Span of Control, Transaction Costs and the Structure of Production Chains[1]

Tomoo Kikuchi,[a] Kazuo Nishimura[b] and John Stachurski[c]

[a] Lee Kuan Yew School of Public Policy, National University of Singapore
[b] Research Institute of Economics and Business Administration, Kobe University
[c] Research School of Economics, Australian National University

September 7, 2017

ABSTRACT. This paper builds coordination costs, transaction costs and other aspects of the theory of the firm into a production chain model with an infinite number of *ex ante* identical producers. The equilibrium determines prices, allocations of productive tasks across firms, firm sizes, and the number of active firms. These prices and allocations match several stylized facts on firm boundaries, vertical integration, and division of the value chain.

JEL Classifications: C62, C63, L11, L16
Keywords: Span of control, firm boundaries, fragmentation, production chains

1. INTRODUCTION

Reflecting on a conversation with an ex-Soviet official wishing to know who was in charge of supplying bread to the city of London, Seabright (2010) observed that "there was nothing naive about his question, because the answer 'nobody is in charge' is, when one thinks about it, astonishingly hard to believe." Seabright's comment highlights the ability of market forces to coordinate a vast number of specialized activities with minimal central planning.

As pointed out by Coase (1937), however, even in free market economies a great deal of top-down planning does in fact take place. The difference is that, rather

[1] This is a revised version of the 2012 working paper "Coase meets Tarski: New Insights from Coase's Theory of the Firm," KIER Working Paper No 828. The authors thank Michal Fabinger, Michihiro Kandori, Debraj Ray, Tom Sargent, Makoto Yano and Shenghao Zhu for helpful discussions and suggestions. The input of Giuseppe Moscarini and three anonymous referees is also gratefully acknowledged. Denis Tan, Ngoc Han Huynh Khanh, Chenghan Hou, Qingyin Ma and Alex Olssen provided excellent research assistance. We received financial support from JSPS Grant-in-Aid for Scientific Research (23000001 and 23330063), Singapore Ministry of Education Academic Research Fund Tier 1 R-122-000-140-112 and ARC Award DP120100321.

Email: `spptk@nus.edu.sg`, `nishimura@rieb.kobe-u.ac.jp`, `john.stachurski@anu.edu.au`

than bureaucrats in government departments, most of these planners are managers, working in entities called *firms*. The operation of free market economies blends planning within firms and decentralized production coordinated by prices. Firms are "islands of conscious power in the ocean of unconscious cooperation."[2]

The islands to which Coase refers exhibit striking diversity. For example, in 2011, Royal Dutch Shell operated in over 80 countries, had annual revenue exceeding the GDP of 150 nations, and paid its CEO 35 times the salary received by the President of the United States. In the same year, the total number of employees at Walmart exceeded the population of all but 4 US cities. In addition to such giants, tens of millions of smaller firms operate around the world.[3]

What forces shape the number and size of large firms and the multitude of smaller ones? The pioneering work of Coase (1937) addressed at least some of these questions by analyzing the trade-off between intra-firm coordination costs, which tend to inhibit firm growth, and external transaction costs, which encourage it. Subsequent researchers analyzed firm size and firm boundaries by considering the effects of imperfect information, incentive and agency problems, incomplete contracts, property rights, decision rights, and the microfoundations of transaction costs.[4]

In this paper we study the determinants of firm size, firm heterogenity and allocation of tasks across firms in the context of production chains coordinated by prices. As in Coase (1937), firms along the chain have nontrivial size because of transaction costs associated with using the market. Entrepreneurs and managers can sometimes coordinate production at a lower cost within the firm. A countervailing force, referred to by Coase (1937) as "diminishing returns to management," prevents firms expanding without limit.[5] The boundary of the firm is determined by the point at which the cost of organizing another productive task within the firm is equal to the cost of acquiring a similar input or service through the market.

[2]This phrase from Coase's essay is originally due to Robertson (1923).

[3]Sources: United States Census Bureau and the Forbes Global 2000 List.

[4]Well known contributions include Jensen and Meckling (1976), Williamson (1979, 1981), Klein et al. (1978), Hart and Moore (1990), Holmstrom and Milgrom (1994) and Grossman and Helpman (2002). Recent surveys include Aghion and Holden (2011) and Bresnahan and Levin (2012).

[5]Well known studies based on this assumption include Lucas (1978) and Becker and Murphy (1992). For Coase, diminishing returns to management were driven by the huge informational requirements associated with large planning problems, leading to "mistakes" and misallocation of resources. The challenges associated with coordinating production through top-down planning were emphasized by Hayek (1945), who highlighted the difficulty of utilizing knowledge not held in its totality by any one individual. Later authors have identified additional causes of high average management costs in larger firms, such as communication costs or free-riding, shirking and other incentive problems. See, for example, Garicano (2000), Jensen and Meckling (1976), Holmstrom and Milgrom (1994), McAfee and McMillan (1995)), Milgrom and Roberts (1992), Geanakoplos and Milgrom (1991), and Meagher (2003).

The firms in the model act competitively. Responding to prices, firms decide how much to produce in-house and how much to subcontract to upstream partners. In this way, *ex ante* identical firms arrange themselves into a production chain with several distinct features. One is that value added is higher the further down the production chain firms are located. A second is that, when firms are allowed to optimally choose the number of subcontractors, downstream firms have the largest number of subcontracting partners, while upstream firms are smaller and have fewer. These results are broadly consistent with the stylized facts on production chains.

As an ancillary result, the model also produces a prediction for the size distribution of firms. While the model lacks the intrinsic heterogeneity required to accurately match observed data, it can yield the kind of heavy tailed firm size distribution that has been noted by many studies.[6] This is interesting because the mapping from identical agents into a power law firm size distribution occurs purely as a result of market interactions and the kinds of frictions identified by the modern theory of the firm, without requiring exogenous shocks or inherent power law features.[7]

One way to understand our model is to observe that firms along the supply chain are in some senses analogous to the layers of management within an organization studied in Garicano (2000). Under this correspondence, the range of tasks performed by a given firm maps to the size or problem solving capability of each layer of management, while trade between firms maps to communication between layers. Our finding that value added and the range of tasks per firm are higher the further we move downstream in the production chain can be seen as partly analogous to the finding in Garicano (2000) that optimal arrangement of management layers has a pyramidal structure, consistent with observed intra-firm organization. Another analogy is that in Garicano (2000), a rise in communication costs between layers tends to increase the size of individual layers in order to economize on communication. Similarly, in our study, a rise in transaction costs between firms increases the size of firms, as firms produce more in-house in order to economize on transaction costs.

There are, however, obvious differences between our work and Garicano (2000). For example, in our study of production chains, prices play an essential role, partly by coordinating production and partly due to the fact that transaction costs are proportional to price. These multiplicative costs compound along the value chain from upstream to downstream, since earlier transaction costs are accommodated in the value of each subsequent transaction. We show that an increase in transaction costs

[6] See, e.g., Axtell (2001), Luttmer (2007), Lucas (1978) or Rossi-Hansberg and Wright (2007).

[7] In this respect, our work connects with Geerolf (2016), who uses the framework provided by Garicano (2000) to study Pareto distributions in labor income, showing that power law distributions can be generated by power production functions.

leads to higher prices, larger firms and a smaller number of firms along one chain in equilibrium.[8]

The multiplicative transaction costs in our model capture the idea that many such costs associated with inter-firm trade are sensitive to the value of a given transaction. Firms will spend more on acquiring information, writing and negotiating contracts, providing insurance, currency hedging or bribes for a transaction that costs $1,000,000 than for a transaction that costs $1,000. At the same time, we acknowledge that some transaction costs are fixed or additive rather than proportional, and for this reason we also study a richer model with both kinds of transaction costs in section 5. In this extension we also allow for the realistic possibility that firms have multiple upstream partners. Thus, not only the relative size of firms but also the number of suppliers per firm is determined endogenously at every level of the industry. This provides a rich network structure that is well suited to the study of modern multi-firm production chains. Although much of our analysis is restricted to simulation, key ideas from the baseline model, such as larger downstream firms in terms of value added, are also found in the extended model.

There are also connections between our paper and the work of Garicano and Rossi-Hansberg (2006), who follow Garicano (2000) in considering layers of management. They show in the appendix that organization of the firm can be decentralized, so that outcomes are the result of market interactions. There is also a numerical component of the study that allows the number of layers to be endogenous. On the other hand, unlike in our model, the number of management layers is constrained by fixed costs associated with each layer. In our model the number of firms is constrained by transaction costs, as envisaged by Coase (1937). These transaction costs compound along the value chain, a fact that magnifies their weight in equilibrium outcomes.

Another related paper in this vein is Caliendo and Rossi-Hansberg (2012), who also extend the model of Garicano (2000), adding a demand side in order to study the impact of trade liberalization on organizations. Among other things, the paper is notable for the fact that the number of management layers is endogenous, just as the number of firms in one production chain is endogenous in our model. At the same time, the proportional transaction costs in our model have no direct counterpart in Caliendo and Rossi-Hansberg (2012), where, as in Garicano and Rossi-Hansberg (2006), the number of layers is constrained by fixed costs associated with each layer. As a result, the cost function in Caliendo and Rossi-Hansberg (2012) evaluated at the optimal management structure is kinked and non-convex. In contrast, prices in

[8] Another way that our model differs from Garicano (2000) is that in Garicano (2000) the number of layers of management is either fixed or infinite. In our model the number of firms is finite and endogenous. This endogeneity allows us to study how transaction and coordination costs impact on all salient aspects of the production chain, including the number of firms.

our model are strictly convex and continuously differentiable. This enables us to obtain sharp results concerning uniqueness of equilibria and other properties, such as a connection between the derivative of equilibrium prices and a classical idea on firm boundaries stated verbally by Coase (1937).

A separate literature connected to our work studies production chains and the trade-offs that determine size and productivity of the chain. An early contribution is the O-ring theory of development presented in Kremer (1993), which analyzes the fragility of production when links in the chain can fail. A more recent example is Levine (2012), which analyzes production chains with Leontief technology and possibly correlated failure probabilities at each stage. He finds that higher failure probabilities lead to shorter chains. Our work differs in that the length of chains is restricted not by failures but rather by transaction costs.

Our work on production chains also connects to studies of fragmentation of production and supply chains found in recent work on international trade. For example, Costinot (2009), Grossman and Rossi-Hansberg (2012), Costinot et al. (2013) and Antràs and Chor (2013) all consider sequential production over a continuum of tasks with a large number of firms. Of these, the most similar to our paper is Costinot et al. (2013) which studies global supply chains. It shares many features with our model, including sequential production, competitive behavior and a potentially sophisticated pattern of vertical specialization. Their model nonetheless differs from ours in that in-house production costs are proportional to the number of tasks performed. Heterogeneity in outcomes is driven not by the trade-off between transaction costs and diminishing returns to management, as in our model, but rather by intrinsic heterogeneity associated with differing productivity levels across countries.

Finally, we obtain a recursive representation of equilibrium prices and allocations that relates to earlier research connecting recursive methods and static problems. One example is Garicano and Rossi-Hansberg (2006), who use recursive methods to study their firm organization problem. Others include Lucas and Rossi-Hansberg (2002) and Hsu et al. (2014), who apply dynamic programming techniques to spacial location problems. Our recursive methods contain a number of innovations, made necessary by the fact that the operator that generates equilibrium prices as a fixed point is in general expansive rather than contractive. As a consequence, our methods deviate from the traditional approach in significant ways. For example, the equilibrium price function is the fixed point of an operator, but its iterates converge to the equilibrium price function only when the initial condition is chosen to match certain shape restrictions.

