Transaction Costs, Span of Control and Competitive Equilibrium[1]

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

[a] Centre on Asia and Globalization, National University of Singapore

[b] Research Institute of Economics and Business, Kobe University

[c] Research School of Economics, Australian National University

August 27, 2014

ABSTRACT. The firm size distribution exhibits important regularities across countries, time and levels of aggregation. To date, attempts to explain these regularities have assumed either exogenous heterogeneity or exogenous idiosyncratic shocks. In this paper we pursue an alternative line of analysis, starting with fundamental ideas from the theory of the firm and examining their aggregate implications in a competitive setting with sequential production and symmetric costs. Key stylized facts for the firm size distribution are replicated in equilibrium, despite the absence of any intrinsic heterogeneity.

JEL Classifications: L11, L16

Keywords: Transaction costs, span of control, distribution of firms

1. INTRODUCTION

While the size distribution of business firms varies over time and space, as well as with the measure of size adopted and the level of aggregation, it also shows significant regularities. For example, the distribution almost always exhibits a high degree of skewness and a long right hand tail. Aggregated data covering firms from

[1]The authors thank Thibault Fally, Debraj Ray, Tom Sargent, Makoto Yano and Shenghao Zhu for helpful discussions, as well as Ricardo Lagos, an associate editor and two anonymous referees for many important comments and suggestions. Denis Tan, Ngoc Han Huynh Khanh, Chenghan Hou, Qingyin Ma and Alex Olssen provided excellent research assistance. We acknowledge financial support from JSPS Grant-in-Aid for Scientific Research (23000001 and 23330063), Singapore Ministry of Education Academic Research Fund 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`

all sectors of the economy suggests that the tail obeys a power law, while for manufacturing firms the tail is somewhat lighter, and well approximated by either a lognormal or gradually truncated lognormal. Such regularities were first documented and analyzed in detail by Gibrat [18], and have since been studied extensively.[2]

The problem of developing a consistent and nontrivial theory of the size distribution of business firms has attracted the interest of many economists. In developing such a theory, one of the main difficulties is to obtain a distribution having the features observed in the data without already imposing these characteristics as assumptions. For example, the elegant model of Lucas [31] predicts a Pareto distribution for firm size, but only after assuming a Pareto distribution for managerial talent. Similarly, Cabral and Mata [9] develop a model based on financial constraints that matches certain features of the data, but the source and distribution of these constraints is exogenous to the model.

An alternative approach to the study of firm size distributions is to model individual firm size dynamics via exogenous, idiosyncratic shocks. This line of work begins with Gibrat [18], who assumes proportional, scale independent firm growth to obtain a lognormal distribution in cross-section. His work was followed by contributions from Simon and Bonini [40], Luttmer [32], Rossi-Hansberg and Wright [38] and others, and is closely connected to research on power laws for other variables such as productivity, city sizes or wealth [14, 5, 16].

In this paper we develop a third line of analysis, without any form of exogenous heterogeneity, stochastic or otherwise. We introduce a model of decentralized production that draws on some of the foundational ideas from the modern theory of the firm and examines their aggregate implications in a competitive equilibrium setting. The model is populated with infinitely many ex-ante identical price taking firms, only some of which choose to be active in equilibrium. The resulting firm size distribution replicates a number of regularities described above, despite the fact that all agents are ex-ante identical and no idiosyncratic shocks are present. Instead all ex-post heterogeneity is generated purely through profit maximizing behavior coordinated by price signals.

[2] See, for example, Simon and Bonini [40], Stanley *et al.* [41], Axtell [3], Cabral and Mata [9], Gupta *et al.* [22] and Rossi-Hansberg and Wright [38].

One benefit of this approach is that it connects the theory of the firm size distribution with some of the fundamental topics from the theory of the firm; in particular, optimal firm size, optimal firm boundaries and the make-or-buy decision. This make-or-buy decision involves a trade off that sits at the heart of organizational economics and the theory of the firm: While markets can and do coordinate production, the existence of frictions associated with search, contracting and other trading activities impose a cost on market participants. On the other hand, when firms eschew the market (i.e., make rather than buy) they are confronted with internal coordination costs associated with imperfect information, costly monitoring, loss of incentives and so on. In other words, both forms of coordination are costly, and the choice of firm boundaries (and hence firm size) requires optimal choices in the face of these costs.[3]

In the model developed below we study these choices from an aggregate perspective. A single final good is produced through the sequential completion of a large range of tasks, or stages of production. The total number of firms involved in implementing these tasks and their size and location in the production chain is endogenous. A nondegenerate distribution of firms arises from the fact that, in equilibrium, upstream and downstream firms face different trade offs in terms of their make-or-buy decision. In particular, equilibrium prices imply that the relative marginal cost of backwards integration into input production is lower for downstream firms than for upstream firms.

In essence, the force behind this differential is the fact that coordination is costly. The presence of market frictions means that while firms can mitigate internal coordination costs by trading inputs, they cannot offset them entirely. One way to view this is that each shift downstream along the stages of production requires not only additional direct costs but more coordination in aggregate (whether implemented internally or through the market). Hence aggregate costs associated with production up to a given stage rise faster than in proportion to the range of tasks needed to reach that stage. This leads to a growing marginal cost associated with market transactions as the product moves downstream. As a consequence, the profit maximizing size for downstream firms is larger. These ideas are clarified in section 5.

1.1. **Related Literature.** The modern theory of optimal firm size begins with the work of Coase [10], who argued that firms exist because there are transaction costs

[3] See, e.g., [10, 28, 44, 21, 23, 25, 34, 36]. The literature review below gives more details.

associated with using the market, and hence entrepreneurs and managers can sometimes coordinate production at a lower cost within the firm. On the other hand, firms do not expand without limit due to "diminishing returns to management." The boundary of the firm is then determined by the point at which the cost of organizing another productive task within the firm is equal to the marginal cost of acquiring a similar input or service through the market [10, p. 395].

Many researchers have expanded on and clarified these ideas. At a fundamental level, most of their theories relate to frictions associated with bounded rationality and imperfect information. For example, bounded rationality and uncertainty prevent firms from writing complete contracts over the goods and services they wish to trade. This leads to the threat of hold-up and ex-post bargaining, increasing the relative cost of market based transactions (e.g., Klein *et al.* [28], Williamson [44], Grossman and Hart [21], Hart and Moore [23]).

Imperfect information and bounded rationality also drive the kinds of internal coordination costs that increase with firm scope, such as the informational cost of coordinating specialists (Becker and Murphy [4]), private information, agency problems and costly monitoring (Jensen and Meckling [27], Holmstrom and Milgrom [25], [34]), costs of communication, bureaucracy and hierarchy (Milgrom and Roberts [36], Geanakoplos and Milgrom [15], Meagher [35]) and so on.[4]

This paper differs from the preceding line of work in two respects. First, rather than investigating microfoundations or studying the unified theory of the firm, we take features such as transaction costs and internal coordination costs as given. In this sense our paper is not intended as a contribution to the theory of the firm *per se*. Rather, we seek to analyze the aggregate implications of these phenomena in equilibrium.[5] Second, the great majority of the formal literature investigates firm decisions in a setting with a small number of firms—usually one or two. Such a

[4] The literature is large and we cite only representative examples. Thoughtful surveys and additional references can be found in Bresnahan and Levin [8], Gibbons [17], Lafontaine and Slade [29] and Aghion and Holden [1].

[5] Why is aggregate analysis important? Consider the effect of an exogenous decrease in transaction costs, say, perhaps as the result of a new technology. In a model of bilateral firm interaction, this will affect the actions of the two firms in the model, which is where the analysis ends. In reality, however, this change in firm behavior will affect prices, which will affect the choices of other firms in the production chain, which will feed back into prices and so on. By capturing and resolving these equilibrium effects, it becomes possible to uncover aggregate implications of the theory of the firm, and through them study the structure of production and consequent firm size distribution.

setting is inappropriate for modeling the firm size distribution. In our paper the number of firms is endogenous and often large.

There are contributions from the theory of the firm literature that treat make-or-buy decision problems in an equilibrium setting with a large number of firms, and hence come closer to the environment that we study. A prominent example is Grossman and Helpman [19]. In the spirit of Coase [10] and Grossman and Hart [21], search frictions and contractual incompleteness make market based trade in intermediate goods expensive, and these costs must be weighed against the extra governance costs associated with integration. In equilibrium, production takes place either in vertically integrated firms or in pairs of specialized companies. While the paper provides careful microfoundations of different coordination costs, the fact that there are only three distinct firm types (vertically integrated firms, specialist input producers and specialist final good producers) makes the model less suitable for investigating the firm size distribution.

A more sophisticated picture of fragmentation of production and supply chains can be found in recent work on outsourcing and international trade. For example, Costinot [11], Grossman and Rossi-Hansberg [20] and Costinot *et al.* [12] all consider sequential production over a continuum of tasks with a large number of firms. In particular, Costinot *et al.* [12] shares many features with our model, including sequential production, competitive behavior and a potentially sophisticated pattern of vertical specialization. On the other hand, it differs in the sense that the heterogeneity observed in equilibrium (where rich countries specialize on downstream production) is driven to some extent by intrinsic heterogeneity (countries have exogenous technological characteristics). This is a natural assumption for the authors, since, as with other papers on international trade, the primary interest is in the *implications* of cross-country variation for patterns of trade and production. As discussed above, the model in this paper assumes no heterogeneity in technological or any other characteristics.

