# KIER DISCUSSION PAPER SERIES

## KYOTO INSTITUTE
## OF
## ECONOMIC RESEARCH

Discussion Paper No.828


“Coase meets Tarski: New Insights from Coase’s Theory of the Firm”


Tomoo Kikuchi, Kazuo Nishimura and John Stachurski



August 2012

KYOTO UNIVERSITY

KYOTO, JAPAN

# Coase meets Tarski: New Insights from Coase's Theory of the Firm*

Tomoo Kikuchi,[†] Kazuo Nishimura[‡] and John Stachurski[§]

August 27, 2012

**Abstract**

This paper formulates a model embedding the key ideas from Ronald Coase's famous essay on the theory of the firm in a simple competitive equilibrium setting with an arbitrary number of firms. The model studies the structure of production when transaction costs and diminishing returns to management are treated as given. In addition to recovering Coase's main insights as equilibrium conditions, the model yields many new predictions on prices, firm boundaries and division of the value chain.

**Keywords:** Transaction costs, vertical integration, production chains
**JEL Classification:** D02, D21, L11, L23

## 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

*The authors are indebted to Chris Jones, Takashi Kamihigashi, Pedro Gomis Porqueras and Makoto Yano for helpful comments and discussions, and to Alex Olssen for excellent research assistance.

†Department of Economics, National University of Singapore. ecstk@nus.edu.sg

‡Institute of Economic Research, Kyoto University. nishimura@kier.kyoto-u.ac.jp

§(Corresponding author) Research School of Economics, The Australian National University. john.stachurski@anu.edu.au

about it, astonishingly hard to believe." Indeed, the ability of market forces to coordinate many specialized activities and channel their collective output into the final goods demanded by consumers has fascinated economists for centuries. In considering this phenomenon in his celebrated essay on the nature of the firm, Ronald Coase (1937) pushed the analysis in a striking new direction: Given the efficiency of market coordination, combined with the fact that planned interactions within firms can, at least in principle, also be coordinated through the market, why, asked Coase, do firms exist at all? What need is there for these "islands of conscious power in the ocean of unconscious cooperation"?[1]

These islands may be vast or small. 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 more than the president of the United States. In the same year, the total number of employees at Wal Mart exceeded the population of all but 4 US cities. In addition to such giants, tens of millions of smaller firms operate around the world.[2] What accounts for the existence of the giant firms, and the multitude of smaller ones? What forces shape their individual sizes and their distribution? And if the market provides an efficient substitute for intra-firm coordination, then why are some CEO salaries measured in tens of millions?

The modern approach to understanding these phenomena begins with Coase (1937), who sought to provide a theory of the firm that was both realistic and tractable "by two of the most powerful instruments of economic analysis developed by Marshall: the idea of margin and that of substitution, together giving the idea of substitution at the margin." He argued that for firms to exist there must be costs associated with using the market (i.e., transaction costs), and that entrepreneurs and managers must be able to substitute away from the market by coordinating production at a lower cost within the firm. On the other hand, since firms do not expand without limit, a countervailing force must be present. Coase referred to this force as "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 cost of acquiring a similar input through the market.

Subsequent to Coase's analysis, many economists have expanded on and developed the theory of the firm. Researchers have considered the effect of imperfect information, incentive and agency problems, incomplete contracts, property rights, decision rights, and the microfoundations of transaction costs. Major contributions were made by Jensen and Meckling (1976), Williamson (1979, 1981), Klein, Crawford and Alchian (1978), Grossman and Hart (1986), Hart and Moore (1990), Holmstrom and Milgrom (1994), Grossman and Helpman (2002) and many others. These studies have been immensely valuable in building an understanding of different attributes and functions of the firm. Up to date surveys and references can be found in Aghion and Holden (2011) and Bresnahan and

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

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

Levin (2012).

In this paper our objective is different. Rather than extending or critiquing the underlying ideas in Coase's theory, our aim is to show that Coase's basic framework has a great deal more to give. We begin by embedding the essential features of his verbal analysis in a competitive model with transaction costs and an arbitrary number of firms. By framing the choice problem of firms recursively, we reduce the model to a single functional equation for equilibrium prices. We prove that this functional equation has a unique solution, and that this solution identifies a well-defined disintegration of the value chain coordinated through the price system. The model recovers Coase's original insights on the boundaries of individual firms in the form of equations derived from first order conditions. More importantly, it also allows us to study the implications of Coase's ideas for the structure of production along the value chain, as determined by the interactions between the choices of individual firms and the equilibrium set of prices. These interactions result in a set of additional predictions regarding the relationship between upstream and downstream firms, the equilibrium impact of transaction and management costs and the distribution of firms.

The first version of our model involves sequential production over a continuum of tasks. Economists have used this structure to address a variety of questions. Recent examples include Grossman and Rossi-Hansberg (2008, 2012), Costinot (2009), Antràs and Chor (2011) and Costinot, Vogel and Wang (2011). Perhaps the most similar to our paper is Costinot, Vogel and Wang (2011), which studies global supply chains. Their model is nonetheless different in that firms are heterogeneous, and that the production processes is affected by "mistakes." In our model mistakes are absent, firms are ex ante identical, and all ex post heterogeneity is generated as part of the equilibrium. In addition, our analytical techniques, based on recursive subdivision of tasks, give us a very detailed picture of prices over the continuum, and allow us to consider such extensions as production with multiple upstream partners (section 4).

The rest of the paper is structured as follows. Section 2 describes the model. The equilibrium is analyzed in section 3. Section 4 considers possible extensions, while section 5 looks at implications and predictions. Section 6 concludes.

## 2 The Model

Following Coase (1937), our focus is supply side and partial equilibrium. To start the analysis we consider a single good that is produced through the sequential completion of a large number of tasks. (More complex production structures are treated in section 4.) In order to provide a sharper marginal analysis, we model the processing stages as a continuum from zero to one. For example, at stage 0.9 the good is 90% complete.

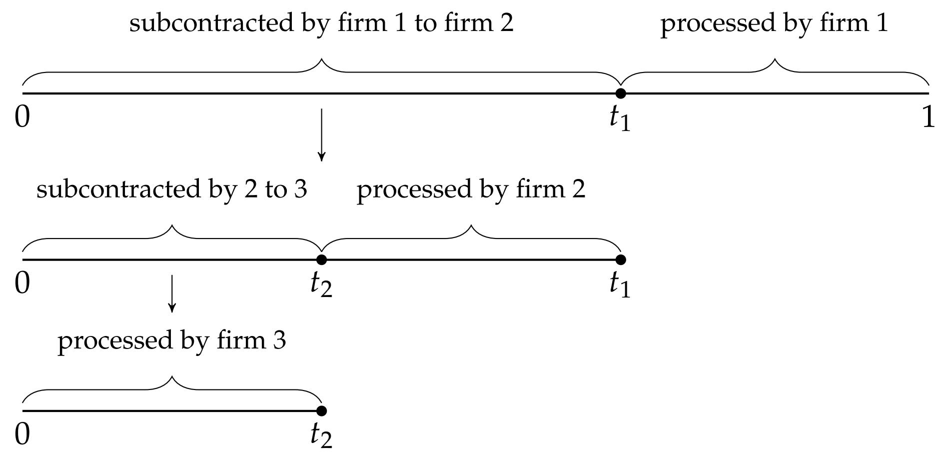



Figure 1: Disintegration of the value chain

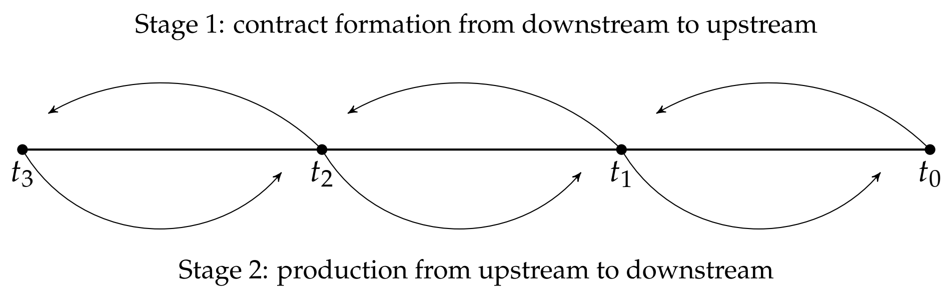



Figure 2: Contracts and Production

## 2.1 The Production Chain

The processing stages are indexed by $t \in [0, 1]$, with $t = 0$ indicating that no tasks have been undertaken and $t = 1$ indicating that production of one unit of the good is complete. Allocation of tasks between firms is determined by subcontracting. The subcontracting scheme is illustrated in figure 1. In this example, 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 good at processing stage $t_1$. 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 $t_3 = 0$.

Figure 1 already suggests the recursive nature of the decision problem for each firm. In choosing how many processing stages to subcontract, each successive firm will face es-

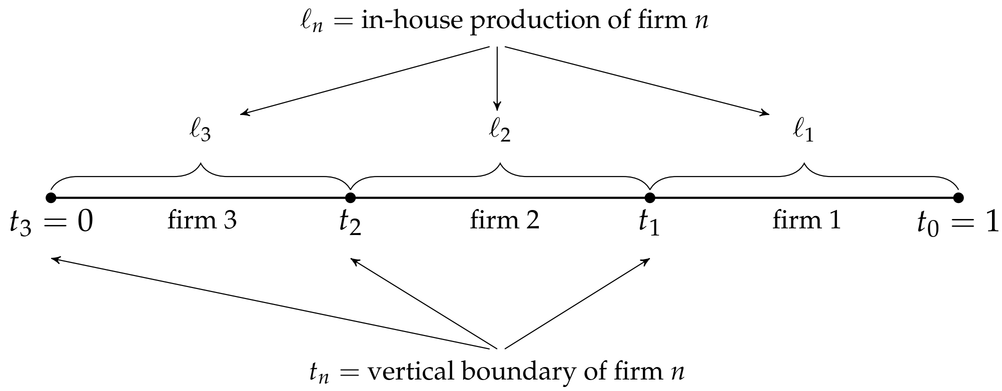



Figure 3: Notation

sentially 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 it. Figure 2 emphasizes the fact that contracts evolve backwards from 1 down to 0 (downstream to upstream), and production then unfolds in the opposite direction. In the present example, 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. Figure 3 serves to clarify notation. The amount of in-house production carried out by firm $n$ is denoted by $\ell_n$. Despite the continuum formulation of the model, we will also refer to $\ell_n$ as the "number" of tasks carried out by firm $n$.