Details of the model structure can be found in section 2. Section 3 defines the equilibrium and gives existence and uniqueness results. Section 4 considers implications for the structure of the value chain and the distribution of firms. Section 5

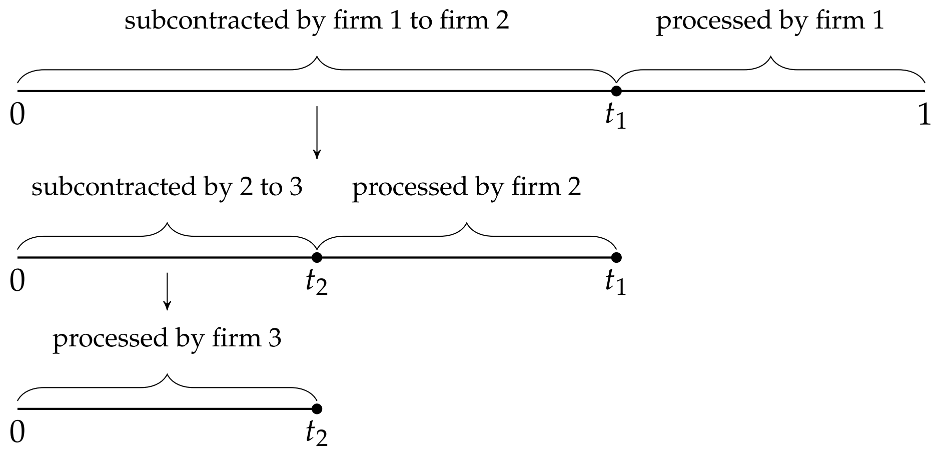


FIGURE 1. Recursive allocation of production tasks

extends our results to the case of multiple upstream partners. Section 6 concludes. Proofs are deferred to the appendix, while code for simulations can be found at `https://github.com/jstac/production_chains`.

2. THE MODEL

We begin by studying production of a single unit of a final good. We start with a linearly ordered production chain, where the good is produced through the sequential completion of a large number of processing stages. On an intuitive level, we can think of movement from one processing stage to the next as requiring a single specialized task. At the same time, to provide a sharper marginal analysis, we model the processing stages as a continuum. In particular, the stages are indexed by $t \in [0,1]$, with $t = 0$ indicating that no tasks have been undertaken and $t = 1$ indicating that the good is complete.

2.1. **The Production Chain.** Allocation of tasks among firms is endogenous. The key ideas can be understood in terms of the subcontracting scheme illustrated in figure 1. In this example, an arbitrary firm—henceforth, firm 1—receives a contract to sell one unit of the completed good to a final buyer. Firm 1 then forms a contract with firm 2 to purchase the partially completed good at stage t_1, with the intention of implementing the remaining $1 - t_1$ tasks in-house (i.e., processing from stage t_1 to stage 1). Firm 2 repeats this procedure, forming a contract with firm 3 to purchase the good at stage t_2. In the example in figure 1, firm 3 decides to complete the chain, selecting $t_3 = 0$.

At this point, production unfolds in the opposite direction (i.e., from upstream to downstream). First, firm 3 completes processing stages from $t_3 = 0$ up to t_2 and transfers the good to firm 2. Firm 2 then processes from t_2 up to t_1 and transfers the good to firm 1, who processes from t_1 to 1 and delivers the completed good to the final buyer. In what follows, the length of the interval of stages carried out by firm i is denoted by ℓ_i. We refer to ℓ_i as the *range* of tasks carried out by firm i. Figure 2 serves to clarify notation.

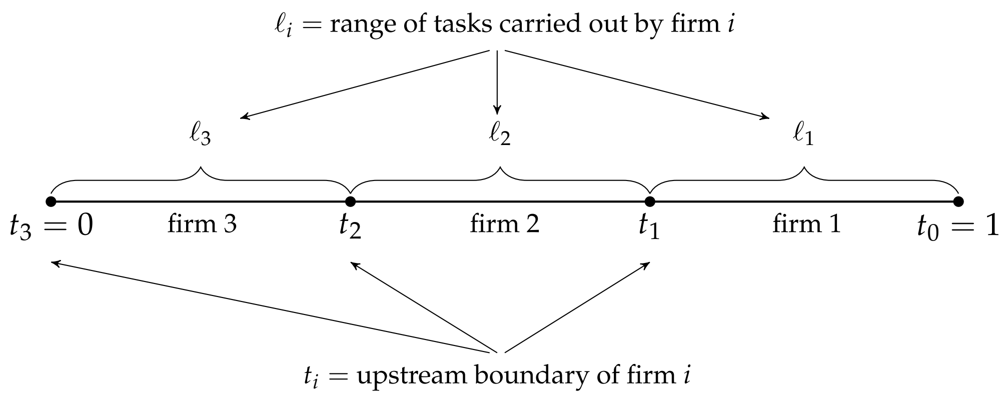


FIGURE 2. Notation

Notice that each firm chooses only its *upstream* boundary, treating its downstream boundary as given. In other words, it chooses how far to integrate backwards into input production. The benefit of this formulation is that it implies a recursive structure for the decision problem for each firm: In choosing how many processing stages to subcontract, each successive firm faces essentially the same decision problem as the firm above it in the chain, with the only difference being that the decision space is a subinterval of the decision space for the firm above. We exploit this recursive structure in our study of equilibrium.

2.2. **In-House Production Costs.** We study allocation of tasks in the presence of what Coase (1937) referred to as diminishing returns to management, that is, rising costs per task when a firm expands the *range* of productive activities implemented within its boundaries and coordinated by its managers. Rising costs per task can be thought of as driven by the expanding informational requirements associated with larger planning problems, leading to progressively higher management costs, incentive problems and misallocation of resources. As in Becker and Murphy (1992), we represent these ideas by taking the cost of carrying out ℓ tasks in-house to be $c(\ell)$, where c is increasing and strictly convex. We also assume that c is continuously differentiable, with $c(0) = 0$ and $c'(0) > 0$. Thus, average cost per task rises with the range of tasks performed in-house. These assumptions also imply that c is strictly increasing.[9]

2.3. **Transaction Costs.** Diminishing returns to management makes in-house production expensive, favoring small firms and external procurement. However, as pointed out by Coase (1937) and reiterated by many authors since, there is a countervailing force that acts against infinite subdivision of firms: the existence of transaction costs associated with buying and selling through the market. One example is the cost of negotiating, drafting monitoring and enforcing contracts with suppliers. Other commonly cited transaction costs include search frictions, transaction fees, taxes, bribes and theft associated with transactions, bargaining and information costs, and the costs of assessing credit worthiness and reliability (see, e.g., Coase (1937); Williamson (1979); North (1993); Blume et al. (2009)).

Transaction costs are represented as a wedge between the buyer's and seller's prices. (Our convention is that the phrase "transaction costs" refers only to transactions that take place through the market, rather than within the firm.) In our model it matters little whether the transaction cost is borne by the buyer, the seller or both (see section 3.3). Hence we assume that the cost is borne only by the buyer. In particular, when two firms agree to a trade at face value v, the buyer's total outlay is δv, where $\delta > 1$. The seller receives only v, and the difference is paid to agents outside the model.

3. Equilibrium

The next step is to define a notion of equilibrium for the production chain. In doing so we assume that all firms are *ex ante* identical and act as price takers. There is a countable infinity of firms $i = 1, 2, \ldots$ There are no fixed costs or barriers to entry.

3.1. **Definition of Equilibrium.** Throughout the paper, an *allocation* is a nonnegative sequence $\boldsymbol{\ell} = \{\ell_i\}_{i \in \mathbb{N}}$ with only finitely many nonzero elements. Recalling figure 2, an allocation $\boldsymbol{\ell}$ defines a division of tasks across firms, with ℓ_i being the range of task implemented by the i-th firm. If $\ell_i = 0$ then firm i is understood to be inactive. We always assume that firms enter in order, with firm 1 being the furthest downstream. This is a labeling convention that confers no special privileges.

[9] The cost function c is assumed to already represent current management best practice, in the sense that no further rearrangement of management structure or internal organization can obtain a lower cost of in-house production. Also note that the cost of carrying out $\ell = s - t$ tasks depends only on the difference $s - t$ rather than s directly. In other words, all tasks are homogeneous. While extensions might consider other cases, our interest is in equilibrium prices and choices of firms in the case where tasks are *ex ante* identical.

An allocation $\boldsymbol{\ell}$ is called *feasible* if $\sum_{i \geqslant 1} \ell_i = 1$. Feasibility means that the entire production process is completed by finitely many firms. Given a feasible allocation $\boldsymbol{\ell}$, let $\{t_i\}$ represent the corresponding transaction stages, defined by

$$t_0 = s \quad \text{and} \quad t_i = t_{i-1} - \ell_i. \tag{1}$$

In particular, t_{i-1} is the downstream boundary of firm i and t_i is its upstream boundary (as in figure 2).

Firms face a price function p, which is a map from $[0,1]$ to $\mathbb{R}_+$, with $p(t)$ interpreted as the price of the good at processing stage t. Since the i-th firm purchases the good at stage t_i, sells it at stage t_{i-1}, and undertakes the remaining ℓ_i tasks in-house, its total costs are its processing costs $c(\ell_i)$ plus gross input cost $\delta p(t_i)$. As transaction costs are incurred only by the buyer, its profits are

$$\pi_i = p(t_{i-1}) - c(\ell_i) - \delta p(t_i). \tag{2}$$

Definition 3.1. Given a price function p and a feasible allocation $\boldsymbol{\ell} = \{\ell_i\}$, let $\{t_i\}$ be the corresponding firm boundaries and let $\{\pi_i\}$ be corresponding profits, as defined in (1) and (2). The pair $(p, \boldsymbol{\ell})$ is called an *equilibrium* for the production chain if

1. $p(0) = 0$,
2. $\pi_i = 0$ for all i, and
3. $p(s) - c(s-t) - \delta p(t) \leqslant 0$ for any pair s, t with $0 \leqslant s \leqslant t \leqslant 1$.

Condition 1 is a zero profit condition for suppliers of initial inputs, which implies that $p(0)$ is equal to the cost of producing these inputs. To simplify notation, we assume that this cost is zero.

Condition 2 states that all active firms make zero profits. Free entry and the infinite fringe of competitors rule out positive profits for incumbents, since any incumbent could be replaced by a member of the competitive fringe filling the same role in the production chain. Profits are never negative in equilibrium because firms can freely exit.

Condition 3 ensures that no firm in the production chain has an incentive to deviate. It also ensures that no inactive firms can enter and extract positive profits.

3.2. **Existence of Equilibrium.** In this section we provide a constructive proof of existence, meaning that the equilibrium is shown to exist and that methods for computing it are also provided. In section 4.2 we will show that this same equilibrium is also unique across a large class of candidate solutions.

To begin, consider the operator T mapping $p \colon [0,1] \to \mathbb{R}_+$ to Tp via

$$Tp(s) = \min_{t \leqslant s} \{c(s-t) + \delta p(t)\} \quad \text{for all} \quad s \in [0,1]. \tag{3}$$

Here and below, the restriction $0 \leqslant t$ in the minimum is understood. The operator T is analogous to a Bellman operator. Under this analogy, p corresponds to a value function and δ to a discount factor. Since $\delta > 1$, the map T is not a contraction in any obvious metric, however, and $T^n p$ diverges for many choices of p, even when continuous and bounded.[10] Nevertheless, there exists a domain on which T is well-behaved: the set of convex increasing continuous functions $p\colon [0,1] \to \mathbb{R}$ such that $c'(0)s \leqslant p(s) \leqslant c(s)$ for all $0 \leqslant s \leqslant 1$. We denote this set of functions by $\mathscr{P}$.

Theorem 3.1. *Under our assumptions the following statements are true:*

1. *T maps $\mathscr{P}$ into itself.*
2. *T has a unique fixed point in $\mathscr{P}$, denoted below by p^*.*
3. *For all $p \in \mathscr{P}$ we have $T^k p \to p^*$ uniformly as $k \to \infty$.*

The significance of T and its fixed point p^* is that, as we now show, there exists an allocation $\boldsymbol{\ell}^*$ such that $(p^*, \boldsymbol{\ell}^*)$ is an equilibrium for the production chain in the sense of definition 3.1. To construct this allocation, we begin by introducing the *equilibrium choice function*

$$t^*(s) := \operatorname*{arg\,min}_{t \leqslant s} \{c(s-t) + \delta p^*(t)\}. \tag{4}$$

By definition, $t^*(s)$ is the cost minimizing upstream boundary for a firm that is contracted to deliver the good at stage s and faces the price function p^*. Since p^* lies in $\mathscr{P}$ and since c is strictly convex, it follows that the right-hand side of (4) is continuous and strictly convex in t, and hence the minimizer $t^*(s)$ exists and is uniquely defined.