More details on the structure of our model can be found in section 2. Section 3 gives our definition of equilibrium and our fundamental existence and uniqueness results. In section 4 we introduce a recursive perspective, which leads to computational techniques and additional insights on the nature of equilibrium prices and allocations. Section 5 considers implications for the structure of the value chain and the distribution of firms. Section 6 concludes.

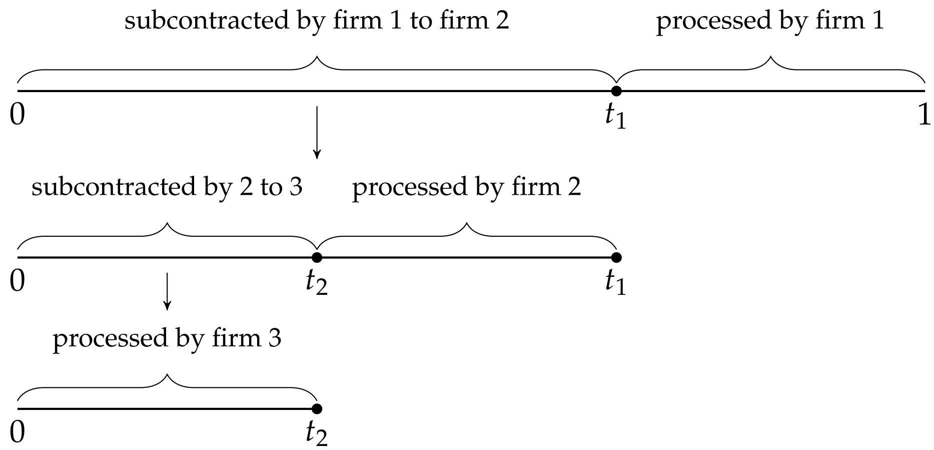


FIGURE 1. Recursive allocation of production tasks

2. THE MODEL

We begin by studying production of a single unit of a final good. Our focus is on 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 takes place via subcontracting. The subcontracting scheme is 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" or "number" of tasks carried out by firm i. Figure 2 serves to clarify notation.

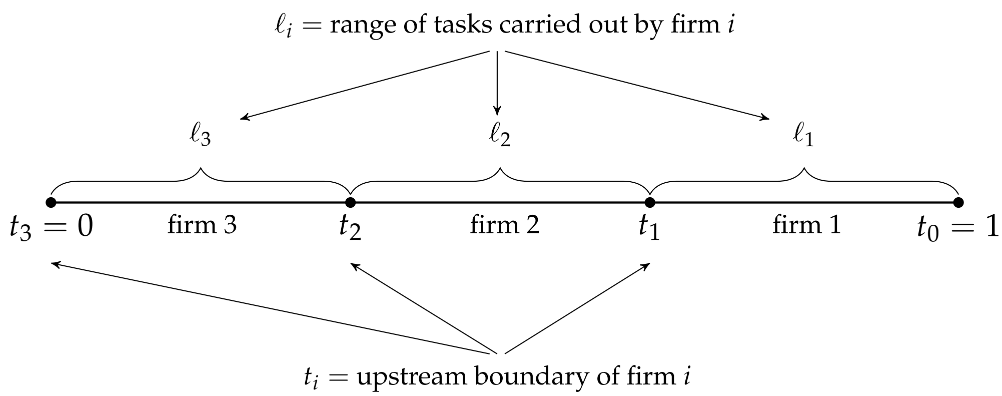


FIGURE 2. Notation

Notice that by construction each firm chooses only its *upstream* boundary, treating its downstream boundary as given. In essence, 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 will face 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. As shown in section 4, this recursive structure provides powerful tools for the study of equilibrium.

2.2. **In House Production Costs.** We study allocation of tasks in the presence of what Coase [10] 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.[6] As in Becker and Murphy [4, p. 1142], 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. Thus, average cost per task rises with the number of tasks performed in-house. We also assume that c is continuously differentiable, with $c(0) = 0$ and $c'(0) > 0$. These assumptions imply that c is strictly increasing.[7]

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 [10] 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. (Suppose for example that contracts are complete but not free.) 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 [10, 44, 37, 7].

Here we follow Arrow [2] in simply regarding transaction costs 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 footnote 12 below). Hence we assume that the cost is borne only by the buyer. In particular, when two firms agree to a trade at face

[6]Influential studies based on this assumption include Coase [10], Lucas [31] and Becker and Murphy [4]. 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 [24], 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 free-riding, shirking and other incentive problems. See the literature review in section 1.1 for details.

[7]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 ℓ tasks $\ell = s - t$ 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 base case where tasks are ex-ante identical.

value v, the buyer's total outlay is taken to be δv, where $\delta > 1$. The seller receives only v, and the difference is paid to agents outside the model.[8]

3. Equilibrium

The next step is to define a notion of equilibrium for the production chain. In doing so we assume throughout that all firms are ex-ante identical and act as price takers, and that active firms are surrounded by an countably infinite fringe of competitive firms ready to step in on the buyer or seller side should it be profitable to do so. There are no fixed costs or barriers to entry and production involves no uncertainty. As a result, no firm makes strictly positive profits and no holdup or other ex-post opportunistic behavior occurs.

3.1. **Definition of Equilibrium.** Throughout the paper, an *allocation* is just a non-negative sequence $\boldsymbol{\ell} := \{\ell_i\}_{i \in \mathbb{N}} \in \mathbb{R}_+^\infty$. An allocation implicitly defines a division of tasks across firms, with ℓ_i being the range of task implemented by the i-th firm (see figure 2). An allocation $\boldsymbol{\ell}$ is called *feasible* if there exists some number n such that $\sum_{i=1}^\infty \ell_i = \sum_{i=1}^n \ell_i = 1$. Feasibility means that the entire production process is completed with only finitely many firms. Finiteness of the allocation is fundamental in our model because production will not take place unless the number of active firms is finite.

If $\ell_i = 0$ then firm i is understood to be inactive. To simplify notation we always assume that firms enter in order, starting with firm 1. In particular, if $\boldsymbol{\ell}$ is feasible and n firms are active then $\ell_i > 0$ for $i \leq n$ and $\ell_i = 0$ for $i > n$. This assumption confers no special privileges on earlier firms and costs no generality in our results, as we can always relabel firms as required.

As price takers, firms are confronted by a *price function*, which is function p from $[0, 1]$ to $\mathbb{R}_+$. The value $p(t)$ is interpreted as the price of the good at processing stage t. Given the price function and the allocation we can compute the profits of the i-th firm. Recalling the notation in figure 2, this firm enters a contract to sell the good at stage t_{i-1} and purchase at stage t_i. It undertakes the remaining ℓ_i tasks in house.

[8] Our proportional representation δv models those transaction costs that tend to rise with the face value of the transaction, such as insurance, contracts, sales taxes, trade credit, currency hedging, search costs, bribes and so forth. In reality firms also face other transaction costs that are more naturally thought of as constant in face value (such as transportation). For the sake of simplicity, we ignore these additive costs.

Its total costs are given by the sum of its processing costs $c(\ell_i)$ and the gross input cost $\delta p(t_i)$. Since transaction costs are incurred only by the buyer (see section 2.3), profits are

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

To exclude trivial cases, we will call a price function p *nontrivial* if, for every $s \in (0,1]$, 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 $\sum_{i=1}^{n} \ell_i = s$ and

$$p(t_{i-1}) - c(\ell_i) - \delta p(t_i) \geq 0 \quad \text{for} \quad i = 1, \ldots, n.$$

Here $\{t_i\}$ is the transaction stages corresponding to $\boldsymbol{\ell}$, defined by $t_0 = s$ and $t_{i+1} = t_i - \ell_{i+1}$. In essence, the meaning is that $p(s)$ is large enough to support production up to stage s. At a lower price one firm will always lose money, and hence such a price cannot be observed in equilibrium.

Definition 3.1. Let $\mathcal{E}$ be the set of all pairs $(p, \boldsymbol{\ell})$, where p is a nontrivial price function and $\boldsymbol{\ell}$ is a feasible allocation. An *equilibrium* for the production process is a pair $(p, \boldsymbol{\ell}) \in \mathcal{E}$ such that

1. $p(0) = 0$,
2. $p(t_{i-1}) - c(\ell_i) - \delta p(t_i) = 0$ for $i = 1, \ldots, n$, and
3. given any other feasible allocation $\boldsymbol{\ell}' \neq \boldsymbol{\ell}$, we have

$$p(t'_{i-1}) - c(\ell'_i) - \delta p(t'_i) \leq 0 \quad \text{for all} \quad i = 1, \ldots, n'$$

with strict inequality for at least one i.

In conditions 2 and 3, n and n' are the finite number of active firms implicit in the definition of feasible allocations $\boldsymbol{\ell}$ and $\boldsymbol{\ell}'$. The sequence of transaction stages $\{t_i\}$ is used in the definition just to simplify notation, and has the usual meaning $t_0 = 1$ and $t_{i+1} = t_i - \ell_{i+1}$. In other words, $\{\ell_i\}$ and $\{t_i\}$ satisfy the relationship shown in figure 2. The relationship between $\{\ell'_i\}$ and $\{t'_i\}$ is analogous.