## 2.2 Management and Transaction Costs

We represent the cost of carrying out $\ell$ tasks in-house by $c(\ell)$, where $0 \leq \ell \leq 1$ and $c(0) = 0$. One of the fundamental forces in Coase's theory of the firm is "diminishing returns to management" (Coase, 1937, p. 395). In what follows, diminishing returns to management is incorporated by assuming that $c$ is strictly convex, implying that average cost per task rises with the quantity of tasks performed in-house. For modeling purposes it is not necessary to specify the mechanisms behind diminishing returns to management. However, for the sake of concreteness it might be helpful to think of tasks as being performed by specialized workers, and consider the costs associated with coordinating workers specialized in different tasks (see Becker and Murphy, 1992). Section 5 provides a detailed discussion.[3]

[3] In our model the cost of carrying out $\ell$ tasks $\ell = s - t$ depends only on the difference $s - t$ rather than $s$ directly. In other words, all tasks are homogeneous. Extensions might consider other cases, such as where

Diminishing returns to management favors sourcing inputs externally over in-house production. Absent a countervailing force, the "equilibrium" size of firms may be infintesimally small. In Coase's analysis, the countervailing force is provided by market transaction costs. Section 5 gives a detailed discussion of transaction costs. One example is the cost of negotiating, monitoring and enforcing contracts with suppliers (think of complete but costly contracts). For now we follow Arrow (1969) by simply regarding transaction costs as a wedge between the buyer's and seller's prices. As will be shown in section 4.1, it matters little from a qualitative perspective whether the transaction cost is borne by the buyer, the seller or both. Hence we assume that the cost is borne only by the buyer. In particular, when the partially processed good is purchased at price $p$, the buyer's total outlay is equal to $\delta p$ with $\delta > 1$. The transaction cost $(\delta - 1)p$ is paid to agents outside the model.

Collect all assumptions on $\delta$ and $c$ together, we assume that $\delta > 1$ and $c$ is strictly convex and continuously differentiable with $c(0) = 0$, $c'(0) > 0$ and $\delta c'(0) \leq c'(1)$. The last two conditions are only mildly restrictive and yield tight characterizations of the equilibrium. Since $c$ is convex and $c'(0) > 0$, the function $c$ is strictly increasing.

## 2.3 Profit Maximization

We assume that all firms are ex ante identical and act as price takers. Contracts are complete, information is perfect, and active firms are surrounded by an infinite number of competitive firms ready to step in on either the buyer or the seller side should it be profitable to do so. There are no fixed costs or barriers to entry. As a result, no holdup occurs in our model.

Let $p(t)$ represent the price of the good completed up to stage $t$. (In particular, $p(1)$ is the price of the final completed good.) To begin the process of determining prices, consider a firm who enters a contract to supply the good at stage $s \in (0, 1]$, and purchase the good at stage $t \leq s$. The firm undertakes the remaining $\ell = s - t$ tasks in house, and its total costs are given by the sum of its processing costs $c(s - t)$ and the gross input cost $\delta p(t)$. Recalling that transaction costs are paid only by the buyer, profits are $p(s) - c(s - t) - \delta p(t)$. If $t$ is chosen to minimize costs, then profits become

$$\pi(s) = p(s) - \min_{t \leq s} \{\delta p(t) + c(s - t)\}. \tag{1}$$

Here and below, the restriction $0 \leq t$ in the minimum is understood.

upstream tasks are more routine and hence cheaper. Here our interest is in equilibrium prices and choices of firms in the base case where tasks are ex-ante identical.

# 3 Equilibrium

The model can now be closed by a zero profit and boundary condition. This section describes equilibrium prices and the resulting vertical production structure.

## 3.1 Equilibrium Prices

Competition forces firm profits to zero, implying that $\pi(s)$ in (1) equals zero for all $s \in (0,1]$. This places a restriction on $p(s)$ for $s \in (0,1]$. The remaining value $p(0)$ can be regarded as the revenue of firms that supply the initial inputs to production. We impose a zero profits condition in this sector as well, implying that $p(0)$ is equal to the cost of producing these inputs. To simplify notation, we assume that this cost is zero, and hence the boundary condition is $p(0) = 0$. Formally, we say that a function $p\colon [0,1] \to \mathbb{R}_+$ satisfies the *equilibrium price equation* if both $p(0) = 0$ and

$$p(s) = \inf_{t \leq s} \{\delta p(t) + c(s-t)\} \quad \text{for all} \quad s \in [0,1]. \tag{2}$$

Evidently the zero function is a solution to the equilibrium price equation. A nonzero solution $p^*$ to the equilibrium price equation is called an *equilibrium price function*. Existence of an equilibrium price function can be established via the Knaster-Tarski fixed point theorem. To see this, let $\mathscr{P}$ be the set of functions $p$ from $[0,1]$ to $\mathbb{R}$ such that $c'(0)s \leq p(s) \leq c(s)$ for all $0 \leq s \leq 1$, and let $T$ be the operator defined over $p \in \mathscr{P}$ by

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

It is not hard to verify that $T$ maps $\mathscr{P}$ into itself (see lemma 7.1 in section 7.2). As $c'(0)$ is assumed to be strictly positive, no element of $\mathscr{P}$ is the zero function, and hence any fixed point of $T$ is a (nontrivial) equilibrium price function. While $T$ is not a contraction mapping in any obvious metric, it is monotone increasing on $\mathscr{P}$ (i.e., order preserving) when $\mathscr{P}$ is endowed with the usual partial order ($p \leq q$ if $p(s) \leq q(s)$ for all $s$) because if $p \leq q$, then for any $s \in [0,1]$ we have $\inf_{t \leq s} \{\delta p(t) + c(s-t)\} \leq \inf_{t \leq s} \{\delta q(t) + c(s-t)\}$, and hence $Tp(s) \leq Tq(s)$. Since $\mathscr{P}$ is a complete lattice, the Knaster-Tarski fixed point theorem ensures the existence of a fixed point in $\mathscr{P}$.

Although the proceeding argument yields existence of an equilibrium price function $p^*$, we also wish to know whether the solution is unique over some suitable class of functions, what properties it possesses, and if numerical methods can be found to compute it in a consistent fashion. An equally important question is whether or not, given the solution, production is actually realized. In particular, given the pricing function, it is conceivable that, in the manner of Zeno's paradox, the backwards contracting problem never terminates. For example, if firm $i$ implements $\ell_i = 2^{-i}$ tasks for all $i$, then the $(n+1)$-th firm contracts at stage $t_n = 1 - \sum_{i=1}^{n} \ell_i = 2^{-n}$. Since this value is always strictly positive, no termination occurs in finite time. Such behavior must be ruled out in order to verify existence of a realistic equilibrium.

Regarding the set $\mathscr{P}$, the upper and lower bound functions $s \mapsto c(s)$ and $s \mapsto c'(0)s$ defining $\mathscr{P}$ have a natural interpretation as representing prices when transaction costs are very high and very low respectively. Regarding the upper bound function, it is intuitive that if $\delta$ is very large, then transaction costs will be prohibitively expensive and a single firm will implement the whole process in-house. Their cost at stage $s$ will be $c(s)$. Given zero profits this is equal to the price $p(s)$. Thus, the upper bound function represents the price when transaction costs are very large and only one firm operates. On the other hand, as $\delta \downarrow 1$, strict positivity of $c$ will cause the number of firms to increase without limit, and the amount of in-house production $\ell$ at each firm will decrease to zero. Average costs at each firm will be $c(\ell)/\ell \to c'(0)$, and the price at stage $s$ will be the integral $c'(0)s$. Hence the lower bound function represents the price when transaction costs are arbitrarily low.

## 3.2 Properties of the Solution

Our first result on properties of the solution establishes uniqueness of the equilibrium price function given the two primitives $c$ and $\delta$, and shows that the price function is always strictly convex and continuously differentiable. (Here and below, proofs are deferred until section 7.) In the statement of the next theorem, $\mathscr{P}_0$ is defined to be the set of functions in $\mathscr{P}$ that are convex and $K$-Lipschitz for $K := c'(1)$.[4] Also, given an equilibrium price function $p^*$, we let the optimal choices be defined by

$$t^*(s) := \operatorname*{arg\,min}_{t \leq s} \{\delta p^*(t) + c(s-t)\} \quad \text{and} \quad \ell^*(s) := s - t^*(s), \tag{4}$$

For example, $\ell^*(s)$ is the optimal amount of in-house production for a firm that is contracted to deliver the good at stage $s$.

**Theorem 3.1.** *The equilibrium pricing equation has a unique solution in $\mathscr{P}_0$. The solution $p^*$ is strictly convex, strictly increasing and continuously differentiable, with*

$$(p^*)'(s) = c'(\ell^*(s)) \quad \textit{for all} \quad s \in (0,1). \tag{5}$$

*Moreover, the functions $t^*$ and $\ell^*$ are well-defined and single-valued on $[0,1]$. Both are increasing and $K$-Lipschitz with $K := 1$.*

The first order condition for (4) is

$$\delta(p^*)'(t^*(s)) = c'(s - t^*(s)). \tag{6}$$

[4] $F\colon X \to \mathbb{R}$ is called $K$-Lipschitz if $|F(a) - F(b)| \leq K|a-b|$ for any $a, b \in X$.

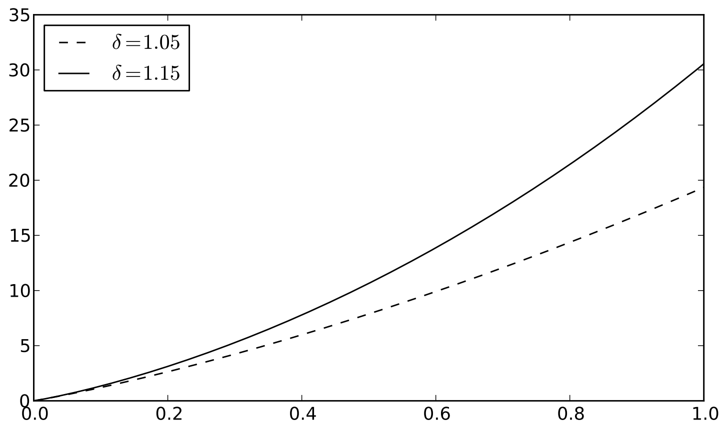



Figure 4: Equilibrium price functions when $c(x) = e^{\theta x} - 1$

Equation (6) is a restatement of the fundamental marginal condition derived verbally by Coase (1937). It gives the point $t^*(s)$ at which the marginal cost of in-house production is equal to the cost of market acquisition. This condition determines the upstream boundary of the firm. Thus, the "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).

The real benefit of representing Coase's theory mathematically is that it allows us to state more. For example, combining (6) with the envelope condition (5) gives an "Euler" equation of the form

$$\delta c'(\ell^*(t^*(s))) = c'(\ell^*(s)). \tag{7}$$

While (6) is a marginal condition for decisions within firms, (7) connects relative costs between any two adjacent firms. It predicts that marginal in-house production cost at a given firm is equal to that of its upstream partner multiplied by gross transaction cost.

A fast, globally convergent algorithm for computing $p^*$ is provided in section 7.1. Two equilibrium price functions are shown in figure 4. For both we set $c(x) = e^{\theta x} - 1$ with $\theta = 10$. The dashed line corresponds to $\delta = 1.05$, while the solid line is for $\delta = 1.15$. The strict convexity of $p^*$ is evident in the figure. This strict convexity is caused not only by strictly diminishing returns to management, but also by positive transaction costs. Intuitively, if transaction costs are zero (i.e., $\delta = 1$), then $p^*(s) = c'(0)s$ (see the last paragraph of section 3.1), which is only weakly convex. Strict convexity arises because the presence of transaction costs prevents firms from eliminating diminishing returns to management by infinite subdivision of tasks across firms. More generally, higher transaction costs cause more curvature in $p^*$ for given $c$, as is evident in figure 4.