We can use t^* to construct an equilibrium allocation as follows: recalling that firm 1 sells the completed good at stage $s = 1$, its optimal upstream boundary is $t^*(1)$. Hence firm 2's optimal upstream boundary is $t^*(t^*(1))$. Continuing in this way produces the sequence $\{t_i^*\}$ defined by

$$t_0^* = 1 \quad \text{and} \qquad t_i^* = t^*(t_{i-1}). \tag{5}$$

The sequence ends when a firm chooses to complete all remaining tasks. We label this firm (and hence the number of firms in the chain) as n^*. More precisely

$$n^* := \inf\{i \in \mathbb{N} : t_i^* = 0\}. \tag{6}$$

The task allocation corresponding to (5) is given by $\ell_i^* := t_{i-1}^* - t_i^*$ for all i. Below, p^* is called the *equilibrium price function* and $\boldsymbol{\ell}^*$ is called the *equilibrium allocation*. The next theorem justifies this terminology.

Theorem 3.2. *The value n^* in (6) is well-defined and finite, the allocation $\boldsymbol{\ell}^* = \{\ell_i^*\}$ is feasible, and the pair $(p^*, \boldsymbol{\ell}^*)$ is an equilibrium for the production chain.*

[10] For example, if $p \equiv 1$, then $T^n p = \delta^n 1$, which diverges to $+\infty$.

Full proofs of theorems 3.1–3.2 are given in the appendix. Much of the insight can be obtained by observing that, as a fixed point of T, the equilibrium price function must satisfy

$$p^*(s) = \min_{t \leqslant s} \{c(s-t) + \delta p^*(t)\} \quad \text{for all} \quad s \in [0,1]. \tag{7}$$

From this equation it is clear that p^* satisfies part 3 of definition 3.1. Moreover, the equilibrium upstream boundary for firm i is the minimizer in (7) when s is its downstream boundary, so profits are zero for all incumbent firms. Hence part 2 of definition 3.1 is satisfied. Part 1 of the definition is immediate from the fact that $p^* \in \mathscr{P}$, whence we obtain $p^*(0) \leqslant c(0) = 0$.

3.3. **Comments and Examples.** Some comments on the preceding results are in order. First, equation (7) illustrates why it matters little whether we place transaction costs on the buyer side, the seller side, or both. For example, suppose that, in addition to the existing buyer side transaction cost, the seller faces a transaction costs parameterized by γ. In particular, the seller receives only fraction $\gamma < 1$ of any sale. The profit function then becomes $\pi(s,t) = \gamma p(s) - c(s-t) - \delta p(t)$. Minimizing over $t \leqslant s$ and setting profits to zero yields $p(s) = \min_{t \leqslant s}\{c(s-t)\gamma^{-1} + \delta\gamma^{-1} p(t)\}$, which is analogous to (7). Nothing substantial has changed, since $\delta/\gamma > 1$, and since c/γ inherits from c all the properties of the cost function stated in section 2.2.

A second comment on the preceding results is that, as shown in the proof of theorem 3.1, the convergence in part 3 of that theorem occurs in finite time from every element of $\mathscr{P}$, and the required number of iterates can be calculated *ex ante*. In particular, if $\bar{s} := \sup\{s \in (0,1] : c'(s) \leqslant \delta c'(0)\}$, then $T^k p = p^*$ whenever $p \in \mathscr{P}$ and $k \geqslant 1/\bar{s}$. Hence we can compute p^* with a high degree of accuracy.

Two equilibrium price functions computed using this method are shown in figure 3. For the initial condition we chose $p = c$. In both cases we used the exponential cost function $c(\ell) = e^{\theta\ell} - 1$ with $\theta = 10$. The dashed line corresponds to $\delta = 1.02$, while the solid line is for $\delta = 1.2$. Not surprisingly, prices shift up pointwise with each rise in transaction costs.[11]

Figure 4 shows the corresponding equilibrium allocations, in addition to the prices. The vertical lines are firm boundaries, computed via (5) using the equilibrium price function, with the latter obtained using the methods just discussed. The top subfigure shows the price function and firm boundaries when $\delta = 1.02$ and c is as in the previous figures. The bottom subfigure shows the same information when $\delta = 1.2$.

[11]At each step of the iteration, we calculated $Tp(s)$ on a grid of 500 points $s_i \in [0,1]$. We then constructed an approximation to Tp using piecewise linear interpolation over the grid $\{s_i\}$ and computed values $\{Tp(s_i)\}$. We then set p equal to the resulting piecewise linear function and moved to the next iteration. The code can be found at `https://github.com/jstac/production_chains`.

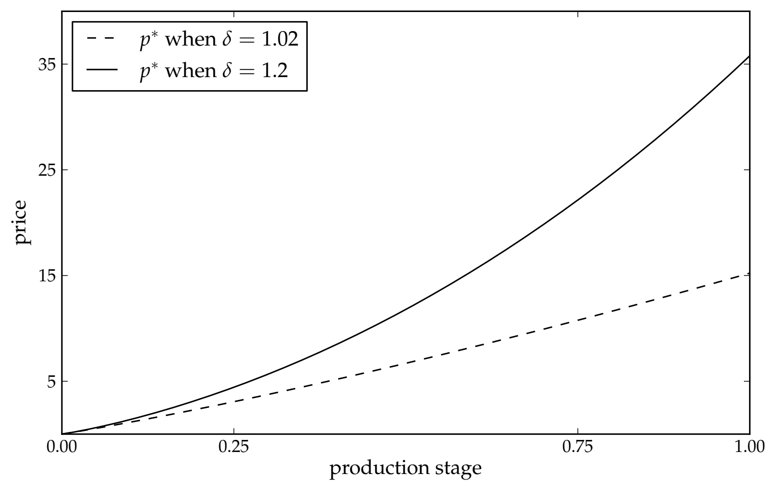


FIGURE 3. Equilibrium price functions when $c(\ell) = e^{\theta \ell} - 1$

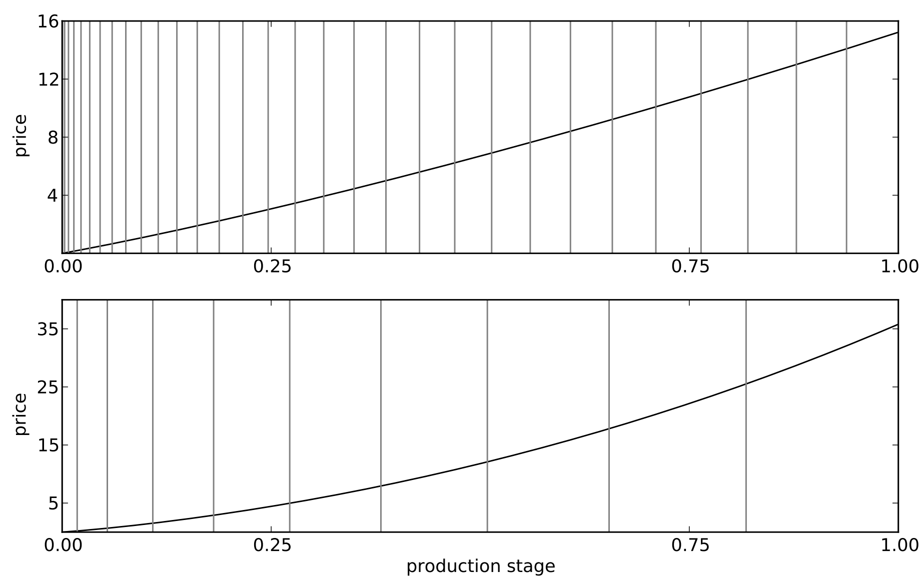


FIGURE 4. Firm boundaries for $\delta = 1.02$ (top) and $\delta = 1.2$ (bottom)

The preceding discussion gives us the price of a single unit of the final good. The final good market is also competitive, and $p^*(1)$ is the amount the most downstream producer must be compensated in order to make zero profits. Assuming the production process described above can be replicated any number of times without affecting factor prices, $p^*(1)$ is also the long run equilibrium price in the market for the final good.

4. Equilibrium Properties

In this section we discuss properties of the equilibrium and their implications.

4.1. **Properties of the Price Function.** Since p^* lies in $\mathscr{P}$, theorem 3.1 implies it is increasing and convex. The next result strengthens these findings and adds an alternative representation of p^* that provides additional intuition.

Proposition 4.1. *The equilibrium price function p^* is is strictly convex and strictly increasing, with $sc'(0) \leqslant p^*(s) \leqslant c(s)$ for all s in $[0,1]$. Moreover, for each such s,*

$$p^*(s) = \min \left\{ \sum_{i=1}^{\infty} \delta^{i-1} c(\ell_i) \ : \ \{\ell_i\} \in \mathbb{R}_+^{\infty} \text{ and } \sum_{i=1}^{\infty} \ell_i = s \right\}. \tag{8}$$

When $s = 1$, the minimizer in (8) is equal to the equilibrium allocation $\boldsymbol{\ell}^*$.

The fact that p^* is strictly increasing is not surprising, given that c is itself strictly increasing. The strict convexity of p^* is discussed further in section 4.4. The upper bound $p^*(s) \leqslant c(s)$ holds because a single firm can always implement the entire process up to stage s, at cost $c(s)$. The lower bound in $p^*(s) \geqslant sc'(0)$ can be thought of as a cost attained when $\delta = 1$ and firms enter without limit.

To understand what the right-hand side of (8) represents, let p be any price function and let $\boldsymbol{\ell}$ be any allocation that produces s, in the sense that $\sum_{i=1}^{n} \ell_i = s$. Let $\{t_i\}$ be the corresponding transaction stages.[12] If all firms make zero profits then $p(t_{i-1}) = c(\ell_i) + \delta p(t_i)$ for $i = 1, \ldots, n$. Iterating on this equation gives

$$p(t_0) = c(\ell_1) + \delta c(\ell_2) + \ldots + \delta^{n-1} c(\ell_n) + \delta^n p(t_n).$$

Since $t_0 = s$ and $t_n = 0$, if we assume as in the definition of equilibrium that $p(0) = 0$, then we have $p(s) = c(\ell_1) + \delta c(\ell_2) + \ldots + \delta^{n-1} c(\ell_n)$, which is the expression minimized in (8). Hence $p^*(s)$ is the minimal cost of producing s when all firms make zero profits and transaction costs cannot be avoided.

4.2. **Uniqueness of Equilibrium.** Let us agree to call a price function p *nontrivial* if, for each $s > 0$, it can support a chain of firms that produce the good up to stage s while each receiving nonnegative profit. In other words, there exists an allocation $\boldsymbol{\ell}$ such that $\sum_i \ell_i = s$ and $p(t_{i-1}) - c(\ell_i) - \delta p(t_i) \geqslant 0$ for all i. At a lower price one firm will always lose money, and hence such a price cannot be observed in equilibrium.

Let $\mathcal{E}$ be all pairs $(p, \boldsymbol{\ell})$ where p is a nontrivial price function and $\boldsymbol{\ell}$ is a feasible allocation. The following result is proved in the appendix.

Theorem 4.2. *The pair $(p^*, \boldsymbol{\ell}^*)$ is the unique equilibrium for the production chain in $\mathcal{E}$.*

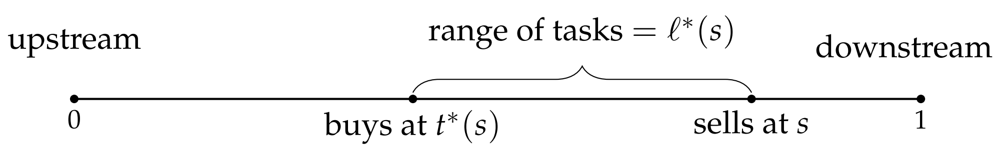
[12] That is, $t_0 = s$ and $t_{i+1} = t_i - \ell_{i+1}$.

The intuition is as follows. First, if p is any nontrivial price function, then p must be large enough to sustain nonnegative profits. By comparison, the equilibrium p^* sustains exactly zero profits for incumbents, leading to $p^* \leqslant p$. Conversely, if p is also an equilibrium price function for some allocation, then p eliminates all profit opportunities, and hence is less than the feasible price function p^*.

4.3. **Marginal Conditions.** We can develop some additional insights on the behavior of firms by examining marginal conditions associated with the equilibrium. As a first step, let

$$\ell^*(s) := s - t^*(s),$$

which is the cost minimizing range of in-house tasks for a firm with downstream boundary s:



The function ℓ^* is plotted for $\delta = 1.2$ and $\delta = 1.02$ in figure 5. Other parameters are the same as for figure 3. Observe that $\ell^*(s)$ increases with s. The next result shows that this is always true, as well as connecting ℓ^* to the derivatives of p^* and c.

