}}, \left(\frac{\hat{q}}{w}\right)^{\frac{1}{\alpha}}\right)\right] \zeta w^{-\zeta-1} dw} e^{-(m_{\hat{t}} - m_t) k^{-\zeta} \int_0^\infty \left[1 - F_{\hat{t}}\left(\frac{\hat{q}}{w}\right)^{\frac{1}{\alpha}}\right] \zeta w^{-\zeta-1} dw} \\ &= F_{t,\hat{t}}(q,\hat{q})^{k^{-\zeta}} \end{aligned}$$

Define $\tau(x) \equiv -\frac{1}{\theta_t} \log F_{t,\hat{t}}(1,x)$ so that $F_{t,\hat{t}}(q,\hat{q}) = e^{-\theta_t q^{-\zeta} \tau(\hat{q}/q)} = F_t(q)^{\tau(\hat{q}/q)}$.

Then using the definition of τ, we have

$$\begin{aligned}
\tau(x) &= \frac{m_t}{\theta_t}\int_0^\infty \left[1 - F_{t,\hat{t}}\left(\left(\frac{1}{z}\right)^{\frac{1}{\alpha}}, \left(\frac{x}{z}\right)^{\frac{1}{\alpha}}\right)\right]\zeta z^{-\zeta-1}dz + \frac{m_{\hat{t}} - m_t}{\theta_t}\int_0^\infty \left[1 - F_{\hat{t}}\left(\frac{x}{z}\right)^{\frac{1}{\alpha}}\right]\zeta z^{-\zeta-1}dz \\
&= \frac{m_t}{\theta_t}\int_0^\infty \left[1 - e^{-\theta_t \tau\left(x^{\frac{1}{\alpha}}\right) z^{\zeta/\alpha}}\right]\zeta z^{-\zeta-1}dz + \frac{(m_{\hat{t}} - m_t)}{\theta_t}\int_0^\infty \left[1 - e^{-\theta_{\hat{t}}\left(x^{\frac{1}{\alpha}}\right)^{-\zeta} z^{\zeta/\alpha}}\right]\zeta z^{-\zeta-1}dz \\
&= \frac{m_t}{\theta_t}\left(\theta_t \tau\left(x^{\frac{1}{\alpha}}\right)\right)^\alpha \Gamma(1-\alpha) + \frac{(m_{\hat{t}} - m_t)}{\theta_t}\left(\theta_{\hat{t}}\left(x^{\frac{1}{\alpha}}\right)^{-\zeta}\right)^\alpha \Gamma(1-\alpha) \\
&= \tau\left(x^{\frac{1}{\alpha}}\right)^\alpha + \left(\frac{\theta_{\hat{t}}}{\theta_t} - \left(\frac{\theta_{\hat{t}}}{\theta_t}\right)^\alpha\right) x^{-\zeta}
\end{aligned}$$

■

F.1 Increase in Efficiency of a Technique

How much does the efficiency of an individual single technique improve between t and $\hat{t}$? Among techniques that deliver efficiency q at t, let $I_{t,\hat{t}}(x|q)$ be the fraction that deliver efficiency no greater than xq at $\hat{t}$.

Claim 5.

$$I_{t,\hat{t}}(x|q) = \frac{\int_0^{(q/z_0)^{1/\alpha}} \frac{1}{q_s^\alpha} H'\left(\frac{q}{q_s^\alpha}\right) \frac{\partial F_{t,\hat{t}}\left(q_s, q_s x^{1/\alpha}\right)}{\partial q} dq_s}{\int_0^{(q/z_0)^{1/\alpha}} \frac{1}{\tilde{q}_s^\alpha} H'\left(\frac{q}{\tilde{q}_s^\alpha}\right) dF_t(\tilde{q}_s)} \tag{19}$$

Proof.

The probability that the efficiency of a single technique increases by the proportion x between t and $\hat{t}$ is the same as the probability that the associated supplier's efficiency increases by the proportion $x^{1/\alpha}$.

