Productivity and Credibility in Industry Equilibrium*

Michael Powell
Northwestern University

June 2016

Abstract

I analyze a model of production in a competitive environment with heterogeneous firms. Efficient production requires individuals within the organization to take noncontractible actions for which rewards must be informally promised rather than contractually assured. The credibility of such promises originates from a firm's future competitive rents. In equilibrium, heterogeneous firms are heterogeneously constrained, and competitive rents are inefficiently concentrated at the top. I explore several policy and empirical implications of this result.

Keywords: relational contracting, productivity, competition

JEL classifications: D24, L14, L22

*Powell: Kellogg School of Management, mike-powell@kellogg.northwestern.edu. This paper is a revised chapter of my MIT Ph.D. thesis. I am indebted to Robert Gibbons for his advice. This paper has benefitted from the suggestions of Daron Acemoglu, Abhijit Banerjee, Daniel Barron, David Berger, Nick Bloom, Dan Cao, Ricardo Caballero, Cheng Chen, Sebastian di Tella, Glenn Ellison, Matthias Fahn, Luis Garicano, Jean-Paul L'Huillier, Hongyi Li, Niko Matouschek, Iuliana Pascu, Michael Peters, Andrea Prat, Luis Rayo, Alp Simsek, Jeroen Swinkels, Chad Syverson, Steven Tadelis, Heiwai Tang, Eric Van den Steen, Xiao Yu Wang, and seminar participants at Caltech, Chicago, Columbia, Duke, HKU, LSE, MIT, Northwestern, NYU, Queen's, Rochester, Stanford, Toronto, UPenn, and USC. I thank MIT Sloan's Program on Innovation in Markets and Organizations and the National Science Foundation for financial support.

1 Introduction

To make a firm more productive, managers have to figure out how to get a given set of people with a given set of resources to work together more effectively. This can be done through harder work—asking for more personal sacrifice on the part of the workers. It can be done through smarter work—putting in place more effective management practices. And it can be done through improving internal resource allocation—ensuring that the right people have the right resources for the job at hand. Getting people to make sacrifices, getting them to cooperate with new managerial initiatives, and getting them to use the firm's resources appropriately requires that they be rewarded for doing so. However, many of these objectives and whether they have been met are not easily describable to third-party enforcers. Instead, firms have to rely, at least in part, on informal promises of rewards.[1] A firm's ability to improve its productivity is therefore constrained by its ability to make credible promises. In this paper, I explore the question of why some firms are able to put in place effective practices and others are not by studying how credibility originates in a model of competition among heterogeneous firms.

I model credibility as self enforcement in a repeated game (Bull, 1987; MacLeod and Malcomson, 1989; Levin, 2003) between a firm's owner and a team of managers. The owner allocates resources to each manager. She would like each manager to utilize those resources appropriately, but formal contracts are unavailable. She can promise to reward the manager for utilizing the resources, but she lacks commitment. In a one-shot game, the owner would never pay the reward, forward-looking managers will squander the firm's resources, and they will not be allocated any resources to begin with. A long-lived firm, however, can credibly promise future rewards, since failure to uphold such promises can put the future of the firm at stake: the firm's future competitive rents can be used as collateral in the firm's promises.

Explicitly modeling the source of competitive rents is therefore important for understanding the opportunities possessed by individual firms. As in Lucas (1978), output is sold into a competitive product market, which consists of many firm owners of heterogeneous ability, and production exhibits decreasing returns to scale. These features imply that firms of different total factor productivity will coexist in equilibrium. Moreover, firms of different *marginal* productivity will coexist in equilibrium, even though all firms face the same factor prices: heterogeneous firms will be heterogeneously constrained, and therefore there will be misallocation of production. The credibility

[1]See Malcomson (2013) for a survey on the importance of informal agreements for motivating effort; Gibbons and Henderson (2013) on how productivity-enhancing managerial practices rely on informal agreements; and Bloom, Sadun, and Van Reenen (2012) on how lack of trust constrains decentralization and therefore productivity.

necessary to sustain decentralization is determined by each firm's potential future competitive rents. Competitive rents, credibility, firms' decentralization levels, and therefore firms' productivity levels are jointly determined in industry equilibrium. The model offers two sets of results.

First, by augmenting a standard Neoclassical production function with noncontractible managerial decisions, the model provides a framework for connecting the Bloom, Sadun, and Van Reenen (2015) view of good managerial practices as a technology with the observed heterogeneity in such practices. This approach highlights the scarcity of credibility as a barrier to the spread of such practices and illustrates how the distribution of good managerial practices depends on the underlying environment in which firms operate. The model delivers patterns that are consistent with three sets of facts:

1. Firm organization and firm performance: more-decentralized firms tend to be more productive (Bloom, Sadun, and Van Reenen, 2012) and better able to cope with economic downturns (Aghion, Bloom, Lucking, Sadun, and Van Reenen, 2015), yet there is a lot of heterogeneity in how decentralized firms are (Bloom, Sadun, and Van Reenen, 2009).

2. Productivity dynamics and firm organization: within-firm productivity is procyclical (Bartelsman and Doms, 2000), productivity dispersion tends to be countercyclical (Kehrig, 2015), and less-decentralized firms are more susceptible to economic downturns (Aghion, Bloom, Lucking, Sadun, and Van Reenen, 2015).

3. Cross-country productivity dispersion and firm organization: in lower-income countries, there is more productivity dispersion (Hsieh and Klenow, 2009; Bartelsman, Haltiwanger, and Scarpetta, 2013) and a thicker left tail of less-productive (Hsieh and Klenow, 2009), less-decentralized (Bloom, Sadun, and Van Reenen, 2012) firms.

Second, because competitive rents serve as collateral, their allocation matters for efficiency. In equilibrium, a high-ability owner expects to earn high levels of competitive rents in the future, which increases her ability to decentralize today. This positive-feedback loop is limited by decreasing returns to scale, but it nevertheless results in aggregate inefficiencies: competitive rents are allocated too progressively. High-ability firms overproduce, imposing first-order pecuniary externality losses on low-ability firms. The competitive equilibrium is therefore constrained-inefficient, and policies that redistribute profits away from the most profitable firms, such as an excise tax with an exemption for small firms, may increase overall welfare.

Related Literature This paper is related to the literature on the determinants of the large and persistent differences in productivity levels across producers (for a survey, see Syverson (2011)), and it is methodologically related to Board and Meyer-ter-Vehn (2015), which augments an efficiency-wage model with on-the-job search and shows that wage and productivity dispersion emerges in a stationary industry equilibrium with ex ante identical firms, and this wage dispersion motivates workers. In my model, the credibility of firms' incentive schemes are constrained by the competitive rents they expect to earn. In equilibrium, ex ante heterogeneity leads some firms to be more constrained than others, and I focus on the implications of these differences.

Also closely related are Chassang (2010) and Gibbons and Henderson (2013), which argue that productivity differences are due to differences in (ex ante identical) firms' success in developing efficient relational contracts. I assume that all firms succeed in implementing optimal relational contracts and show that optimal relational contracts can amplify existing differences. My analysis does not address firm dynamics, unlike Chassang (2010), Ellison and Holden (2014), and Li, Matouschek, and Powell (forthcoming). It provides a theory of steady-state dispersion in organizational practices and productivity, not a theory of the process that leads to it.

The pecuniary externality that I identify is related to work by Schroth (2016), which highlights an intertemporal pecuniary externality in the context of financial intermediation with limited commitment and shows that policy interventions that increase rents in the future (by capping bank size) can be welfare-enhancing by trading off future distortions against the current distortions experienced by growing banks. In contrast, my analysis highlights a steady-state pecuniary externality that large firms impose on small firms and shows that policy interventions that redistribute rents from large firms to small firms can be welfare-enhancing.

Further, this paper contributes to the literatures on firm governance in industry equilibrium (Grossman and Helpman, 2002; Gibbons, Holden, and Powell, 2012; Legros and Newman, 2013; Alfaro, Conconi, Fadinger, and Newman, forthcoming; and Wu, forthcoming) and on the aggregate implications of contractual incompleteness (Caballero and Hammour, 1998; Francois and Roberts, 2003; Martimort and Verdier, 2004; Cooley, Marimon, and Quadrini, 2004; and Acemoglu, Antras, and Helpman, 2007). My analysis is most similar to Acemoglu, Antras, and Helpman (2007), which examines the impact of unresolved hold-up on technology adoption. In contrast, I explore how firms' ability to resolve contractual incompleteness using relational contracts varies with underlying firm characteristics and with the competitive environment.

Finally, this paper is related to the literature on misallocation and economic growth (Banerjee and Duflo, 2005; Jeong and Townsend, 2007; Restuccia and Rogerson, 2008; and Hsieh and

Klenow, 2009), which has argued that cross-country differences in the efficiency of resource allocation across firms can explain a substantial portion of the differences in GDP per capita. How to improve resource allocation across firms depends on why resources were not allocated efficiently to begin with. Several recent papers in the macro tradition (Banerjee and Moll, 2010; Buera, Kaboski, and Shin, 2011; Midrigan and Xu; and Moll, 2014) focus on the role of underdeveloped financial markets. Other explanations include heterogeneous markups that distort relative output prices (Peters, 2013), adjustment costs (Asker, Collard-Wexler, and De Loecker, 2014), and size-dependent public policies (Guner, Ventura, and Xu, 2008; and Garicano, Lelarge and Van Reenen, forthcoming). My model provides an alternative and complementary mechanism that generates persistent misallocation in a perfectly competitive environment with no adjustment costs or credit rationing.

2 The Model

There is a unit mass of firms, indexed by $i \in [0, 1]$, each run by a risk-neutral owner (she) who is the residual claimant. Output requires managers (he), who must be given resources to be productive. The managers' utilization of these resources is noncontractible. Managers are homogeneous and risk-neutral, and they are on the long side of the market, so that in equilibrium, they will receive no rents. It will be notationally convenient, but not consequential, to assume that each firm draws potential managers from its own firm-specific pool, so I make this assumption.

Play is infinitely repeated, and I denote the period by $t = 0, 1, 2, \ldots$. All players share a common discount factor $\delta < 1$. Output is homogeneous across firms and sold into a competitive product market. Aggregate demand is stationary, $D_t(p_t) = D(p_t)$, where p_t is the output price in period t, smooth, downward-sloping, it satisfies $\lim_{p \to 0} D(p) = \infty$ and $\lim_{p \to \infty} D(p) = 0$, and it is generated by consumers who have quasilinear preferences.

Stage Game At the beginning of the stage game, owner i decides whether to pay a fixed cost of production, F. If she pays the fixed cost, she then decides on a mass M_{it} of managers to whom to make an offer. She offers each manager $m \in [0, M_{it}]$ a triple $(r_{itm}, s_{itm}, b_{itm})$, where r_{itm} is a level of discretionary resources she allocates to manager m, s_{itm} is a noncontingent payment made upon acceptance, and b_{itm} is a reward she promises to pay manager m if and only if he utilizes all the resources he has been allocated.

Each manager m then decides whether to accept this proposal or to reject it in favor of an outside opportunity that yields exogenous utility $W > 0$. If manager m accepts, the owner transfers resources r_{itm} to him, and he chooses a resource-utilization level $\hat{r}_{itm} \geq 0$ and keeps the remaining resources, $r_{itm} - \hat{r}_{itm}$, which he values dollar-for-dollar. This utilization choice is commonly observed, and the owner subsequently decides whether to pay manager m a reward of b_{itm}. Output for firm i is then realized and sold at price p_t.