Condition 1 is a zero profit condition for suppliers of initial inputs (e.g., raw materials). In particular, the value $p(0)$ is the revenue of firms that supply the initial inputs to production. The zero profits condition in this sector implies that $p(0)$ is equal to the cost of producing these inputs. To simplify notation, we assume that this cost is zero, and hence $p(0) = 0$.

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 exactly the same role in the production chain. Moreover, profits are never negative in equilibrium, since firms can freely exit. Zero profits also mean that no active firm has an incentive to exit.

Condition 3 has two main implications. To understand them, observe first that whenever condition 3 holds we must have

$$p(s) - c(s-t) - \delta p(t) \leq 0 \tag{1}$$

for any pair s, t with $t \leq s \leq 1$. Since active firms achieve zero profits, this means that no active firm has an incentive to deviate in terms of either location or number of stages processed in house. In particular, all active firms are maximizing profits.

The second implication of condition 3 is stability of the equilibrium against entry by non-incumbents, either individually or as a coalition. In particular, provided that they act as price takers, any alternative chain of firms would find at least one firm with strictly negative profits. Such a chain cannot function as a noncooperative equilibrium. Moreover, since all other firms would make nonpositive profits, the total sum of profits would be negative. This rules out the formation of a coalition based on side payments.

3.2. **Existence and Uniqueness.** In this section we show that exactly one equilibrium exists for this economy and provide an explicit representation. To begin we define p^* as a function on $[0,1]$ by

$$p^*(s) := \min \left\{ \sum_{i=1}^{\infty} \delta^{i-1} c(\ell_i) \;:\; \{\ell_i\} \in \mathbb{R}_+^{\infty} \text{ and } \sum_{i=1}^{\infty} \ell_i = s \right\}. \tag{2}$$

The domain of the minimization problem is not finite dimensional because it is not obvious a priori that an optimal sequence will have only finitely many positive values, or, even if this is the case, what the number of nonzero values will be. In fact, since the domain is infinite dimensional, some degree of effort is required to show that a solution exists. However, it turns that a solution does exist under our assumptions, and that the minimizer is unique:

Proposition 3.1. *For each $s \in [0,1]$, the value $p^*(s)$ in (2) is finite and well defined, and the minimizer $\boldsymbol{\ell} = \{\ell_i\}$ exists and is unique. Moreover, $\{\ell_i\}$ is decreasing and has only finitely many nonzero entries.*

The proof is deferred to section 7. To understand what $p^*(s)$ 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 the transaction stages $\{t_i\}$ be defined from $\boldsymbol{\ell}$ in the usual way.[9] If all firms make zero profits then $p(t_{i-1}) = c(\ell_i) + \delta p(t_i)$ for $i = 1, \dots, n$. 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 that we're minimizing in (2). Hence the value $p^*(s)$ in (2) can be thought of as the minimal cost of producing s when all firms make zero profits. In other words, it is the solution to a planner problem where the planner cannot avoid transaction costs. Not surprisingly, it also yields an equilibrium for our model. In fact it is the only equilibrium. These facts are clarified in our next result. In the statement of the theorem, p^* is the function defined in (2) and $\boldsymbol{\ell}^*$ is the minimizer in (2) when $s = 1$. $\mathcal{E}$ is as specified in definition 3.1.

Theorem 3.2. *There exists exactly one equilibrium in $\mathcal{E}$ and that equilibrium is $(p^*, \boldsymbol{\ell}^*)$.*

The proof is in section 7. In all of what follows, the equilibrium number of firms corresponding to $\boldsymbol{\ell}^*$ is denoted by n^*. That n^* is finite follows from proposition 3.1.[10]

From the first order conditions for (2), we see that, at the equilibrium allocation $\boldsymbol{\ell}^* = \{\ell_i^*\}$, active firms that are adjacent satisfy

$$\delta\, c'(\ell_{i+1}^*) = c'(\ell_i^*). \tag{3}$$

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.

The preceding discussion gives us the price of a single unit of the final good in question. 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

[9] That is, $t_0 = s$ and $t_{i+1} = t_i - \ell_{i+1}$. As in figure 2 but replacing $t_0 = 1$ with $t_0 = s$.

[10] The proof of this fact depends on our assumption that $c'(0) > 0$.

without affecting factor prices, $p^*(1)$ is also the long run equilibrium price in the market for the final good. Note that, in view of (2), we have $p^*(1) = \sum_{i=1}^{n^*} \delta^{i-1} c(\ell_i^*)$.

Example 3.1. 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 (3) 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 7.13) that the equilibrium number of firms is

$$n^* = \left\lfloor 1/2 + (1 + 8\theta / \ln \delta)^{1/2} / 2 \right\rfloor, \tag{4}$$

where $\lfloor a \rfloor$ is the largest integer less than or equal to a.

In (4), observe that n^* is decreasing in δ. This is not surprising, since higher transaction costs encourage less use of the market and more in-house production. In other words, firms get larger. Since the number of tasks does not change, larger firms implies less firms.

At the same time, n^* is increasing in θ. To see why, observe that $\theta = c''(\ell)/c'(\ell)$, 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. A Recursive View

Macroeconomics has been revolutionized by the application of recursive methods to the study of dynamic competitive economies [30, 42]. Recursive methods lend themselves to simple computation and shine new light on the equilibria. While the model considered in this paper is not dynamic, it does have a recursive structure (recall section 2.1). It turns out that, by exploiting this structure, it becomes possible to adapt many of the fundamental recursive methods developed in the field of economic dynamics to the equilibrium problem treated in section 3.

4.1. **The Price Function as a Fixed Point.** As a first step, we introduce an operator T that maps a given function $p \colon [0,1] \to \mathbb{R}_+$ to Tp via

$$Tp(s) = \min_{t \leq s} \{ c(s-t) + \delta p(t) \} \quad \text{for all} \quad s \in [0,1]. \tag{5}$$

Here and below, the restriction $0 \leq t$ 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. At the same time, in the standard dynamic programming setting, the fundamental theory is driven by contraction mapping arguments, which in turn depend on the assumption that the discount factor is strictly less than one. Here, however, the "discount factor" δ is strictly greater than one, and T in (5) is not a contraction in any obvious metric. Indeed, $T^n p$ diverges for many choices of p, even when continuous and bounded. For example, if $p \equiv 1$, then $T^n p = \delta^n 1$, which diverges to $+\infty$.

Nevertheless, it turns out that there exists a domain on which T is very well-behaved. The restricted domain in question is the set of convex increasing continuous functions $p \colon [0,1] \to \mathbb{R}$ such that $c'(0)s \leq p(s) \leq c(s)$ for all $0 \leq s \leq 1$. We denote this set of functions by $\mathscr{P}$. In the statement of the next theorem, p^* is as usual the equilibrium price function defined in (2).

Theorem 4.1. *Under our assumptions the following statements are true:*

1. *The operator T defined in* (5) *maps $\mathscr{P}$ into itself.*
2. *T has exactly one fixed point in $\mathscr{P}$, and that fixed point is p^*.*
3. *For all $p \in \mathscr{P}$ we have $T^k p \to p^*$ uniformly as $k \to \infty$.*

In fact, as the proofs show, the convergence in part 3 of the theorem occurs in finite time from every element of $\mathscr{P}$. Moreover, a uniform upper bound on the number of iterations necessary can be calculated from the primitives. In particular, if we let

$$\bar{s} := \sup\{s \in (0,1] : c'(s) \leq \delta c'(0)\},$$

then $p \in \mathscr{P}$ implies $T^k p = p^*$ whenever $k \geq 1/\bar{s}$.

Two equilibrium price functions are shown in figure 3, computed by iterating with T from initial condition $p = c$. In both cases we use the exponential cost function from example 3.1 with $\theta = 10$. The dashed line corresponds to $\delta = 1.02$, while the solid line is for $\delta = 1.2$. Not surprisingly, when transaction costs rise so do prices.[11]

4.2. **Implications for the Price Function.** One important implication of theorem 4.1 is that p^* is an element of $\mathscr{P}$, and hence by definition is increasing, convex and satisfies certain other properties. Following essentially the same procedure used in the literature on recursive economic dynamics, we can extend and strengthen

[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.

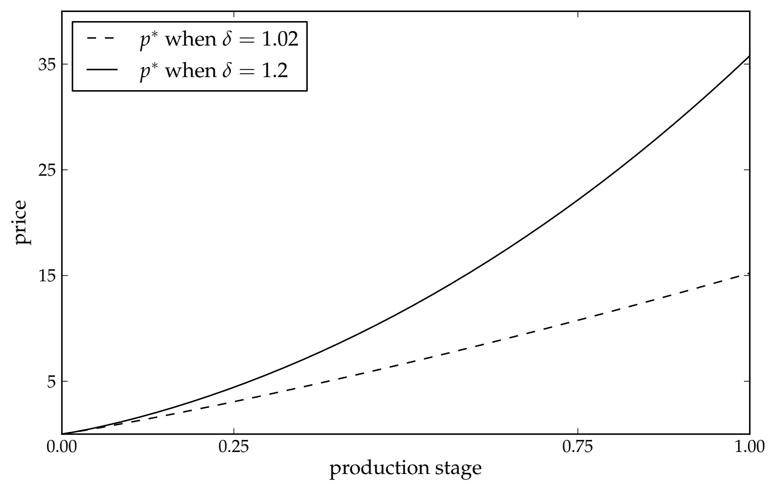


FIGURE 3. Equilibrium price functions when $c(\ell) = e^{\theta \ell} - 1$

these results. For example, continuous differentiability of p can be established by a simple variation on the approach routinely used to prove differentiability of the value function [6]. The next proposition summarizes what we know.