## 3.3 Structure of Production

Given the pricing function $p^*$, the equilibrium structure of the vertical production chain introduced in figures 1–3 is determined recursively: Firm 1 receives the order for the final good at price $p^*(1)$. Letting $t^*$ be as defined in (4), firm 1 subcontracts to firm 2 at $t_1 := t^*(1)$, firm 2 subcontracts to firm 3 at $t_2 := t^*(t_1)$, and, in general, firm $n$ subcontracts to firm $n+1$ at

$$t_n := t^*(t_{n-1}) \quad \text{with} \quad t_0 = 1. \tag{8}$$

The states $t_1, t_2, \ldots$ constitute the vertical boundaries of the firms. (For completeness we set $t_0 := 1$. See figure 3.) In-house production by firm $n$ is given by $\ell_n := t_{n-1} - t_n$. Figure 5 shows computed firm boundaries $\{t_n\}$ represented by vertical bars. The cost function is the same as in figure 4, while transaction costs are $\delta = 1.02$ (top) and $\delta = 1.20$ (bottom). The sequence $\{\ell_n\}$ is the distance between adjacent vertical bars in figure 5, with $n$ increasing from right to left. Notice that higher transaction costs lead to a smaller number of larger firms. Note also that, using the definition of $\ell_n$ and (7) we have $\delta c'(\ell_{n+1}) = c'(\ell_n)$. The following result is fundamental:

**Theorem 3.2.** *The production process always completes with a finite number of firms.*

Theorem 3.2 is understood to mean that if $c$ and $\delta$ are taken as given and $\{t_n\}$ is the sequence defined by (8), then there exists an $N^* \in \mathbb{N}$ such that $t_{N^*} = 0$.[5] The integer $N^*$ represents the equilibrium number of firms.

As suggested by figures 4 and 5, when transaction costs rise, prices rise and the equilibrium number of firms decreases. To state this formally, let $\delta_a$ and $\delta_b$ be in $(1, \infty)$, let $p_a^*$ be the equilibrium price function for $\delta_a$, let $N_a^*$ be the equilibrium number of firms, and let $p_b^*$ and $N_b^*$ be defined analogously.

**Proposition 3.1.** *If $\delta_a \leq \delta_b$, then $p_a^* \leq p_b^*$ pointwise on $[0,1]$ and $N_a^* \geq N_b^*$.*

The inverse relationship between transaction costs and $N^*$ can also be seen in the following example, which treats a specific cost function.

**Example 3.1.** Let $c(x) = e^{\theta x} - 1$ with $\theta > 0$. Since $c''(x)/c'(x) = \theta$, this parameter represents curvature, and hence diminishing returns to management. From the condition $\delta c'(\ell_{n+1}) = c'(\ell_n)$ we obtain the recursion $\ell_{n+1} = \ell_n - \ln \delta / \theta$, relating the amount of in-house production by neighboring firms in terms of the two primitives $\delta$ and $\theta$. Using this equation, the constraint $\sum_{n=1}^{N^*} \ell_n = 1$ and some algebra, we can show that the equilibrium number of firms is

$$N^* = \lfloor \left(1 + \sqrt{1 + 8\theta / \ln \delta}\right) / 2 \rfloor. \tag{9}$$

[5] Alternatively, $N^* := \inf\{N \in \mathbb{N} : (t^*)^N(1) = 0\}$.

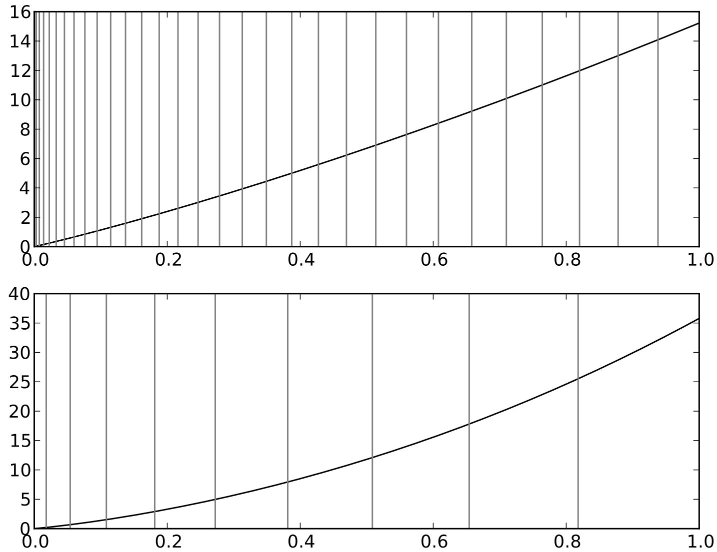



Figure 5: Firm boundaries for $\delta = 1.02$ (top) and $\delta = 1.20$ (bottom)

A proof is given in lemma 7.10 below. Here $\lfloor a \rfloor$ is the largest integer less than or equal to $a$. As expected, the number of firms is increasing in $\theta$ (more rapid diminishing returns to management implies a larger number of relatively small firms, and greater use of the market) and decreasing in $\delta$ (see figure 5).

Let $v_n := p^*(t_{n-1}) - p^*(t_n)$, the value added of firm $n$. The next results on prices and the structure of production follow from theorems 3.1 and 3.2.

**Proposition 3.2.** $p^*(1) = \sum_{i=1}^{N^*} \delta^{i-1} c(\ell_i)$.

**Proposition 3.3.** *$\ell_{n+1} \leq \ell_n$ and $v_{n+1} \leq v_n$ for all $n$ in $1, \ldots, N^* - 1$.*

Proposition 3.2 confirms that in equilibrium the price of the final good is the sum of costs incurred. Proposition 3.3 is more significant. It states that the number of tasks performed in-house is larger the further downstream the firm is in the value chain (as can be seen in figure 5), and the same is true for the amount of value added each firm contributes to the final value of the good. The increase in in-house production as we move downstream is due to the fact that the value of the good increases as more processing stages are completed, and with this increase comes a proportional rise in transaction costs (intuitively, more valuable goods entail more costly contracts). As a result, downstream firms face higher transaction costs. To economize on these costs they produce more in-house. The second inequality in proposition 3.3 follows from the first and convexity of $p^*$.

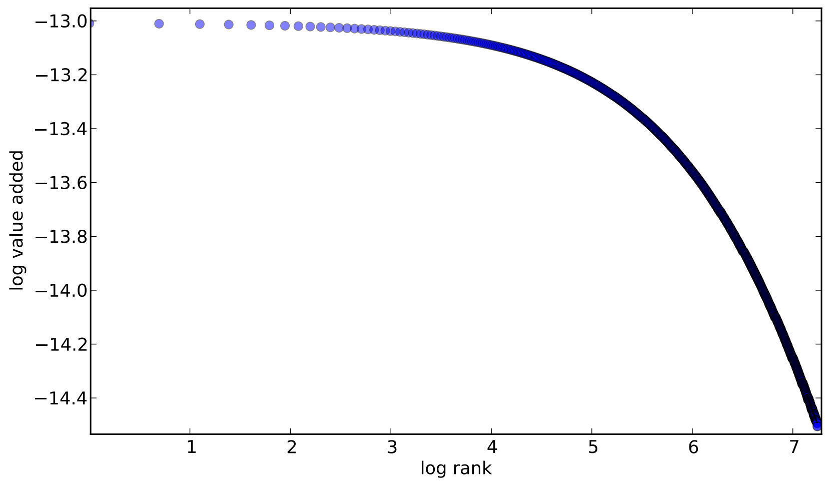



Figure 6: Size-rank plot (logscale)

Proposition 3.3 indicates that, although firms are ex-ante identical, in equilibrium they will organize into a nondegenerate distribution, with firm size measured by value added increasing from upstream to downstream. The shape of the distribution of firm sizes depends on the primitives $c$ and $\delta$. Typically it is right-skewed. The rank-size plot for firm size measured by value added is shown in figure 6. Each point corresponds to one firm in the vertical structure generated by primitives $c(x) = \exp(x^2) - 1$ and $\delta = 1.00075$. In equilibrium there are 1402 firms. The vertical axis is the log of value added, and the horizontal axis is the log of firm rank by value added. (Firms are ranked from largest to smallest in terms of value added.) The rank-size distribution shows significant curvature, although it is almost linear for smaller firms. (In fact, for these parameters, a linear regression of the data restricted to the smallest decile of firms has a very close fit and slope of -1.02. A slope of -1 corresponds to Zipf's law.)

# 4 Extensions

## 4.1 Alternative Costs

Until now we have assumed that the transaction cost is borne entirely by the buyer, while claiming that this assumption costs little in terms of generality. To see why, suppose instead that the transaction cost is borne by both the buyer and the seller in some given proportions. (For example, the cost of drafting and negotiating contracts might be borne by both sides of the transaction.) In particular, suppose as before that when the buyer purchases the good at price $p(t)$ her total outlay is $\delta p(t)$ with $\delta > 1$, and suppose in addition that the revenue net of contract costs received by the seller is $\gamma p(t)$ for some $\gamma \in (0,1)$. The profit function in section 2.3 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

$$p(s) = \min_{t \leq s} \left\{ \frac{\delta}{\gamma} p(t) + \frac{c(s-t)}{\gamma} \right\}. \tag{10}$$

This equation has the same form as (2). Since $\delta/\gamma > 1$ and $c/\gamma$ inherits from $c$ all the properties of the cost function stated in section 2.2, the preceding results go through, and qualitative properties of the solution are the same.

## 4.2 Multiple Partners

So far we have assumed that production is linearly sequential, and firms contract with only one parter. In reality, most vertically integrated firms have multiple upstream partners. For example, in 2004 Toyota group had 168 direct parts suppliers. These primary suppliers themselves had 5,437 direct suppliers, and in turn these secondary suppliers had 41,703 tertiary suppliers (Tsuji, 2004).

As a result of our solution strategy based on recursive subdivision of tasks, our model can easily be generalized to include multiple upstream partners. To begin, consider the tree in figure 7. As before, firm $n$ choose an interval $\ell_n$ of tasks to perform in-house, and subcontracts the remainder. In this case, however, each downstream firm subcontracts to two upstream partners. We can also consider more general cases, where each downstream firm subcontracts has $k$ upstream partners. If a firm contracts to supply the good at stage $s$, chooses a quantity $\ell \leq s$ to produce in-house, and then divides the remainder $s - \ell$ equally across $k$ upstream partners, then profits will be given by revenue $p(s)$ minus input costs $\delta\, k\, p((s-\ell)/k)$ minus in-house production costs $c(\ell)$. That is,

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

Setting profits to zero, letting $t := s - \ell$ and minimizing with respect to $t$ yields the functional equation

$$p(s) = \inf_{t \leq s} \{\delta\, k\, p(t/k) + c(s-t)\}.$$

This equation is an immediate generalization of (2). Similar techniques can be employed to show the existence and uniqueness of a solution $p^*$, to compute $p^*$, and to compute from $p^*$ the optimal choices of firms and the resulting distribution of firms.