Technology and Profits Owners have heterogeneous ability, which I denote by φ_i. The realized distribution of ability is given by the distribution function Φ, which is absolutely continuous. Given mass M_{it} of managers who choose utilization levels $\hat{r}_{it} \equiv \{\hat{r}_{itm}\}_{m \in M_{it}}$, firm i's production in period t is given by

$$y_i\left(\hat{r}_{it}, M_{it}\right) = \varphi_i \left(\int_0^{M_{it}} \hat{r}_{itm}^{\frac{\theta}{1-\theta}} dm \right)^{1-\theta}.$$

Utilization levels across managers are substitutes. I assume $\theta < 1/2$, which ensures that the unconstrained problem has a solution, and this solution can be characterized by the firm's first-order conditions. In period t, if owner i pays all promised rewards, her profits are

$$\pi_i\left(\hat{r}_{it}, M_{it}, p_t\right) = p_t y_i\left(\hat{r}_{it}, M_{it}\right) - \int_0^{M_{it}} \left(r_{itm} + s_{itm} + b_{itm}\right) dm - F.$$

As a benchmark, Section 3 analyzes the model under the assumption that utilization choices are contractible, which eliminates the need to use relational incentives. Section 4 examines the relational-incentives case.

Equilibrium Industry equilibrium has to specify (a) for each firm i, the complete plan for the relationship between the owner and her managers and (b) how these plans within each firm aggregate up to determine industry-wide variables.

To describe the former, I define a **relational contract** for firm i as a complete contingent plan for its relationships with its managers, which specifies management choices $\{M_{it}\}_t$, offers $(r_{itm}, s_{itm}, b_{itm})_{tm}$, and utilization choices $\{\hat{r}_{itm}\}_{tm}$ as a function of the history of past play within the firm as well as the history of output prices up to, and including, date t. A relational contract is **self-enforcing** if it describes a subgame-perfect equilibrium of the game between owner i and her managers. Note that I am implicitly assuming that firm i's actions can depend on firm j's actions only inasmuch as the latter affect output prices p_t. An **optimal relational contract** for firm i is a self-enforcing relational contract that yields higher expected profits for firm i than any other self-enforcing relational contract.

The notion of industry equilibrium that I will use is one in which all firms conjecture the same price sequence and choose optimal relational contracts, and this conjectured price sequence clears the output market in each period. Formally, a **rational-expectations equilibrium (REE)** is a set of sequences of prices $\{p_t\}_t$, management $\{M_{it}\}_{it}$, offers $(r_{itm}, s_{itm}, b_{itm})_{itm}$, and utilization choices $\{\hat{r}_{itm}\}_{itm}$ such that at each time t :

1. Given promised reward b_{itm} and resources r_{itm}, manager m for firm i optimally chooses $\hat{r}_{itm} = r_{itm}$.

2. Given the conjectured price sequence $\{p_t\}_t$, owner i optimally chooses management levels $\{M_{it}\}_t$ and makes offers $(r_{itm}, s_{itm}, b_{itm})_{tm}$.

3. p_t clears the output market in period t.

A rational-expectations equilibrium is a **stationary REE** if prices are constant: $p_t = p$.

3 Complete-Contracts Benchmark

Production described in the previous section differs from standard Neoclassical production in two respects. First, managers are not passive after accepting employment, since they make resource-utilization choices. Second, these utilization choices are noncontractible. To isolate the implications of the first of these assumptions, I derive the firm's optimal production decisions when utilization choices are contractible.

If utilization choices are contractible, the owner can directly choose each manager's utilization level and use the payment, s_{itm} to pin him to his participation constraint. Because there are no intertemporal linkages in the problem, each firm solves its profit-maximization problem period-by-period. Given a price level p_t, owner i chooses M_{it}, and $\{r_{itm}\}_m$ to solve the following problem.

$$\max_{\{r_{itm}, s_{itm}\}_m, M_{it}} p_t \varphi_i \left(\int_0^{M_{it}} r_{itm}^{\frac{\theta}{1-\theta}} dm \right)^{1-\theta} - \int_0^{M_{it}} (r_{itm} + s_{itm})\, dm - F \tag{1}$$

subject to each manager's participation constraint: $s_{itm} \geq W$.

At the optimum, managers' participation constraints hold with equality. Additionally, since $\theta < 1/2$, the firm's problem is concave in $\{r_{itm}\}_m$, and since managers are symmetric, any optimal solution must satisfy $r_{itm} = r_{it}$ for all m. The production function therefore collapses into a Cobb-Douglas production function, but the costs are not linear in (r_{it}, M_{it}), since they depend on the total amount of resources allocated to managers, $r_{it} M_{it}$. The firm's problem becomes

$$\max_{r_{it}, M_{it}} p_t \varphi_i r_{it}^{\theta} M_{it}^{1-\theta} - (W + r_{it})\, M_{it} - F.$$

There will be a shutdown value of ability, φ_S, for which all firms with ability $\varphi_i < \varphi_S$ optimally choose not to produce. The solution to the firm's problem is summarized in the following proposition. Define $r^{FB} \equiv W\theta / (1-2\theta)$.

Proposition 1. If $\varphi_i < \varphi_S = F^{\theta} \left(r^{FB}\right)^{1-2\theta} / (\theta p)$, firm i does not produce. If $\varphi_i \geq \varphi_S$, firm i chooses $r^{FB}(\varphi_i, p_t) = r^{FB}$ and

$$M^{FB}(\varphi_i, p_t) = (\theta p_t \varphi_i)^{1/\theta} \left(r^{FB}\right)^{(\theta-1)/\theta}.$$

Equilibrium total factor productivity for a firm with ability φ_i is

$$A_i^{FB}(\varphi_i) = y_i / \left(M_i^{1-\theta}\right) = \varphi_i \left(r^{FB}\right)^{\theta}.$$

Proof of Proposition 1. See appendix.

First, observe that r^{FB} does not depend on φ_i or on p_t. When resource utilization is contractible, higher-ability firms or firms that face higher output prices produce more by hiring more managers rather than by allocating more resources to each manager. To see why this is true, note that output can be written $\varphi_i R_{it}^{\theta} M_{it}^{1-2\theta}$, where R_{it} is the total resources allocated to managers. Under this parametrization, costs are linear in R_{it} and M_{it}, so this transformed problem is a standard profit-maximization problem with a Cobb-Douglas production function, and the solution therefore has constant expenditure shares: $R^{FB} / \left(WM^{FB}\right) = r^{FB}/W = \theta/(1-2\theta)$ is independent of the quantity produced. Next, the solution to the period-t problem does not depend on variables from any other period, and demand is stationary, so output prices will be constant. A stationary REE is then a price level p and a vector of firm-level choices $\{r_i, M_i\}_i$ such that these choices are optimal in each period given the price level, and the price level clears the output market in each period. It is straightforward to verify that a stationary REE exists, is unique, and it is Pareto-efficient. It is also worth noting that firm i's equilibrium total factor productivity depends only on the firm's ability, φ_i, and the first-best level of resource utilization, r^{FB}. In particular, equilibrium total factor productivity does not depend on equilibrium output prices p. This result will stand in contrast to the results of the next section, where managers' utilization choices are noncontractible.

4 Relational Incentive Contracts

I now turn to the heart of the model and assume that resource utilization is noncontractible. The owner would like to provide incentives for her managers to utilize resources, but she can only do so by making a promise that she will pay pre-specified rewards if her managers choose particular utilization levels. She lacks commitment and therefore would prefer to renege on these rewards once utilization choices have been made. However, she may use her firm's future competitive rents as a partial commitment device. Her ability to do so depends on the clarity with which her failure to pay promised rewards is communicated to her pool of managers. Throughout, I make the stark assumption that deviations are perfectly observable.

Assumption 1. A firm's pool of managers commonly observes allocated resources and utilization choices of individual managers and whether they were paid their promised rewards.

The upshot of Assumption 1 is that the totality of a firm's future competitive rents can be used as collateral in its promises. In Section 6, I discuss how weaker observability further constrains firms, serving effectively as a reduction in the firm's discount factor.

4.1 Dynamic Enforcement

This section provides conditions under which utilization levels $\{\hat{r}_{itm}\}_{tm}$ are sustainable as part of a relational contract. Given a price sequence $\{p_t\}_t$, I will say that $\{\hat{r}_{itm}\}_{tm}$ is **sustainable** if there exists a self-enforcing relational contract in the game between owner i and her managers in which each manager m utilizes $\hat{r}_{itm}$ at time t. Further, I will say that a relational contract involves **grim-trigger punishment** if, off the equilibrium path, managers are never allocated resources, managers reject all offers and utilize no resources, and the owner never pays any bonuses, and I will say that a relational contract is a **full-utilization relational contract** if $r_{itm} = \hat{r}_{itm}$ for all t and m. The appendix describes the necessary and sufficient conditions for a relational contract to be self-enforcing and shows that, in order to characterize the set of implementable $\{\hat{r}_{itm}\}_{tm}$, we can focus on a simple class of equilibria.

Lemma 1. If $\{\hat{r}_{itm}\}_{tm}$ is sustainable, it is sustainable by a full-utilization relational contract with grim-trigger punishment.

Proof of Lemma 1. See appendix.

Lemma 1 has two parts. It first shows that it is without loss to focus on grim-trigger punishment, since such punishment constitutes an optimal penal code. The second part shows that it is without loss to focus on full-utilization relational contracts, since any payments made to manager m through allowing manager m to divert resources can alternatively be made ex ante through the salary portion of his payment. Proposition 2 provides necessary and sufficient conditions for $\{r_{itm}\}_{tm}$ to be sustainable.

Proposition 2. Given $\{p_t\}_t$, utilization choices $\{r_{itm}\}_{tm}$ are sustainable if and only if

$$\int_0^{M_{it}} r_{itm} dm \leq \sum_{\tau=t+1}^{\infty} \delta^{\tau-t-1} \left[p_\tau \varphi_i \left(\int_0^{M_{it}} r_{itm}^{\frac{\theta}{1-\theta}} dm \right)^{1-\theta} - \int_0^{M_{it}} (W + r_{itm})\, dm - F \right],$$

for each t.

Proof of Proposition 2. See appendix.

Proposition 2 provides an expression for the dynamic enforcement constraint and shows that the amount of resources that can be utilized in any equilibrium is bounded by the firm's expected future profits. These expected future profits in turn depend on the future price sequence $\{p_\tau\}_{\tau=t+1}^{\infty}$, which I will endogenize in the next subsection. An optimal relational contract therefore maximizes the firm's profits subject to this dynamic enforcement constraint.

4.2 Rational-Expectations Equilibrium

Throughout the rest of the paper, I will focus on stationary REEs. The following proposition establishes existence and uniqueness of a stationary REE. The proof is constructive and forms the basis of the analysis in the next section.

Proposition 3. There exists a unique stationary REE.

Proof of Proposition 3. Suppose all firms conjecture price sequence $p_t = p$ for all t. I will show that aggregate supply is well-defined and stationary. Fix a firm i, and assume all other firms choose stationary offers $(r_{jtm}, s_{jtm}, b_{jtm}) = (r_{jm}, s_{jm}, b_{jm})$ and constant management levels $\{M_{jt}\} = \{M_j\}$. Further, suppose firm i chooses constant management levels $\{M_{it}\} = \{M_i\}$. From firm i's perspective, the environment is stationary. By the same argument as in Levin (2003), firm i can replicate any optimal relational contract with a stationary relational contract. Thus, $(r_{itm}, s_{itm}, b_{itm}) = (r_{im}, s_{im}, b_{im})$, which in turn rationalizes the firm's choice of a constant management sequence. This implies a constant aggregate production sequence, which yields aggregate supply $Y(p)$.