Proposition 4.3. *Both t^* and ℓ^* are increasing and Lipschitz continuous, while p^* is continuously differentiable at all $s \in (0,1)$ with derivative*

$$(p^*)'(s) = c'(\ell^*(s)). \tag{9}$$

Equation (9) follows from $p^*(s) = \min_{t \leqslant s} \{c(s-t) + \delta p^*(t)\}$ and the envelope theorem. It is analogous to a standard result from optimal growth theory, which states that the derivative of the value function is equal to the marginal utility of optimal consumption. The monotonicity of t^* and ℓ^* are discussed further below.

A related equation is the first order condition for $p^*(s) = \min_{t \leqslant s} \{c(s-t) + \delta p^*(t)\}$, the minimization problem for a firm with upstream boundary s, which is

$$\delta (p^*)'(t^*(s)) = c'(s - t^*(s)). \tag{10}$$

This condition matches a marginal condition expressed verbally by Coase; that is, "a firm will tend to expand until the costs of organizing an extra transaction within the firm become equal to the costs of carrying out the same transaction by means of an exchange on the open market..." (Coase, 1937, p. 395).

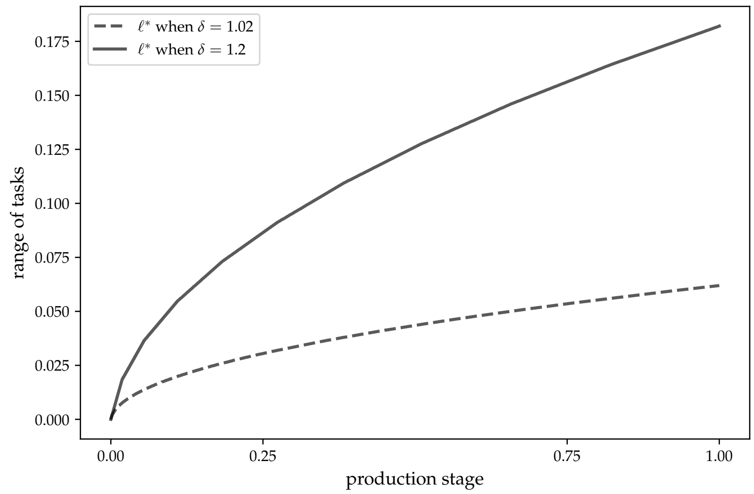


FIGURE 5. Optimal range of in-house tasks as a function of s

Combining (9) and (10) and evaluating at $s = t_i$, we see that active firms that are adjacent satisfy

$$\delta \, c'(\ell^*_{i+1}) = c'(\ell^*_i). \tag{11}$$

In other words, the marginal in-house cost per task at a given firm is equal to that of its upstream partner multiplied by gross transaction cost. This expression can be thought of as a "Coase–Euler equation," which determines inter-firm efficiency by indicating how two costly forms of coordination (markets and management) are jointly minimized in equilibrium.

Example 4.1. Let c have the exponential form $c(\ell) = e^{\theta\ell} - 1$. Note that $c(0) = 0$ and $c'(0) > 0$ as required. From the Coase-Euler equation (11) we have $\ell_{i+1} = \ell_i - \ln\delta/\theta$. Using this equation, the constraint $\sum_{i=1}^{n^*} \ell_i = 1$ and some algebra, it can be shown (cf., lemma 7.12) that the equilibrium number of firms is

$$n^* = \left\lfloor 1/2 + (1 + 8\theta/\ln\delta)^{1/2}/2 \right\rfloor, \tag{12}$$

where $\lfloor a \rfloor$ is the largest integer less than or equal to a. The value n^* is decreasing in δ because higher transaction costs encourage less use of the market and more in-house production. In other words, firms get larger. Since the range of tasks does not change, larger firms imply less firms. At the same time, n^* is increasing in θ because $\theta = c''(\ell)/c'(\ell)$, so θ parameterizes curvature of c, and hence the intensity of diminishing returns to management. More intense diminishing returns to management encourages greater use of the market, and hence a larger number of smaller firms.

4.4. **Upstreamness and Firm Size.** One implication of the model that deserves further investigation is that, in terms of the range of tasks that they implement, downstream firms are always larger than upstream firms. Indeed, since c' is increasing and $\delta > 1$, equation (11) implies that $\ell^*_{i+1} < \ell^*_i$ for any pair of adjacent active firms i and $i+1$. Similarly, proposition 4.3 shows that the optimal choice function $s \mapsto \ell^*(s)$ is increasing in s. The monotone relationship between firm size and downstreamness can be seen in figure 4.

The same monotone relationship between upstream and downstream firms observed in the sequence $\{\ell^*_i\}$ also holds for other measures of firm size. For example, recall that value added for firm i is $v_i := p^*(t_{i-1}) - p^*(t_i)$. Observe that $\ell^*_{i+1} > \ell^*_i$ can also be written as $t_i - t_{i+1} < t_{i-1} - t_i$. Since p^* is increasing and convex, it follows that $v_{i+1} < v_i$. In other words, value added shares the monotone relationship between upstream and downstream possessed by $\{\ell^*_i\}$.

To gain a better understanding of this monotone relationship in terms of the choice problem faced by firms, consider the first order condition (10), which can be expressed here by saying that firms choose their downstream boundary to equalize the marginal cost of in-house production and the marginal cost of inputs at a given stage along the value chain. Since the equilibrium price function is strictly convex, the marginal cost of these inputs rises as we move from upstream to downstream. Hence downstream firms choose to do more in-house.

From the preceding discussion it is clear that strict convexity of p^* is essential to size increasing with downstreamness. The equilibrium price function is strictly convex partly because transaction costs that are sensitive to value compound as we go from upstream to downstream. At the same time, diminishing returns to management also matter, since they motivate firms to divide the value chain in the first place. (If c is linear, say, then a single firm will complete the value chain, and the relative size of firms cannot be discussed.)

Let's consider how the prediction that firm size increases with downstreamness compares with the data. A partial answer can be obtained by using indices developed in Fally (2012) and Antràs et al. (2012). These indices describe industries' position in vertical production chains by exploiting information about relationships in input-output tables. Fally and Hillberry (2015) calculate these indices for manufacturing industries in 9 Asian countries and the US and obtain negative correlation between upstreamness and value added content (Fally and Hillberry, 2015, p. 41). These findings are consistent with our theoretical prediction that firm size tends to be smaller upstream than downstream.

Of course the predictions of our model concern individual production chains, while the regressions described above are across production chains. Table 1 makes an attempt to address production chains using aggregated data. The table is constructed

Industry	stage 1	stage 2	stage 3
Automobile	0.2346	0.1353	0.0201
Light truck	0.2113	0.1377	0.0205
Nonupholstered furniture	0.4736	0.0319	0.0101
Upholstered furniture	0.3891	0.0423	0.0099

TABLE 1. Maximum value added by stage

from the 2002 Bureau of Economic Analysis input-output tables. It shows data for the 4 most downstream industries in the table, where downstreamness is computed using the same index as in Antràs et al. (2012). Each row of the table traces back the value added at every production stage associated with one dollar of spending on the industry in question. Stage 1 is value added by the industry itself. Stage 2 is the maximum of the value added associated with this same dollar of spending by all of the direct input industries, and stage 3 is the maximum of the value added by all of the direct inputs to these inputs. Value added increases with downstreamness, coinciding with the prediction stated above.[13]

4.5. **Comparative Statics.** Variations in transaction costs shift equilibrium outcomes monotonically in the directions intuition suggests. In particular, a rise in transaction costs causes prices to rise, the size of firms to increase, and the equilibrium number of firms to fall. The next proposition gives details:

Proposition 4.4. *If $\delta_a \leqslant \delta_b$, then $p_a^* \leqslant p_b^*$, $\ell_i^a < \ell_i^b$ for all active firms and $n_b^* \leqslant n_a^*$.*

Here p_a^* is the equilibrium price function for transaction cost δ_a, p_b^* is that for δ_b, and so on. Compare the upper and lower panels in figure 4 to see these results in a simulation.

5. MULTIPLE UPSTREAM PARTNERS

So far we have assumed that production is linearly sequential, in the sense that firms contract with only one partner. In reality, most vertically integrated firms have multiple upstream partners. Generalization of the model to include multiple upstream partners is possible when we adopt a recursive approach. In the process we also include additive transaction costs, so as to achieve a more realistic framework and explore robustness of our findings in this setting.

[13]Case studies of production chains also tend to find that downstream firms are larger. See, for example, Kimura (2002) and Subrahmanya (2008).

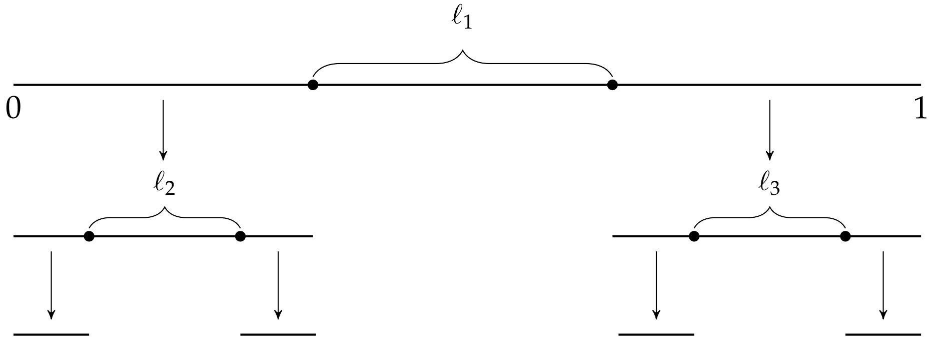


FIGURE 6. Subcontracting when $k = 2$

To develop intuition, consider the tree in figure 6. Here each firm subcontracts to two upstream partners. As drawn, firm 1 chooses the set ℓ_1 for in-house production and subcontracts the remaining two intervals on the left and right to two different upstream partners (the shape of the division is chosen only to simplify the diagram). Similarly, the upstream partners choose to implement intervals ℓ_2 and ℓ_3 respectively, and subcontract the remainder. The process of subdivision repeats until the firms at some lower level choose to implement all remaining tasks.

More generally, suppose that each firm has k upstream partners. Suppose that a firm contracts to supply the good at stage s, chooses a quantity $\ell \leqslant s$ to produce in-house, and then divides the remainder $s - \ell$ equally across k upstream partners. In this case, profits will be given by revenue $p(s)$ minus input costs $\delta\, k\, p((s-\ell)/k)$, in-house production costs $c(\ell)$, and the additive transaction costs mentioned above. We assume that, for each firm, additive transaction costs are an increasing convex function $g(k)$ of the number of transactions k. Thus, profits are

$$\pi(s, \ell) = p(s) - \delta\, k\, p((s-\ell)/k) - c(\ell) - g(k).$$

If we take k as fixed, then, setting profits to zero, letting $t := s - \ell$ and minimizing with respect to t yields the equation

$$p(s) = \min_{t \leqslant s}\{c(s-t) + g(k) + \delta\, k\, p(t/k)\}.$$

If $g(1) = 0$, then this equation is an immediate generalization of (7).

However, once multiple partners are allowed, the assumption that the number of partners is fixed becomes both unsatisfactory and counterfactual. For this reason we allow each firm to choose k at each stage, leading to the functional equation

$$p(s) = \min_{k \in \mathbb{N},\, t \leqslant s}\{c(s-t) + g(k) + \delta\, k\, p(t/k)\}. \tag{13}$$

No ad hoc upper bound is imposed on the number of partners.

The fact that $\delta > 1$ means that a fixed point approach based around contractions is problematic, just as it was in the single partner case. Here, additive costs and endogenous multiple partners lead to further complications. We can, however, establish the following, which implies the existence of a price function satisfying (13). In the statement of the theorem, $\mathscr{F}$ is all nonnegative increasing $p\colon [0,1] \to \mathbb{R}$ such that $p(0) = 0$.