Proposition 4.2. *The equilibrium price function p^* is the unique function in $\mathscr{P}$ satisfying*

$$p^*(s) = \min_{t \leq s} \{c(s-t) + \delta p^*(t)\} \quad \text{for all} \quad s \in [0,1]. \tag{6}$$

In addition, p^* is strictly convex, strictly increasing, continuously differentiable, and satisfies $c'(0)s \leq p^*(s) \leq c(s)$ for all $s \in [0,1]$.

The first statement in the proposition is that the equilibrium price function defined in (2) satisfies—and in fact is essentially defined by—the recursive expression in (6). To interpret this expression, recall from section 2.1 that each firm treats its downstream boundary s as given, while its upstream boundary t is a choice variable. Since t appears only in the firm's costs, the problem of profit maximization is equivalent to cost minimization. If t is chosen to minimize costs, then profits are $p(s) - \min_{t \leq s} \{c(s-t) + \delta p(t)\}$. Thus (6) tells us that a firm with downstream boundary s makes exactly zero profits when it chooses its upstream boundary optimally. Since p^* is continuous and strictly convex, as is c, the set of upstream boundaries that attains zero profits for the firm will be a singleton, and buying at any other location will incur a loss.[12]

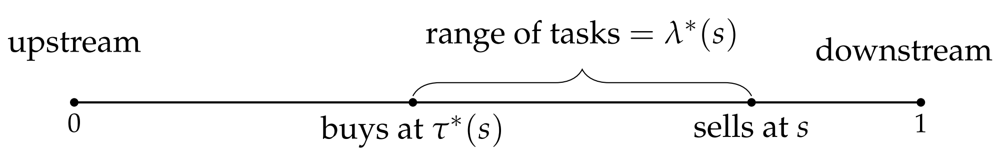
[12] Equation (6) also helps us to understand why it matters little whether we place transaction costs on just the buyer side, just the seller side, or both. For example, suppose now that, in addition to

Regarding the second statement in proposition 4.2, the fact that p^* is strictly increasing is not surprising, giving that c is itself strictly increasing. The strict convexity of p^* is more subtle, and is discussed in detail in section 5.1. The intuition behind the upper bound $p^*(s) \leq c(s)$ is that a single firm can always implement the entire process up to stage s, at cost $c(s)$. This is an upper bound on what the equilibrium allocation will deliver. The intuition behind the lower bound $p^*(s) \geq sc'(0)$ is as follows: Observe that since c is strictly convex, average cost per task $c(\ell)/\ell$ decreases as ℓ gets smaller. If we consider the situation when transaction costs are zero (i.e., $\delta = 1$), this property will encourage firms to enter without limit. If there are n firms involved in producing up to stage s, each producing the small quantity s/n, then, since $\delta = 1$, the aggregate cost of producing to stage s is just $n\, c(s/n)$, and $n\, c(s/n) = s\, c(s/n)/(s/n) \to sc'(0)$ as $n \to \infty$. Hence $sc'(0)$ can be viewed as a lower bound for costs and therefore prices.

4.3. **A Recursive Approach to Allocations.** The equilibrium allocation $\boldsymbol{\ell}^*$ can also be calculated using recursive methods. To do so, we introduce the "choice" functions

$$\tau^*(s) := \operatorname*{arg\,min}_{t \leq s} \{c(s-t) + \delta p^*(t)\} \quad \text{and} \quad \lambda^*(s) := s - \tau^*(s). \tag{7}$$

By definition, $\tau^*(s)$ is the cost minimizing upstream boundary for a firm that is contracted to deliver the good at stage s and faces the equilibrium price function p^*. Similarly, $\lambda^*(s)$ is the cost minimizing range of in-house tasks given p^*. The next figure illustrates:



That these functions are well-defined and single-valued follows from the convexity and continuity results on p^* established in proposition 4.2. The function λ^* is plotted for $\delta = 1.2$ and $\delta = 1.02$ in figure 4. Other parameters are the same as for figure 3.

---

the existing buyer side transaction cost, the seller face 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 $t \leq s$ and setting profits to zero yields $p(s) = \min_{t \leq s}\{c(s-t)\gamma^{-1} + \delta\gamma^{-1} p(t)\}$, which is analogous to (6). 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.

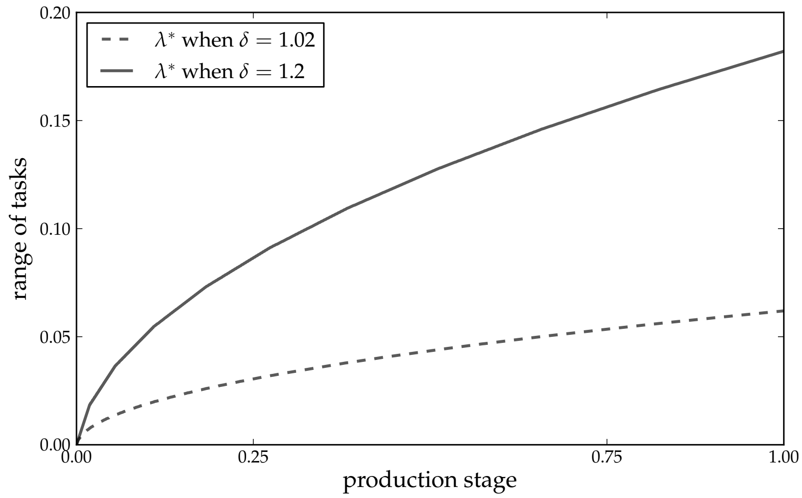


FIGURE 4. Optimal range of in-house tasks as a function of s

We see that $\lambda^*(s)$ increases with production stage s, and that a rise in transaction costs leads to a pointwise increase in $\lambda^*(s)$, so that larger transaction costs means larger firms at all stages of the production chain. We return to both points later on.

Consider the following recursive approach to constructing an efficient production chain. By definition, firm 1 sells the completed good at stage $s = 1$. Hence its optimal range of tasks is $\lambda^*(1)$, and, assuming a suitable partner is available, it buys at $1 - \lambda^*(1)$. By definition this is equal to $\tau^*(1)$. In other words, $\tau^*(1)$ is the optimal upstream boundary for firm 1. If we repeat this process for firm 2, we find that firm 2's optimal range of tasks is $\lambda^*(\tau^*(1))$, and its optimal upstream boundary is $\tau^*(\tau^*(1))$. Continuing in this way produces an allocation of tasks across firms. Formally, the allocation is defined by

$$t_0 = 1, \quad t_i := \tau^*(t_{i-1}) \quad \text{and} \quad \ell_i := \lambda^*(t_{i-1}). \tag{8}$$

It remains however to determine the relationship between this allocation and the equilibrium allocation $\boldsymbol{\ell}^*$. Not surprisingly, it turns out that they are the same. The next proposition has the details.

Proposition 4.3. *The allocation $\{\ell_i\}$ defined recursively in (8) is identical to $\boldsymbol{\ell}^*$.*

In other words, implementing the recursive procedure described above produces the equilibrium allocation of firms. Figure 5 shows firm boundaries computed in this way. The top subfigure shows the price function and firm boundaries when

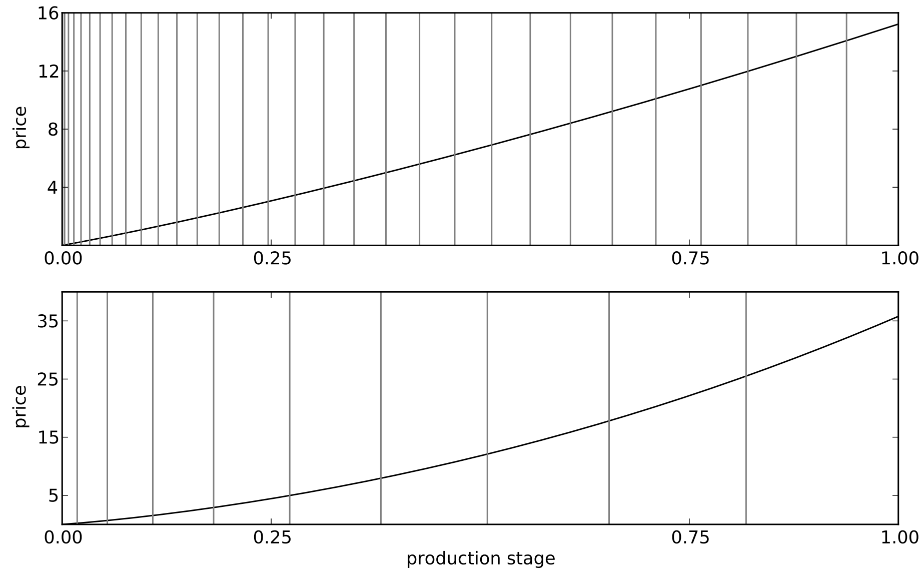


FIGURE 5. Firm boundaries for $\delta = 1.02$ (top) and $\delta = 1.2$ (bottom)

$\delta = 1.02$ and c is as in the previous figures. The boundaries are represented by vertical bars. The bottom subfigure shows the same information when $\delta = 1.2$.