For a supplier with efficiency q_s at t, the probability that a single downstream technique delivers efficiency no greater than q at t is $H(q/q_s^\alpha)$, with density $\frac{1}{q_s^\alpha} H\left(\frac{q}{q_s^\alpha}\right)$. Bayes rule implies that the conditional density of suppliers' efficiency among techniques that deliver efficiency q at t is

$$\frac{\frac{1}{q_s^\alpha} H'\left(\frac{q}{q_s^\alpha}\right) F_t'(q_s)}{\int_0^{(q/z_0)^{1/\alpha}} \frac{1}{\tilde{q}_s^\alpha} H'\left(\frac{q}{\tilde{q}_s^\alpha}\right) dF_t(\tilde{q}_s)}$$

Among suppliers with efficiency q_s at t, the fraction with efficiency no greater than $q_s x^{1/\alpha}$ at $\hat{t}$ is $\frac{1}{F_t'(q_s)} \frac{\partial F_{t,\hat{t}}\left(q_s, q_s x^{1/\alpha}\right)}{\partial q}$, where the $\frac{\partial F_{t,\hat{t}}(\cdot,\cdot)}{\partial q}$ is the partial derivative with respect to its first argument.

To find, I, we integrate the probability that the supplier's efficiency improved by no more than $x^{1/\alpha}$ over the possible values of the supplier's efficiency. ■

Claim 6. *With the functional forms, $I(x|q) = \tau(x) + \frac{1}{\zeta}\tau'(x)\,x$.*

Proof. First, we use the functional forms and take the limit as $z_0 \to 0$ to get

$$\frac{\frac{1}{q_s^\alpha} H'\left(\frac{q}{q_s^\alpha}\right)}{\int_0^{(q/z_0)^{1/\alpha}} \frac{1}{\tilde{q}_s^\alpha} H'\left(\frac{q}{\tilde{q}_s^\alpha}\right) dF_t(\tilde{q}_s)} \to \frac{q_s^{\alpha\zeta}}{\int_0^\infty \tilde{q}_s^{-(1-\alpha)\zeta-1} e^{-\theta_t \tilde{q}_s^{-\zeta}} d\tilde{q}_s} = \frac{q_s^{\alpha\zeta}}{\theta_t^\alpha \Gamma(1-\alpha)}$$

Note that this expression is independent of the efficiency delivered by the technique.

Second, we have that with the functional forms we have $F_{t,\hat{t}}(q,\hat{q}) = e^{-\theta_t q^{-\zeta}\tau(\hat{q}/q)}$, which means the derivative with respect to its first argument is $\frac{\partial F_{t,\hat{t}}(q,\hat{q})}{\partial q} = \left[\tau\left(\frac{\hat{q}}{q}\right) + \frac{1}{\zeta}\tau'\left(\frac{\hat{q}}{q}\right)\frac{\hat{q}}{q}\right]\theta_t \zeta q^{-\zeta-1} e^{-\theta_t q^{-\zeta}\tau(\hat{q}/q)}$. Thus we have

$$\frac{\partial F_{t,\hat{t}}\left(q_s, q_s x^{1/\alpha}\right)}{\partial q} = \left[\tau\left(x^{1/\alpha}\right) + \frac{1}{\zeta}\tau'\left(x^{1/\alpha}\right)x^{1/\alpha}\right]\theta_t \zeta q_s^{-\zeta-1} e^{-\theta_t q_s^{-\zeta}\tau\left(x^{1/\alpha}\right)}$$

We can simplify the term in brackets by using $\tau\left(x^{1/\alpha}\right)^\alpha = \tau(x) - \left(\frac{\theta_{\hat{t}}}{\theta_t} - \left(\frac{\theta_{\hat{t}}}{\theta_t}\right)^\alpha\right)x^{-\zeta}$ and its derivative, $\tau\left(x^{1/\alpha}\right)^{\alpha-1}\tau'\left(x^{1/\alpha}\right)x^{1/\alpha-1} = \tau'(x) - \left(\frac{\theta_{\hat{t}}}{\theta_t} - \left(\frac{\theta_{\hat{t}}}{\theta_t}\right)^\alpha\right)(-\zeta)\,x^{-\zeta-1}$, so this becomes

$$\frac{\partial F_{t,\hat{t}}\left(q_s, q_s x^{1/\alpha}\right)}{\partial q} = \left[\tau(x) + \frac{1}{\zeta}\tau'(x)\,x\right]\tau(x^{1/\alpha})^{1-\alpha}\theta_t \zeta q_s^{-\zeta-1} e^{-\theta_t q_s^{-\zeta}\tau\left(x^{1/\alpha}\right)}$$

Finally, eqution 1 can be written as

$$\begin{aligned} I(x|q) &\to \int_0^\infty \frac{q_s^{\alpha\zeta}}{\theta_t^\alpha \Gamma(1-\alpha)}\left[\tau(x) + \frac{1}{\zeta}\tau'(x)\,x\right]\tau(x^{1/\alpha})^{1-\alpha}\theta_t \zeta q_s^{-\zeta-1} e^{-\theta_t q_s^{-\zeta}\tau\left(x^{1/\alpha}\right)} dq_s \\ &= \tau(x) + \frac{1}{\zeta}\tau'(x)\,x \end{aligned}$$