The remaining task is to find the constant price sequence consistent with supply and demand in each period. Aggregate supply is upward-sloping, since future competitive rents, and hence today's output, are increasing in p for all firms. Further, it is continuous, since Φ is absolutely continuous. Since aggregate demand has an infinite choke price and is decreasing, smooth, and asymptotes to 0, existence and uniqueness of such a price p follows.■

I pause to comment briefly on uniqueness of equilibrium. First, given constant prices, the game played within a particular firm is a repeated game, and there may therefore exist many self-enforcing relational contracts. However, choice of a sub-optimal relational contract within a firm is ruled out by the definition of REE. Moreover, as in Levin (2003), the assumption that transfers are unrestricted ensures that the maximal surplus generated within a firm is independent of the distribution of that surplus—in principle, the surplus could be allocated to the managers, but this would constitute a suboptimal relational contract from the firm's perspective. In Section 6, I discuss two additional issues related to uniqueness: first, the structure of wage and bonus payments is indeterminate, and second, even if all firms choose optimal relational contracts, there may exist a nonstationary REE with price cycles.

4.3 Equilibrium Optimal Relational Contracts

This subsection characterizes optimal relational contracts in the stationary REE. The proof of Proposition 3 charts a road map for constructing the stationary REE: (1) fix output prices $p_t = p$ and solve for each firm's optimal stationary relational contract, (2) aggregate the production of individual firms to generate the industry supply curve $Y(p)$, and (3) solve for the equilibrium price p^* that satisfies $Y(p^*) = D(p^*)$. In this section, I will drop the i subscript and work directly with firm ability φ.

Since $\theta < 1/2$, production is concave in individual utilization levels. Since managers are symmetric, any optimal relational contract will involve $r_m = r$ for all m. At the steady state, per-period profits for a firm with ability φ are given by

$$\pi(r, M) = p\varphi r^{\theta} M^{1-\theta} - (W + r) M - F.$$

In an optimal relational contract, firms maximize their per-period profits subject to their pooled dynamic-enforcement constraint. That is, each firm takes p as given and solves

$$\max_{r,M} \pi(r, M) \tag{2}$$

subject to

$$(1 - \delta) rM \leq \delta\pi(r, M). \tag{3}$$

In the formulation of the production function, if all managers choose the same effort levels, production exhibits decreasing returns in M. In Propositions A1 and A2 in the appendix, I show that the discount factor the firm faces can be interpreted as an effective discount factor that combines pure time preferences, monitoring technology on the part of the firm, and aspects of the quality of formal contracting institutions. Better monitoring and better formal contracting institutions correspond to higher values of δ.

Define cutoffs $\varphi_L(p) \equiv \delta^{-\theta}\varphi_S(p)$ and $\varphi_H(p) \equiv (2-1/\delta)^{-\theta}\varphi_S(p)$.The next proposition characterizes the solution to the constrained problem (2) subject to (3).

Proposition 4. There exists a weakly increasing function $\mu^*(\varphi, p)$ satisfying $\mu^*(\varphi, p) = 0$ for all $\varphi < \varphi_L(p)$ and $\mu^*(\varphi, p) = 1$ for all $\varphi \geq \varphi_H(p)$ such that the solution to the constrained problem satisfies

$$r^*(\varphi, p)/r^{FB} = M^*(\varphi, p)/M^{FB}(\varphi, p) = \mu^*(\varphi, p),$$

Equilibrium total factor productivity is $A(\varphi, p) = \mu^*(\varphi, p)^\theta A^{FB}(\varphi)$.

Proof of Proposition 4. See appendix.

At the constrained optimum, both of the firm's choice variables are proportional to their first-best values with the same constant of proportionality, given by $\mu^*(\varphi, p)$. The firm's solution can therefore be characterized by the function $\mu^*(\varphi, p)$ and its first-best solution. The exact expression for $\mu^*(\varphi, p)$ is in the appendix, and Figure 1 below characterizes $\mu^*(\varphi, p)$ as a function of φ for a fixed price level p. To compare the solutions of the constrained and unconstrained problems, define $\mu^{FB}(\varphi, p)$ to be equal to 0 if $\varphi < \varphi_S(p)$ and equal to 1 if $\varphi \geq \varphi_S(p)$. When resource utilization is not contractible, future competitive rents are a determinant of the firm's current productivity. Higher-ability firms have higher future competitive rents and are therefore less constrained in equilibrium. These considerations introduce three regions relative to the complete-contracts benchmark. For $\varphi_S(p) \leq \varphi < \varphi_L(p)$, the firm should produce but is unable to. For $\varphi_L(p) \leq \varphi < \varphi_H(p)$, the dynamic enforcement constraint is binding, and the firm is unable to produce efficiently. For $\varphi \geq \varphi_H(p)$, the firm is unconstrained and therefore produces at first-best efficiency.

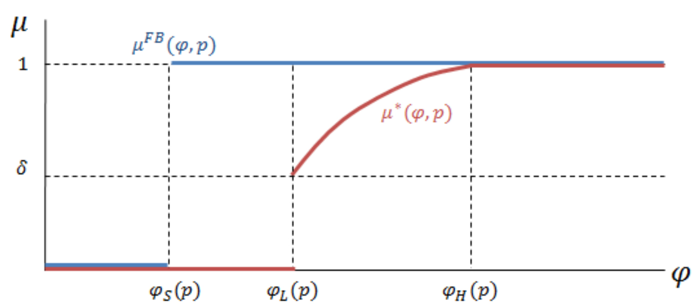

Figure 1: This figure depicts the complete-contracts solution (blue line) and the constrained-optimal solution (red line) as a function of firm ability for a given output price.

Consequently, equilibrium total factor productivity for a firm with ability φ is proportional to its first-best total factor productivity. That is, $A(\varphi, p) = \mu^*(\varphi, p)^\theta A^{FB}(\varphi, p)$. A firm's total factor productivity depends on the effective discount factor it faces and is therefore increasing in the quality of monitoring technology and in the strength of formal contracting institutions. In addition, total factor productivity is jointly determined with the equilibrium price—a firm's production possibilities set is endogenous to market conditions, unlike in the standard Neoclassical model of the firm.

In the complete-contracts competitive equilibrium, firms of heterogeneous total factor productivity coexist in equilibrium. However, because firms are unconstrained and face identical factor prices, firms' marginal productivities are equalized. Relative to this benchmark, if resource utilization is not contractible, firms of heterogeneous marginal productivity coexist in equilibrium: high-ability firms are less constrained, implying a lower marginal productivity at the optimum. Such firms are able to sustain greater levels of resource utilization and therefore will have higher total factor productivity.

So far, the analysis in this subsection held output prices constant and derived firm-level production. Given price p, a firm of ability φ produces $y^*(\varphi, p)$. If all firms expect the same constant price p, then aggregate supply is given by

$$Y(p) \equiv \int_{\varphi_L(p)}^{\infty} y^*(\varphi, p)\, d\Phi(\varphi),$$

where $\varphi_L(p)$ is the cutoff value of ability such that $\varphi < \varphi_L(p)$ implies that a firm of ability φ will not produce in equilibrium. The cutoff $\varphi_L(p)$ is continuous and decreasing in p: if prices are higher, future competitive rents are higher, so firms with lower ability will be able to produce. Further, $y^*(\varphi, p)$ is increasing in p: unconstrained firms choose to produce more if prices are higher, and constrained firms are able to produce more, because their future competitive rents are higher. Therefore, aggregate supply, $Y(p)$ is increasing in p. Equilibrium prices, p^*, in the unique stationary REE therefore solve $D(p^*) = Y(p^*)$ in each period.

5 Properties of Rational-Expectations Equilibrium

Relational incentive contracts generate a mechanism through which future profits determine current productivity, resulting in firm-level income effects. In this section, I explore the consequences of this observation both in terms of its efficiency consequences (Section 5.1) and the comparative statics it implies (Section 5.2). In Section 7, I discuss the extent to which these comparative statics are consistent with several sets of facts.

5.1 Efficiency

Since a firm's future profits serve as an input into its current production, the distribution of profits across heterogeneous firms affects the overall efficiency of production. In this section, I show that profits are not distributed efficiently. To build intuition for the nature and cause of the inefficient profit distribution, let $\pi^*(\varphi, p^*, F)$ denote the solution to (2) subject to (3) for a firm with ability φ when equilibrium prices are p^*, and let $\lambda^*(\varphi, p^*, F)$ denote the shadow cost of (3) at the optimum. By the envelope theorem,

$$d\pi^*/d(-F) = 1 + \lambda^*(\varphi, p^*, F).$$

In addition to the static effect on per-period profits, a reduction in fixed costs increases future profits, which increases the firm's credibility and allows it to decentralize more. The dynamic effect is greater the more constrained the firm is, and by Proposition 4, lower-φ firms are more constrained, so $\lambda^*(\varphi, p^*, F)$ is decreasing in φ. Higher-ability firms are less constrained in equilibrium and therefore benefit less from an increase in future profits.

In principle, a social planner could improve upon the competitive-equilibrium allocation. To understand why, suppose the support of Φ is $[0, \infty)$, so that for any price level p, there will be a positive mass of unconstrained firms with $\varphi_i > \varphi_H(p) + \zeta$ for some small but positive ζ and a positive mass of constrained firms. Consider a persistent excise tax rate of τ levied on all firms with $\varphi_i > \varphi_H(p) + \zeta$. Let $T(\tau)$ be the tax revenues generated by this tax scheme, and define p^τ to solve $D(p^\tau) = Y(p^\tau; \tau)$, where $Y(p; \tau)$ is the resulting aggregate supply function. Total per-period welfare is

$$\begin{aligned}\mathcal{W}(\tau) \quad = \quad & \underbrace{\int_{\varphi_L(p^\tau)}^{\varphi_H(p^\tau)+\zeta} \pi^*(\varphi, p^\tau; 0)\, d\Phi(\varphi)}_{\text{Producer surplus of untaxed firms}} + \underbrace{\int_{\varphi_H(p^\tau)+\zeta}^{\infty} \pi^*(\varphi, p^\tau; \tau)\, d\Phi(\varphi)}_{\text{Producer surplus of taxed firms}} \qquad (4) \\ & + \underbrace{\int_{p^\tau}^{\infty} D(p)\, dp}_{\text{Consumer surplus}} + \underbrace{T(\tau)}_{\text{Revenues}},\end{aligned}$$

where $\pi^*(\varphi, p; \tau)$ is the equilibrium per-period profits a firm with ability φ receives if prices are p, and it faces a tax rate τ, so that the effective prices it receives are $(1-\tau)p$.

Proposition 5. Suppose the support of Φ is $[0, \infty)$. Under the tax scheme described above, $\mathcal{W}'(0) > 0$.

Proof of Proposition 5. See appendix.

The idea of the proof is that a small excise tax on unconstrained firms increases the output price, which induces a transfer from consumers to constrained firms. Statically, this is just a transfer, but dynamically, this increase in profits relaxes the dynamic enforcement constraint of constrained firms and increases their efficiency. Proposition 5 highlights the source of the market inefficiency: high-ability firms overproduce and impose a negative pecuniary externality on low-ability firms with first-order welfare consequences. This pattern of inefficiencies implies that competitive rents are allocated too progressively.

Proposition 5 suggests that when formal contracting institutions are weak, a small excise tax with an exemption for small firms may improve aggregate welfare by boosting the profits of low-ability firms. In contrast, a policy that directly subsidizes small firms does not lead to unambiguous gains. For example, a small-business tax credit funded by a nondistortionary head tax would increase the profits of such firms by more than the monetary cost of the subsidy, but it would increase aggregate output and reduce output prices. This price effect reduces profits of those firms that do not receive the subsidy, which may lead to further losses if they are constrained in equilibrium.