Theorem 5.1. *The set of $p \in \mathscr{F}$ satisfying (13) for all $s \in [0,1]$ is a complete lattice.*

Analogous to section 3, let T be the operator on $\mathscr{F}$ defined by

$$Tp(s) = \min_{k \in \mathbb{N},\, t \leqslant s} \{c(s-t) + g(k) + \delta\, k\, p(t/k)\}.$$

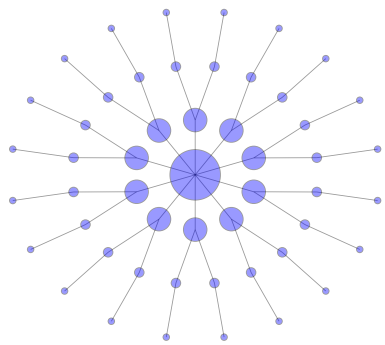
To compute prices and the optimal choices of firms, we adapt the recursive procedure outlined in section 3. The equilibrium price function p^* is obtained by iterating with T, starting at initial condition $p_0 = c$. Given p^*, we then obtain the production network recursively using a method analogous to the one described in section 3.2 (here including optimal choice of the number of upstream partners at each step).

Figure 7 explores some of the networks that arise for different parameter values.[14] In these networks, each node represents a firm, and the node size is proportional to the value added of that firm. The most downstream firm is the firm at the center of the network. The additive transaction cost function is $g(k) = \beta(k-1)$ for a given constant β, while $c(\ell) = \ell^\theta$ for different values of θ. Here $g(k)$ is proportional to $k-1$ so that $g(1) = 0$, as discussed above. Details of all computations can be found at `https://github.com/jstac/production_chains`.

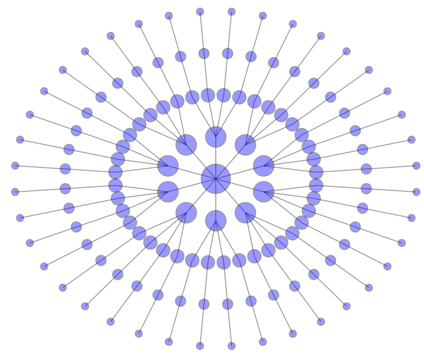
The way that the network changes with the parameters is in line with the intuition arising from the discussion so far. Comparing networks (A) and (B), firms respond to a reduction in multiplicative transaction costs by forming a deeper network with more layers. The number of firms roughly doubles, from 51 to 101. Comparing networks (A) and (C), the change is a lower additive transaction cost in (C), which encourages more subcontractors at each level of the network, as well as a larger number of firms. Comparing (A) and (D), the difference is in greater curvature in the internal coordination cost function c, which also encourages more outsourcing.

There are two results that are robust across all of our simulations. First, downstream firms have greater value added (as can be seen by comparing the relative size of the nodes in figure 7). This is consistent with the one partner case and with stylized facts about upstreamness and firm size, as discussed in section 4. Second, the equilibrium

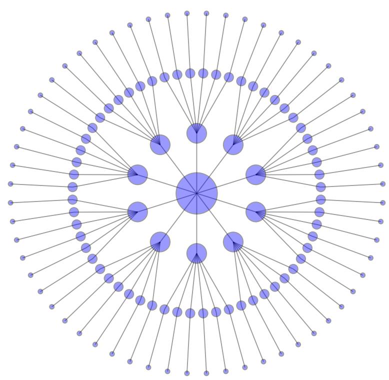
[14] In the parameterizations below we allow $c'(0) = 0$, despite our earlier assumption in the linear setting that $c'(0) > 0$. The reason is that, in the current setting, additive transaction costs prevent firms from becoming too small, and hence yield an equilibrium with a finite number of firms.



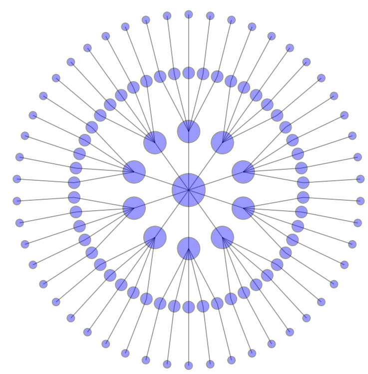
(A) $\beta = 0.001, \delta = 1.15, \theta = 1.1$



(B) $\beta = 0.001, \delta = 1.05, \theta = 1.1$



(C) $\beta = 0.0001, \delta = 1.15, \theta = 1.1$



(D) $\beta = 0.001, \delta = 1.15, \theta = 1.2$

FIGURE 7. Production networks with mulitple upstream partners

number of upstream partners is weakly (and often strictly) decreasing as we move from downstream to upstream (from the center to the edges). This prediction is also broadly consistent with the literature on production chains. For example, Kimura (2002) finds in a study of the Japanese machinery industry that the number of direct affiliates gets larger with the firm size, and that the probability of working as a subcontractor declines as firms get larger.[15]

The production chain model given above generates not only a network but also a nontrivial firm size distribution. Lacking both capital and heterogeneity, our model

[15] It is also noteworthy that, as the co-editor pointed out, the second layer always has 10 firms despite changes in parameters. Thus, shocks to fundamentals have effects on its structure that are mostly felt down the supply chain and die out up the chain.

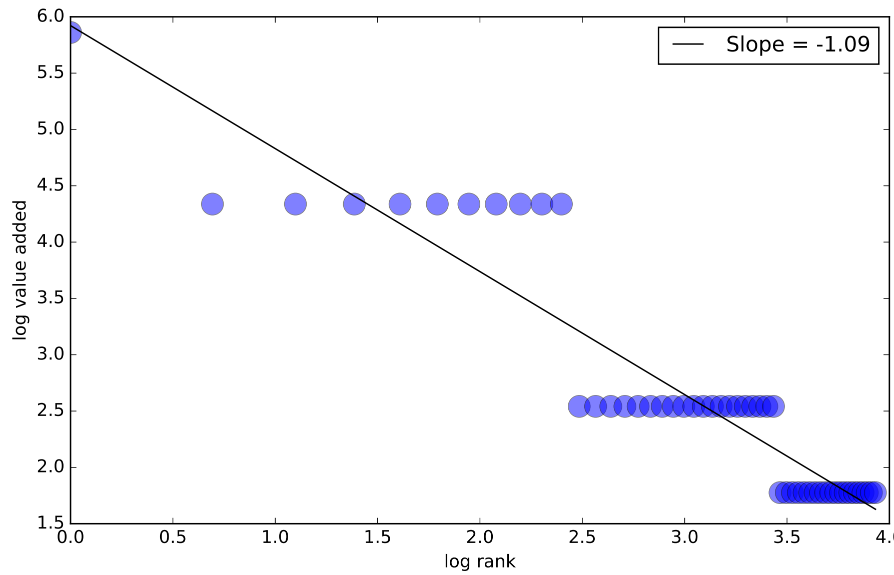


FIGURE 8. A size-rank plot of the distribution of firms sizes

is too stylized to match all features of the firm size distribution. Nonetheless, it is worth considering whether or not the distributions generated by the model match any significant empirical regularities.

While the observed size distribution of business firms varies over time and space, as well as with the measure of size adopted and the level of aggregation, one well known regularity is a power law in the right tail, as has been observed in many studies. An early example is Axtell (2001), who finds evidence of Zipf's law in aggregate US firm size data (a form of power law with a unit coefficient). Since then many economists have constructed theories that generate power laws in the firm size distribution from more basic primitives.[16]

To make statements about the firm size distribution we use value added as our measure of size.[17] The rank-size plot in figure 8 shows log size by value added against log rank in the distribution, with the generated data taken from the same parameters found in (A) of figure 7. The data look artificial due to the stylized nature of

[16] See, e.g., Lucas (1978), Luttmer (2007) and, more recently, Geerolf (2016).

[17] We can also make predictions regarding firm size in terms of employees if we specify productivity. For example, let m be the number of employees in a given firm, and suppose that m is related to the range of tasks ℓ via $\ell = f(m)$. Conversely, for a given range of tasks ℓ, the required number of employees is $m = f^{-1}(\ell)$. Assume that f is strictly increasing and strictly concave with $f(0) = 0$, and that $c(\ell) = wm = wf^{-1}(\ell)$ for some wage rate w. Then c satisfies the assumptions we imposed on it in section 2.2. The value $m_i = c(\ell_i)/w$ is then equal to the number of employees in the i-th firm.

the model and the linear fit is not tight, but the slope is close to -1, which corresponds to Zipf's law. The heavy tail can also be seen by observing that the largest observation is more than 6 standard deviations greater than the mean.

6. CONCLUSION

In this paper we embedded several ideas from Coase (1937) into a competitive product market with an infinite number of identical price-taking firms. By developing an approach to the equilibrium problem based on recursive subdivision of tasks, we showed how to obtain prices, actions of firms and vertical division of the value chain. This allowed us to investigate individual firm boundaries and the vertical structure of production. We analyzed the relative size of firms, the relationship between upstream and downstream firms, the overall distribution of firms, and the relationship between diminishing returns to management, transaction costs and the properties of the vertical production chain.

We showed in particular how the equilibrium price function and the structure of the production chain vary with transaction and internal coordination costs. We derived a first order condition that corresponds to the marginal condition determining firm boundaries stated verbally by Coase (1937). This permits the trade-off associated with the make-or-buy decision to be investigated by analyzing how changes to policy or technology show up in the marginal costs and benefits of integration as quantified by the first order condition. We also added an Euler equation relating costs and hence sizes of adjacent firms.

On a technical front, we provided a recursive formulation and computational methods that are likely to have applications to other fields involving sequential production or allocation of tasks, such as offshoring by multinationals or division of labor with failure probabilities or other frictions. In addition, the model presented above was a baseline model in most dimensions, with perfect competition, perfect information, identical firms and identical tasks. These assumptions can potentially be weakened. The effect of altering contract structures could also be investigated, as could the various possibilities for determining upstream partners in section 5.

The model has several interesting empirical implications, one of which is that, in a setting with multiple upstream partners, the number of upstream partners decreases with upstreamness. A second prediction is the positive relationship between downstreamness and value added caused by increasing marginal costs of transactions as semi-processed goods move from upstream to downstream. The same prediction also holds when value added is replaced by number of tasks performed in-house, and, while this quantity is typically unobservable, it can potentially be proxied by the number of distinct specialist occupations, or the expenditure on managers used to coordinate these specialists. A more detailed model and firm level data will shed further light on these relationships.

7. Appendix

This appendix collects proofs. We start by focusing on fixed points of the operator T defined in (3). We pursue a direct proof of existence, uniqueness and convergence.

7.1. **Preliminaries.** Our first result shows that T preserves convexity.

Lemma 7.1. *If $p \in \mathscr{P}$, then Tp is strictly convex.*

Proof. Pick any $0 \leqslant s_1 < s_2 \leqslant 1$ and any $\alpha \in (0,1)$. Let $t_i := \arg\min_{t \leqslant s_i} \{\delta p(t) + c(s_i - t)\}$ for $i = 1, 2$, and $t_3 := \alpha t_1 + (1-\alpha) t_2$. Given that $t_i \leqslant s_i$ we have $0 \leqslant t_3 \leqslant \alpha s_1 + (1-\alpha) s_2$, and hence

$$Tp(\alpha s_1 + (1-\alpha)s_2) \leqslant \delta p(t_3) + c(\alpha s_1 + (1-\alpha)s_2 - t_3).$$

The right-hand side expands out to

$$\delta p[\alpha t_1 + (1-\alpha)t_2] + c[\alpha s_1 - \alpha t_1 + (1-\alpha)s_2 - (1-\alpha)t_2].$$

Using convexity of p and strict convexity of c, we obtain $Tp(\alpha s_1 + (1-\alpha)s_2) < \alpha Tp(s_1) + (1-\alpha)Tp(s_2)$, which is strict convexity Tp. □

Lemma 7.2. *Let $p \in \mathscr{P}$ and let t_p and ℓ_p be the optimal responses, defined by*

$$t_p(s) := \arg\min_{t \leqslant s} \{\delta p(t) + c(s-t)\} \quad \textit{and} \quad \ell_p(s) := s - t_p(s). \tag{14}$$

If s_1 and s_2 are any two points with $0 < s_1 \leqslant s_2 \leqslant 1$, then

(1) *both $t_p(s_1)$ and $\ell_p(s_1)$ are well defined and single-valued.*
(2) *$t_p(s_1) \leqslant t_p(s_2)$ and $t_p(s_2) - t_p(s_1) \leqslant s_2 - s_1$.*
(3) *$\ell_p(s_1) \leqslant \ell_p(s_2)$ and $\ell_p(s_2) - \ell_p(s_1) \leqslant s_2 - s_1$.*

Proof. Since $t \mapsto \delta p(t) + c(s_1 - t)$ is continuous and strictly convex (by convexity of p and strict convexity of c), and since $[0, s_1]$ is compact, existence and uniqueness of $t_p(s_1)$ and $\ell_p(s_1)$ must hold. Regarding the claim that $t_p(s_1) \leqslant t_p(s_2)$, let $t_i := t_p(s_i)$. Suppose instead that $t_1 > t_2$. We aim to show that, in this case,

$$\delta p(t_1) + c(s_2 - t_1) < \delta p(t_2) + c(s_2 - t_2), \tag{15}$$

which contradicts the definition of t_2.[18] To establish (15), observe that t_1 is optimal at s_1 and $t_2 < t_1$, so

$$\delta p(t_1) + c(s_1 - t_1) < \delta p(t_2) + c(s_1 - t_2).$$

$$\therefore \quad \delta p(t_1) + c(s_2 - t_1) < \delta p(t_2) + c(s_1 - t_2) + c(s_2 - t_1) - c(s_1 - t_1).$$

Given that c is strictly convex and $t_2 < t_1$, we have

$$c(s_2 - t_1) - c(s_1 - t_1) < c(s_2 - t_2) - c(s_1 - t_2).$$

[18] Note that $t_1 < s_1 \leqslant s_2$, so t_1 is available when t_2 is chosen.