4.4. **Marginal Conditions.** We can develop some additional insights on the behavior of firms by examining the marginal conditions associated with the equilibrium. For example, note that, in view of the convexity and differentiability results obtained in proposition 4.2, interior solutions to (7) are characterized by the first order condition

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

This condition formalizes a fundamental marginal condition that formed the essence of Coase's theory of the firm, and has provided intuition for much of the subsequent research. 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, [10, p. 395]).

The next theorem collects some useful results about τ^* and λ^*.

Theorem 4.4. *Both τ^* and λ^* are increasing and Lipschitz continuous everywhere on $[0, 1]$. Moreover, for $s \in (0, 1)$, the derivative of p^* satisfies*

$$(p^*)'(s) = c'(\lambda^*(s)). \tag{10}$$

Equation (10) follows from $p^*(s) = \min_{t \le s} \{c(s-t) + \delta p^*(t)\}$ and the envelope theorem. The monotonicity of τ^* and λ^* is not entirely obvious from our assumptions. It implies that firms will make different choices depending on their location in the production chain. We return to this point below.

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. For completeness we state and prove these results in section 7 (see proposition 7.12).

5. The Distribution of Firms

In this section we study the implications of the model for the structure of industry equilibrium and the size distribution of firms.

5.1. **The Structure of the Equilibrium.** One aspect of the equilibrium evident in figure 5 is that, in terms of the range of tasks that they implement, downstream firms are larger than upstream firms. This is a robust feature of the equilibrium, as can be seen from (3). Indeed, since c' is increasing and $\delta > 1$, for (3) to hold we must always have $\ell^*_{i+1} < \ell^*_i$ for any pair of adjacent active firms i and $i+1$. In fact, as discussed below, the same monotone relationship holds for other, more traditional, measures of firm size.

To gain a better understanding of this monotone relationship in terms of the choice problem faced by firms, consider the first order condition (9), which can be expressed here by saying that a firm with downstream boundary s chooses ℓ to solve $\delta (p^*)'(s - \ell) = c'(\ell)$. By proposition 4.2, the equilibrium price function p^* is strictly convex, and hence $(p^*)'(s - \ell)$ is strictly increasing in s. It follows that the optimal choice of ℓ is also strictly increasing in s. (*Strict* convexity of p^* is important here because if p^* was linear then the optimal choice of ℓ wouldn't depend on s.) Larger values of s are closer to the final good. Hence the fact that downstream firms are larger than upstream firms is a direct consequence of the strict convexity of p^*.

Why is the equilibrium price function always strictly convex? On the surface, strict convexity is driven by strictly diminishing returns to management, which implies that costs and hence prices rise at a rate greater than proportional to the range of tasks accomplished. However, division of tasks between firms mitigates diminishing returns to management—in effect, substituting market coordination for costly in-house coordination. In the extreme case where transaction costs are absent, firms will enter without limit, with each successive subdivision of task further reducing industry-wide production costs. In fact the limiting price function as $\delta \to 1$ is linear in s, as described in section 4.2. On the other hand, provided that $\delta > 1$, this limit is not attained, and p^* is strictly convex.

To summarize, strictly diminishing returns to management and strictly positive transaction costs together imply that the full "social" cost of production (i.e., the combination of direct production costs and transaction costs) to a given stage increases in more than proportion to the range of tasks, which in turn drives the strict convexity of p^* we observe in equilibrium, and hence the rising *marginal* cost of sourcing through the market as we move from upstream to downstream. As a result, downstream firms choose to implement a large range of tasks in house. This in turn generates the monotone relationship $\ell^*_{i+1} < \ell^*_i$ between the i-th firm and its upstream partner.

5.2. **Other Measures of Firm Size.** The same monotone relationship between upstream and downstream firms observed in the sequence $\{\ell^*_i\}$ also holds for other, more tangible, measures of firm size. To begin, consider value added, which for firm i is given by $v_i := p^*(t_{i-1}) - p^*(t_i)$. 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 make a prediction regarding firm size in terms of employees, we need to move beyond our reduced form specification of costs. As a simple case, let m be the number of employees in a given firm, and suppose that the production function in terms of range of tasks is $\ell = f(m)$ where f is strictly increasing and strictly concave with $f(0) = 0$. It follows that, for a given range of tasks ℓ, the required number of employees is $m = f^{-1}(\ell)$. For the sake of simplicity, suppose that costs are proportional to employees. That is, $c(\ell) = wm = wf^{-1}(\ell)$ for some wage rate w. It is straightforward to verify that c satisfies the assumptions we imposed on it above. In this setting, the number of employees in firm i is just $m_i = c(\ell_i)/w$. Strict monotonicity in the sequence $\{m_i\}$ is then inherited from the sequence $\{\ell_i\}$.[13]

5.3. **The Firm Size Distribution.** It is clear from the proceeding discussion that the model considered in this paper generates a nondegenerate firm size distribution, whether firm size is measured by range of tasks, value added or employment. The same is true of revenue. It remains to see how realistic these distributions are relative to what economists have learned from the data. In this section we recall some of the most important facts concerning the firm size distribution and compare them to the predictions from the model.

Several observed features of the firm size distribution have remained stable over time and space. For example, it has been well documented that firm size distributions are almost always heavily skewed, with a long right tail. Typically the median is significantly larger than the mean. Moreover, this is true regardless of whether size is measured by employment, assets or revenue, and independent of whether we consider individual industries or aggregate data [40, 3, 38, 41, 22, 26].

The firm size distribution generated by our model matches these basic features. For example, figure 6 shows the distribution of firms by value added (top), employment (middle) and revenue (bottom) when $c(\ell) = \exp(10\ell^2) - 1$ and $\delta = 1.01$. The equilibrium number of firms under this parameterization is 336. (In the calculation of employment we adopt the set up discussed in the previous section and set $w = 1$.) Skewness and a long right tail are clearly evident in figure 6. The mean is more than four times the median in all of these measures at the same parameter set, while Pearson's skewness coefficient is close to 2. Alternative parameterizations change these values but produce qualitatively similar results.

It is also of interest to compare the tail properties of these distributions with those found in the data. The tail of the observed firm size distribution is most often modeled either with some variation on the lognormal distribution [18, 9, 22] or a power law [40, 3, 32, 33]. Recent work has shown that tail properties vary significantly across industries and countries, albeit with some interesting regularities [38, 22]. For manufacturing firms, the size of the right hand tail is typically smaller than would be suggested by the lognormal distribution [41, 22]. This can be seen in the Zipf plot in figure 7, which compares size against rank for the set of North American

[13] While it is beyond the scope of the present paper to investigate these predictions empirically, we note that indices of downstreamness do exist in the literature. See, for example, [13].

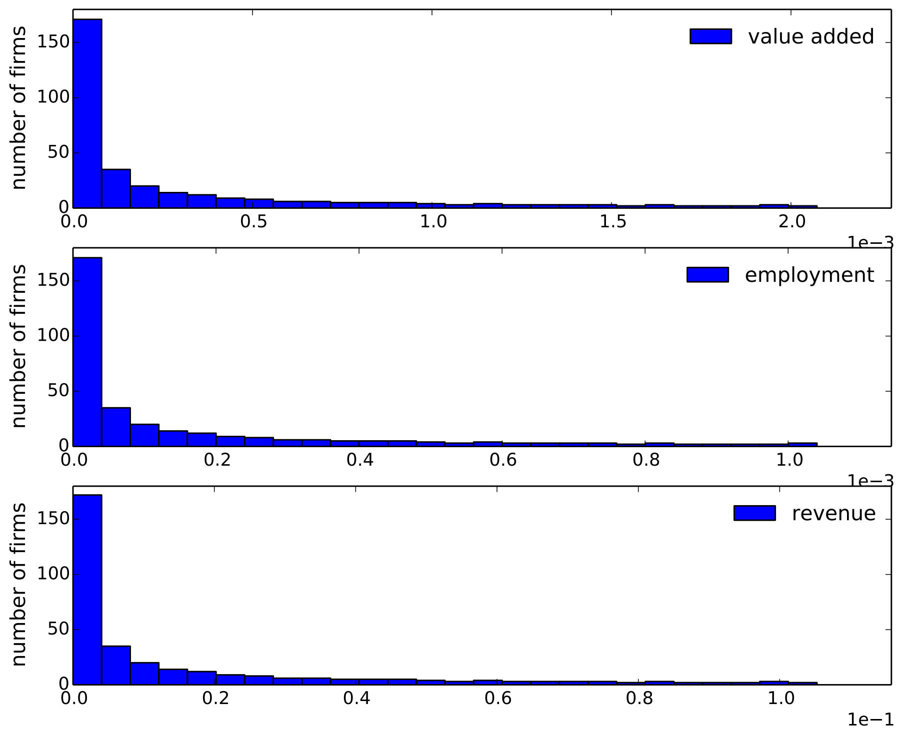


FIGURE 6. Distribution of firms sizes

manufacturing firms reporting positive revenue in 2012 in the Compustat database. The vertical axis is revenue in 2012 US dollars and the horizontal axis is the rank of the firm reporting that revenue (rank i indicates i-th largest firm). Both axes are log scaled. The lognormal approximation shows samples from a lognormal distribution fitted to the firm data using maximum likelihood.[14]