Moreover, policies that concentrate profits among high-ability firms can potentially have negative effects. For example, industrial policy that favors "national champions" tends to favor already-successful firms in an industry, shifting profits away from more-constrained firms. Reducing export barriers tends to favor the higher-ability firms willing to incur the costs of setting up foreign distribution networks. Such firms will expand and drive up domestic factor prices, therefore reducing the profits of more-constrained firms. If the countries pursuing such industrial policies or reducing export barriers have poor formal contracting institutions, this concentration of profits could have negative effects on aggregate productivity.

A full treatment of optimal corporate taxation in the presence of credibility constraints is beyond the scope of this paper, but one implication of the analysis here, in contrast to classical results on optimal-tax theory (Diamond and Mirrlees, 1971), is that taxing the output of a subset of firms may lead to an increase in total surplus. This is because, in the Neoclassical model of production that Diamond and Mirrlees (and the ensuing literature) study, absent any distortionary taxes on production, aggregate production is carried out efficiently.

5.2 Comparative Statics

This section derives several comparative statics, which form the basis for the model's empirical implications. Recall that, given constant output prices p, in an optimal relational contract, the TFP for a firm with ability φ is $A(\varphi, p, F) = \varphi \mu^* (\varphi, p, F)^\theta \left(r^{FB}\right)^\theta$. Proposition 6 shows that the sensitivity of TFP to future competitive rents is greater for firms of lower ability.

Proposition 6. $\log A^*(\varphi, p, F)$ is increasing in φ and p and decreasing in F and for $\varphi > \varphi'$, $\log A^*(\varphi, p, F) - \log A^*(\varphi', p, F)$ is decreasing in p and increasing in F.

Proof of Proposition 6. See appendix.

Section 7.2 explores the implication of Proposition 6 for the dynamics of the productivity distribution in the context of a persistent shock to aggregate demand. Proposition 7 shows that, holding output prices constant, an increase in the strength of formal contracting institutions—which by Proposition A2, I capture by an increase in δ—increases firm productivity and output, but especially so for low-φ firms.

Proposition 7. $\log y^*(\varphi, p, F, \delta)$ and $\log A^*(\varphi, p, F, \delta)$ are increasing in δ and exhibit decreasing differences in (φ, δ).

Proof of Proposition 7. See appendix.

This result is in line with Johnson, McMillan, and Woodruff's (2002) finding that "... entrepreneurs who say the courts are effective have measurably more trust in their trading partners..." and Laeven and Woodruff's (2007) finding that firms operating in Mexican states with stronger legal environments are more productive than those operating in states with weaker legal environments. This proposition suggests that an increase in δ will lead to a convergence in the productivity distribution among existing firms. However, it will also potentially lead to the entry of low-ability firms. Moreover, all firms will produce more when δ, so output prices must fall: if p^δ solves $D(p^\delta) = Y(p^\delta)$ when the discount factor is δ, then p^δ is decreasing in δ. This price reduction leads to a net reduction in production of unconstrained firms, since the increase in δ does not allow them to produce more. Provided that the price effects are not too large, the effects identified in Proposition 7 hold even after allowing for price adjustments, as Proposition 8 shows.

Proposition 8. Suppose that $\hat{\delta} < 1/2$ and either (a) φ has a log-convex distribution and $|\varepsilon_{D,p}| > 1$ or (b) φ has a log-concave distribution and $|\varepsilon_{D,p}| < 1$. Then $Var\left(A^*\left(\varphi, p^\delta, \delta\right)\middle|\, \varphi \geq \varphi_L^\delta\right)$ is greater for $\delta = \hat{\delta}$ than for $\delta = 1$.

Proof of Proposition 8. See appendix.

To illustrate these effects, Figure 2 below shows simulated output and productivity for an economy with complete contracts (corresponding to $\delta = 1$) and for an economy with no formal contracts (corresponding to $\delta = 1/3$). For the purposes of these simulations, I set $\theta = 0.2$, $W = 1$, $F = 0.02$, $D(p) = p^{-3/2}$, and I assumed that φ has a Pareto distribution with lower bound 0.3 and a shape parameter of 5.

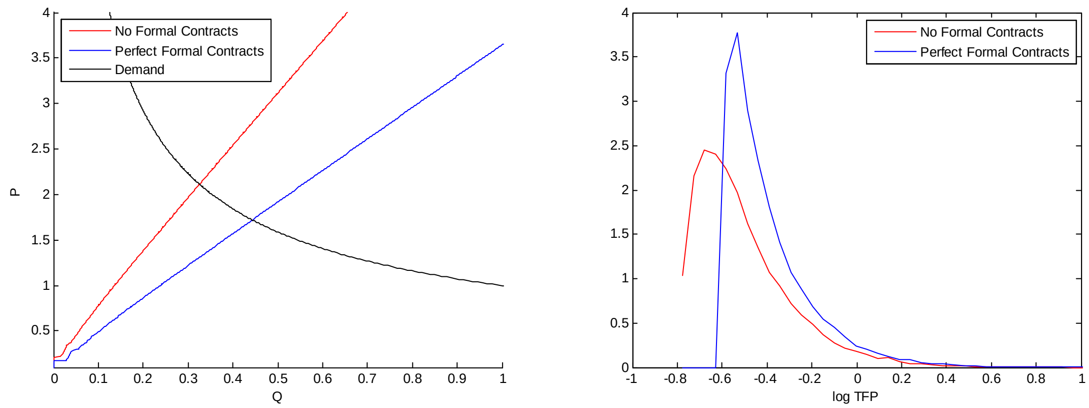

Figure 2: The left panel of this figure plots the aggregate inverse supply functions for an economy with no formal contracts (red) and for an economy with perfect formal contracts (blue) as well as the inverse demand curve (black). Equilibrium prices are lower in the economy with perfect formal contracts. The right panel plots the distribution of log TFP for all firms that produce in equilibrium for an economy with no formal contracts (red) and for an economy with perfect formal contracts (blue).

The left panel plots the aggregate inverse supply functions for the $\delta = 1$ and the $\delta = 1/3$ economies and illustrates that aggregate supply is greater at all price levels for the $\delta = 1$ economy, so equilibrium prices are lower. The right panel plots the distribution of $\log A^*$ for these two economies. For the $\delta = 1$ economy, there is a sharp cut-off corresponding to φ_S at the equilibrium prices. For the $\delta = 1/3$ economy, the productivity distribution is a leftward shift, which is more concentrated at the left tail of the distribution. Moreover, more-unproductive firms continue to produce, because output prices are higher.

6 Discussion of the Model

I now discuss several of the model's assumptions and the features of its solution.

Monopolistic competition and free entry: The main model maintains the assumption of perfect competition in a single-good product market, but similar results would be obtained in a model with monopolistic competition, where φ_i is a function of the size of the market for the product variety that firm i produces. Also, I fixed the mass of firms in the economy to be 1, but the mass of firms that produce in equilibrium is of course endogenous. The model can be extended to allow for endogenous entry in which a firm can incur a sunk cost to draw an ability $\varphi_i \sim \Phi$, as in the Hopenhayn (1992) framework, and the resulting mass of entrants would be determined by an ex ante indifference condition. Allowing for entry in this manner does not qualitatively change the results but would allow for an analysis of how entry barriers impact the productivity distribution.

Resource diversion versus effort: The model is qualitatively similar to one in which the owner asks each manager to exert observable effort at a private cost. Other than differences in interpretation (i.e., such a model would speak to effort provision rather than decentralization), the only difference would be that the physical incidence of effort costs would fall on the managers, so their wages would have to be correspondingly higher in order to compensate them for these costs.

Imperfectly observable bonus payments: I make the stark assumption that future managers commonly observe previous resource utilization and the firm's bonus payments. In an organization with internal labor markets, next period's managers come from the ranks of this period's workers, so this assumption does not require strong external monitoring. Additionally, perfect observability can be relaxed to all-or-nothing public monitoring, which I show in Proposition A1 in the appendix. Weakening the observability assumption in this way amounts to a decrease in the discount factor in the model, which tightens firms' dynamic enforcement constraints.

Time structure of payments to managers: Costless transfers ensure that the maximal surplus generated within a firm in each period is independent of the distribution of that surplus, which implies that, while the optimal mass of managers and resources allocated to managers is uniquely pinned down, the time structure of payments to managers is not. Any stream of payments with the following three properties is consistent with optimal relational contracts: (1) managers' participation constraints bind in their first period of employment, (2) they are satisfied in all future periods, and (3) firms earn nonnegative profits in each period. One alternative way of implementing the constrained-optimal choices of the firm is through the use of efficiency wages (i.e., high contingent wages paid at the beginning of the following period) rather than bonus payments.

Two-point REE price cycles: Even if all firms choose optimal relational contracts, there may exist nonstationary REEs with price cycles. To understand why, suppose all firms believe output prices will cycle between a pair of prices p_L in odd periods and $p_H > p_L$ in even periods. In even periods, the future looks relatively grim, as prices will be low in the following period. This weak outlook constrains firms' resource utilization in even periods, which reduces aggregate supply and therefore is consistent with today's high prices. A similar argument in odd periods establishes that this two-point price cycle is consistent with equilibrium. If all firms are unconstrained in the unique stationary REE, there do not exist two-point REE price cycles. Simulation results suggest that when this condition is not satisfied, there are typically a continuum of two-point REE price cycles. Such nonstationary REEs likely have similar efficiency properties as and analogous comparative statics to the stationary REE, but studying their further properties is an interesting theoretical direction to pursue.

7 Productivity Dispersion and Dynamics and Firm Organization

There has been recent empirical work providing evidence on the importance of future rents and relational contracts for the efficiency of trading relationships, consistent with the model's main mechanism. McMillan and Woodruff (1999) provide cross-sectional evidence consistent with the idea that relationship-specific rents foster the use of trade credit among nonstate manufacturers in Vietnam. Macchiavello and Morjaria (2015) provide a novel way of identifying the value of buyer-seller relationships between Kenyan rose exporters and foreign buyers and find that trade is constrained by the value of relationships. Macchiavello and Morjaria (2016) employ an instrumental-variables strategy and show that a decrease in the value of relationships between Rwandese coffee farmers and mills leads to a decrease in the efficiency of production and the quality of output. Evidence regarding the importance of future rents and relational contracts within firms for firm-level productivity has been less systematic (see Gibbons and Henderson (2012) for several specific examples), largely owing to a paucity of standardized transaction-level data within firms.

In this section, I discuss how the further implications of the model's main mechanism (Propositions 6, 7, and 8) connect to three sets of facts. The first set of facts relates to the relationship between firm organization and firm performance. The second set of facts relates to the relationship between productivity dynamics over the business cycle and firm organization, and the third set of facts concerns cross-country productivity dispersion and firm organization. For each set of facts, there are typically existing (and certainly alternative hypothetical) explanations, so there is significant scope for more work establishing the direct relevance of the model's main mechanism within each domain.

Firm Organization and Firm Performance In the last couple decades, empirical research on firm-level productivity has consistently shown that there are large and persistent productivity differences across firms within industries (see Syverson (2011) for a recent survey). Recent empirical work has made the case that these performance differences may be driven by differences in the quality of management practices across these firms. Bloom, Sadun, and Van Reenen (2012) highlight the importance of decentralization for productivity, showing that greater decentralization improves firm productivity, and Aghion, Bloom, Lucking, Sadun, and Van Reenen (2015) show that more-decentralized firms are better able to cope with economic downturns.