Combining this with the last inequality yields (15).

Next we show that $\ell_1 \leqslant \ell_2$, where $\ell_1 := \ell_p(s_1)$ and $\ell_2 := \ell_p(s_2)$. In other words, $\ell_i = \arg\min_{\ell \leqslant s_i}\{\delta p(s_i - \ell) + c(\ell)\}$. The argument is similar to that for t_p, but this time using convexity of p instead of c. To induce the contradiction, we suppose that $\ell_2 < \ell_1$. As a result, we have $0 \leqslant \ell_2 < \ell_1 \leqslant s_1$, and hence ℓ_2 was available when ℓ_1 was chosen. Therefore,

$$\delta p(s_1 - \ell_1) + c(\ell_1) < \delta p(s_1 - \ell_2) + c(\ell_2),$$

where the strict inequality is due to the fact that minimizers are unique. Rearranging and adding $\delta p(s_2 - \ell_1)$ to both sides gives

$$\delta p(s_2 - \ell_1) + c(\ell_1) < \delta p(s_2 - \ell_1) - \delta p(s_1 - \ell_1) + \delta p(s_1 - \ell_2) + c(\ell_2).$$

Given that p is convex and $\ell_2 < \ell_1$, we have

$$p(s_2 - \ell_1) - p(s_1 - \ell_1) \leqslant p(s_2 - \ell_2) - p(s_1 - \ell_2).$$

Combining this with the last inequality, we obtain

$$\delta p(s_2 - \ell_1) + c(\ell_1) < \delta p(s_2 - \ell_2) + c(\ell_2),$$

contradicting optimality of ℓ_2.[19]

To complete the proof of lemma 7.2, we also need to show that $t_p(s_2) - t_p(s_1) \leqslant s_2 - s_1$, and similarly for ℓ. Starting with the first case, we have

$$t_p(s_2) - t_p(s_1) = s_2 - \ell_p(s_2) - s_1 + \ell_p(s_1) = s_2 - s_1 + \ell_p(s_1) - \ell_p(s_2).$$

As shown above, $\ell_p(s_1) \leqslant \ell_p(s_2)$, so $t_p(s_2) - t_p(s_1) \leqslant s_2 - s_1$, as was to be shown. Finally, the corresponding proof for ℓ_p is obtained in the same way, by reversing the roles of t_p and ℓ_p. This concludes the proof of lemma 7.2. □

Recall the constant $\bar{s}$ defined in section 3.3, existence of which follows from the conditions in section 2 and the intermediate value theorem. Regarding $\bar{s}$ we have the following lemma, which states that the best action for a firm subcontracting at $s \leqslant \bar{s}$ is to implement all stages up to s (i.e., to start at stage 0).

Lemma 7.3. *If $p \in \mathscr{P}$, then $s \leqslant \bar{s}$ if and only if $\min_{t \leqslant s}\{\delta p(t) + c(s-t)\} = c(s)$.*

Proof. First suppose that $s \leqslant \bar{s}$. Seeking a contradiction, suppose there exists a $t \in (0, s]$ such that $\delta p(t) + c(s-t) < c(s)$. Since $p \in \mathscr{P}$ we have $p(t) \geqslant c'(0)t$ and hence $\delta p(t) \geqslant \delta c'(0)t \geqslant c'(\bar{s})t$. Since $s \leqslant \bar{s}$, this implies that $\delta p(t) \geqslant c'(s)t$. Combining these inequalities gives $c'(s)t + c(s-t) < c(s)$, contradicting convexity of c.

Suppose on the other hand that $\inf_{t \leqslant s}\{\delta p(t) + c(s-t)\} = c(s)$. We claim that $s \leqslant \bar{s}$, or, equivalently $c'(s) \leqslant \delta c'(0)$. To see that this is so, observe that since $p \in \mathscr{P}$ we have $p(t) \leqslant c(t)$, and hence $c(s) \leqslant \delta p(t) + c(s-t) \leqslant \delta c(t) + c(s-t)$, for all $t \leqslant s$.

$$\therefore \quad \frac{c(s) - c(s-t)}{t} \leqslant \frac{\delta c(t)}{t} \qquad \forall\, t \leqslant s.$$

Taking limits we get $c'(s) \leqslant \delta c'(0)$ as claimed. □

[19] Note that $0 \leqslant \ell_1 \leqslant s_1 \leqslant s_2$, so ℓ_1 is available when ℓ_2 is chosen.

Lemma 7.4. *Let $p \in \mathscr{P}$ and let ℓ_p be as in* (14). *If $s \geqslant \bar{s}$, then $\ell_p(s) \geqslant \bar{s}$. If $s > 0$, then $\ell_p(s) > 0$.*

Proof. By lemma 7.2, ℓ_p is increasing, and hence if $\bar{s} \leqslant s \leqslant 1$, then $\ell_p(s) \geqslant \ell_p(\bar{s}) = \bar{s} - t_p(\bar{s}) = \bar{s}$. By lemma 7.3, if $0 < s \leqslant \bar{s}$, then $\ell_p(s) = s - t_p(s) = s > 0$. □

Lemma 7.5. *If $p \in \mathscr{P}$, then Tp is differentiable on $(0, 1)$ with $(Tp)' = c' \circ \ell_p$.*

Proof. Fix $p \in \mathscr{P}$ and let t_p be as in (14). Fix $s_0 \in (0, 1)$. By Benveniste and Scheinkman (1979), to show that Tp is differentiable at s_0 it suffices to exhibit an open neighborhood $U \ni s_0$ and a function $w \colon U \to \mathbb{R}$ such that w is convex, differentiable, satisfies $w(s_0) = Tp(s_0)$ and dominates Tp on U. To this end, observe that, in view of lemma 7.4, we have $t_p(s_0) < s_0$. Now choose an open neighborhood U of s_0 such that $t_p(s_0) < s$ for every $s \in U$. On U, define $w(s) := \delta p(t_p(s_0)) + c(s - t_p(s_0))$. Clearly w is convex and differentiable on U, and satisfies $w(s_0) = Tp(s_0)$. To see that $w(s) \geqslant Tp(s)$ when $s \in U$, observe that if $s \in U$ then $0 \leqslant t_p(s_0) \leqslant s$, and

$$Tp(s) = \min_{t \leqslant s} \{\delta p(t) + c(s - t)\} \leqslant \delta p(t_p(s_0)) + c(s - t_p(s_0)) = w(s).$$

As a result, Tp is differentiable at s_0 with $(Tp)'(s_0) = w'(s_0) = c'(\ell_p(s_0))$. □

Lemma 7.6. *Let $p \in \mathscr{P}$ and let t_p and ℓ_p be as defined in* (14). *If p is a fixed point of T, then $\delta\, c'(\ell_p(t_p(s))) = c'(\ell_p(s))$ for all $s > \bar{s}$.*

Proof. Since p is a fixed point of T it follows from lemma 7.5 that p is differentiable and $p'(s) = c'(\ell_p(s))$. Moreover, since $s > \bar{s}$, lemma 7.3 implies that $t_p(s) > 0$, and hence the optimal choice in the definition of $t_p(s)$ is interior. Thus the first order condition associated with the definition holds, which is $\delta p'(t_p(s)) = c'(\ell_p(s))$. Combining these two equalities gives lemma 7.6. □

Lemma 7.7. *The operator T defined in* (3) *maps $\mathscr{P}$ into itself.*

Proof. Let p be an arbitrary element of $\mathscr{P}$. To see that $Tp(s) \leqslant c(s)$ for all $s \in [0, 1]$, fix $s \in [0, 1]$ and observe that, since $p \in \mathscr{P}$ implies $p(0) = 0$, the definition of T implies $Tp(s) \leqslant \delta p(0) + c(s + 0) = c(s)$. Next we check that $Tp(s) \geqslant c'(0)s$ for all $s \in [0, 1]$. Picking any such s and using the assumption that $p \in \mathscr{P}$, we have $Tp(s) \geqslant \inf_{t \leqslant s}\{\delta c'(0)t + c(s - t)\}$. By $\delta > 1$ and convexity of c, we have $\delta c'(0)t + c(s - t) \geqslant c'(0)t + c(s - t) \geqslant c'(0)t + c'(0)(s - t) = c'(0)s$. Therefore $Tp(s) \geqslant \inf_{t \leqslant s} c'(0)s = c'(0)s$.

It remains to show that Tp is continuous, convex and monotone increasing. That Tp is convex was shown in lemma 7.1. Regarding the other two properties, let ℓ_p and t_p be as defined in (14). By the results in lemma 7.2, these functions are increasing and (Lipschitz) continuous on $[0, 1]$. Since $Tp(s) = \delta p(t_p(s)) + c(\ell_p(s))$, it follows that Tp is also increasing and continuous. □

Lemma 7.8. *If $p, q \in \mathscr{P}$, then $T^n p = T^n q$ whenever $n \geqslant 1/\bar{s}$.*

Proof. The proof is by induction. First we argue that $T^1 p = T^1 q$ on the interval $[0, \bar{s}]$. Next we show that if $T^k p = T^k q$ on $[0, k\bar{s}]$, then $T^{k+1} p = T^{k+1} q$ on $[0, (k+1)\bar{s}]$. Together these two facts imply the claim in lemma 7.8.

To see that $T^1 p = T^1 q$ on $[0, \bar{s}]$, pick any $s \in [0, \bar{s}]$ and recall from lemma 7.3 that if $h \in \mathscr{P}$ and $s \leqslant \bar{s}$, then $Th(s) = c(s)$. Applying this result to both p and q gives $Tp(s) = Tq(s) = c(s)$. Hence $T^1 p = T^1 q$ on $[0, \bar{s}]$ as claimed. Turning to the induction step, suppose now that $T^k p = T^k q$ on $[0, k\bar{s}]$, and pick any $s \in [0, (k+1)\bar{s}]$. Let $h \in \mathscr{P}$ be arbitrary, let $\ell_h(s) := \arg\min_{t \leqslant s}\{\delta h(t) + c(s-t)\}$ and let $t_h(s) := s - \ell_h(s)$. By lemma 7.4, we have $\ell_h(s) \geqslant \bar{s}$, and hence $t_h(s) \leqslant s - \bar{s} \leqslant (k+1)\bar{s} - \bar{s} \leqslant k\bar{s}$. In other words, given arbitrary $h \in \mathscr{P}$, the optimal choice at s is less than $k\bar{s}$. Since this is true for $h = T^k p$, we have

$$T^{k+1} p(s) = \min_{t \leqslant s}\{c(s-t) + \delta T^k p(t)\} = \min_{t \leqslant k\bar{s}}\{c(s-t) + \delta T^k p(t)\}.$$

Using the induction hypothesis and the preceding argument for $h = T^k q$, this is equal to

$$\min_{t \leqslant k\bar{s}}\{c(s-t) + \delta T^k q(t)\} = \min_{t \leqslant s}\{c(s-t) + \delta T^k q(t)\} = T^{k+1} q(s).$$

We have now shown that $T^{k+1} p = T^{k+1} q$ on $[0, (k+1)\bar{s}]$. The proof is complete. □

Lemma 7.9. *The operator T has one and only one fixed point in $\mathscr{P}$.*

Proof. To show existence, let $n \geqslant 1/\bar{s}$ and fix any $p \in \mathscr{P}$. In view of lemma 7.8, we have $T^n(Tp) = T^n p$. Equivalently, $T(T^n p) = T^n p$. In other words, $T^n p$ is a fixed point of T. Regarding uniqueness, let p and q be two fixed points of T in $\mathscr{P}$, and let $n \geqslant 1/\bar{s}$. In view of lemma 7.8, we have $p = T^n p = T^n q = q$. □

Our next step is to show that the unique fixed point of T in $\mathscr{P}$ is precisely the minimum value function defined in (8). In the statement of the result, t_p^j is the j-th composition of t_p with itself, and t_p^0 is the identity.