■

F.2 Conditional Distributions

This section gives equations for the conditional distributions, R^{same}, R^{unused}, and R^{new}. These can be characterized as follows:

Claim 7. *The conditional distributions satisfy*

1. $R^{same}(\hat{q}|q) = I\left(\frac{\hat{q}}{q}|q\right)$
2. $R^{unused}(\hat{q}|q) = e^{-M_t \int_0^q \left[1 - I\left(\frac{\hat{q}}{\tilde{q}}|\tilde{q}\right)\right] dG_t(\tilde{q})}$
3. $R^{new}(\hat{q}) = e^{-(M_{\hat{t}} - M_t)[1 - G_{\hat{t}}(\hat{q})]}$

Proof. That $R^{same}(\hat{q}|q) = I\left(\frac{\hat{q}}{q}|q\right)$ follows from the definition of I. The expression for $R^{new}(\hat{q})$ can be derived in the same way as the derivation of as the cross sectional distribution. Turning to R^{unused}, consider a single technique with efficiency no greater than q at t. What is the probability that it delivers efficiency no greater than $\hat{q}$ at $\hat{t}$? The density of efficiency at t among such techniques is $\frac{G_t'(\tilde{q})}{G_t(q)}$, so the probability that the technique delivers efficiency no greater than $\hat{q}$ at $\hat{t}$ is $\frac{\int_0^q I(\hat{q}/\tilde{q}|\tilde{q}) dG_t(\tilde{q})}{G_t(q)}$.

Among firms with efficiency $\tilde{q}$ at t, the fraction with n techniques at t can be computed using Bayes rule:

$$\Pr\left(n \text{ techniques given efficiency } q\right) = \frac{\frac{e^{-M_t} M_t^n}{n!} n G_t(q)^{n-1} G_t'(q)}{\sum_{k=1}^{\infty} \frac{e^{-M_t} M_t^k}{k!} k G_t(q)^{k-1} G_t'(q)} = \frac{\frac{e^{-M_t} M_t^{n-1}}{(n-1)!} G_t(q)^{n-1}}{F_t(q)}$$

The numerator is the product the probability that a firm discovered n techniques and the probability that most efficient of these delivers efficiency q. The denominator is the sum of this quantity over the possible number of techniques.

For a firm with n techniques and efficiency q at t, the probability that the $n-1$ that are not used at t all deliver efficiency no greater than $\hat{q}$ at $\hat{t}$ is $\left(\frac{\int_0^q I(\hat{q}/\tilde{q}|\tilde{q}) dG_t(\tilde{q})}{G_t(q)}\right)^{n-1}$. Therefore among all firms with efficiency q

at t, the probability that no unused techniques deliver efficiency greater than $\hat{q}$ at t is

$$\begin{aligned}
R^{unused}(\hat{q}|q) &= \sum_{n=1}^{\infty} \frac{\frac{e^{-M_t} M_t^{n-1}}{(n-1)!} G_t(q)^{n-1}}{F_t(q)} \left(\frac{\int_0^q I(\hat{q}/\tilde{q}|\tilde{q}) dG_t(\tilde{q})}{G_t(q)} \right)^{n-1} \\
&= \sum_{n=1}^{\infty} \frac{\frac{e^{-M_t} M_t^{n-1}}{(n-1)!} \left(\int_0^q I(\hat{q}/\tilde{q}|\tilde{q}) dG_t(\tilde{q}) \right)^{n-1}}{F_t(q)} \\
&= \frac{e^{-M_t \left[1 - \int_0^q I(\hat{q}/\tilde{q}|\tilde{q}) dG_t(\tilde{q})\right]}}{F_t(q)} \\
&= e^{-M_t \int_0^q [1 - I(\hat{q}/\tilde{q}|\tilde{q})] dG_t(\tilde{q})}
\end{aligned}$$