The observation that decentralization improves firm productivity naturally begs the question of why different firms within the same industry choose different levels of decentralization—and indeed, there is a lot of heterogeneity in how decentralized firms are (see Bloom, Sadun, and Van Reenen, 2009, figure 4). The model in this paper preserves the causal interpretation that decentralization increases productivity and profitability, but it highlights the scarcity of credibility as a barrier to further decentralization for constrained firms.

Productivity Dynamics and Firm Organization Macroeconomic evidence, dating back to at least Hultgren (1960) suggests that aggregate productivity is pro-cyclical.[2] Bartelsman and Doms (2000) decompose the changes in aggregate productivity into between- and within-firm productivity changes over the period of 1977-1987 and find that the within-firm component is procyclical. Baily, Bartelsman, and Haltiwanger (2001) decompose the within-firm component further and show that productivity is more pro-cyclical for firms with worse long-run prospects, a finding which is also consistent with recent work by Kehrig (2015), which shows that productivity dispersion is greater during recessions than during booms, and the change in productivity dispersion is driven primarily by the left tail of the productivity distribution: during recessions, there are more unproductive firms, and during booms, there are fewer. Moreover, as mentioned above, Aghion, Bloom, Lucking, Sadun, and Van Reenen (2015) show that in the most recent global recession, firms that were less decentralized prior to the recession saw the greatest decline in productivity during the recession.

In summary, three key facts regarding productivity dynamics and firm organization are: (1) within-firm productivity changes are pro-cyclical, (2) these changes are concentrated in the left tail of the distribution, and (3) productivity declined more for less-decentralized firms in the most recent recession. Fact 1 is inconsistent with standard Neoclassical growth models in which firm-level productivity is exogenous and with efficiency-wage models, which predict countercyclical within-firm productivity. Fact 2 is inconsistent with models of costly labor adjustment and with models in which aggregate fluctuations are driven by independent and exogenous firm-level productivity shocks, as both would predict procyclical productivity dispersion.

All three of these facts are consistent with Proposition 6. To see why, compare the stationary REEs in two economies:[3] a low-demand economy in which aggregate demand is given by $D_L(p)$ and a high-demand economy in which $D_H(p) > D_L(p)$ for all p. Denote the stationary REE price levels in the low- and high-demand economies by p_L^* and p_H^*. It will necessarily be the case that $p_H^* > p_L^*$,

[2] With the exception of the most recent two recessions in the U.S. (See Gali and van Rens, 2014).

[3] The comparison of steady states can be viewed as the limiting case of a model with an aggregate demand state that follows a finite-state Markov process with persistence. Introducing such persistent demand fluctuations into this model is straightforward but involved. Optimal relational contracts in an environment with a Markovian public state variable are sequentially optimal (see Proposition 5 in Barron and Powell (2016)) and can therefore be replicated by a Markovian relational contract.

because aggregate demand is strictly decreasing, and aggregate supply is strictly increasing. For all φ such that firms with ability φ operate in both economies, Proposition 6 shows that we will have $A(\varphi, p_H^*) \geq A(\varphi, p_L^*)$. That is, all firms will have weakly higher TFP in the high-demand economy than in the low-demand economy. Further, by Proposition 6, $A(\varphi, p_H^*) - A(\varphi, p_L^*)$ is larger for firms with lower levels of φ. If we interpret this exercise of unexpectedly (and permanently) changing aggregate demand as the dawning of a boom or a bust, the model predicts pro-cyclical within-firm productivity changes. Additionally, these productivity changes will be primarily centered around low-ability, less-decentralized firms. The implications of Proposition 6 are therefore consistent with all three facts.

Cross-Country Firm Organization and Productivity Dispersion Hsieh and Klenow (2009) and Bartelsman, Haltiwanger, and Scarpetta (2013) document both substantial dispersion in within-country productivity, controlling for industry composition, and greater productivity dispersion in lower-income countries. Moreover, Hsieh and Klenow (2009, Figure 1) shows that the left tail of the productivity distribution is much thicker in India than in the United States. In addition, Bloom, Sadun, and Van Reenen (2012) show that there are substantial cross-country differences in the average level of firm decentralization. In this section, I connect these cross-country facts to cross-country differences in the strength of formal contracting institutions.

Proposition A2 in the appendix shows that stronger formal contracts can complement relational contracts, leading to improvements in firm-level productivity. With stronger formal contracting institutions, credibility becomes relatively less important for sustaining decentralization, consistent with the positive correlation between Kaufmann, Kraay, and Mastruzzi's (2006) country-level measure of "rule of law" and Bloom, Sadun, and Van Reenen's (2012) measure of decentralization in organizations. Since in equilibrium, low-ability firms are more constrained, stronger formal contracting institutions disproportionately benefit such firms by allowing them to decentralize more.

In addition, Proposition 8 predicts that in countries with stronger formal contracting institutions, productivity dispersion will be lower, and the distribution of productivity will have a thinner left tail. In order to examine this prediction, I use country-level measures of labor-productivity dispersion from Bartelsman, Haltiwanger, and Scarpetta (2013),[4] and a measure of the quality of formal contracting institutions ("Rule of Law") from Kaufmann, Kraay, and Mastruzzi (2006).[5]

[4] Bartelsman, Haltiwanger, and Scarpetta (2013) construct a harmonized database (standardizing definitions for meaningful cross-country comparisons) that covers 24 industrial and emerging economies from the 1990s.

[5] This commonly used measure in the international trade literature is an aggregate survey indicator "measuring the extent to which agents have confidence in and abide by the rules of society, and in particular the quality of contract enforcement... ." The values I use are from 2005.

Figure 3 plots labor productivity dispersion against the measure of formal contracting institutions and confirms that countries with higher measures of formal contracting institutions tend to have less productivity dispersion. The results are qualitatively similar using either of Djankov, La Porta, Lopez-de-Silanes, and Shleifer's (2003) measures of the quality of court enforcement.

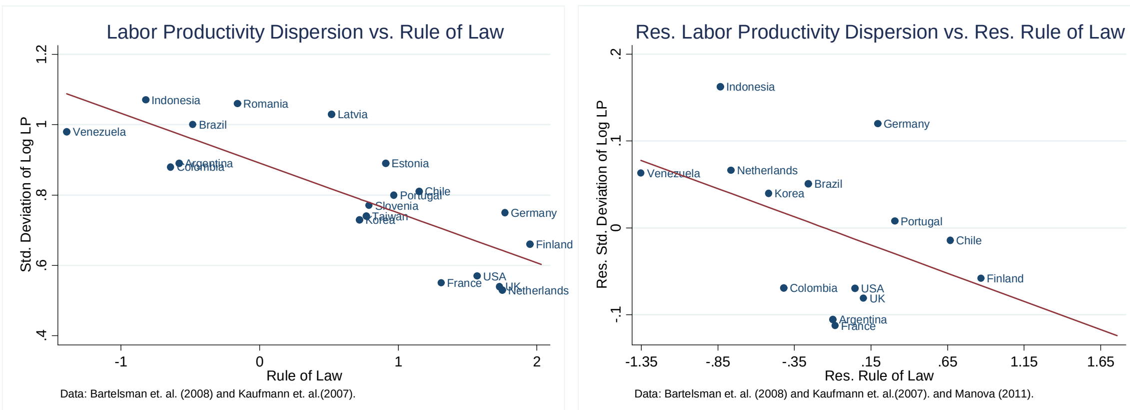

Figure 3. The left panel plots Bartelsman et al.'s (2013) labor-productivity dispersion measure against Kaufmann et al.'s (2006) "Rule of Law" measure. The right panel plots residuals from a regression of Bartelsman et al.'s (2003) labor-productivity dispersion on Manova's (2013) measure of "Private Credit" against residuals from a regression of Kaufmann et al.'s (2006) "Rule of Law" measure on "Private Credit."

There are, of course, other potential explanations for this cross-country pattern, including differences in the quality of capital markets, differences in product-market competition, and differences in the underlying ability distribution across countries. To explore whether the relationship between productivity dispersion and Rule of Law is driven entirely by correlation between the quality of formal contracting institutions and capital markets, I used Manova's (2013) measure of "Private Credit," which proxies for the quality of capital markets.[6] The right panel of Figure 3 shows that the relationship between "Rule of Law" and productivity dispersion is robust to controlling for this measure. This result does not suggest that firm-level financial constraints and differences in capital markets are not important factors driving productivity dispersion differences but rather that other factors also appear to be important. Further, there may be important interactions between credit-market imperfections and a firm's inability to make credible promises to its workers (see, for example, Barron and Li (2016)).

[6] "Private credit" is the amount of credit by banks and other financial intermediaries to the private sector as a share of GDP during the years 1985-1995.

It is also worth observing that, in countries with weak formal contracting institutions, an entrepreneur who is constrained by the credibility of her promises may pursue alternative strategies for relaxing this constraint. For example, she may hire managers with whom she interacts more frequently, such as relatives. The upshot is that family firms are likely to be more prevalent in countries with poor formal contracting institutions, consistent with evidence from La Porta, Lopez-de-Silanes, and Shleifer (1999). Though privately—and potentially socially—beneficial, such alternative firm-level policies do not eliminate the inefficiency of the competitive equilibrium. Firms may overemploy trustworthy family members, even if they are not a good fit for the job; further, a skilled entrepreneur may lack the familiar connections necessary to profitably expand her enterprise to its optimal size.

8 Conclusion

This paper explores the determinants of the credibility of a firm's promises and its role in determining a firm's organization and productivity. In equilibrium, different firms may be differentially constrained. The implications of this observation are consistent with three sets of facts related to firm organization, firm productivity, and how both vary across countries and over time. Further, the allocation of profits matters for firm-level productivity, and profits tend to be inefficiently concentrated at the top in a competitive equilibrium, suggesting that policies that redistribute profits towards less-profitable firms may increase overall welfare.

The fact patterns described in Section 7, while potentially of independent interest, are only indirect tests of the theory. The underlying causal mechanisms involved are: (1) an increase in expected future profits increases current productivity and (2) the effect of an increase in future profits on current productivity is decreasing in future profits. An important future direction for the results in this paper is establishing direct evidence of these mechanisms.

This paper has focused on the distortions that arise in the steady state of an economy. Taking a more dynamic view, if we think of firm growth requiring noncontractible investments by a firm's managers, then the rate at which a firm grows may be limited by its medium-run profitability. Small, but productive, firms may be unable to grow, and as a result, there may be inefficiently slow industrial churn in countries with weaker formal contracting institutions. Such a model may be able to generate results consistent with the recent Hsieh and Klenow (2014) facts on firm growth.