In figure 8 the same figure is generated, but now for data simulated from our model. The parameters are the same as for figure 6. Thus the data is identical to the last panel of figure 6, except that it is rescaled so that the largest observation has the same value as the largest observation in the empirical data set. This scaling has no effect on the shape of the Zipf plot. Comparing figures 7 and 8, we see that the data sets share a number of characteristics, being close to lognormal for medium to

[14]Here manufacturing firms are those with Standard Industrial Classification codes in the range 2000–3999. We include all firms reporting positive revenue in the sample, a total of 2,118 firms

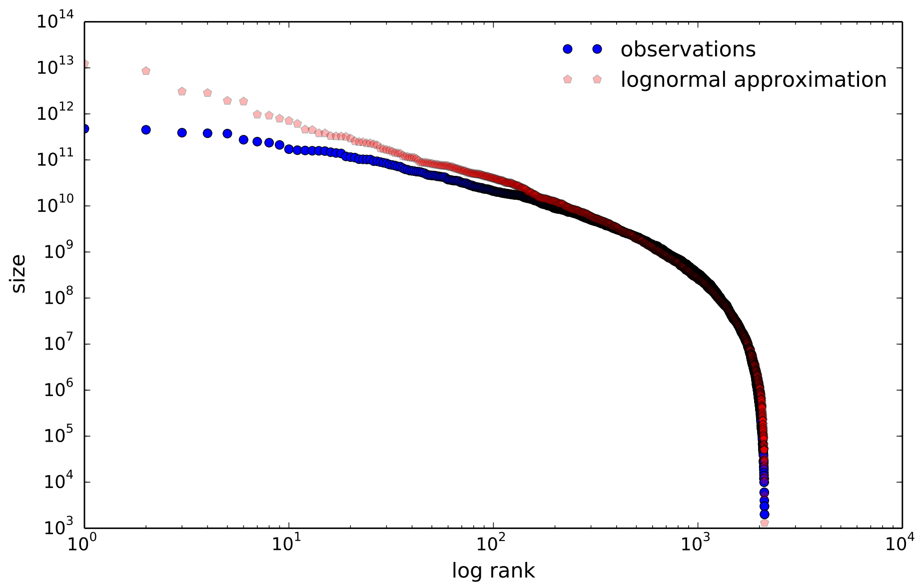


FIGURE 7. Empirical Zipf plot for SIC 2000–3999 firms

smaller firms, but with the largest firms about one order of magnitude smaller than predicted by the fitted lognormal.[15]

6. CONCLUSION

This paper examines some fundamental ideas from the theory of the firm and their aggregate implications in terms of the structure of production and the firm size distribution. Despite the lack of intrinsic heterogeneity or exogenous shocks, the model generates an equilibrium firm size distribution that replicates a number of stylized facts. The distribution is nondegenerate, with a multitude of small firms and relatively few large firms. The distribution is highly skewed, with a long right-hand tail. The distribution as a whole is well represented by a gradually truncated lognormal distribution. This fits well with past observations and data from the manufacturing sector.

The paper also connects three important strands of the economics literature. First, we connect the theory of the firm size distribution with the modern theory of the

[15]Further improvements in fit would presumably require a model with additional realistic features, such a varying capital intensities, plus some understanding of horizontal integration. The significance of human and physical capital intensities for firm size distributions was emphasized by Rossi-Hansberg and Wright [38].

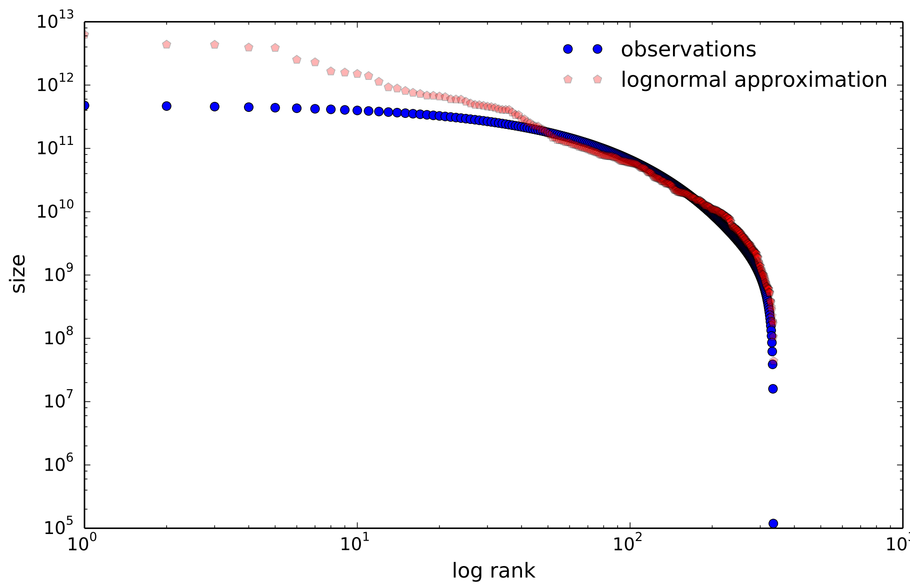


FIGURE 8. Zipf plot for simulated data

optimal size of business firms. Second, by framing decision problems of firms in a recursive manner, where each successive firm faces the same decision as the previous firm in the chain but on a reduced subset of the stages of production, we connect the equilibrium problem with the theory of recursive methods for economic dynamics. This allows us to exploit a variety of standard recursive methods to compute and investigate the structure of the equilibrium.

As a byproduct of the analysis of the firm size distribution, we also recover Coase's key insight on the boundaries of firms, but now in the form of a first order condition rather than a purely verbal description. This holds out the possibility that the trade-off associated with the make-or-buy decision can be investigated from a new perspective, where the equilibrium impact of changes to policy or technology show up in the marginal costs and benefits of integration as quantified by the first order condition. The model also generates a strong prediction regarding the relationship between upstream and downstream firms. In particular, firm size increases with downstreamness as a result of increasing marginal costs of transactions as the semi-processed good moves from upstream to downstream.

The recursive techniques for characterizing and computing equilibrium introduced in this paper are likely to have applications in other fields involving sequential production, such as international trade, or production chains with failure probabilities and other frictions. In addition, the model presented above is a baseline model in all dimensions, with perfect competition, perfect information, identical firms and homogeneous tasks. These assumptions can be weakened. The effect of altering contract structures and supply possibilities could also be investigated. For example, it might be possible to develop a recursive approach to generalizations including multiple upstream partners. Such topics are left to future research.

7. PROOFS

This section collects all proofs. We start by focusing on the properties of the operator T defined in (5). Existence of a fixed point in $\mathscr{P}$ can be established using well-known order-theoretic or topological fixed point results. Here we pursue a more direct proof, which simultaneously yields 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 $0 \leq s_1 < s_2 \leq 1$ and any $\alpha \in (0,1)$. Let $t_i := \arg\min_{t \leq s_i} \{\delta p(t) + c(s_i - t)\}$ for $i = 1, 2$, and $t_3 := \alpha t_1 + (1-\alpha) t_2$. Given that $t_i \leq s_i$ we have $0 \leq t_3 \leq \alpha s_1 + (1-\alpha) s_2$, and hence

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

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*

$$t_p(s) := \arg\min_{t \leq s} \{\delta p(t) + c(s-t)\} \quad \text{and} \quad \ell_p(s) := s - t_p(s). \tag{11}$$

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

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

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 $t_p(s_1) \leq t_p(s_2)$, let $t_i := t_p(s_i)$. Suppose instead that $t_1 > t_2$. We aim to show that, in this case,

$$\delta p(t_1) + c(s_2 - t_1) < \delta p(t_2) + c(s_2 - t_2), \tag{12}$$

which contradicts the definition of t_2.[16] To establish (12), 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

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

Combining this with the last inequality yields (12).

Next we show that $\ell_1 \leq \ell_2$, where $\ell_1 := \ell_p(s_1)$ and $\ell_2 := \ell_p(s_2)$. In other words, $\ell_i = \arg\min_{\ell \leq s_i} \{\delta p(s_i - \ell) + c(\ell)\}$. 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 $0 \leq \ell_2 < \ell_1 \leq s_1$, 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

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

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.[17]

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

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

To complete the proof of lemma 7.2, we also need to show that $t_p(s_2) - t_p(s_1) \leq s_2 - s_1$, 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, $\ell_p(s_1) \leq \ell_p(s_2)$, so $t_p(s_2) - t_p(s_1) \leq s_2 - s_1$, 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 4.1, 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 $s \leq \bar{s}$ is to implement all stages up to s (i.e., to start at stage 0).

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

Proof. First suppose that $s \leq \bar{s}$. 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 $p(t) \geq c'(0)t$ and hence $\delta p(t) \geq \delta c'(0)t = c'(\bar{s})t$. Since $s \leq \bar{s}$, this implies that $\delta p(t) \geq c'(s)t$. Combining these inequalities gives $c'(s)t + c(s-t) < c(s)$, contradicting convexity of c.