■

Claim 8. *With the functional forms, these are given by*

1. $R^{same}(\hat{q}|q) = \tau\left(\frac{\hat{q}}{q}\right) + \frac{1}{\zeta}\tau'\left(\frac{\hat{q}}{q}\right)\frac{\hat{q}}{q}$
2. $R^{unused}(\hat{q}|q) = e^{-\theta_{\hat{t}}\hat{q}^{-\zeta}\left[\frac{m_t}{m_{\hat{t}}} - 1\right] - \theta_t q^{-\zeta}[\tau(\hat{q}/q) - 1]}$
3. $R^{new}(\hat{q}) = e^{-\theta_{\hat{t}}\hat{q}^{-\zeta}\left[1 - \frac{m_t}{m_{\hat{t}}}\right]}$

Proof. The expression for R^{same} follows directly from the expression for I. For R^{new} we have

$$R^{new}(\hat{q}|q) = e^{-(M_{\hat{t}} - M_t)[1 - G_{\hat{t}}(\hat{q})]} = e^{-M_{\hat{t}}[1 - G_{\hat{t}}(\hat{q})]\frac{M_{\hat{t}} - M_t}{M_{\hat{t}}}} \to e^{-\theta_{\hat{t}}\hat{q}^{-\zeta}\frac{m_{\hat{t}} - m_t}{m_{\hat{t}}}}$$

To find R^{unused}, we first have that with the functional forms, $\lim_{z_0 \to 0} M_t G_t'(\tilde{q}) = \theta_t \zeta \tilde{q}^{-\zeta-1}$. Thus the integral

$$\begin{aligned}
M_t \int_0^q [1 - I(\hat{q}/\tilde{q}|\tilde{q})] \, dG_t(\tilde{q}) \quad &\to \quad \int_0^q [1 - I(\hat{q}/\tilde{q})] \, \theta_t \zeta \tilde{q}^{-\zeta-1} d\tilde{q} \\
&= \quad \theta_t \hat{q}^{-\zeta} \int_{\hat{q}/q}^{\infty} [1 - I(u)] \, \zeta u^{\zeta-1} d\tilde{q}
\end{aligned}$$

We can substitute in the expression for I to get

$$\begin{aligned}
\int_{\hat{q}/q}^{\infty} [1 - I(u)] \, \zeta u^{\zeta-1} d\tilde{q} &= \int_{\hat{q}/q}^{\infty} \left[1 - \tau(u) - \frac{1}{\zeta}\tau'(u)u\right] \zeta u^{\zeta-1} d\tilde{q} \\
&= \int_{\hat{q}/q}^{\infty} [1 - \tau(u)] \, \zeta u^{\zeta-1} - \tau'(u) u^{\zeta} d\tilde{q} \\
&= [1 - \tau(u)] \, u^{\zeta} \Big|_{\hat{q}/q}^{\infty}
\end{aligned}$$

To find $\lim_{u\to\infty} \frac{1-\tau(u)}{u^{\zeta}}$, note that we can rewrite equation (18) as

$$\frac{\tau(u)-1}{u^{-\zeta}} = \frac{\tau(u^{1/\alpha})^{\alpha}-1}{\left[\tau(u^{1/\alpha})-1\right]^{\alpha}}\left[\frac{\tau(u^{1/\alpha})-1}{(u^{1/\alpha})^{-\zeta}}\right]^{\alpha} + \left(\frac{\theta_{\hat{t}}}{\theta_t} - \left(\frac{\theta_{\hat{t}}}{\theta_t}\right)^{\alpha}\right)$$

Since $\lim_{u\to\infty} \frac{\tau(u^{1/\alpha})^{\alpha}-1}{\left[\tau(u^{1/\alpha})-1\right]^{\alpha}} = 0$, we have $\lim_{u\to\infty} \frac{1-\tau(u)}{u^{\zeta}} = \left(\frac{\theta_{\hat{t}}}{\theta_t}\right)^{\alpha} - \frac{\theta_{\hat{t}}}{\theta_t}$. Putting these pieces together, we have

$$R^{unused}(\hat{q}|q) \to e^{-\theta_t \hat{q}^{-\zeta}\left\{\left(\frac{\theta_{\hat{t}}}{\theta_t}\right)^{\alpha} - \frac{\theta_{\hat{t}}}{\theta_t} - [1-\tau(\hat{q}/q)](\hat{q}/q)^{\zeta}\right\}} = e^{-\left[\frac{m_t}{m_{\hat{t}}}-1\right]\theta_{\hat{t}}\hat{q}^{-\zeta} - \theta_t q^{-\zeta}[\tau(\hat{q}/q)-1]}$$

■

F.3 Surplus

The average surplus among firms with efficiency q is

$$S(q) = q^{\varepsilon-1} + M\int_{z_0}^{\infty}\int_0^{\infty}\left[S\left(\max\{\tilde{q}, zq^{\alpha}\}\right) - S(\tilde{q})\right] d\tilde{F}(\tilde{q})dH(z)$$