Appendix

Solution to the Model

Proposition 1. If $\varphi_i < \varphi_S = F^\theta \left(r^{FB}\right)^{1-2\theta} / (\theta p)$, firm i optimally does not produce. If $\varphi_i \geq \varphi_S$, firm i optimally chooses $r^{FB}(\varphi_i, p_t) = r^{FB}$ and

$$M^{FB}(\varphi_i, p_t) = (\theta p_t \varphi_i)^{1/\theta} \left(r^{FB}\right)^{(\theta-1)/\theta}.$$

Equilibrium total factor productivity for a firm with ability φ_i is given by

$$A_i^{FB}(\varphi_i, p_t) = y_i / \left(M_i^{1-\theta}\right) = \varphi_i \left(r^{FB}\right)^\theta.$$

Proof of Proposition 1. Let $R = rM$. The firm's problem is to

$$\max_{R,M} p\varphi R^\theta M^{1-2\theta} - WM - R - F.$$

Suppose the firm produces. Taking first-order conditions and solving for the firm's optimal choices of R and M gives the expressions for $r^{FB}(\varphi_i, p_t)$ and $M^{FB}(\varphi_i, p_t)$. The firm's profits, if it produces, are therefore

$$(\theta p \varphi)^{1/\theta} \left(r^{FB}\right)^{(2\theta-1)/\theta} - F,$$

which are nonnegative if and only if $\varphi \geq \varphi_S$.■

Self-Enforcing Relational Incentive Contracts Fix a firm i. Take a bounded price sequence $\{p_t\}_t$ as given and suppress the i subscript. Define $o_{tm} = (r_{tm}, s_{tm}, b_{tm})$ to be an offer made to manager m, where if $M_t \leq \bar{M}$ for some $\bar{M}$ large, then $o_{tm} = (0,0,0)$ for all $m \in (M_t, \bar{M}]$, and let $o_t = \{o_{tm}\}_{m\in[0,\bar{M}]}$. Define $d_{tm} \in \{0,1\}$ to be manager m's acceptance decision in period t, where $d_{tm} = 1$ if manager m accepted offer o_{tm}, and let $d_t = \{d_{tm}\}_{m\in[0,\bar{M}]}$. Define $\hat{r}_{tm} \in [0, r_{tm} d_{tm}]$ to be manager m's utilization decision in period t, where the upper bound is zero if $d_{tm} = 0$, and let $\hat{r}_t = \{\hat{r}_{tm}\}_{m\in[0,\bar{M}]}$. Define $\hat{b}_{tm} \in \mathbb{R}_+$ to be the bonus payment made from the owner to manager m in period t, and let $\hat{b}_t = \left\{\hat{b}_{tm}\right\}_{m\in[0,\bar{M}]}$. Denote a public history at the end of period t by $h^t = \left(o_0, d_0, \hat{r}_0, \hat{b}_0, \ldots, o_t, d_t, \hat{r}_t, \hat{b}_t\right)$, and let H^t be the set of public histories. For the public histories within periods, denote $h_o^t = h^t \amalg o_t$, $h_d^t = h_o^t \amalg \{d_t\}$, $h_{\hat{r}}^t = h_d^t \amalg \{\hat{r}_t\}$, and $h_{\hat{b}}^t = h^{t+1}$, and denote the sets of such within-period public histories by H_o^t, H_d^t, and $H_{\hat{r}}^t$.

A **relational contract** is a strategy profile σ consisting of a strategy σ_i for owner i and σ_m for each manager m, where each strategy maps public histories to feasible actions. Denote continuation play at history h_x^t by $\sigma|\, h_x^t$. A relational contract is **self-enforcing** if it describes a subgame-perfect equilibrium of the game, and we will say that a self-enforcing contract **sustains** utilization choices $\{\hat{r}_{tm}\}_{m\in[0,\bar{M}]}$ if, on the equilibrium path, each manager m utilizes $\hat{r}_{tm}$ in period t. A set of utilization choices $\{\hat{r}_{tm}\}_{m\in[0,\bar{M}]}$ is **sustainable** if there exists a self-enforcing relational contract that sustains it. A self-enforcing relational contract involves **grim-trigger punishment** if, off the equilibrium path, $o_{tm} = (0,0,0)$ and $d_{tm} = \hat{r}_{tm} = \hat{b}_{tm} = 0$ for all m.

Given a relational contract, payoffs are

$$u_{tm}\left(h^{t-1}\right) = (1-\delta)\left(d_{tm}\left(s_{tm} + \hat{b}_{tm}\left(\hat{r}_{tm}\right) + r_{tm} - \hat{r}_{tm}\right) + (1-d_{tm})\,W\right) + \delta u_{t+1,m}\left(h^{t}\right)$$

and

$$\pi_t\left(h^{t-1}\right) = (1-\delta)\left(p_t y_t\left(\hat{r}_{tm}, M_t\right) - \int_0^{M_t}\left(r_{tm} + s_{tm} + \hat{b}_{tm}\right)d_{tm} - F\right) + \delta\pi_{t+1}\left(h^{t}\right)$$

Let $\mathcal{E}_t$ be the set of self-enforcing payoffs in period t. A relational contract is self-enforcing if (i) the owner and all managers $m \in [0, M_t]$ are willing to initiate the contract: $u_{tm}\left(h^{t-1}\right) \geq W$ and $\pi_t\left(h^{t-1}\right) \geq 0$ for all m and all t, (ii) each manager m is willing to choose $\hat{r}_{tm}$ in each t:

$$\hat{r}_{tm} \in \underset{0 \leq \hat{r} \leq r_{tm}}{\operatorname{argmax}}\,(1-\delta)\left(\hat{b}_{tm}\left(\hat{r}_{tm}\right) + r_{tm} - \hat{r}\right) + \delta u_{t+1,m}\left(h^{t}\right),$$

(iii) the owner is willing to make the discretionary payments

$$\hat{b}_t\left(h_{\hat{r}}^{t-1}\right) \in \underset{\hat{b}_{tm}}{\operatorname{argmax}} - (1-\delta)\int_0^{M_t}\hat{b}_{tm} + \delta\pi_{t+1}\left(h^{t}\right),$$

and (iv) continuation payoffs are self-enforcing: $\pi_{t+1}\left(h^{t}\right)$ and $u_{t+1,m}\left(h^{t}\right)$ are in $\mathcal{E}_{t+1}$. I will refer to (i) as the participation constraints, (ii) as the incentive constraints, (iii) as the no-reneging constraint, and (iv) as the self-enforcement constraint.

Lemma 1. If $\{\hat{r}_{itm}\}_{m \in [0,\bar{M}],t}$ is sustainable, there is a self-enforcing relational contract with grim-trigger punishment and full utilization.

Proof of Lemma 1. I will first show that if $\{\hat{r}_{itm}\}_{m \in [0,\bar{M}],t}$ is sustainable, there is a self-enforcing relational contract with grim-trigger punishment that sustains it. Then, I will show that if there is a relational contract with grim-trigger punishment that sustains $\{\hat{r}_{itm}\}_{m \in [0,\bar{M}],t}$, there is also one with with full utilization that sustains $\{\hat{r}_{itm}\}_{m \in [0,\bar{M}],t}$. For the first part, suppose σ sustains $\{\hat{r}_{itm}\}_{m \in [0,\bar{M}],t}$. Then if we replace all off-equilibrium continuation payoffs with 0, the participation constraints will still hold, the incentive constraints and no-reneging constraint will continue to hold and may in fact be slack, and since off-equilibrium grim-trigger play is a SPNE of the stage game, the off-equilibrium continuation payoffs are also self-enforcing.

For the second part, suppose there is a self-enforcing relational contract with grim-trigger punishment σ that sustains $\{\hat{r}_{itm}\}_{m \in [0,\bar{M}],t}$, and suppose $r_{itm} > \hat{r}_{itm}$ for some m and some t. Consider an alternative strategy $\tilde{\sigma}$ that is the same as σ but for which $\tilde{r}_{itm} = \hat{r}_{itm}$ and $\tilde{s}_{itm} = s_{itm} + r_{itm} - \hat{r}_{itm}$. This alternative strategy yields the same equilibrium payoffs for all players and does not affect any of the other constraints, so $\tilde{\sigma}$ is also an equilibrium strategy that sustains $\{\hat{r}_{itm}\}_{m \in [0,\bar{M}],t}$.■

Proposition 2. Utilization choices $\{r_{itm}\}_{m \in [0,\bar{M}],t}$ are sustainable if and only if

$$\int_0^{M_{it}} r_{itm}dm \leq \sum_{\tau=t+1}^{\infty}\delta^{\tau-t-1}\left[p_\tau\varphi_i\left(\int_0^{M_{it}} r_{itm}^{\frac{\theta}{1-\theta}}dm\right)^{1-\theta} - \int_0^{M_{it}}\left(W + r_{i\tau m}\right)dm - F\right] \tag{5}$$

for each t.

Proof of Proposition 2. By Lemma 1, it is without loss of generality to consider full-utilization self-enforcing relational contracts with grim-trigger punishment. Suppose manager m believes the owner will pay reward b_{itm} if and only if he chooses $\hat{r}_{itm}$. Then he will choose $\hat{r}_{itm}$ rather than his maximal reneging temptation of $\hat{r}_{itm} = 0$ if and only if

$$(1-\delta)\, b_{itm} + \delta u_{i,t+1,m}\left(h^t\right) \geq (1-\delta)\, r_{itm} + \delta W.$$

If managers choose $\{r_{itm}\}_{m\in[0,M_{it}]}$, the owner will pay bonuses $\{b_{itm}\}_{m\in[0,M_{it}]}$ if

$$-(1-\delta)\int_0^{M_{it}} b_{itm} + \delta \pi_{t+1}\left(h^t\right) \geq 0.$$

Pooling these constraints together provides a necessary condition for $\{r_{itm}\}_{t,m}$ to be sustained:

$$(1-\delta)\int_0^{M_{it}} r_{itm} dm \leq \delta\left[\pi_{t+1}\left(h^t\right) + \int_0^{M_{it}} \left(u_{i,t+1,m}\left(h^t\right) - W\right)\right].$$

This condition is also sufficient, since if it is satisfied, there exists a set of $\{b_{itm}\}_{tm}$ that satisfy the firm's no-reneging constraint and each manager's incentive constraint. Finally, since future rents must come from future profits, this condition is satisfied if an only if

$$(1-\delta)\int_0^{M_{it}} r_{itm} dm \leq (1-\delta)\sum_{\tau=t+1}^{\infty}\delta^{\tau-t-1}\left[p_\tau\varphi_i\left(\int_0^{M_{it}} r_{itm}^{\frac{\theta}{1-\theta}} dm\right)^{1-\theta} - \int_0^{M_{it}}\left(W + r_{i\tau m}\right) dm - F\right],$$

which is the desired result.■

Proposition 4. In this model, the solution to the constrained problem satisfies

$$r^*\left(\varphi,p\right)/r^{FB} = M^*\left(\varphi,p\right)/M^{FB}\left(\varphi,p\right) = \mu^*\left(\varphi,p\right),$$

where $0 \leq \mu^*(\varphi,p) \leq 1$ is (weakly) increasing in p and δ and (weakly) decreasing in W. Further,

$$\mu^*\left(\varphi\right) = \begin{cases} 1 & \varphi \geq \varphi_H \\ \delta\left(1 + \left(1 - \left(\varphi_L/\varphi\right)^{1/\theta}\right)^{1/2}\right) & \varphi_L \leq \varphi < \varphi_H \\ 0 & \varphi < \varphi_L, \end{cases}$$

where

$$\varphi_L = \frac{F^\theta}{\theta p}\left(\frac{1}{\delta}\right)^\theta \left(r^{FB}\right)^{1-2\theta},$$

and

$$\varphi_H = \begin{cases} \frac{F^\theta}{\theta p}\left(\frac{\delta}{2\delta-1}\right)^\theta \left(r^{FB}\right)^{1-2\theta} & \delta \geq 1/2 \\ \infty & \delta < 1/2. \end{cases}$$

Proof of Proposition 4. Throughout this proof, I drop the i subscript for the firm. Proposition 3 allows us to focus on the stationary problem. Manager symmetry and decreasing returns to utilization imply that $r_m = r$ for all $m \in [0, M]$. The firm's problem is then

$$\max_{r,M} p\varphi r^{\theta} M^{1-\theta} - (W + r) M - F$$

subject to

$$(1-\delta) rM \leq \delta \left(p\varphi r^{\theta} M^{1-\theta} - (W + r) M - F \right).$$