Figure 8 shows the results of these computations for $k = 3$ under two parameterizations. In these networks, each node represents a firm, and the node size is proportional to the value added of that firm. As with the $k = 1$ case, downstream firms (i.e., firms towards

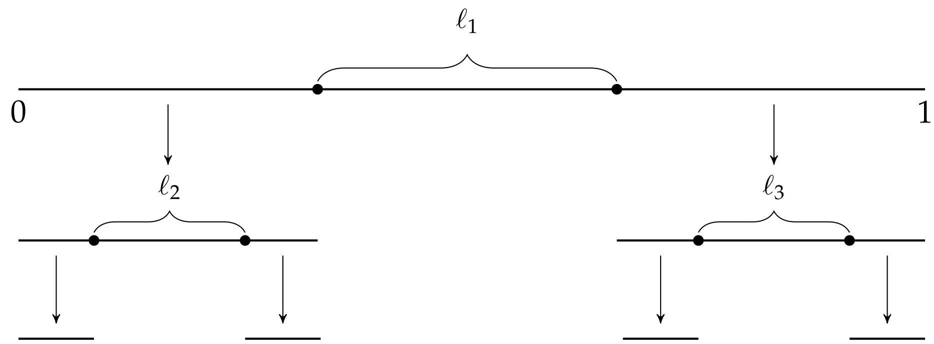



Figure 7: Subcontracting when $k = 2$

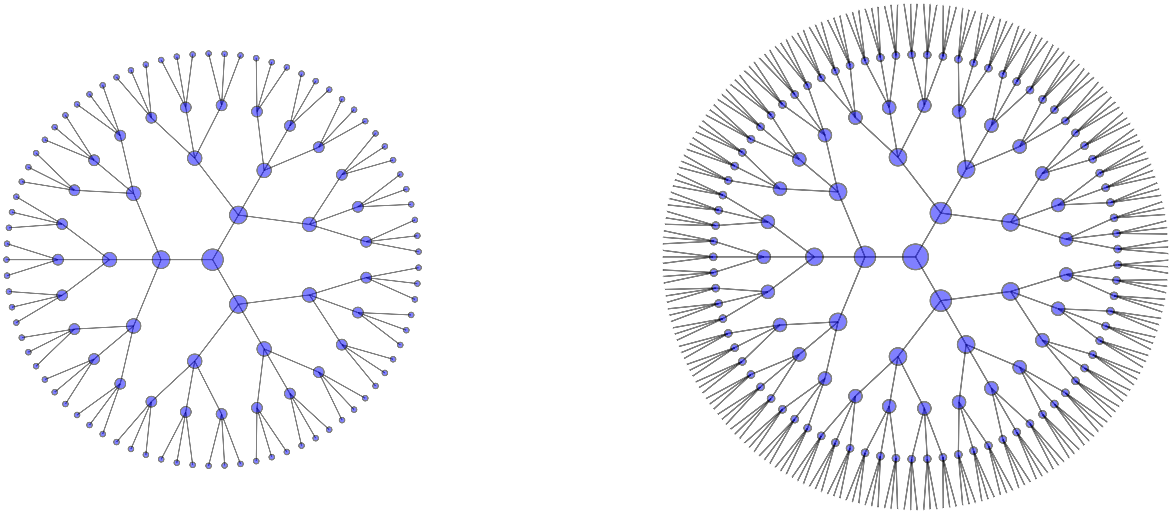
the center of the network) have greater value added. In figure 8 the nonlinear relationship between transaction costs and the number of firms is visible. A small reduction in transaction costs increases the number of firms from 121 to 364.

# 5 Discussion

It is easy to ignore the role of transaction costs when modeling production. They are typically small relative to input costs, and tend to create the kinds of frictions that make for awkward modeling. In this connection, our model helps to illustrate how transaction costs can be modeled in a tractable way, and, more importantly, how these small costs combine to have large effects that are fundamental in determining the structure and overall cost of production. The reason is that for a fine division of the value chain or large production network combining many firms, these small costs are incurred many times, and hence the impact of small changes is first order. This first order effect can be seen in the exponential terms in the expression $p^*(1) = \sum_{i=1}^{N^*} \delta^{i-1} c(\ell_i)$ given in proposition 3.2, and in the comparison of production networks shown in figure 8. As a numerical example, a simple computation using our model shows that if $c(x) = \exp(x^2) - 1$ and $\delta = 1.01$, then a 1% rise in $\delta$ leads to a 40% fall in the number of firms and a 90% increase in the price of the final good.[6]

How large are transaction costs in practice? It turns out that the size varies widely across industries and countries. For example, Gabre-Madhin (2001) found in a study of

[6] It is worth noting that the expression $p^*(1) = \sum_{i=1}^{N^*} \delta^{i-1} c(\ell_i)$ does not in fact imply that prices are exponential in $\delta$. The reason is that when $\delta$ changes firms reoptimize, leading to a change in the sequence $\{\ell_i\}$. In particular, higher transaction costs are mitigated by a reduction in the number of firms.



$\delta = 1.010$ $\delta = 1.009$

Figure 8: Production network with $c(x) = e^x - 1$ and $k = 3$

Ethiopian grain markets that the cost of searching for prices in terms of opportunity cost of labor and working capital amounted to 19% of trader costs. In modern grain markets linked by computers, the cost of discovering spot prices is much lower. Similarly, while some industries find transportation costs low enough to form widely dispersed production networks, for other producers these costs may be prohibitive. For example, BBC News reported that in Mogadishu in November 2004, a 50km drive from the airstrip to the city involved passing seven checkpoints, each run by a different militia, and at each of these "border crossings" all lorries had to pay a fee ranging from \$3 to \$300, depending on the value of the goods being carried.[7] In such settings, simple production networks with fewer transactions will be favored over larger ones and productivity will be lower.

The size of transaction costs in developing economies was a major concern of North (1993), who, preempting growth theory's current preoccupation with institutions as a fundamental source of cross-country variation in productivity, highlighted the fact that transaction costs were a function of institutions and social norms, that these institutions and norms varied greatly across countries, and that the impact of transaction costs on production could be extremely large. In his words, "it is the exchanges that don't occur because of the high cost of transacting (and therefore producing) that are the real underlying source of poverty of third world economies" (North, 1993, p. 11). Whether this claim is true or not, it seems very likely that transaction costs in developing economies are a major impediment to the development of complex production networks, and to the

[7]Reported by Joseph Winter on BBC News, 17th Nov. 2004.

adoption of modern technologies that require such networks.

On a theoretical level, recognition of the systemic impact of transaction costs dates back to Adam Smith (1776). Smith famously argued that the division of labor is limited by the extent of the market. Pushing the analysis further, he also noted that the extent of the market is itself limited by transportation costs (Smith, 1776, p. 31). This was an early acknowledgment of the fundamental role played by transaction costs in limiting the specialization and the division of labor. Smith's insights in this direction have been extended by authors such as Houthakker (1956), Yang and Ng (1993) and Becker and Murphy (2002). Our model is clearly connected, although the modeling techniques are very different and the focus is on division of tasks across firms.

The most commonly cited transaction costs include transportation and transaction fees, search, bargaining and information costs, costs of assessing credit worthiness and reliability, and the costs associated with negotiating, writing, monitoring and enforcing contracts. Discussion of these different costs can be found in references such as Coase (1937), Williamson (1979, 1981), Arrow (1969) and North (1993). Both Coase and North emphasized contracts as a major component of transaction costs (see, e.g., North, 1993, p. 6). Costly contracts fit naturally with Coase's theory of the firm (and our model) because their burden can be substantially reduced by vertical integration (albeit at the expense of incurring other kinds of costs).

Overall, the predictions of our model vis-a-vis transaction costs are broadly consistent with empirical and case studies in the literature. For example, Sandefur (2010) noted that the dramatic market reforms that occurred in Ghana in the 1980's were followed by a significant fall in average firm size as measured by employment, from 19 to 9 in the fifteen years from 1987. The reforms (market deregulation, enforcement of contracts and stricter penalties for bribe extraction) all suggest a reduction in transaction costs. While Sandefur presented this contraction in firm sizes as a puzzle, in our model it is a prediction (see, e.g., figures 5 and 8). A similar fall in average firm size following market reforms was reported in Slovenia (Polanec, 2004).

Besides transaction costs, the other fundamental force in Coase's theory of the firm is diminishing returns to management. Our model sets the processing stages as the continuum $[0, 1]$, but on an intuitive level we can think of a discrete and finite number of stages, and imagine that progressing from one stage to the next involves a unique task carried out by a specialized worker. The cost of implementing and coordinating such tasks tends to grow at rate greater than proportional to the number of tasks. This problem was emphasized by Robinson (1934) and Hayek (1945), who highlighted the difficulty of utilizing knowledge not held in its totality by any one individual (unless that same knowledge is coordinated through the market via prices). An excellent discussion on the cost of coordinating specialized works is given in Becker and Murphy (1992), who cite communication problems, increased conflict in larger teams, opportunistic behavior, principal-agent conflicts, free-riding and general management costs associated with coordination. Becker and Murphy (1992) argue that this cost is a more important determinant of the division of labor than the extent of the market.[8]

In our model, the rate at which internal coordination costs grow depends on the degree of convexity or curvature of $c$. This curvature is parameterized as $\theta$ in example 3.1. A decrease in curvature leads at the industry level to fewer transactions and a smaller number of larger firms. The reason is that the market and prices become relatively more expensive as a means of coordinating production. Such predictions are hard to confirm empirically, since many innovations that change internal coordination costs (e.g, computer networks) will tend to affect transaction costs as well (through lower search costs and so on). On the other hand, some major management innovations have reduced the cost of coordinating specialists without affecting transaction costs. Examples include moving assembly lines and the multi-divisional firm structure. As expected, such innovations have tended to spur the growth of larger firms. A classic reference on management innovations and their affect on industry structure is Chandler (1997).

An interesting prediction of our model is the differences between upstream and downstream firms in equilibrium. Our model predicts that downstream firms will be larger both in the sense that they implement a larger number of tasks in house and they provide greater value added (proposition 3.3). This prediction needs to be interpreted with caution, since our model relates to the production of a single unit of a single good, and includes no discussion of industry equilibrium, horizontal integration or likely differences between upstream and downstream firms in terms of fixed costs or capital requirements. Careful study of our prediction would require firm level data and the ability to control for these differences. Nevertheless, it is interesting to look at preliminary findings based on industry level data.

To this end, figure 9 compares value added against downstreamness. Data is taken from the 2002 Bureau of Economic Analysis input-output tables and the NBER CES data set. The points in the figure correspond to 173 individual industries that match directly in the IO and CES data sets. The horizontal axis shows downstreamness as calculated from the index developed by Fally (2012) and Antràs *et al.* (2012).[9] The vertical axis measures value added per plant. (While data on value added per firm would be more useful, such information is not available in these data sets.) Although the data is noisy, it shows the

[8]Other important references on coordination costs include Simon (1947), Arrow (1974, 1975), Williamson (1975), Lucas (1978), Blau and Meyer (1987), Demsetz (1988), Geanakopolis and Milgrom (1991), Rasmusen and Zenger (1990), Bolton and Dewatripont (1994), Holmstrom and Milgrom (1994), McAfee and McMillan (1995) and Garicano (2000).

[9]Fally (2012) calculates $N_i$, the total number of sequential stages required for production of good $i$ (in his study, "good" and "industry" are identified), and $D_i$, the number of remaining stages before reaching final demand. Our measure of downstreamness is $N_i/(N_i + D_i - 1)$.