Claim 9. *Using the functional forms,* $\lim_{z_0\to 0} S(q) = q^{\varepsilon-1} + \frac{mq^{\zeta\alpha}}{\theta}\frac{1}{1-\alpha}\frac{\varepsilon-1}{\zeta}Q^{\varepsilon-1}$

Proof. Consider the surplus created by all downstream techniques. Using the functional forms, this can be written as

$$\begin{aligned}
M\int_{z_0}^{\infty}\int_0^{zq^{\alpha}}\left[S\left(zq^{\alpha}\right) - S(\tilde{q})\right] d\tilde{F}(\tilde{q})dH(z) &= mz_0^{-\zeta}\int_{z_0}^{\infty}\int_0^{zq^{\alpha}}\left[S\left(zq^{\alpha}\right) - S(\tilde{q})\right] d\tilde{F}(\tilde{q})z_0^{\zeta}\zeta z^{-\zeta-1}dz \\
&= mq^{\zeta\alpha}\int_{z_0q^{\alpha}}^{\infty}\int_0^{w}\left[S(w) - S(\tilde{q})\right] d\tilde{F}(\tilde{q})\zeta w^{-\zeta-1}dw \\
&\to mq^{\zeta\alpha}\int_0^{\infty}\int_0^{w}\left[S(w) - S(\tilde{q})\right] dF(\tilde{q})\zeta w^{-\zeta-1}dw
\end{aligned}$$

Let $\kappa \equiv \theta\int_0^{\infty}\int_0^{w}\left[S(w) - S(\tilde{q})\right] dF(\tilde{q})\zeta w^{-\zeta-1}dw$ so that the surplus can be written as

$$S(q) = q^{\varepsilon-1} + \frac{mq^{\zeta\alpha}}{\theta}\kappa \tag{20}$$

To derive an expression for κ, we can use use the expression for $S(q)$ and plug this into the definition of κ.

$$\begin{aligned}
\kappa &= \theta \int_0^\infty \int_0^w S(w) dF(\tilde{q}) \zeta w^{-\zeta-1} dw - \theta \int_0^\infty \int_0^w S(\tilde{q}) dF(\tilde{q}) \zeta w^{-\zeta-1} dw \\
&= \theta \int_0^\infty S(w) F(w) \zeta w^{-\zeta-1} dw - \theta \int_0^\infty \int_{\tilde{q}}^\infty \zeta w^{-\zeta-1} dw S(\tilde{q}) dF(\tilde{q}) \\
&= \theta \int_0^\infty S(w) F(w) \zeta w^{-\zeta-1} dw - \theta \int_0^\infty \tilde{q}^{-\zeta} S(\tilde{q}) dF(\tilde{q})
\end{aligned}$$

Plugging in the expression for S from equation (20) and integrating gives

$$\begin{aligned}
\kappa &= \theta^{\frac{\varepsilon-1}{\zeta}} \Gamma\left(1 - \frac{\varepsilon-1}{\zeta}\right) + \frac{m\Gamma(1-\alpha)}{\theta^{1-\alpha}} \kappa + \theta^{\frac{\varepsilon-1}{\zeta}} \Gamma\left(2 - \frac{\varepsilon-1}{\zeta}\right) + \frac{m\Gamma(2-\alpha)}{\theta^{1-\alpha}} \kappa \\
&= \frac{\varepsilon-1}{\zeta} \Pi^0 + \alpha\kappa
\end{aligned}$$

Thus $\kappa = \frac{1}{1-\alpha} \frac{\varepsilon-1}{\zeta} Q^{\varepsilon-1}$, so the surplus of a firm is ■

F.3.1 Surplus for a Technique

Claim 10. *With the functional forms, relative to the average across all firms, the surplus generated by a new upstream technique for a firm with efficiency q is $(1-\alpha)\frac{\left(\theta q^{-\zeta}\right)^{1-\frac{\varepsilon-1}{\zeta}}}{\Gamma(2-\frac{\varepsilon-1}{\zeta}}+\alpha\frac{\left(\theta q^{-\zeta}\right)^{1-\alpha}}{\Gamma(2-\alpha}$ Relative to the average across all firms, the surplus generated by a new downstream technique for a firm with efficiency q is $\frac{\lambda}{\theta} q^{\zeta\alpha}$*