Suppose the firm is constrained at the optimum. Define $M(r)$ such that the constraint holds with equality. The unconstrained problem is then

$$\max_{r} \frac{1-\delta}{\delta} M(r) r.$$

Taking first-order conditions, the firm chooses r such that $\frac{M'(r)}{M(r)} r = -r$. Implicitly differentiating the constraint with respect to r and substituting the resulting expression into the first-order condition yields

$$\frac{py(r^*, M^*)}{M^*} = \frac{W}{1-2\theta} \tag{6}$$

and we know from the constraint that

$$\frac{py(r^*, M^*)}{M^*} = \left(W + \frac{r}{\delta} \right) + \frac{F}{M^*}. \tag{7}$$

(6) implies

$$M^*(r^*) = \left(\frac{1-2\theta}{W} \right)^{1/\theta} (p\varphi)^{1/\theta} r^*,$$

and substituting this into (7), we have that r^* solves a quadratic equation. The linearity of $M^*(r^*)$ results from the assumption that production is constant returns to scale in (r, M). Without this assumption, r^* would be the solution to a nonlinear equation. If we define φ_L as in the statement of the proposition, the solution to this quadratic equation is

$$r^*/r^{FB} = \delta \left(1 + \left(1 - (\varphi_L/\varphi)^{1/\theta} \right)^{1/2} \right) \equiv \mu.$$

The optimal management choice is linear in r^*, so it also satisfies $M^*/M^{FB} = \mu$. It is then easy to show that the constraint is binding for $\varphi < \varphi_H$. For $\varphi \geq \varphi_H$, the solution to the constrained problem is the same as the solution to the unconstrained problem.■

Broadening the Interpretation of the Discount Factor

The main model considers two extreme contracting assumptions (perfect formal contracts in Section 3 and no formal contracts in Section 4) and perfect public observability of all actions. In this appendix, I show that extending the model to incorporate a form of imperfect formal contracts and all-or-nothing public monitoring of the owner's actions is tantamount to changing the discount factor in the main model.

All-or-nothing public monitoring Let $B_{it} = \int b_{itm} dm$ be the total bonus payments made by the owner in period t, and let ρ_{it} be a publicly observed variable that is equal to B_{it} with probability q and equal to $\emptyset$ with probability $1-q$. Suppose that at the end of each period, each manager observes ρ_{it}, but not any individual b_{itm}. Then the following proposition is true.

Proposition A1. $\{r_{itm}\}$ is sustainable if and only if

$$\left(1-\tilde{\delta}\right) \int r_{itm} dm \leq \tilde{\delta} \pi_i,$$

where $\tilde{\delta} = \delta q / \left(1 - \delta + \delta q\right)$.

Proof of Proposition A1. Lemma 1 continues to hold in this setting, so manager m's incentive constraint in period t is $b_{itm} \geq r_{itm}$, and owner i's no-reneging constraint is

$$-\left(1-\delta\right) \int b_{itm} + \delta \pi_i \geq \delta \left[\left(1-q\right) \pi_i + q \cdot 0\right].$$

All managers' incentive constraints and owner i's no-reneging constraint can be satisfied if and only if

$$\left(1-\delta\right) \int r_{itm} \leq \delta q \pi_i,$$

which is equivalent to the condition given in the statement of the proposition.■

Imperfect Formal Contracts Here, I propose a modification of the main model to incorporate formal contracts that are complementary to relational contracts, and in which stronger formal contracts are tantamount to a higher discount factor in the main model. Suppose a third-party enforcer observes r_{itm} and $\hat{r}_{itm}$ but will only enforce a contract that punishes manager m if deviations are at least $(1-\omega)$-egregious, for $\omega \in [0,1]$: given δ_{itm}, the owner can write an enforceable short-term formal contract that conditions on the event $1\left\{\hat{r}_{itm} \geq \omega r_{itm}\right\}$ and, in particular, pays $-\infty$ if $\hat{r}_{itm} < \omega r_{itm}$. Enforcement is otherwise costless. I refer to ω as the **quality of formal contracting institutions**.

Restrict attention to **full-utilization relational contracts**, in which any choice $\hat{r}_{itm} < r_{itm}$ is viewed as a deviation, which results in punishment. In contrast to the no formal-contracts model in Section 4, restricting to full-utilization relational contracts is consequential: relaxing this restriction enables the firm to achieve first-best utilization levels by setting $r_i = r^{FB}/\omega$, allowing each manager to choose $\hat{r}_i = \omega r_i = r^{FB}$ and keep the remaining $\left(1-\omega\right) r^{FB}/\omega$. I make this restriction for tractability purposes. I also assume that a management team operating to the letter of a formal contract yields no more profits than the firm could realize if it simply shut down.[7] The firm's outside option is therefore independent of the strength of formal contracting institutions.

Proposition A2. $\{r_{itm}\}$ is sustainable if and only if

$$\left(1-\tilde{\delta}\right) \int r_{itm} dm \leq \tilde{\delta} \pi_i,$$

[7] For example, because managers choose both $\hat{r}_{im}$ and a noncontractible binary effort $e \in \{0,1\}$ at cost $e \cdot \varepsilon$ such that the effective resources used in production are $e \cdot \hat{r}_{im}$. For $\varepsilon > 0$, formal contracts alone cannot motivate any manager to choose $e = 1$.

where $\tilde{\delta} = \delta / (\delta + (1-\omega)(1-\delta))$.

Proof of Proposition A2. Suppose a manager has been allocated resources r_{itm}. If he utilizes less than ωr_{itm}, he will receive $-\infty$, so his maximal reneging temptation is to walk away with $(1-\omega) r_{itm}$ resources. The bonus he receives therefore must satisfy $b_{itm} \geq (1-\omega) r_{itm}$. The owner's dynamic enforcement constraint is the same as in the main model:

$$(1-\delta) \int b_{itm} \leq \delta \pi_i.$$

All managers' incentive constraints and owner i's no-reneging constraint can therefore be satisfied if and only if

$$(1-\delta)(1-\omega) \int r_{itm} \leq \delta \pi_i,$$

which is equivalent to the condition given in the statement of the proposition.■

Structure of Inefficiencies

Proposition 5. Suppose the support of Φ is $[0, \infty)$. Under the tax scheme described in (4), $\mathcal{W}'(0) > 0$.

Proof of Proposition 5. At $\tau = 0$ and p^0, a marginal increase in τ leads to a reduction in production. In order for markets to clear, prices must increase. Thus, $\left.\frac{dp^\tau}{d\tau}\right|_{\tau=0,p^0} > 0$. We proceed by examining the effects of a marginal increase in τ from $\tau = 0$ on the four expressions in $\mathcal{W}(\tau)$. Since consumers have quasilinear preferences, the effect of a change in taxes on consumers is straightforward:

$$\left.\frac{d}{d\tau} \int_{p^\tau}^{\infty} D(p)\, dp \right|_{\tau=0,p^0} = -D\left(p^0\right) \left.\frac{dp^\tau}{d\tau}\right|_{\tau=0,p^0}.$$

Let $T(\varphi;\tau) = \pi^*(p^\tau, \varphi; 0) - \pi^*(p^\tau, \varphi; \tau) - O\left(\tau^2\right)$ denote the revenues that the tax scheme generates from a firm of ability φ, so

$$\int_{\varphi_L(p^\tau)}^{\varphi_H(p^\tau)+\zeta} \pi^*(p^\tau, \varphi; 0)\, d\Phi(\varphi) + \int_{\varphi_H(p^\tau)+\zeta}^{\infty} \pi^*(p^\tau, \varphi; \tau)\, d\Phi(\varphi) + T(\tau) = \int_{\varphi_L(p^\tau)}^{\infty} \pi^*(p^\tau, \varphi; 0)\, d\Phi(\varphi) - O\left(\tau^2\right).$$

Next, using Leibniz's rule,

$$\left.\frac{d}{d\tau} \int_{\varphi_L(p^\tau)}^{\infty} \pi^*(p^\tau, \varphi; 0)\, d\Phi(\varphi) \right|_{\tau=0,p^0} = \left(S\left(p^0\right) + \Delta + E\left[\chi\left(p^0, \varphi; 0\right) \middle| \varphi \geq \varphi_L\left(p^0\right)\right]\right) \left.\frac{dp^\tau}{d\tau}\right|_{\tau=0,p^0}$$

where $\Delta > 0$ and $\chi\left(p^0, \varphi; 0\right) > 0$ and is decreasing in μ. Finally, since $S\left(p^0\right) = D\left(p^0\right)$,

$$\mathcal{W}'(0) = \left(\Delta + E\left[\chi\left(p^0, \varphi; 0\right) \middle| \varphi \geq \varphi_L\left(p^0\right)\right]\right) \left.\frac{dp^\tau}{d\tau}\right|_{\tau=0,p^0} > 0,$$

which establishes the claim.■

Comparative Statics

Notation. Let $\underline{\mu}(\delta)=\delta$ and $\xi(\varphi,p,F,\delta)=\frac{\underline{\mu}(\delta)-\frac{1}{2}\mu(\varphi,p,F,\delta)}{\mu(\varphi,p,F,\delta)-\underline{\mu}(\delta)}$. The following definitions will be useful in what follows.

$$
\begin{aligned}
\mu(\varphi,p,F,\delta) &= \delta+\delta\left(1-\frac{F}{\delta}\left(\frac{\left(r^{FB}\right)^{1-2\theta}}{\theta p\varphi}\right)^{1/\theta}\right)^{1/2} \\
y^{FB}(\varphi,p) &= \varphi^{\frac{1}{\theta}}(\theta p)^{\frac{1-\theta}{\theta}}\left(r^{FB}\right)^{-\frac{1-2\theta}{\theta}} \\
Z_\varphi &= \theta\varphi;\ Z_p=\theta p;\ Z_{-F}=F;\ Z_\omega=\delta\xi \\
\Xi &= \xi(1+\xi)(1+2\xi)
\end{aligned}
$$

Remark. For $X\in\{\varphi,p,-F\}$,

$$\frac{\partial\mu}{\partial X}=\frac{\mu\xi}{Z_X},\frac{\partial\mu}{\partial\delta}=\frac{\mu\xi}{Z_\delta}(1+\xi)$$

and for $X\in\{\varphi,p,-F,\delta\}$, $\frac{\partial\xi}{\partial X}=-\frac{\Xi}{Z_X}$. Finally, note that

$$\frac{\partial y^{FB}}{\partial\varphi}=\frac{y^{FB}}{Z_\varphi},\frac{\partial y^{FB}}{\partial p}=(1-\theta)\frac{y^{FB}}{Z_p}.$$

Lemma 2. $\log\mu(\varphi,p,F,\delta)$ is increasing in and exhibits decreasing differences in $(\varphi,p,-F,\delta)$.

Proof of Lemma 2. That $\log\mu$ is increasing in $(\varphi,p,-F,\delta)$ follows from their characterizations in the remark above, since $Z_X>0$ for $X\in\{\varphi,p,-F,\delta\}$ and $\mu,\xi>0$. To examine decreasing differences, we have to check the cross-partials. For $X\in\{\varphi,p,-F\}$ and $Y\in\{\varphi,p,-F\}$, $X\neq Y$,

$$\frac{\partial^2\log\mu}{\partial X\partial Y}=-\frac{\Xi}{Z_YZ_X},$$

and for $X\in\{\varphi,p,-F\}$,

$$\frac{\partial^2\log\mu}{\partial\delta\partial X}=-\frac{\Xi}{Z_YZ_X}(1+2\xi).$$

Since $Z_X>0$ for all $X\in\{\varphi,p,-F,\delta\}$, $\Xi>0$, and $(1+2\xi)>0$, $\frac{\partial^2\log\mu}{\partial X\partial Y}<0$ for all $X\neq Y$ and $X,Y\in\{\varphi,p,-F,\delta\}$.■

Proposition 6. $\log A^*(\varphi,p,F)$ is increasing in φ and p and decreasing in F and for $\varphi>\varphi'$, $\log A^*(\varphi,p,F)-\log A^*(\varphi',p,F)$ is decreasing in p and increasing in F.