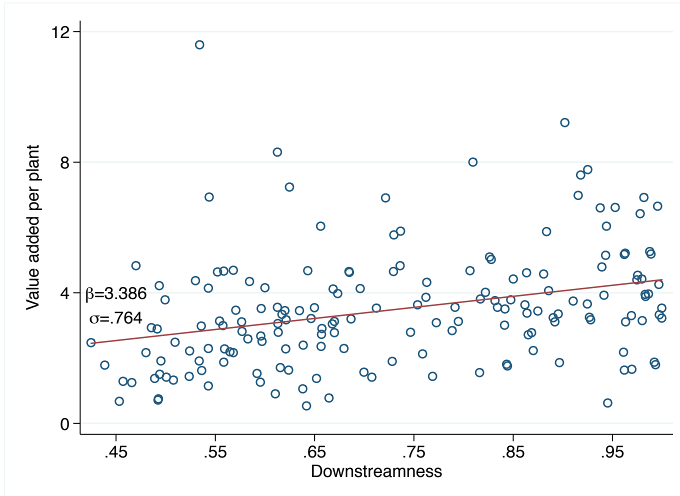



Figure 9: Plant value added by downstreamness

positive correlation between these variables that is essentially consistent with the prediction in our model.

Table 1 is also constructed from the input-output tables to address the relationship between downstreamness and value added, although the methodology is different. The table shows data for the 4 most downstream industries in the table (results for other industries are similar). 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 0 is value added by the industry itself. Stage 1 is the maximum of the value added by all of the direct input industries, and stage 2 is the maximum of the value added by all of the direct inputs to these inputs. For example, if the algorithm that computed the values along one row of the table was applied to either of the production networks in figure 8, the stage 0 value would be the size of the central node (recall that node size measures share of value added in the network), the stage 1 value would be the maximum of the size of the three nodes in the first ring of immediate subcontractors, and the stage 2 value would be the maximum of the size of the 9 nodes in the next ring. As with figure 9, we find that value added is increasing with downstreamness.

| Industry | stage 0 | stage 1 | stage 2 |
|---|---|---|---|
| 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

# 6 Conclusion

This paper proposed a mathematical formulation of Coase's (1937) theory of the firm that permits investigation not just of firm boundaries but also of the vertical structure of production across the industry within a competitive equilibrium setting. By developing a new methodology based on recursive subdivision of tasks combined with fixed point theory, we proved existence of the equilibrium price function, and showed it to be unique, convex and differentiable. We showed that the optimal actions of firms given this price function are unique, well-defined and continuous. We provided a consistent algorithm for computing equilibrium prices and actions. We derived a first order condition that corresponds to the key marginal condition determining firm boundaries stated verbally by Coase (1937), and added an "Euler" equation relating costs of adjacent firms. We analyzed the structure of production, showing that the equilibrium number of firms is finite, and that value added and the number of tasks performed in-house are increasing from upstream to downstream. We investigated the distribution of firms, and the relationship between diminishing returns to management, transaction costs and the properties of the vertical production chain. Some extensions to the model were considered, including production networks with multiple upstream partners.

There paper opens many avenues for future research. The techniques developed in the paper are likely to have applications to other fields, such as offshoring by multinationals, or division of labor with transaction costs, failure probabilities or other frictions. In addition, the model presented above was a baseline model in all dimensions, with perfect competition, perfect information, identical firms and identical tasks, and 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 4.2. On the technical side, the operator $T$ in (3) merits further investigation. It may turn out to be a contraction if the metric is well chosen, or have connections to more common operators or optimization problems. Such analysis might further improve the algorithms for computing prices and actions and lead to additional insights. Finally, the model has a number of testable implications that require investigation.

# 7 Appendix: Computation and Remaining Proofs

## 7.1 Computation

To compute an approximation to the equilibrium pricing function for given $\delta$ and $c$, one possibility is to take a function in $\mathscr{P}$ and iterate with $T$. However, in practice we can only approximate the iterates, and, since $T$ is not a contraction mapping convergence is problematic. On the other hand, as we now show, there is a fast, non-iterative alternative that is guaranteed to converge.

**Algorithm 1** Construction of $p$ from $G = \{0, h, 2h, 3h, \ldots, 1\}$

```
p(0) ← 0
s ← h
while s ≤ 1 do
    evaluate p(s) via equation (11)
    define p on [0, s] by linear interpolation of p(0), p(h), p(2h), ..., p(s)
    s ← s + h
end while
```

Let $G = \{0, h, 2h, 3h, \ldots, 1\}$ for fixed $h$. Given $G$, we define our approximation $p$ to $p^*$ via the recursive procedure in algorithm 1. In the fourth line, the evaluation of $p(s)$ is by setting

$$p(s) = \min_{t \leq s-h} \{\delta p(t) + c(s-t)\}. \tag{11}$$

In line five, the linear interpolation is piecewise linear interpolation of grid points $0, h, 2h, \ldots, s$ and values $p(0), p(h), p(2h), \ldots, p(s)$.

The procedure can be implemented because the minimization step on the right-hand side of (11), which is used to compute $p(s)$, only evaluates $p$ on $[0, s-h]$, and the values of $p$ on this set are determined by previous iterations of the loop. Once the value $p(s)$ has been computed, the following line extends $p$ from $[0, s-h]$ to the new interval $[0, s]$. The process repeats. Once the algorithm completes, the resulting function $p$ is defined on all of $[0, 1]$ and satisfies $p(0) = 0$ and (11) for all $s \in G$ with $s > 0$.

Now consider a sequence of grids $\{G_n\}$, and the corresponding functions $\{p_n\}$ defined by algorithm 1. Let $G_n = \{0, h_n, 2h_n, \ldots, 1\}$ with $h_n = 2^{-n}$. In this setting we have the following result, the proof which is given in section 7.3.

**Theorem 7.1.** *The sequence $\{p_n\}$ converges uniformly to $p^*$.*

In the proof we show that $\{p_n\}$ is monotone decreasing in $n$. Thus, the limiting function $\bar{p} := \lim_n p_n$ exists in $\mathscr{P}$. We show that $p_n$ converges to $\bar{p}$ uniformly, and $\bar{p}$ solves the equilibrium pricing equation. In view of the uniqueness result in theorem 3.1, we then have $\bar{p} = p^*$ and $p_n \to p^*$ uniformly.[10]

[10]Notice that these computational arguments provide a *constructive* proof of the existence of an equilibrium pricing function.

## 7.2 Proofs from Section 3

**Lemma 7.1.** *The operator $T$ defined in (3) maps $\mathscr{P}$ into itself.*

*Proof.* Let $p$ be an arbitrary element of $\mathscr{P}$. To see that $Tp(s) \leq c(s)$ 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 $Tp(s) \geq \inf_{t \leq s} c'(0)s = c'(0)s$. □

By the conditions in section 2.2 and the intermediate value theorem, there exists an $\bar{s} \in (0,1]$ such that $c'(\bar{s}) = \delta c'(0)$. 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 start from stage $t = 0$.

**Lemma 7.2.** *If $p \in \mathscr{P}$, then $s \leq \bar{s}$ if and only if $\inf_{t \leq s}\{\delta p(t) + c(s-t)\} = c(s)$.*

*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 $c$.

Now suppose 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

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

Taking the limit gives $c'(s) \leq \delta c'(0)$. □

**Lemma 7.3.** *The set $\mathscr{P}_0$ defined in theorem 3.1 is nonempty, convex and compact in $b[0,1]$, the Banach space of bounded, real-valued functions on $[0,1]$ endowed with the supremum norm.*

*Proof.* Regarding convexity of $\mathscr{P}_0$, it is straightforward to check that the convex combination of any two convex functions in $b[0,1]$ is again convex, and the convex combination of any two $K$-Lipschitz functions in $b[0,1]$, is again $K$-Lipschitz. Evidently the convex combination of two functions in $\mathscr{P}_0$ is bounded above by $s \mapsto c(s)$ and bounded below by $s \mapsto c'(0)s$. It follows that $\mathscr{P}_0$ is a convex subset of $b[0,1]$. Regarding compactness of $\mathscr{P}_0$, the fact that $\mathscr{P}_0$ has compact closure in $b[0,1]$ follows immediately from the Arzelà-Ascoli theorem.[11] Hence $\mathscr{P}_0$ will be seen to be compact if we can show that it is closed.

[11] The Arzelà-Ascoli theorem implies that $\mathscr{P}_0$ has compact closure in the space of continuous functions on $[0,1]$, but this property implies the same for $\mathscr{P}_0$ as a subset of $b[0,1]$.

Closedness of $\mathscr{P}_0$ in $b[0,1]$ is readily apparent, since uniform limits preserve convexity, the $K$-Lipschitz property and the bounds defining $\mathscr{P}$. $\square$

**Lemma 7.4.** *Let $p \in \mathscr{P}_0$ and define*

$$t_p(s) := \arg\min_{t \leq s}\{\delta p(t) + c(s-t)\} \quad \text{and} \quad \ell_p(s) := s - t_p(s). \tag{12}$$

*Let $s_1$ and $s_2$ be two points with $0 < s_1 \leq s_2 \leq 1$. The following statements are true:*

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.

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

$$\delta p(t_1) + c(s_2 - t_1) < \delta p(t_2) + c(s_2 - t_2), \tag{13}$$

which contradicts the definition of $t_2$.[12]

To establish (13), 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 (13).

Next we shown 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 $\ell_2$ was available when $\ell_1$ was chosen. Therefore,

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

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

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 $\ell_2$.[13]

To complete the proof of lemma 7.4, we also need to show that $t_p(s_2) - t_p(s_1) \leq s_2 - s_1$, and similarly for $\ell$. 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 $\ell_p$ is obtained in the same way, by reversing the roles of $t_p$ and $\ell_p$. This concludes the proof of lemma 7.4. $\square$

**Lemma 7.5.** *Let $p \in \mathscr{P}_0$, let $\ell_p$ be as defined in (12), and let $\bar{s}$ be the point in $(0, 1]$ defined in lemma 7.2. If $s \geq \bar{s}$, then $\ell_p(s) \geq \bar{s}$. If $s > 0$, then $\ell_p(s) > 0$.*

*Proof.* By lemma 7.4, $\ell_p$ 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 7.2, if $0 < s \leq \bar{s}$, then $\ell_p(s) = s - t_p(s) = s > 0$. $\square$

**Lemma 7.6.** *If $p \in \mathscr{P}_0$, 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 (12). 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 exhibit such a function, observe that in view of lemma 7.5, 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 $(Tp)'(s_0) = w'(s_0) = c'(\ell_p(s_0))$. $\square$

[13]Note that $0 \leq \ell_1 \leq s_1 \leq s_2$, so $\ell_1$ is available when $\ell_2$ is chosen.