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

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

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

Lemma 7.4. *Let $p \in \mathscr{P}$ and let ℓ_p be as in (11). If $s \geq \bar{s}$, then $\ell_p(s) \geq \bar{s}$. If $s > 0$, then $\ell_p(s) > 0$.*

Proof. By lemma 7.2, ℓ_p is increasing, and hence if $\bar{s} \leq s \leq 1$, then $\ell_p(s) \geq \ell_p(\bar{s}) = \bar{s} - t_p(\bar{s}) = \bar{s}$. By lemma 7.3, if $0 < s \leq \bar{s}$, 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 (11). Fix $s_0 \in (0,1)$. By [6], 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 exhibit such a function, 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 $w(s) \geq Tp(s)$ when $s \in U$, observe that if $s \in U$ then $0 \leq t_p(s_0) \leq s$, and

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

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* (11)*. 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* (5) *maps $\mathscr{P}$ into itself.*

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

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 (11). 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 $n \geq 1/\bar{s}$.*

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 $s \leq \bar{s}$, 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 $\ell_h(s) := \arg\min_{t \leq s}\{\delta h(t) + c(s-t)\}$ and let $t_h(s) := s - \ell_h(s)$. By lemma 7.4, we have $\ell_h(s) \geq \bar{s}$, and hence $t_h(s) \leq s - \bar{s} \leq (k+1)\bar{s} - \bar{s} \leq k\bar{s}$. 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

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

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

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

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 $n \geq 1/\bar{s}$ 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 $n \geq 1/\bar{s}$. 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 (2). 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 (11). Let p^* be as defined in (2). If s is any point in $[0, 1]$, then*

$$p^*(s) = p(s) = \sum_{i=1}^{\infty} \delta^{i-1} c(\ell_i), \tag{13}$$

where the allocation $\{\ell_i\}$ in (13) 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 $z \leq \bar{s}$. Hence we need only prove that $t_p^n(s) \leq \bar{s}$ 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 $\ell_p(t_p^n(s)) \geq \ell_p(\bar{s}) = \bar{s}$ 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 $i \geq n$ 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 (13).

Now we turn to the first equality in (13). 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 $p^*(s) \leq p(s)$. Thus it remains only to show that $p(s) \leq p^*(s)$ also holds.

To establish this, we will show that our allocation $\{\ell_i\}$ computed from t_p and ℓ_p is the minimizer in (2). For (2), 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

$$\delta^{i-1} c'(\ell_i) = \mu_i + \alpha, \qquad \mu_i \geq 0 \quad \text{and} \quad \mu_i \ell_i = 0 \quad \text{for all} \quad i \in \mathbb{N}. \tag{14}$$

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

$$\delta^{\bar{n}-1}\, c'(\ell_{\bar{n}}) = \delta^{\bar{n}-2}\, c'(\ell_{\bar{n}-1}) = \cdots = \delta\, c'(\ell_2) = c'(\ell_1) = \alpha. \tag{15}$$

Now take any $i \in \{1, \ldots, \bar{n}\}$. Since $\mu_n = 0$, the first equality in (14) follows from (15) 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 $\mu_i = \delta^{i-1} c'(0) - c'(\ell_1) \geq 0$ when $i > \bar{n}$. Since $i > \bar{n}$, it suffices to show that $\delta^{\bar{n}} c'(0) \geq c'(\ell_1)$. In view of (15), this claim is equivalent to $\delta^{\bar{n}} c'(0) \geq \delta^{\bar{n}-1}\, c'(\ell_{\bar{n}})$, or $\delta c'(0) \geq c'(\ell_{\bar{n}})$. Regarding this inequality, recall the definition of $\bar{s}$ as the largest point in $(0, 1]$ satisfying $c'(\bar{s}) \leq \delta c'(0)$. From lemma 7.3 we have $\ell_{\bar{n}} \leq \bar{s}$. Since c' is increasing, we conclude that $\delta c'(0) \geq c'(\ell_{\bar{n}})$.

This completes the proof that the allocation $\{\ell_i\}$ defined in the statement of lemma 7.10 is a minimizer in (2). Hence $p(s) \leq p^*(s)$. We have already shown that the reverse inequality holds, and hence the first equality in (13) 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*

$$p^*(s) := \min \left\{ \sum_{i=1}^{\infty} \delta^{i-1} c(\ell_i) \;:\; \{\ell_i\} \in \mathbb{R}_+^{\infty} \text{ and } \sum_{i=1}^{\infty} \ell_i = s \right\}. \tag{16}$$

If λ^* and τ^* are as defined in (7) then

$$p^*(s) = \min_{t \leq s} \{c(s-t) + \delta p^*(t)\} = c(\lambda^*(s)) + \delta p(\tau^*(s)). \tag{17}$$

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

$$t_0 = s,\ \ell_i = \lambda^*(t_{i-1}) \text{ and } t_i = t_{i-1} - \ell_i \tag{18}$$

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 (16).

Proof. That $p^*(s) = \min_{t \leq s}\{c(s-t) + \delta p^*(t)\}$ follows immediately from lemma 7.10, which tells us that $Tp^* = p^*$. The second equality in (17) is immediate from the definitions of λ^* and τ^*. Moreover, since $p = p^*$ in lemma 7.10, it follows that $\ell_p = \lambda^*$ and $t_p = \tau^*$, 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 (18). As shown in lemma 7.10, this allocation is the minimizer in (16). 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. **Proofs for Main Results.** We are now ready to prove the claims stated in main section of the paper.

Proof of proposition 3.1. All claims in the proposition follow from corollary 7.11. □

Proof of theorem 3.2. Let p^* and $\boldsymbol{\ell}^*$ be as in the statement of the proposition. Let n^* be the smallest $n \in \mathbb{N}$ such that $t_n = 0$ (for existence see corollary 7.11). Let $\{t_i^*\}$ be defined by $t_0^* = 1$ and $t_i^* = t_{i-1}^* - \ell_i^*$. The claim is that $(p^*, \boldsymbol{\ell}^*)$ is the unique equilibrium in $\mathcal{E}$ in the sense of definition 3.1.

That $(p^*, \boldsymbol{\ell}) \in \mathcal{E}$ follows directly from the definitions. That $p^*(0) = 0$ is immediate from (16). That $p^*(t_{i-1}^*) = c(\ell_i^*) + \delta p^*(t_i^*)$ for $i = 1, \ldots, n^*$ is shown in corollary 7.11. Regarding the third part of the definition of equilibrium, let $\boldsymbol{\ell}'$ be any other feasible allocation. By corollary 7.11, the function p^* satisfies $p^*(s) = \min_{t \leq s}\{c(s-t) + \delta p^*(t)\}$ for all s, and hence, at any given i, the inequality $p(t_i') - c(\ell_i') - \delta p(t_{i+1}') \leq 0$ must hold. In fact, since c and p^* are strictly convex (see lemma 7.1), the minimizer in $\min_{t \leq s}\{c(s-t) + \delta p^*(t)\}$ is always unique, and hence $p(t_i') - c(\ell_i') - \delta p(t_{i+1}') < 0$ at any i where the allocations and transaction stages are not equal. Since $\boldsymbol{\ell}' \neq \boldsymbol{\ell}^*$, this must be true at some i. Hence $(p^*, \boldsymbol{\ell})$ is an equilibrium as claimed.

Now let's consider uniqueness. First consider two equilibria in $\mathcal{E}$ with the same price functions p but distinct allocations. By applying condition 3 of the definition to each of the two allocations, it must be that one firm makes negative profits in at least one of these allocations. This contradicts condition 2 of the definition (zero profits). In other words, two equilibria with the same price function must have the same allocation. Hence it suffices to show that if p is an equilibrium price function (i.e., if $(p, \boldsymbol{\ell})$ is an equilibrium for some feasible allocation $\boldsymbol{\ell}$), then $p = p^*$.

To this end, let p be an equilibrium price function. Fix $s \in [0, 1]$. By the definition of equilibrium, p is nontrivial, so there exists an allocation $\boldsymbol{\ell}$ such that $\sum_{i=1}^n \ell_i =$

s and $p(t_{i-1}) \geq c(\ell_i) + \delta p(t_i)$ for $i = 1, \ldots, n$. Iterating on this inequality gives $p(s) \geq c(\ell_1) + \delta c(\ell_2) + \delta^2 c(\ell_3) + \cdots$. It now follows from (16) that $p(s) \geq p^*(s)$. For the reverse inequality, let $\{\ell_i\}$ and $\{t_i\}$ be as defined in (18). 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 (1) holds for all t, s with $t \leq s$. In particular, $p(t_{i-1}) \leq c(\ell_i) + \delta p(t_i)$ for $i = 1, \ldots, n$. Iterating on this inequality from $i = 1$ gives $p(s) = p(t_0) \leq \sum_{i=1}^{n} \delta^{i-1} c(\ell_i)$. In other words, $p(s) \leq p^*(s)$. □

Proof of theorem 4.1. 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. That this fixed point is equal to p^* was shown in lemma 7.10. The claim that $T^n p = p^*$ for all $n \geq 1/\bar{s}$ follows from the preceding result and lemma 7.8. □

Proof of proposition 4.2. The first claim in the proposition follows from the previously established fact that p^* is the unique fixed point of T in $\mathscr{P}$. 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 $(p^*)'(s) = c'(\lambda^*(s))$. Since c and λ^* are continuous, the latter by lemma 7.2, and since $c'(s) > 0$ for all s, the equation $(p^*)'(s) = c'(\lambda^*(s))$ implies that p^* also continuously differentiable and strictly increasing. Finally, the bounds $c'(0)s \leq p^*(s) \leq c(s)$ are immediate from $p^* \in \mathscr{P}$. □

Proof of proposition 4.3. This claim follows directly from corollary 7.11 when we set $s = 1$ in the latter. □

Proof of theorem 4.4. We have already shown that $(p^*)'(s) = c'(\lambda^*(s))$. The claimed properties on τ^* and λ^* are immediate from lemma 7.2. □

7.3. **Comparative Statics.** As discussed in section 4.5, a rise in transaction costs causes prices to rise, the size of firms to increase, and the equilibrium number of firms to fall. Here is the formal statement:

Proposition 7.12. *If $\delta_a \leq \delta_b$, then $p_a^* \leq p_b^*$, $\ell_i^a < \ell_i^b$ for all active firms and $n_b^* \leq n_a^*$.*

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 5 to see a these results in a simulation.