Lemma 7.10. *Let p be the unique fixed point of T in $\mathscr{P}$, and let t_p and ℓ_p be as defined in* (14)*. Let p^* be as defined in* (8)*. If s is any point in $[0, 1]$, then*

$$p^*(s) = p(s) = \sum_{i=1}^{\infty} \delta^{i-1} c(\ell_i), \tag{16}$$

where the allocation $\{\ell_i\}$ in (16) *is defined by $\ell_i := \ell_p(t_p^{i-1}(s))$ for all i.*

Proof. Pick any $s \in [0, 1]$. To see that the second equality is valid, observe from repeated applications of the fixed point property $p = Tp$ and the definitions of t_p and ℓ_p that

$$\begin{aligned} p(s) &= c(\ell_p(s)) + \delta p(t_p(s)) \\ &= c(\ell_p(s)) + \delta c(\ell_p(t_p(s))) + \delta^2 p(t_p^2(s)) \\ &\;\;\vdots \\ &= c(\ell_p(s)) + \delta c(\ell_p(t_p(s))) + \cdots + \delta^{n-1} c(\ell_p(t_p^{n-1}(s))) + \delta^n p(t_p^n(s)). \end{aligned}$$

for any n. Adopting the notation from the statement of the lemma, we can write this as

$$p(s) = \sum_{i=1}^{n} \delta^{i-1} c(\ell_i) + \delta^n p(t_p^n(s)) \quad \text{for all} \quad n \in \mathbb{N}.$$

We next show that $t_p^n(s) = 0$ for sufficiently large n. To see this, observe that, in view of lemma 7.3, we have $t_p(z) = 0$ whenever $z \leqslant \bar{s}$. Hence we need only prove that $t_p^n(s) \leqslant \bar{s}$ for some n. Suppose that this is not true. Then, since ℓ_p is increasing, since $\ell_p(\bar{s}) = \bar{s}$ and since $t_p^n(s) > \bar{s}$ for all n, we must have $\ell_p(t_p^n(s)) \geqslant \ell_p(\bar{s}) = \bar{s}$ for all n. On the other hand, $t_p^n(s) > \bar{s}$ for all n also implies that $\ell_p(t_p^n(s)) \to 0$ as $n \to \infty$. Contradiction.

For $i \geqslant n$ we also have $\ell_i = 0$. Since $p(0) = 0$ (recall the definition of $\mathscr{P}$) we have

$$p(s) = \sum_{i=1}^{n} \delta^{i-1} c(\ell_i) + \delta^n p(t_n) = \sum_{i=1}^{n} \delta^{i-1} c(\ell_i) = \sum_{i=1}^{\infty} \delta^{i-1} c(\ell_i).$$

This completes our proof of the second equality in (16).

Now we turn to the first equality in (16). To simplify notation, let $t_i := t_p^i(s)$. By the definition of $\{\ell_i\}$ and $\{t_i\}$ we have $\sum_{i=1}^{\infty} \ell_i = \sum_{i=1}^{\infty}(t_{i-1} - t_i) = t_0 = s$. As we've just shown that $p(s) = \sum_{i=1}^{\infty} \delta^{i-1} c(\ell_i)$, it follows from the definition of $p^*(s)$ that $p^*(s) \leqslant p(s)$. Thus it remains only to show that $p(s) \leqslant p^*(s)$ also holds.

To establish this, we will show that our allocation $\{\ell_i\}$ computed from t_p and ℓ_p is the minimizer in (8). For (8), given the convexity of c, the Karush-Kuhn-Tucker (KKT) conditions for optimality are necessary and sufficient. The conditions are existence of Lagrange multipliers $\alpha \in \mathbb{R}$ and $\{\mu_n\} \subset \mathbb{R}$ such that

$$\delta^{i-1} c'(\ell_i) = \mu_i + \alpha, \qquad \mu_i \geqslant 0 \quad \text{and} \quad \mu_i \ell_i = 0 \quad \text{for all} \quad i \in \mathbb{N}. \tag{17}$$

To see that this holds, let $\bar{n}$ be the largest n such that $\ell_n > 0$, let $\alpha := c'(\ell_1)$, and let $\mu_i := 0$ for $i = 1, \ldots, \bar{n}$ and $\mu_i := \delta^{i-1} c'(0) - \alpha$ for $i > \bar{n}$. We claim that $(\{\ell_i\}, \alpha, \{\mu_i\})$ satisfies the KKT conditions. To see this, observe that by repeatedly applying lemma 7.6 we obtain

$$\delta^{\bar{n}-1}\, c'(\ell_{\bar{n}}) = \delta^{\bar{n}-2}\, c'(\ell_{\bar{n}-1}) = \cdots = \delta\, c'(\ell_2) = c'(\ell_1) = \alpha. \tag{18}$$

Now take any $i \in \{1, \ldots, \bar{n}\}$. Since $\mu_n = 0$, the first equality in (17) follows from (18) and the second is immediate. On the other hand, if $i > \bar{n}$, then $\ell_i = 0$, and hence $\delta^{i-1} c'(\ell_i) = \delta^{i-1} c'(0) = \mu_i + \alpha$, where the last equality is by definition. Moreover, $\mu_i \ell_i = 0$ as required. Thus is remains only to check that $\mu_i = \delta^{i-1} c'(0) - c'(\ell_1) \geqslant 0$ when $i > \bar{n}$. Since $i > \bar{n}$, it suffices to show that $\delta^{\bar{n}} c'(0) \geqslant c'(\ell_1)$. In view of (18), this claim is equivalent to $\delta^{\bar{n}} c'(0) \geqslant \delta^{\bar{n}-1}\, c'(\ell_{\bar{n}})$, or $\delta c'(0) \geqslant c'(\ell_{\bar{n}})$. Regarding this inequality, recall the definition of $\bar{s}$ as the largest point in $(0, 1]$ satisfying $c'(\bar{s}) \leqslant \delta c'(0)$. From lemma 7.3 we have $\ell_{\bar{n}} \leqslant \bar{s}$. Since c' is increasing, we conclude that $\delta c'(0) \geqslant c'(\ell_{\bar{n}})$.

This completes the proof that the allocation $\{\ell_i\}$ defined in the statement of lemma 7.10 is a minimizer in (8). Hence $p(s) \leqslant p^*(s)$. We already showed that the reverse inequality holds, and hence the first equality in (16) also holds. □

The next result serves mainly to summarize implications and notation.

Corollary 7.11. *Let s be any point in $[0,1]$ and let*

$$p^*(s) := \min\left\{ \sum_{i=1}^{\infty} \delta^{i-1} c(\ell_i) \;:\; \{\ell_i\} \in \mathbb{R}_+^{\infty} \text{ and } \sum_{i=1}^{\infty} \ell_i = s \right\}. \tag{19}$$

If ℓ^* and t^* are as defined in (4) then

$$p^*(s) = \min_{t \leqslant s}\{c(s-t) + \delta p^*(t)\} = c(\ell^*(s)) + \delta p(t^*(s)). \tag{20}$$

Moreover, if we define $\{\ell_i\}$ and $\{t_i\}$ by

$$t_0 = s,\ \ell_i = \ell^*(t_{i-1}) \text{ and } t_i = t_{i-1} - \ell_i \tag{21}$$

then there exists a finite $n \in \mathbb{N}$ such that $t_n = 0$, and $p^*(t_{i-1}) = c(\ell_i) + \delta p^*(t_i)$ for $i = 1, \ldots, n$. Finally, the allocation $\{\ell_i\}$ is the unique minimizer in (19).

Proof. That $p^*(s) = \min_{t \leqslant s}\{c(s-t) + \delta p^*(t)\}$ follows immediately from lemma 7.10, which tells us that $Tp^* = p^*$. The second equality in (20) is immediate from the definitions of ℓ^* and t^*. Moreover, since $p = p^*$ in lemma 7.10, it follows that $\ell_p = \ell^*$ and $t_p = t^*$, and hence the "recursive" allocation $\{\ell_i\}$ defined in lemma 7.10 by $\ell_i := \ell_p(t_p^{i-1}(s))$ for all i is the same allocation defined in (21). As shown in lemma 7.10, this allocation is the minimizer in (19). In the proof of lemma 7.10 it is also shown that $t_n = 0$ for some finite n.

Finally, to see that the minimizer is unique, consider the functional F on the linear space of nonnegative sequences $\mathbb{R}_+^{\infty}$ defined by $F(\boldsymbol{\ell}) = F(\{\ell_i\}) = \sum_{i=1}^{\infty} \delta^{i-1} c(\ell_i)$. Let $\mathcal{L}$ be the set of $\boldsymbol{\ell} \in \mathbb{R}_+^{\infty}$ such that $F(\boldsymbol{\ell}) < \infty$. Evidently any minimizer lives in $\mathcal{L}$. From strict convexity of c it is easy to show that $\mathcal{L}$ is a convex set and F is a strictly convex function on $\mathcal{L}$. Hence the minimizer is unique. □

7.2. Main Results.

We can now prove the claims stated in the main section of the paper.

Proof of theorem 3.1. That T maps $\mathscr{P}$ to itself was shown in lemma 7.7. That T has a unique fixed point in $\mathscr{P}$ was shown in lemma 7.9. Moreover, $T^n p = p^*$ for all $n \geqslant 1/\bar{s}$, as shown in lemmas 7.8–7.9, so the final claim in theorem 3.1 is also true. □

Proof of theorem 3.2. Let p^* and $\boldsymbol{\ell}^*$ be as in the statement of the theorem. That the recursively generated allocation is feasible and the integer n^* in (6) is well defined can be seen from corollary 7.11. To see that $(p^*, \boldsymbol{\ell}^*)$ is an equilibrium for the production chain, observe that $p^*(0) = 0$ by (19), and that the zero profit conditions holds because $p^*(t_{i-1}^*) = c(\ell_i^*) + \delta p^*(t_i^*)$ for $i = 1, \ldots, n^*$, as shown in corollary 7.11. Part 3 of the definition of equilibrium is implied by the fact tht p^* is a fixed point of T. □

Proof of proposition 4.1. Since $p^* \in \mathscr{P}$ and the image under T of any function in $\mathscr{P}$ is strictly convex (lemma 7.1), we see that $p^* = Tp^*$ is strictly convex. Letting $p = p^*$ in lemma 7.5, we see that p^* is differentiable, with $(p^*)'(s) = c'(\ell^*(s))$. Since c and ℓ^* are continuous, the latter by lemma 7.2, and since $c'(s) > 0$ for all s, the equation $(p^*)'(s) = c'(\ell^*(s))$ implies that p^* also continuously differentiable and strictly increasing. The bounds $c'(0)s \leqslant p^*(s) \leqslant c(s)$

are immediate from $p^* \in \mathscr{P}$. That p^* is equal to the right-hand side of (8) was shown in lemma 7.10. $\square$

Proof of theorem 4.2. That $(p^*, \boldsymbol{\ell}^*)$ is an equilibrium in $\mathcal{E}$ follows directly from the definitions, theorem 3.2 and (20). Regarding uniqueness, our first claim is that if p is a nontrivial price function and an equilibrium for some allocation, then $p = p^*$. To see that this is so, fix $s \in [0,1]$. Since p is nontrivial, there exists an allocation $\boldsymbol{\ell}$ such that $\sum_{i=1}^n \ell_i = s$ and $p(t_{i-1}) \geqslant c(\ell_i) + \delta p(t_i)$ for $i = 1, \dots, n$. Iterating on this inequality gives $p(s) \geqslant c(\ell_1) + \delta c(\ell_2) + \delta^2 c(\ell_3) + \cdots$. It now follows from (19) that $p(s) \geqslant p^*(s)$. For the reverse inequality, let $\{\ell_i\}$ and $\{t_i\}$ be as defined in (21). By corollary 7.11 we have $p^*(s) = \sum_{i=1}^n \delta^{i-1} c(\ell_i)$. On the other hand, since p is an equilibrium price function, the inequality in condition 3 of definition 3.1 holds. In particular, $p(t_{i-1}) \leqslant c(\ell_i) + \delta p(t_i)$ for $i = 1, \dots, n$. Iterating on this inequality from $i = 1$ gives $p(s) = p(t_0) \leqslant \sum_{i=1}^n \delta^{i-1} c(\ell_i)$. In other words, $p(s) \leqslant p^*(s)$.