**Lemma 7.7.** *If $p \in \mathscr{P}_0$, then $Tp$ is strictly convex.*

*Proof.* To see this, pick any $0 \leq s_1 < s_2 \leq 1$ and any $\lambda \in (0,1)$. Let

$$t_i := \operatorname*{arg\,min}_{t \leq s_i} \{\delta p(t) + c(s_i - t)\} \qquad \text{for } i = 1, 2,$$

and $t_3 := \lambda t_1 + (1-\lambda)t_2$. It is easy to check that $0 \leq t_3 \leq \lambda s_1 + (1-\lambda)s_2$, and hence

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

The right-hand side expands out to

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

Using convexity of $p$ and strict convexity of $c$, we obtain

$$Tp(\lambda s_1 + (1-\lambda)s_2) < \lambda Tp(s_1) + (1-\lambda)Tp(s_2).$$

In other words, $Tp$ is strictly convex. □

**Lemma 7.8.** *The operator $T$ maps $\mathscr{P}_0$ into itself.*

*Proof.* Let $p \in \mathscr{P}_0$. That $Tp$ satisfies the bounds $c'(0)s \leq p(s) \leq c(s)$ was proved in lemma 7.1. That $Tp$ is convex was shown in lemma 7.7. Finally, $Tp$ is $K$-Lipschitz for $K := c'(1)$ because, by lemma 7.6 and convexity of $c$, we have $(Tp)'(s) \leq c'(\ell_p(s)) \leq c'(1)$. The $c'(1)$-Lipschitz property now follows immediately from the mean value theorem. □

**Lemma 7.9.** *The operator $T$ is continuous as a mapping from $b[0,1]$ to itself.*

*Proof.* That $T$ maps $b[0,1]$ into itself is trivial to prove. To see continuity of the mapping, observe that for $f$ and $g$ in $b[0,1]$ we have $|\inf f - \inf g| \leq \sup |f - g|$, and hence, for any given $s \in [0,1]$ and $p, q \in b[0,1]$,

$$|\inf_{t \leq s} \{\delta p(t) + c(s-t)\} - \inf_{t \leq s} \{\delta q(t) + c(s-t)\}| \leq \delta \sup_{t \leq s} |p - q| \leq \delta \sup_{t \in [0,1]} |p - q|.$$

$$\therefore \quad |Tp(s) - Tq(s)| \leq \delta \sup_{t \in [0,1]} |p - q|.$$

Taking the supremum of the left hand side over $s$, we see that $T$ is $\delta$-Lipschitz as a mapping from $b[0,1]$ to itself. Since every Lipschitz continuous function is continuous, we conclude that $T$ is a continuous self mapping on $b[0,1]$. □

It was shown in section 3 that an equilibrium price function exists. The proof used Tarski's fixed point theorem. We now claim that $p$ has a Lipschitz continuous and convex fixed point. The simplest way to do this is to apply a fixed point theorem to $T$ as a mapping from $\mathscr{P}_0$ to itself. Since $\mathscr{P}_0$ is not a complete lattice, we use Schauder's fixed point theorem instead. (Of course, this makes the earlier application of Tarski's fixed point theorem redundant, but the proof of Tarski's fixed point theorem is simpler, and this is the reason we chose to include it as well.)

**Proposition 7.1.** *The operator $T$ has a fixed point in $\mathscr{P}_0$.*

*Proof.* In lemma 7.9 we saw that $T$ is a continuous self mapping on $b[0,1]$. Moreover, $\mathscr{P}_0$ is a nonempty convex compact subset of $b[0,1]$ (lemma 7.3), and $T$ maps $\mathscr{P}_0$ into itself (lemma 7.8). Schauder's fixed point theorem now implies that $T$ has a fixed point in $\mathscr{P}_0$. □

**Proposition 7.2.** *The operator $T$ has at most one fixed point in $\mathscr{P}_0$.*

*Proof.* Let $p$ and $q$ be two fixed points of $T$ in $\mathscr{P}_0$. The proof is by induction. First we show that there exist an $k \in \mathbb{N}$ such that $p$ and $q$ agree on the interval $[0, 1/k]$. Next we show that if $n \in \mathbb{N}$, $n < k$ and $p$ and $q$ agree on $[0, n/k]$, then they also agree on $[0, (n+1)/k]$. It follows that $p$ and $q$ agree on all of $[0,1]$.

To begin, let $\bar{s}$ be as defined in lemma 7.2. Let $k$ be chosen such that $1/k \leq \bar{s}$. Observe that if $s \leq 1/k$, then $p(s) = q(s)$, because, in light of the definition of $\bar{s}$, if $s \leq \bar{s}$, then

$$p(s) = \min_{t \leq s}\{\delta p(t) + c(s-t)\} = c(s) = \min_{t \leq s}\{\delta q(t) + c(s-t)\} = q(s).$$

Turning to the induction step, let $n < k$ and suppose that $p = q$ on $[0, n/k]$. Fix $s \in [0, (n+1)/k]$. We claim that $p(s) = q(s)$. To see this, first suppose that $s \leq \bar{s}$. In that case $p(s) = q(s) = c(s)$ as shown above. Suppose instead that $s > \bar{s}$. In that case, lemma 7.5 and the definition of $k$ yield $\ell_p(s) \geq \ell_p(\bar{s}) = \bar{s} \geq 1/k$. Therefore $t_p(s) = s - \ell_p(s) \leq s - 1/k \leq (n+1)/k - 1/k = n/k$. Exactly the same argument holds for $q$ as well, so we have $t_p(s) \leq n/k$ and $t_q(s) \leq n/k$. This tells us that when computing the optimal action $t$ for either $p$ or $q$, we can restrict attention to minimizing over $[0, n/k]$. Moreover, by the induction hypothesis, $p$ and $q$ agree on $[0, n/k]$, This leads to our conclusion:

$$p(s) = \min_{t \leq n/k}\{\delta p(t) + c(s-t)\} = \min_{t \leq n/k}\{\delta q(t) + c(s-t)\} = q(s).$$

The proof of uniqueness is complete. □

*Proof of theorem 3.1.* Existence and uniqueness of an equilibrium price function $p^*$ in $\mathscr{P}_0$ follows from propositions 7.1 and 7.2. Since $p^* \in \mathscr{P}_0$ and the image under $T$ of any function in $\mathscr{P}_0$ is strictly convex (lemma 7.7), we see that $Tp^*$ is strictly convex. Since $p^* = Tp^*$, the function $p^*$ itself is strictly convex. A similar argument combined with lemma 7.6 shows that $p^*$ is differentiable, and $(p^*)'(s) = c'(\ell^*(s))$. Since $c'$ is strictly positive on $[0, 1]$, the last equality also shows that $p^*$ is strictly increasing. Moreover, since $\ell^*$ is Lipschitz continuous (lemma 7.4) and $c$ is assumed to be continuously differentiable, the equation $(p^*)'(s) = c'(\ell^*(s))$ implies that $p^*$ is not only differentiable but also continuously differentiable. Finally, the claims in theorem 3.1 regarding $t^*$ and $\ell^*$ are immediate from lemma 7.4. □

*Proof of theorem 3.2.* In view of lemma 7.2 we have $\ell^*(s) = s$ and hence $t^*(s) = 0$ whenever $s \leq \bar{s}$, so it suffices to prove that $t_n \leq \bar{s}$ for some $n$. This must be the case because $\ell^*$ is increasing, and hence the amount of in-house production by a firm contracting at $s \geq \bar{s}$ satisfies $\ell^*(s) \geq \ell^*(\bar{s}) = \bar{s} > 0$. In other words, for firms contracting above $\bar{s}$, each takes a step of length at least $\bar{s}$. In particular, $t_n \leq 1 - n\bar{s}$. □

**Lemma 7.10.** *If $c(x) = e^{\theta x} - 1$, then the equilibrium number of firms is given by (9).*

*Proof of lemma 7.10.* 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 $\ell_{n+1} = \ell_n - r$, and hence $\ell_1 = \ell_n + (n-1)r$. It is easy to check that when $c(x) = e^{\theta x} - 1$, the constant $\bar{s}$ in lemma 7.15 is equal to $r$. Hence $0 < \ell_N \leq r$. Therefore $(N-1)r < \ell_1 \leq Nr$. From $\sum_{n=1}^N \ell_n = 1$ 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

$$\frac{1}{2}\left(-1 + \sqrt{1 + 8/r}\right) < N \leq \frac{1}{2}\left(1 + \sqrt{1 + 8/r}\right).$$

The expression for $N = N^*$ in (9) now follows. □

*Proof of proposition 3.2.* This follows from the recursion $p^*(t_n) = \delta p^*(t_{n+1}) + c(\ell_{n+1})$ and the fact that $p(t_N) = p(0) = 0$. □

*Proof of proposition 3.3.* By theorem 3.1, the function $\ell^*$ is increasing. By construction $t_n \leq t_{n-1}$, and hence $\ell_{n+1} = \ell^*(t_n) \leq \ell^*(t_{n-1}) = \ell_n$. Thus part 1 holds. Part 2 follows from part 1 and convexity of $p^*$. □

*Proof of proposition 3.1.* We begin with the claim that $p_a^* \leq p_b^*$. To construct the proof, let $G = \{0, h, 2h, \dots 1\}$ be a fixed grid, and, for $i = 1, 2$, let $p_i$ be the approximation generated by algorithm 1 and parameter $\delta_i$. We claim that $p_a \leq p_b$. The proof is by induction. First, observe that by (15) we have $p_a(h) = p_b(h) = c(h)$. Given that the functions are constructed by linear approximation between grid points, it follows that $p_a \leq p_b$ on $[0, h]$.

Thus it remains to show that if $j \in \mathbb{N}$ and $p_a \leq p_b$ on $[0, jh]$, then $p_a \leq p_b$ on $[0, (j+1)h]$. To see this, observe that if the induction hypothesis is true, then

$$\begin{aligned} p_a((j+1)h) &= \min_{t \leq jh} \{\delta_a p_a(t) + c((j+1)h - t)\} \\ &\leq \min_{t \leq jh} \{\delta_b p_b(t) + c((j+1)h - t)\} = p_b((j+1)h). \end{aligned}$$

Thus $p_a \leq p_b$ on $[0, (j+1)h]$ as claimed.

Now let $h = h_n = 2^{-n}$, and let $p_a = p_a^n$ and $p_b = p_b^n$. We have shown that $p_a^n \leq p_b^n$. By theorem 7.1 we have $p_i^n \to p_i^*$. Since pointwise inequalities are preserved under limits, it follows that $p_a^* \leq p_b^*$.

As the next step of the proof, we first show that the number of tasks carried out by the most upstream firm decreases when $\delta$ increases from $\delta_a$ to $\delta_b$. Let $\ell_i^a$ be the number of task carried out by firm $i$ when $\delta = \delta_a$, and let $\ell_i^b$ 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 (7) 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 $\delta_b$ is less than $N = N_a^*$ and we are done. Suppose instead that $\ell_N^b > 0$. In view of lemma 7.2, 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 7.2 again, we see that the $N$-th firm completes the good, and hence $N_b^* = N_a^*$. □

## 7.3 Proof of Theorem 7.1

To begin the proof of theorem 7.1, suppose first that $G$ is the fixed grid $0, h, 2h, \ldots, 1$, and $p$ is the function defined in algorithm 1.

**Lemma 7.11.** *The function $p$ is convex on $[0, 1]$.*

*Proof.* Since $p$ is piecewise linear, it suffices to show that for any consecutive $s_0, s_1, s_2$ in $G$ we have $p(s_2) - p(s_1) \geq p(s_1) - p(s_0)$. Equivalently,

$$p(s_1) \leq \frac{1}{2}\{p(s_0) + p(s_2)\}. \tag{14}$$

We prove this first for $s_0 = 0$, and then proceed by induction. When $s_0 = 0$, we have $s_1 = h$ and $s_2 = 2h$, and (14) reduces to the claim that $2p(h) \leq p(2h)$. To verify this, observe first that

$$p(h) = \min_{0 \leq t \leq 0} \{\delta p(t) + c(h-t)\} = c(h), \tag{15}$$

while

$$p(2h) = \min_{t \leq h} \{\delta p(t) + c(2h-t)\}.$$

On $[0, h]$, $p(t)$ is defined by linear interpolation between $p(0)$ and $p(h) = c(h)$, so in particular $p(t) = c(h)(t/h)$. Since $c$ is convex, we then have

$$p(2h) \geq \min_{0 \leq t \leq h} \{c(h)(t/h) + c(2h-t)\} = 2c(h) = 2p(h).$$

We have now confirmed that $p$ is convex on $[0, 2h]$. The next step is to show that if $q$ is convex on $[0, s_1]$, then $q$ is convex on $[0, s_2] = [0, s_1 + h]$. To verify this, we only need to check that (14) holds when $q$ is convex on $[0, s_1]$.