Proof. Let $V^U(q) \equiv \int_q^\infty [S(\hat{q}) - S(q)]\, dG(\hat{q})$ be the expected surplus of a single upstream technique for a firm with efficiency q. As an intermediate step, given the expression for S we have

$$\begin{aligned}
\int_q^\infty S(\hat{q}) \zeta \hat{q}^{-\zeta-1} d\hat{q} &= \frac{q^{\varepsilon-1-\zeta}}{1-\frac{\varepsilon-1}{\zeta}} + \frac{1}{1-\alpha} \frac{\theta^{-\alpha} q^{\zeta\alpha-\zeta}}{\Gamma(1-\alpha)} \frac{1}{1-\alpha} \frac{\varepsilon-1}{\zeta} Q^{\varepsilon-1} \\
&= \frac{\left(q^{-\zeta}\right)^{1-\frac{\varepsilon-1}{\zeta}}}{1-\frac{\varepsilon-1}{\zeta}} + \frac{\theta^{-\alpha}\left(q^{-\zeta}\right)^{1-\alpha}}{\Gamma(1-\alpha)} \frac{1}{1-\alpha} \frac{\varepsilon-1}{\zeta} Q^{\varepsilon-1}
\end{aligned}$$

With this, derive an expression for $MV^U(q)$, using $MG'(q) \to \theta\zeta q^{-\zeta-1}$:

$$\begin{aligned}
MV^U(q) &= \int_q^\infty [S(\hat{q}) - S(q)]\,\theta\zeta q^{-\zeta-1} d\hat{q} = \theta \int_q^\infty S(\hat{q})\zeta q^{-\zeta-1} d\hat{q} - S(q)\theta q^{-\zeta} \\
&= \frac{\frac{\varepsilon-1}{\zeta}}{1-\frac{\varepsilon-1}{\zeta}}\theta q^{\varepsilon-1-\zeta} + \frac{\alpha}{1-\alpha}\frac{\theta^{1-\alpha}q^{\zeta\alpha-\zeta}}{\Gamma(1-\alpha)}\frac{1}{1-\alpha}\frac{\varepsilon-1}{\zeta}Q^{\varepsilon-1} \\
&= \left[(1-\alpha)\frac{\left(\theta q^{-\zeta}\right)^{1-\frac{\varepsilon-1}{\zeta}}}{\Gamma\left(2-\frac{\varepsilon-1}{\zeta}\right)} + \alpha\frac{\left(\theta q^{-\zeta}\right)^{1-\alpha}}{\Gamma(2-\alpha)}\right]\frac{1}{1-\alpha}\frac{\varepsilon-1}{\zeta}Q^{\varepsilon-1}
\end{aligned}$$

Integrating over all values of q gives $\int_0^\infty MV^U(\tilde{q})dF(\tilde{q}) = \frac{1}{1-\alpha}\frac{\varepsilon-1}{\zeta}Q^{\varepsilon-1}$

$$\frac{V^U(q)}{\int_0^\infty V^U(\tilde{q})dF(\tilde{q})} = \frac{MV^U(q)}{\int_0^\infty MV^U(\tilde{q})dF(\tilde{q})} = (1-\alpha)\frac{\left(\theta q^{-\zeta}\right)^{1-\frac{\varepsilon-1}{\zeta}}}{\Gamma\left(2-\frac{\varepsilon-1}{\zeta}\right)} + \alpha\frac{\left(\theta q^{-\zeta}\right)^{1-\alpha}}{\Gamma(2-\alpha)}$$

Let $V^D(q) \equiv \int_{z_0}^\infty \int_0^{qz^\alpha} [S(zq^\alpha) - S(\tilde{q}]\,dF(\tilde{q})dH(z)$ be the expected value of a downstream technique for a firm with efficiency q. Then the value of a downstream technique is

As in the derivation of the expression for $S(q)$, we have the $MV^D(q) \to \frac{m}{\theta}q^{\zeta\alpha}\kappa$. Then for a firm with efficiency q, the value of an upstream technique relative to the average among other firms is

$$\frac{V^D(q)}{\int_0^\infty V^D(\tilde{q})dF(\tilde{q})} = \frac{MV^D(q)}{\int_0^\infty MV^D(\tilde{q})dF(\tilde{q})} \to \frac{\frac{m}{\theta}q^{\zeta\alpha}\kappa}{\int_0^\infty \frac{m}{\theta}\tilde{q}^{\zeta\alpha}\kappa dF(\tilde{q})} = \frac{m}{\theta}q^{\zeta\alpha}$$

■