Now let $\boldsymbol{\ell}^*$ be as above and let $\boldsymbol{\ell}$ be any other feasible allocation. We claim that if $(p^*, \boldsymbol{\ell})$ is an equilibrium, then $\boldsymbol{\ell} = \boldsymbol{\ell}^*$. To see this, suppose to the contrary that $\boldsymbol{\ell}^* = \{\ell_i^*\}$ and $\boldsymbol{\ell} = \{\ell_i\}$ are distinct. Then, by the fact that $\{\ell_i^*\}$ is the unique minimizer in (19) when $s = 1$, we have $p^*(1) = \sum_{i=1}^n \delta^{i-1} c(\ell_i^*) < \sum_{i=1}^n \delta^{i-1} c(\ell_i)$. At the same time, since $(p^*, \{\ell_i\})$ is an equilibrium, the zero profit condition yields $p^*(t_{i-1}) \leqslant c(\ell_i) + \delta p^*(t_i)$ for $i = 1, \dots, n$, and hence, by iterating, $p^*(1) = \sum_{i=1}^n \delta^{i-1} c(\ell_i)$. Contradiction. $\square$

Proof of proposition 4.3. We already showed that $(p^*)'(s) = c'(\ell^*(s))$. The claimed properties on t^* and ℓ^* are immediate from lemma 7.2. $\square$

Proof of proposition 4.4. Let $\delta_a \leqslant \delta_b$. Let T_a and T_b be the corresponding operators. We begin with the claim that $p_a^* \leqslant p_b^*$. It is easy to verify that if $p \in \mathscr{P}$, then $T_a p \leqslant T_b p$ pointwise on $[0,1]$. Since T_a and T_b are order preserving (i.e., $p \leqslant q$ implies $Tp \leqslant Tq$), this leads to $T_a^n p \leqslant T_b^n p$. For n sufficiently large, this states that $p_a^* \leqslant p_b^*$. Next we show that the number of tasks carried out by the most upstream firm decreases when δ increases from δ_a to δ_b. Let ℓ_i^a be the number of task carried out by firm i when $\delta = \delta_a$, and let ℓ_i^b be defined analogously. Let $n = n_a^*$. Seeking a contradiction, suppose that $\ell_n^b > \ell_n^a$. In that case, convexity of c and (11) imply that

$$c'(\ell_{n-1}^b) = \delta_b c'(\ell_n^b) > \delta_a c'(\ell_n^a) = c'(\ell_{n-1}^a).$$

Hence $\ell_{n-1}^b > \ell_{n-1}^a$. Continuing in this way, we obtain $\ell_i^b > \ell_i^a$ for $i = 1, \dots, n$. But then $\sum_{i=1}^n \ell_i^b > \sum_{i=1}^n \ell_i^a = 1$. Contradiction.

Now we can turn to the claim that $n_b^* \leqslant n_a^*$. As before, let $n = n_a^*$, the equilibrium number of firms when $\delta = \delta_a$. If $\ell_n^b = 0$, then the number of firms at δ_b is less than $n = n_a^*$ and we are done. Suppose instead that $\ell_n^b > 0$. In view of lemma 7.3, we have $\delta_a c'(0) \geqslant c'(\ell_n^a)$. Moreover, we have just shown that $\ell_n^a \geqslant \ell_n^b$. Combining these two inequalities and using $\delta_b > \delta_a$, we have $\delta_b c'(0) \geqslant c'(\ell_n^b)$. Applying lemma 7.3 again, we see that the n-th firm completes the good, and hence $n_b^* = n_a^*$. $\square$

Proof of theorem 5.1. Let T be the operator in (5). Evidently T maps $\mathscr{F}$ into itself. By construction, solutions to (13) in $\mathscr{F}$ coincide with fixed points of T. Endow $\mathscr{F}$ with the usual pointwise order, so that $f \leqslant g$ in $\mathscr{F}$ whenever $f(s) \leqslant g(s)$ for all $s \in [0,1]$. With this order, it is easy to verify that $(\mathscr{F}, \leqslant)$ is a complete lattice.[20] Moreover, since k and δ are both nonnegative, T is order preserving on $(\mathscr{F}, \leqslant)$. The claim now follows from the Knaster–Tarski fixed point theorem. □

Lemma 7.12. *If $c(\ell) = e^{\theta\ell} - 1$, then the equilibrium number of firms is given by* (12).

Proof. Let $n = n^*$ be the equilibrium number of firms and let $r := \ln(\delta)/\theta$. From $\delta c'(\ell_{n+1}) = c'(\ell_n)$ we obtain $\ell_{i+1} = \ell_i - r$, and hence $\ell_1 = \ell_n + (n-1)r$. It is easy to check that when $c(\ell) = e^{\theta\ell} - 1$, the constant $\bar{s}$ defined above is equal to r. Applying lemma 7.3 we get $0 < \ell_n \leqslant r$. Therefore $(n-1)r < \ell_1 \leqslant nr$. From $\sum_{i=1}^n \ell_i = 1$ and $\ell_1 = \ell_n + (n-1)r$ it can be shown that $n\ell_1 - n(n-1)r/2 = 1$. Some straightforward algebra now yields $-1 + \sqrt{1 + 8/r} < 2n \leqslant 1 + \sqrt{1 + 8/r}$. The expression for $n = n^*$ in (12) follows. □

References

AGHION, P. AND R. HOLDEN (2011): "Incomplete contracts and the theory of the firm: What have we learned over the past 25 years?" *The Journal of Economic Perspectives*, 25, 181–197.

ANTRÀS, P. AND D. CHOR (2013): "Organizing the global value chain," *Econometrica*, 81, 2127–2204.

ANTRÀS, P., D. CHOR, T. FALLY, AND R. HILLBERRY (2012): "measuring the upstreamness of production and trade flows," *The American Economic Review*, 102, 412–416.

AXTELL, R. L. (2001): "Zipf distribution of US firm sizes," *Science*, 293, 1818–1820.

BECKER, G. S. AND K. M. MURPHY (1992): "The division of labor, coordination costs, and knowledge," *The Quarterly Journal of Economics*, 107, 1137–1160.

BENVENISTE, L. M. AND J. A. SCHEINKMAN (1979): "On the differentiability of the value function in dynamic models of economics," *Econometrica*, 47, 727–732.

BLUME, L. E., D. EASLEY, J. KLEINBERG, AND E. TARDOS (2009): "Trading networks with price-setting agents," *Games and Economic Behavior*, 67, 36–50.

BRESNAHAN, T. F. AND J. D. LEVIN (2012): "Vertical integration and market structure," Tech. rep., National Bureau of Economic Research.

CALIENDO, L. AND E. ROSSI-HANSBERG (2012): "The Impact of Trade on Organization and Productivity," *The Quarterly Journal of Economics*, 127, 1393–1467.

COASE, R. H. (1937): "The nature of the firm," *Economica*, 4, 386–405.

COSTINOT, A. (2009): "On the origins of comparative advantage," *Journal of International Economics*, 77, 255–264.

COSTINOT, A., J. VOGEL, AND S. WANG (2013): "An elementary theory of global supply chains," *The Review of Economic Studies*, 80, 109–144.

FALLY, T. (2012): "On the fragmentation of production in the US," Tech. rep., University of Colorado.

[20]For example, with this order the supremum in $(\mathscr{F}, \leqslant)$ reduces to the usual pointwise supremum of functions, and the pointwise supremum of increasing functions is increasing. Similarly, the pointwise infimum of increasing functions is increasing.

FALLY, T. AND R. HILLBERRY (2015): "A Coasian Model of International Production Chains," Tech. rep., UC Berkeley.

GARICANO, L. (2000): "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy*, 108, 874–904.

GARICANO, L. AND E. ROSSI-HANSBERG (2006): "Organization and Inequality in a Knowledge Economy," *The Quarterly Journal of Economics*, 121, 1383–1435.

GEANAKOPLOS, J. AND P. MILGROM (1991): "A theory of hierarchies based on limited managerial attention," *Journal of the Japanese and International Economies*, 5, 205–225.

GEEROLF, F. (2016): "A Theory of Pareto Distributions," Tech. rep., UCLA.

GROSSMAN, G. M. AND E. HELPMAN (2002): "Integration Versus Outsourcing In Industry Equilibrium," *The Quarterly Journal of Economics*, 117, 85–120.

GROSSMAN, G. M. AND E. ROSSI-HANSBERG (2012): "Task trade between similar countries," *Econometrica*, 80, 593–629.

HART, O. AND J. MOORE (1990): "Property Rights and the Nature of the Firm," *Journal of Political Economy*, 98, 1119–1158.

HAYEK, F. A. (1945): "The use of knowledge in society," *The American Economic Review*, 35, 519–530.

HOLMSTROM, B. AND P. MILGROM (1994): "The firm as an incentive system," *The American Economic Review*, 84, 972–991.

HSU, W.-T., T. J. HOLMES, AND F. MORGAN (2014): "Optimal city hierarchy: A dynamic programming approach to central place theory," *Journal of Economic Theory*, 154, 245–273.

JENSEN, M. C. AND W. H. MECKLING (1976): "Theory of the firm: Managerial behavior, agency costs and ownership structure," *Journal of Financial Economics*, 3, 305–360.

KIMURA, F. (2002): "Subcontracting and the performance of small and medium firms in Japan," in *Small Firm Dynamism in East Asia*, Springer, 163–175.

KLEIN, B., R. G. CRAWFORD, AND A. A. ALCHIAN (1978): "Vertical integration, appropriable rents, and the competitive contracting process," *Journal of Law and Economics*, 297–326.

KREMER, M. (1993): "The O-ring theory of economic development," *The Quarterly Journal of Economics*, 108, 551–575.

LEVINE, D. K. (2012): "Production chains," *Review of Economic Dynamics*, 15, 271–282.

LUCAS, R. E. (1978): "On the size distribution of business firms," *The Bell Journal of Economics*, 9, 508–523.

LUCAS, R. E. AND E. ROSSI-HANSBERG (2002): "On the internal structure of cities," *Econometrica*, 70, 1445–1476.

LUTTMER, E. G. (2007): "Selection, growth, and the size distribution of firms," *The Quarterly Journal of Economics*, 122, 1103–1144.

MCAFEE, R. P. AND J. MCMILLAN (1995): "Organizational diseconomies of scale," *Journal of Economics & Management Strategy*, 4, 399–426.

MEAGHER, K. J. (2003): "Generalizing incentives and loss of control in an optimal hierarchy: the role of information technology," *Economics Letters*, 78, 273–280.

MILGROM, P. AND J. ROBERTS (1992): *Economics, organization and management*, Prentice-Hall International.

NORTH, D. C. (1993): “Institutions, transaction costs and productivity in the long run,” Tech. rep.

ROBERTSON, D. H. (1923): *Control of industry*, Nisbet and Co. Ltd, London.

ROSSI-HANSBERG, E. AND M. L. WRIGHT (2007): “Establishment size dynamics in the aggregate economy,” *The American Economic Review*, 97, 1639–1666.

SEABRIGHT, P. (2010): *The company of strangers: a natural history of economic life*, Princeton University Press.

SUBRAHMANYA, B. (2008): “Industrial subcontracting and structure in Japan: evolution and recent trends,” *Journal of Management History*, 14, 23–38.

WEINTRAUB, G. Y., C. L. BENKARD, AND B. VAN ROY (2011): “Industry dynamics: Foundations for models with an infinite number of firms,” *Journal of Economic Theory*, 146, 1965–1994.

WILLIAMSON, O. E. (1979): “Transaction-cost economics: the governance of contractual relations,” *Journal of Law and Economics*, 22, 233–261.

——— (1981): “The economics of organization: the transaction cost approach,” *American Journal of Sociology*, 87, 548–577.