To see check (14), let $t_i$ be an optimal choice corresponding to $s_i$, in the sense that $t_i$ is a minimizer of $\{\delta p(t) + c(s_i - t)\}$ over $0 \leq t \leq s_i - h$. Note that, by the definition of $t_i$,

$$t_3 := \frac{1}{2}(t_0 + t_2) \leq \frac{1}{2}(s_0 - h + s_2 - h) = s_1 - h.$$

As a result, $t_3$ was available when $t_1$ was chosen, and we have

$$p(s_1) \leq \delta p(t_3) + c(s_1 - t_3).$$

Using convexity of $p$ on $[0, s_1]$ and the fact that $t_0$ and $t_2$ are both less than $s_1$, we have

$$p(t_3) = p(t_0/2 + t_2/2) \leq \frac{p(t_0) + p(t_2)}{2}.$$

Using convexity of $c$, we have

$$c(s_1 - t^*) = c\left(\frac{s_0 - t_0}{2} + \frac{s_2 - t_2}{2}\right) \leq \frac{c(s_0 - t_0) + c(s_2 - t_2)}{2}.$$

Combining the last three inequalities, we obtain

$$p(s_1) \leq \frac{1}{2}\left[\{\delta p(t_0) + c(s_0 - t_0)\} + \{\delta p(t_2) + c(s_2 - t_2)\}\right] = \frac{1}{2}\left[p(s_0) + p(s_2)\right].$$

In other words, (14) is valid. Convexity is now proved. $\square$

Given $s \in [0, 1]$, let $t(s)$ is a minimizer of $\delta p(t) + c(s - t)$ over $[0, s - h]$, and let $\ell(s) = s - t(s)$.

**Lemma 7.12.** *Let $s_1$ and $s_2$ be any two points in $G_n$ with $0 < s_1 \leq s_2$. The following statements are true:*

1. *Both $t(s_1)$ and $\ell(s_1)$ are well defined and unique.*
2. *$t(s_1) \leq t(s_2)$ and $t(s_2) - t(s_1) \leq s_2 - s_1$.*
3. *$\ell(s_1) \leq \ell(s_2)$ and $\ell(s_2) - \ell(s_1) \leq s_2 - s_1$.*

The proof is almost identical to that of lemma 7.4 and hence omitted.

**Lemma 7.13.** *The function $p$ is monotone increasing on $[0,1]$.*

*Proof.* Since $p$ is defined by linear interpolation, it is enough to show that $p(s_1) \leq p(s_2)$ for consecutive grid points $s_1$ and $s_2 = s_1 + h$. The proof is by induction. Evidently the claim is true for $s_1 = 0$, because $p(0) = 0$. So now let $s_1$ be an arbitrary point in $G$, and suppose that $p$ is monotone increasing on $[0, s_1]$. We need to show that $p(s_1) \leq p(s_2)$. To see that this is the case, observe that

$$p(s_i) = \delta p(t(s_i)) + c(\ell(s_i)).$$

By monotonicity of $t$ and $\ell$, we have $\ell(s_1) \leq \ell(s_2)$ and $t(s_1) \leq t(s_2)$. Moreover, $c$ is monotone increasing, $p$ is monotone increasing on $[0, s_1]$, and, by definition, $t(s_2) \leq s_1$. Putting these facts together, we obtain $\delta p(t(s_1)) + c(\ell(s_1)) \leq \delta p(t(s_2)) + c(\ell(s_2))$. Hence, $p(s_1) \leq p(s_2)$, as was to be shown. $\square$

**Lemma 7.14.** *For all $s \in [0,1]$ we have $p(s) \geq c'(0)s$.*

*Proof.* For $s = 0$ the result is obvious, so suppose that $s > 0$. Let $\sigma \leq \min\{s, h\}$. Using this inequality, convexity of $p$ and $c$, and the fact that $p(h) = c(h)$, we obtain $c'(0) \leq c(h)/h = p(h)/h = p(\sigma)/\sigma \leq p(s)/s$. $\square$

**Lemma 7.15.** *Let $s \in G$. If $c'(s) \leq \delta c'(0)$, then $t(s) = 0$.*

*Proof.* Let $s \in G$ with $c'(s) \leq \delta c'(0)$. The value $t(s)$ uniquely solves $\min_{0 \leq t \leq s-h}\{\delta p(t) + c(s-t)\}$. This solution is zero if $c(s) < \delta p(t) + c(s-t)$ when $0 < t \leq s - h$, or, equivalently,

$$\frac{c(s) - c(s-t)}{t} < \frac{\delta p(t)}{t} \quad \text{when } 0 < t \leq s - h.$$

This inequality is valid, because, by strict convexity of $c$, the assumption $c'(s) \leq \delta c'(0)$ and lemma 7.14,

$$\frac{c(s) - c(s-t)}{t} < c'(s) \leq c'(0) \leq \frac{\delta p(t)}{t}$$

for all $t$ with $0 < t \leq s$. $\square$

For the remainder of this section, we adopt the setting of theorem 7.1. In particular, we consider a sequence of grids $\{G_n\}$ $G_n = \{0, h_n, 2h_n, \ldots, 1\}$ with $h_n = 2^{-n}$. The corresponding functions are denoted by $\{p_n\}$.

**Lemma 7.16.** *The sequence $\{p_n\}_{n=1}^{\infty}$ is pointwise monotone decreasing.*

*Proof.* Fix $n \in \mathbb{N}$. The claim is that $p_{n+1}(s) \leq p_n(s)$ for all $s \in [0,1]$. Let $s_1$ and $s_2 = s_1 + h_n$ be two consecutive points in $G_n$, and suppose that $p_{n+1} \leq p_n$ on $[0, s_1]$. We claim that $p_{n+1}(s_2) \leq p_n(s_2)$ also holds. This is sufficient for $p_{n+1}(s) \leq p_n(s)$ on $[s_1, s_2]$, given that (i) $p_n(s)$ is a linear interpolation from $s_1$ to $s_2$, (ii) $p_{n+1}$ is a linear interpolation from $s_1$ to $s' := (s_1 + s_2)/2$ and then $s'$ to $s_2$, and (iii) $p_{n+1}$ is convex.

To show that $p_{n+1}(s_2) \leq p_n(s_2)$, recall that, by the induction hypothesis, we have $p_{n+1} \leq p_n$ on $[0, s_1]$. It follows that

$$\begin{aligned} p_{n+1}(s_2) &= \min_{t \leq s'} \{\delta p_{n+1}(t) + c(s_2 - t)\} \\ &\leq \min_{t \leq s_1} \{\delta p_{n+1}(t) + c(s_2 - t)\} \\ &\leq \min_{t \leq s_1} \{\delta p_n(t) + c(s_2 - t)\} \\ &= p_n(s_2), \end{aligned}$$

as was to be shown. □

**Lemma 7.17.** *The sequence $\{q_n\}_{n=1}^{\infty}$ is uniformly bounded and equicontinuous.*

*Proof.* The statement that $\{p_n\}_{n=1}^{\infty}$ is uniformly bounded means that there exists a constant $M$ independent of $n$ with $\sup_{0 \leq s \leq 1} |p_n(s)| \leq M$. This is clearly true because, on one hand, $p_n$ is nonnegative (see, e.g., lemma 7.14), and on the other hand, since $\{p_n\}$ is monotone decreasing (lemma 7.16) and each $p_n$ is increasing (lemma 7.13), we have $p_n(s) \leq p_1(1)$ for all $s$ and $n$.

It remains to show that $\{p_n\}_{n=1}^{\infty}$ is equicontinuous. Given that $p_n$ is increasing and convex (lemma 7.11), a sufficient condition for equicontinuity is

$$\exists\, M, K \in \mathbb{N} \quad \text{s.t.} \quad \frac{p_n(1) - p_n(1 - h_n)}{h_n} \leq K \text{ for all } n \geq M \tag{16}$$

(In other words, the slope of the function $p_n$ over the last two grid points is bounded independent of $n$. Note that the bound in (16) only has to be checked for $n$ greater than some finite $M$, because finite families of continuous functions on compact sets are always equicontinuous.) In what follows, we simplify notation dropping the $n$ subscript attached to $G$, $p$, $h$, $t$ and $\ell$ ($t$ and $\ell$ being the functions described in lemma 7.12).

To begin the proof, first observe that if $s \in G$, then

$$p(s) - p(s-h) \leq \delta\{p(t(s)) - p(t(s-h))\} + c'(1)h, \tag{17}$$

as follows the fact that, by lemma 7.12 and convexity of $c$,

$$c(\ell(s)) - c(\ell(s-h)) \leq c'(1)\{\ell(s) - \ell(s-h)\} \leq c'(1)\{s - (s-h)\} = c'(1)h$$

It is also the case that if $s \in G$ and $\sigma(s) := \inf\{r \in G : r \geq t(s)\}$, then

$$p(t(s)) - p(t(s-h)) \leq p(\sigma(s)) - p(\sigma(s) - h) \tag{18}$$

This inequality follows from lemma 7.12, which tells us that $t(s) - t(s-h) \leq s - (s-h) = h$, or $t(s-h) \geq t(s) - h$. Hence, applying monotonicity and convexity of $p$, we have $p(t(s)) - p(t(s-h)) \leq p(t(s)) - p(t(s) - h) \leq p(\sigma(s)) - p(\sigma(s) - h)$, which is (18). Combining (17) and (18), we obtain the recursion

$$p(s) - p(s-h) \leq \delta\{q(\sigma(s)) - q(\sigma(s) - h)\} + c'(1)h. \tag{19}$$

In particular, if we define $s_j := \sigma^j(1) = \sigma \circ \sigma \circ \cdots \circ \sigma(1)$, then, from (19),

$$p(s_j) - p(s_j - h) \leq \delta\{p(s_{j+1}) - p(s_{j+1} - h)\} + c'(1)h. \tag{20}$$

Now let $\bar{s}$ be the $s \in (0, 1]$ that satisfies $c'(s) = \delta c'(0)$. The significance of $\bar{s}$ is that, in view of lemma 7.15, we have $\ell(\bar{s}) = \bar{s}$, and $t(\bar{s}) = 0$. Moreover, since $\ell$ is increasing, if $s_j \geq \bar{s}$, then $\ell(s_j) \geq \bar{s}$. As a consequence, provided that $s_j \geq \bar{s}$, we have

$$s_{j+1} \leq t(s_j) + h = s_j - \ell(s_j) + h \leq s_j - \bar{s} + h.$$

Starting at $s_0 = 1$ and iterating on this inequality, we obtain

$$s_j \leq 1 - j(\bar{s} - h) \tag{21}$$

Recall that $h = h_n$ in fact depends on $n$, but $\bar{s}$ does not (see lemma 7.15). It follows that there exists an $M \in \mathbb{N}$ such that $\bar{s} - h_M > 0$. Applying (21), we can then take $J$ independent of $n$ with $s_J = 0$. Iterating backwards on (20) starting at $s_0 = 1$, we have

$$p(1) - p(1-h) \leq \delta^J\{p(s_J) - p(s_J - h)\} + c'(1)h\sum_{i=1}^{J}\delta^i = c'(1)h\sum_{i=1}^{J}\delta^i.$$

Dividing through by $h$ gives (16), and equicontinuity is established. □

Since $\{p_n\}_{n=1}^{\infty}$ is monotonically decreasing and bounded below by zero, the function $\bar{p} := \lim_{n\to\infty} p_n$ is well-defined.