Proof of proposition 7.12. Let $\delta_a \leq \delta_b$. Let T_a and T_b be the corresponding operators. We begin with the claim that $p_a^* \leq p_b^*$. It is easy to verify that if $p \in \mathscr{P}$, then $T_a p \leq T_b p$ pointwise on $[0,1]$. Since T_a and T_b are order preserving (i.e., $p \leq q$ implies $Tp \leq Tq$), this leads to $T_a^n p \leq T_b^n p$. For n sufficiently large, this states that $p_a^* \leq p_b^*$. 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 (3) 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 $i = 1, \ldots, n$. 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 $n_b^* \leq n_a^*$. 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 $\delta_a c'(0) \geq c'(\ell_n^a)$. Moreover, we have just shown that $\ell_n^a \geq \ell_n^b$. Combining these two inequalities and using $\delta_b > \delta_a$, we have $\delta_b c'(0) \geq c'(\ell_n^b)$. Applying lemma 7.3 again, we see that the n-th firm completes the good, and hence $n_b^* = n_a^*$. ☐

Lemma 7.13. *If $c(\ell) = e^{\theta \ell} - 1$, then the equilibrium number of firms is given by* (4).

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 $0 < \ell_n \leq r$. Therefore $(n-1)r < \ell_1 \leq nr$. 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 $-1 + \sqrt{1 + 8/r} < 2n \leq 1 + \sqrt{1 + 8/r}$. The expression for $n = n^*$ in (4) follows. ☐

References

[1] Philippe Aghion and Richard Holden. Incomplete contracts and the theory of the firm: What have we learned over the past 25 years? *The Journal of Economic Perspectives*, 25(2):181–197, 2011.

[2] Kenneth J Arrow. *The organization of economic activity*, volume 1, pages 47–64. US Govt Printing Office Washington, DC, 1969.

[3] Robert L Axtell. Zipf distribution of US firm sizes. *Science*, 293(5536):1818–1820, 2001.

[4] Gary S Becker and Kevin M Murphy. The division of labor, coordination costs, and knowledge. *The Quarterly Journal of Economics*, 107(4):1137–1160, 1992.

[5] Jess Benhabib, Alberto Bisin, and Shenghao Zhu. The distribution of wealth and fiscal policy in economies with finitely lived agents. *Econometrica*, 79(1):123–157, 2011.

[6] Lawrence M Benveniste and Jose A Scheinkman. On the differentiability of the value function in dynamic models of economics. *Econometrica*, 47(3):727–732, 1979.

[7] Lawrence E Blume, David Easley, Jon Kleinberg, and Eva Tardos. Trading networks with price-setting agents. *Games and Economic Behavior*, 67(1):36–50, 2009.

[8] Timothy F Bresnahan and Jonathan D Levin. Vertical integration and market structure. Technical report, National Bureau of Economic Research, 2012.

[9] Luis Cabral and Jose Mata. On the evolution of the firm size distribution: Facts and theory. *American Economic Review*, 93(4):1075–1090, 2003.

[10] Ronald H Coase. The nature of the firm. *Economica*, 4(16):386–405, 1937.

[11] Arnaud Costinot. On the origins of comparative advantage. *Journal of International Economics*, 77(2):255–264, 2009.

[12] Arnaud Costinot, Jonathan Vogel, and Su Wang. An elementary theory of global supply chains. *The Review of Economic Studies*, 80(1):109–144, 2013.

[13] Thibault Fally. On the fragmentation of production in the US. University of Colorado, mimeo, 2011.

[14] Xavier Gabaix. Zipf's law for cities: an explanation. *Quarterly Journal of Economics*, 114(3):739–767, 1999.

[15] John Geanakoplos and Paul Milgrom. A theory of hierarchies based on limited managerial attention. *Journal of the Japanese and International Economies*, 5(3):205–225, 1991.

[16] Christian Ghiglino. Random walk to innovation: why productivity follows a power law. *Journal of Economic Theory*, 147(2):713–737, 2012.

[17] Robert Gibbons. Four formal (izable) theories of the firm? *Journal of Economic Behavior & Organization*, 58(2):200–245, 2005.

[18] Robert Gibrat. *Les inégalités économiques*. Recueil Sirey, 1931.

[19] Gene M Grossman and Elhanan Helpman. Integration versus outsourcing in industry equilibrium. *The Quarterly Journal of Economics*, 117(1):85–120, 2002.

[20] Gene M Grossman and Esteban Rossi-Hansberg. Task trade between similar countries. *Econometrica*, 80(2):593–629, 2012.

[21] Sanford J Grossman and Oliver D Hart. The costs and benefits of ownership: A theory of vertical and lateral integration. *The Journal of Political Economy*, 94(4):691–719, 1986.

[22] Hari M Gupta, José R Campanha, Daniela R de Aguiar, Gabriel A Queiroz, and Charu G Raheja. Gradually truncated log-normal in USA publicly traded firm size distribution. *Physica A*, 375(2):643–650, 2007.

[23] Oliver Hart and John Moore. Property rights and the nature of the firm. *Journal of Political Economy*, 98(6):1119–1158, 1990.

[24] Friedrich A Hayek. The use of knowledge in society. *The American Economic Review*, 35(4):519–530, 1945.

[25] Bengt Holmstrom and Paul Milgrom. The firm as an incentive system. *The American Economic Review*, 84(4):972–991, 1994.

[26] Yuji Ijiri and Herbert A Simon. Interpretations of departures from the Pareto curve firm-size distributions. *The Journal of Political Economy*, 82(2):315–331, 1974.

[27] Michael C Jensen and William H Meckling. Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics*, 3(4):305–360, 1976.

[28] Benjamin Klein, Robert G Crawford, and Armen A Alchian. Vertical integration, appropriable rents, and the competitive contracting process. *Journal of Law and Economics*, pages 297–326, 1978.

[29] Francine Lafontaine and Margaret Slade. Vertical integration and firm boundaries: the evidence. *Journal of Economic Literature*, 45(3):629–685, 2007.

[30] Lars Ljungqvist and Thomas J Sargent. *Recursive macroeconomic theory*. 3 edition, 2012.

[31] Robert E Lucas. On the size distribution of business firms. *The Bell Journal of Economics*, 9(2):508–523, 1978.

[32] Erzo GJ Luttmer. Selection, growth, and the size distribution of firms. *The Quarterly Journal of Economics*, 122(3):1103–1144, 2007.

[33] Erzo GJ Luttmer. Technology diffusion and growth. *Journal of Economic Theory*, 147(2):602–622, 2012.

[34] R Preston McAfee and John McMillan. Organizational diseconomies of scale. *Journal of Economics & Management Strategy*, 4(3):399–426, 1995.

[35] Kieron J Meagher. Generalizing incentives and loss of control in an optimal hierarchy: the role of information technology. *Economics Letters*, 78(2):273–280, 2003.

[36] Paul Milgrom and John Roberts. *Economics, organization and management*. Prentice-Hall International, 1992.

[37] Douglass C North. Institutions, transaction costs and productivity in the long run. *Washington University in St. Louis, mimeo*, 1993.

[38] Esteban Rossi-Hansberg and Mark LJ Wright. Establishment size dynamics in the aggregate economy. *The American Economic Review*, 97(5):1639–1666, 2007.

[39] Patrick W Schmitz. Information gathering, transaction costs, and the property rights approach. *American Economic Review*, 96(1):422–434, 2006.

[40] Herbert A Simon and Charles P Bonini. The size distribution of business firms. *The American Economic Review*, 48(4):607–617, 1958.

[41] Michael Stanley, Sergey V Buldyrev, Shlomo Havlin, Rosario N Mantegna, Michael A Salinger, and H Eugene Stanley. Zipf plots and the size distribution of firms. *Economics*

Letters, 49(4):453–457, 1995.

[42] Nancy Stokey and Robert E Lucas. *Recursive Methods in Economic Dynamics (with EC Prescott)*. Harvard University Press, 1989.

[43] Gabriel Y Weintraub, C Lanier Benkard, and Benjamin Van Roy. Industry dynamics: Foundations for models with an infinite number of firms. *Journal of Economic Theory*, 146(5):1965–1994, 2011.

[44] Oliver E Williamson. Transaction-cost economics: the governance of contractual relations. *Journal of Law and Economics*, 22(2):233–261, 1979.