**Lemma 7.18.** *The function $\bar{p}$ is continuous, and $\{p_n\}_{n=1}^{\infty}$ converges to $\bar{p}$ uniformly.*

*Proof.* Lemma 7.17 and the Arzelà -Ascoli theorem imply that $\{p_n\}_{n=1}^{\infty}$ has a uniformly convergent subsequence. Since $\{p_n\}_{n=1}^{\infty}$ is monotone decreasing and converges pointwise to $\bar{p}$, the entire sequence converges uniformly to $\bar{p}$. Since continuity is preserved under uniform limits, the function $\bar{p}$ is continuous. □

**Theorem 7.2.** *The function $\bar{p}$ is a solution to the equilibrium price equation (2).*

*Proof.* Evidently $\bar{p}(0) = 0$. In view of lemma 7.14, $\bar{p}$ is not the zero function (i.e., nontrivial). It remains only to show that (2) holds. Since both $\bar{p}$ and $c$ are continuous, the left-hand side and right-hand side of (2) are both continuous in $s$ (the right-hand side by the theorem of the maximum). Since continuous functions that agree on a dense subset of $(0,1]$ must agree everywhere on $(0,1]$ it suffices to show that (2) holds for all $s > 0$ in the dyadic rationals $\cup_n G_n$. We now fix such $s \in \cup_n G_n$ and show that (2) holds.

Since $s \in \cup_n G_n$, there exists an $N_1 \in \mathbb{N}$ such that $s \in G_n$ whenever $n \geq N_1$. For such $n$ we have $s \in G_n$ and $s > 0$, implying that (11) holds. In particular,

$$p_n(s) = \min_{t \leq s - h_n} \{\delta p_n(t) + c(s-t)\},$$

and hence

$$\bar{p}(s) = \lim_{n\to\infty} p_n(s) = \lim_{n\to\infty} \min_{t \leq s-h_n} \{\delta p_n(t) + c(s-t)\}.$$

It is therefore sufficient for the theorem to establish that this expression agrees with the right-hand side of (2). In other words, we aim to show that

$$\lim_{n\to\infty} \min_{t \leq s-h_n} \{\delta p_n(t) + c(s-t)\} = \min_{t \leq s} \{\delta \bar{p}(t) + c(s-t)\}. \tag{22}$$

Here the left-hand side is the limit of $p_n(s)$, which is monotonically decreasing. Regarding the right-hand side, we write $g(t) = \delta \bar{p}(t) + c(s-t)$. Fixing $\epsilon > 0$, the problem is then to show that

$$\exists\, N \in \mathbb{N} \text{ s.t. } n \geq N \implies \min_{t \leq s-h_n} \{\delta p_n(t) + c(s-t)\} < \min_{t \leq s} g(t) + \epsilon. \tag{23}$$

Since $p_n \to \bar{p}$ uniformly, we choose $N_2 \in \mathbb{N}$ such that $n \geq N_2$ implies $\sup_x (p_n(x) - \bar{p}(x)) < \epsilon/2$. For such $n$, we have

$$\min_{t \leq s-h_n} \{\delta p_n(t) + c(s-t)\} \leq \min_{t \leq s-h_n} \{\delta p(t) + \epsilon/2 + c(s-t)\}.$$

To summarize,

$$n \geq \max\{N_1, N_2\} \implies \min_{t \leq s-h_n} \{\delta p_n(t) + c(s-t)\} \leq \min_{t \leq s-h_n} g(t) + \epsilon/2.$$

Since $g$ is continuous on $[0,s]$ and $h_n \downarrow 0$, we can choose an $N_3$ such that

$$n \geq N_3 \implies \min_{t \leq s - h_n} g(t) \leq \min_{t \leq s} g(t) + \epsilon/2.$$

Combining these last two inequalities, we have established that (23) holds when $N := \max\{N_1, N_2, N_3\}$. □

# References


[1] Aghion, Philippe, and Richard Holden, "Incomplete Contracts and the Theory of the Firm: What Have We Learned in the Past 25 Years," *Journal of Economic Perspectives* 25:2 (2011), 181–197.

[2] Antràs, Pol, Davin Chor, "Organizing the Global Value Chain," mimeo (2011).

[3] Antràs, Pol, Davin Chor, Thibault Fally, and Russell Hillberry, "Measuring the Upstreamness of Production and Trade Flows," *American Economic Review Papers and Proceedings* 102:3 (2012), 412–416.

[4] Arrow, Kenneth J., "The Organization of Economic Activity: Issues Pertinent to the Choice of Market versus Non-market Allocations" (pp. 47–64) in *Analysis and Evaluation of Public Expenditures: The PPP System, Volume 1* (Washington, DC: Government Printing Office, 1969)

[5] Arrow, Kenneth J., *The Limits of Organization* (New York, NY: Norton, 1974).

[6] Arrow, Kenneth J., "Vertical Integration and Communication," *Bell Journal of Economics* 6:1 (1975), 173–83.

[7] Becker, Gary S., and Kevin M. Murphy, "The Division of Labor, Coordination Costs, and Knowledge," *Quarterly Journal of Economics* 107:4 (1992), 1137–1160.

[8] Benveniste, Lawrence M., and Jose A. Scheinkman, "On the Differentiabilityof the Value Function in Dynamic Models of Economics," *Econometrica* 47:3 (1979) 727–32.

[9] Blau, Peter M., and Marshall W. Meyer, *Bureaucracy in Modern Society* (New York, NY: Random House, 1987).

[10] Bolton, Patrick, and Mathias Dewatripont, "The Firm as a Communication Network," *Quarterly Journal of Economics* 109:4 (1994) 809–839.

[11] Bresnahan, Timothy F., and Jonathan D. Levin, "Vertical Integration and Market Structure," NBER working paper no. 17889 (2012).

[12] Chandler, Alfred D., *The Visible Hand: The Managerial Revolution in American Business* (Harvard University Press, 1977)

[13] Coase, Ronald H., "The Nature of the Firm," *Economica* 4:16 (1937), 386–405.

[14] Costinot, Arnaud, "On the Origins of Comparative Advantage," *Journal of International Economics* 77:2 (2009), 255–264.

[15] Costinot, Arnaud, Jonathan Vogel, and Su Wang, "An Elementary Theory of Global Supply Chains," NBER working paper no. 16936 (2011).

[16] Demsetz, Harold, "Theory of the Firm Revisisted," *Journal of Law and Economics* 4:1 (1988), 141–162.

[17] Fally, Thibault, "On the Fragmentation of Production in the U.S.," mimeo (2012).

[18] Gabre-Madhin, Eleni Z., "Market Institutions,Transaction Costs, and Social Capital in the Ethiopian Grain Market," International Food Policy Research Institute Research Report 124.

[19] Garicano, Luis, "Hierarchies and the Organization of Knowledge in Production," *Journal of Political Economy* 108:5 (2000), 874–904.

[20] Geanakoplos, John, and Paul Milgrom, "A Theory of Hierarchies Based on Limited Managerial Attention," *Journal of the Japanese and International Economies* 5:3 (1991), 205–225.

[21] Grossman, Gene, and Elhanan Helpman, "Integration versus Outsourcing in Industry Equilibrium," *Quarterly Journal of Economics* 117:1 (2002), 85–120

[22] Grossman, Gene, and Esteban Rossi-Hansberg, "Trading Tasks: A Simple Theory of Offshoring," *American Economic Review* 98:5 (2008), 1978–1997.

[23] Grossman, Gene, and Esteban Rossi-Hansberg, "Task Trade between Similar Countries," *Econometrica* 80:2 (2012), 593–629.

[24] Grossman, Sandford, and Oliver Hart, "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy* 94:4 (1986), 691–719.

[25] Hart, Oliver, and John Moore, "Property Rights and the Nature of the Firm," *Journal of Political Economy* 98:6 (1990), 1119–1158.

[26] Hayek, Friedrich A., "The Use of Knowledge in Society," *American Economic Review* 35:4 (1945), 519–530.

[27] Holmstrom, Bengt, and Paul Milgrom, "The Firm as an Incentive System," *American Economic Review* 84:4 (1994), 972–991.

[28] Houthakker, Hendrik S., "Economics and Biology: Specialization and Speciation," *Kyklos* 9, 181–189.

[29] Jensen, Michael C., and William H. Meckling, "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure," *Journal of Financial Economics* 3:4 (1976), 305–360.

[30] Klein, Benjamin, Robert Crawford, and Armen Alchian, "Vertical Integration, Appropriable Rents, and the Competitive Contracting Process," *Journal of Law and Economics* 21:2 (1978), 297–326.

[31] Lucas, E. Robert, "On the Size Distribution of Business Firms," *Bell Journal of Economics* 9:2 (1978), 508-523.

[32] McAfee, Preston and John McMillan, "Organizational Diseconomies of Scale," *Journal of Economics and Management Strategy* 4:3 (1995), 399–426.

[33] North, Douglass C., "Institutions, Transaction Costs and Productivity in the Long Run," Washington University in St. Louis, Economics working paper archive no. 9309004 (1993).

[34] Polanec, Sašo, "On the Evolution of Size and Productivity in Transition: Evidence from Slovenian Manufacturing Firms," KU Leuven LICOS Centre for Transition Economics, Discussion Paper 154/2004(2004).

[35] Rasmusen, Eric B., and Todd Zenger, "Diseconomies of Scale in Employment Contracts," *Journal of Law, Economics and Organization* 6:1 (1990), 65–92.

[36] Robertson, Dennis H., *Control of Industry* (London: Nisbet and Co., 1923).

[37] Sandefur, Justin, "On the Evolution of the Firm Size Distribution in an African Economy," The Centre for the Study of African Economies, Oxford University, Working Paper 2010-05 (2010).

[38] Seabright, Paul, *The Company of Strangers: a Natural History of Economic Life* (Princeton, NJ: Priceton University Press, 2010)

[39] Simon, Robert A., *Administrative Behavior, a Study of Decision Making Processes in Administrative Organization* (New York, NY: The Macmillan Co., 1947)

[40] Smith, Adam, *The Wealth of Nations* (1776, reprinted by University of Chicago Press in 1976)

[41] Tsuji, Masasugu, "Automotive Clusters in Japan: The Aichi Region," mimeo, OSIPP, Osaka University (2004).

[42] Williamson, Oliver E., "Transaction-Cost Economics: The Governance of Contractual Relations," *Journal of Law and Economics* 22:2 (1979), 233–261.

[43] Williamson, Oliver E., "The Economics of Organization: The Transaction Cost Approach," *American Journal of Sociology* 87:3 (1981), 548–577.

[44] Williamson, Oliver E., *Markets and Hierarchies: Analysis and Antitrust Implications* (New York, NY: Free Press, 1975).

[45] Yang, Xiaokai and Yew-Kwang Ng, *Specialization and Economic Organization* (Amsterdam: North-Holland, 1993).