Proof of Proposition 6. Since $\log A^*=\log\varphi+\theta\log\mu^*+\theta\log r^{FB}$ and μ^* is increasing in $(\varphi,p,-F)$ by Lemma 2, $\log A^*$ is increasing in $(\varphi,p,-F)$. Since μ is the only term that contains interactions, $\log A^*$ exhibits decreasing differences in $(\varphi,p,-F)$ if $\log\mu$ exhibits decreasing differences in $(\varphi,p,-F)$, which it does by Lemma 2.■

Proposition 7. $\log y^*(\varphi,p,F,\delta)$ and $\log A^*(\varphi,p,F,\delta)$ are increasing in δ and exhibit decreasing differences in (φ,δ).

Proof of Proposition 7. Note that

$$\begin{aligned}\log A^* &= \log \varphi + \theta \log \mu^* + \theta \log r^{FB} \\ \log y^* &= \log y^{FB} + \log \mu^*.\end{aligned}$$

y^{FB} does not depend directly on δ. Since μ is increasing in δ, $\log A^*$ and $\log y^*$ are increasing in δ. The only terms in $\log A^*$ and $\log y^*$ that depend both on φ and ω are the $log\mu$ term. We know from Lemma 2 that $\log \mu$ exhibits decreasing differences in (φ, δ). The proposition then follows.■

Industry-Equilibrium Comparative Statics

This section provides a proof of Proposition 8. It proceed first by establishing four lemmas. Lemma 3 connects the equilibrium price response to properties of the industry supply and demand curves. Lemma 4 shows that when price effects are small (large), the equilibrium ability cutoff is decreasing (increasing) in δ. Lemma 5 shows that the lowest observed productivity level will be lower when δ is lower. Lemma 6 shows that the slope of TFP with respect to ability is higher when δ is lower, if $\delta < 1/2$. These lemmas are used in the proof of Proposition 8. Define $Y(p, \delta)$ to be the aggregate supply curve given discount factor δ.

Lemma 3. Let p^δ solve $D\left(p^\delta\right) = Y\left(p^\delta, \delta\right)$. Then

$$\left.\frac{dp^\delta}{d\delta}\frac{\delta}{p}\right|_{p^\delta} = -\theta \frac{\frac{\partial Y}{\partial p} p + D}{\frac{\partial Y}{\partial p} p + \left|\varepsilon_{D,p}\right| D}.$$

Proof of Lemma 3. If we totally differentiate the market-clearing condition and rearrange, we get

$$\frac{dp^\delta}{d\delta}\frac{\delta}{p} = -\frac{\frac{\partial Y}{\partial \delta}\delta}{\frac{\partial Y}{\partial p} p + \left|\frac{dD}{dp}\right| p}.$$

We now seek to derive a relationship between $\frac{\partial Y}{\partial \delta}\delta$ and $\frac{\partial Y}{\partial p} p$. Denote by $g(\varphi)$ the pdf of Φ. Supply is

$$Y(p, \delta) = \int_{\varphi_L(p,\delta)}^{\infty} y^*(\varphi, p, \delta)\, g(\varphi)\, d\varphi$$

and therefore, using Leibniz's rule,

$$\begin{aligned}\frac{\partial Y}{\partial p} &= -\frac{\partial \varphi_L}{\partial p} y^*(\varphi_L, p, \delta)\, g(\varphi_L) + \int_{\varphi_L}^{\infty} \frac{\partial y^*}{\partial p} g(\varphi)\, d\varphi \\ \frac{\partial Y}{\partial \delta} &= -\frac{\partial \varphi_L}{\partial \delta} y^*(\varphi_L, p, \delta)\, g(\varphi_L) + \int_{\varphi_L}^{\infty} \frac{\partial y^*}{\partial \delta} g(\varphi)\, d\varphi\end{aligned}$$

Recall that $\varphi_L(p, \delta) = \delta^{-\theta}\varphi_S(p)$, so that

$$\begin{aligned}\frac{\partial \varphi_L}{\partial p}\frac{p}{\varphi_L} &= -1; \frac{\partial \varphi_L}{\partial \delta} = -\theta \\ \frac{\partial y^*}{\partial p} &= \frac{y^*}{p}\left(\frac{1+\xi}{\theta} - 1\right); \frac{\partial y^*}{\partial \delta} = \frac{y^*}{\delta}(1+\xi).\end{aligned}$$

We therefore get

$$\begin{aligned}
\frac{\partial Y}{\partial p}p &= \varphi_L y^*\left(\varphi_L, p, \delta\right) g\left(\varphi_L\right) + \int_{\varphi_L}^{\infty}\left(\frac{1+\xi}{\theta}-1\right) y^* g\left(\varphi\right) d\varphi \\
\frac{\partial Y}{\partial \delta}\frac{\delta}{\theta} &= \varphi_L y^*\left(\varphi_L, p, \delta\right) g\left(\varphi_L\right) + \int_{\varphi_L}^{\infty}\frac{1+\xi}{\theta} y^* g\left(\varphi\right) d\varphi.
\end{aligned}$$

Finally, note that

$$\frac{\partial Y}{\partial \delta}\frac{\delta}{\theta} - \frac{\partial Y}{\partial p}p = S\left(p, \delta\right).$$

At $p = p^\delta$, $D\left(p^\delta\right) = S\left(p^\delta, \delta\right)$, so the result follows.$\blacksquare$

Lemma 4. Let $\varphi_L^\delta \equiv \delta^{-\theta}\varphi_S\left(p^\delta\right)$. If $|\varepsilon_{D,p}| > 1$, then $\varphi_L^\delta > \varphi_S\left(p^1\right)$. If $|\varepsilon_{D,p}| < 1$, then $\varphi_L^\delta < \varphi_S\left(p^1\right)$.

Proof of Lemma 4. We know that $\varphi_L\left(p^\delta, \delta\right) = \delta^{-\theta}\varphi_S\left(p^\delta\right)$ and

$$\frac{d\varphi_S\left(p^\delta\right)}{d\delta} = -\frac{\varphi_S\left(p^\delta\right)}{p^\delta}\frac{dp^\delta}{d\delta}.$$

Using this result and the result from Lemma 3,

$$\frac{d\varphi_L}{d\delta} = \frac{\varphi_L\left(p^\delta\right)}{\delta}\theta\left(\frac{D\left(1-|\varepsilon_{D,p}|\right)}{\frac{\partial Y}{\partial p}p + |\varepsilon_{D,p}|\,D}\right).$$

By definition, $\varphi_S\left(p^1\right) = \varphi_L^1$, and by the fundamental theorem of calculus

$$\varphi_L^1 = \varphi_L^\delta + \int_\delta^1 \frac{d\varphi_L}{d\delta}d\delta.$$

This is less than φ_L^δ if $|\varepsilon_{D,p}| > 1$, so that $\frac{d\varphi_L}{d\delta} < 0$ for all δ, and it is greater than φ_L^δ if $|\varepsilon_{D,p}| < 1$, so that $\frac{d\varphi_L}{d\delta} > 0$ for all δ.$\blacksquare$

Lemma 5. $\varphi_L^\delta \mu\left(\varphi_L^\delta\left(p^\delta\right)\right)^\theta \left(r^{FB}\right)^\theta < \varphi_S\left(p^1\right)\left(r^{FB}\right)^\theta$.

Proof of Lemma 5. We know that $\mu\left(\varphi_L^\delta\right) = \delta$ and $\varphi_L^\delta = \delta^{-\theta}\varphi_S\left(p^\delta\right)$, which implies that

$$\varphi_L^\delta \mu\left(\varphi_L^\delta\left(p^\delta\right)\right)^\theta \left(r^{FB}\right)^\theta = \varphi_S\left(p^\delta\right)\left(r^{FB}\right)^\theta < \varphi_S\left(p^1\right)\left(r^{FB}\right)^\theta,$$

where in the last inequality, I used the fact that $p^1 < p^\delta$, and therefore $\varphi_S\left(p^1\right) > \varphi_S\left(p^0\right)$.$\blacksquare$

Lemma 6. Suppose $\delta < 1/2$. $\varphi\mu\left(\varphi, p, F, \delta\right)^\theta$ increases faster than φ for $\varphi \geq \varphi_L^\delta$.

Proof of Lemma 6. We know that

$$\frac{d}{d\varphi}\left(\varphi - \varphi\mu^\theta\right) = 1 - \mu^\theta\left(1+\xi\right).$$

If this expression is negative, then $\frac{d}{d\varphi}\varphi < \frac{d}{d\varphi}\varphi\mu^{\theta}$. Note that

$$\frac{\partial}{\partial\varphi}\mu^{\theta}\left(1+\xi\right) = -\xi\left(1+\xi\right)\left(\frac{1-\theta+2\xi}{\theta\varphi}\right)\mu^{\theta} < 0,$$

so $1-\mu^{\theta}\left(1+\xi\right)$ is minimized at φ_H (and all $\varphi > \varphi_H$), where it equals $\left(\frac{1}{2}-\delta\right)/\left(1-\delta\right)$, which is positive since $\delta < 1/2$.■

Proposition 8. Suppose $\delta < 1/2$ and either (a) φ has a log-convex distribution and $|\varepsilon_{D,p}| > 1$ or (b) φ has a log-concave distribution and $|\varepsilon_{D,p}| < 1$. Then $Var\left(A^{*}\left(\varphi,p^{\delta},F,\delta\right)\middle|\,\varphi \geq \varphi_L^{\delta}\right)$ is greater for $\delta < 1/2$ than for $\delta = 1$.

Proof of Proposition 8. From Lemma 6, we know that $\left|\frac{dA^{\delta}}{d\varphi}\right| < \left|\frac{dA^{1}}{d\varphi}\right|$ for all $\varphi \geq \varphi_L^{\delta}$. Tang and See (2009) show that if f and g are functions of a random variable and $|f| < |g|$ almost everywhere, then $Var\left(f\right) < Var\left(g\right)$. This result implies that

$$Var\left(A^{\delta}\left(\varphi\right)\middle|\,\varphi \geq \varphi_L^{\delta}\right) > Var\left(A^{1}\left(\varphi\right)\middle|\,\varphi \geq \varphi_L^{\delta}\right) = \left(r^{FB}\right)^{2\theta} Var\left(\varphi\middle|\,\varphi \geq \varphi_L^{\delta}\right).$$

If $|\varepsilon_{D,p}| > 1$ and φ is log-convex, then we have that $Var\left(\varphi|\,\varphi \geq k\right)$ is increasing in k (see Burdett (1996)), and the result follows since $\varphi_L^{\delta} > \varphi_S^{1}$. If $|\varepsilon_{D,p}| < 1$ and φ is log-concave, then $Var\left(\varphi|\,\varphi \geq k\right)$ is decreasing in k and the result follows since $\varphi_L^{\delta} < \varphi_S^{1}$.■

Remark. In fact the Burdett (1996) result shows that a sufficient condition for $Var\left(\varphi|\,\varphi \geq k\right)$ to be increasing in k is that the triple cumulative integration of φ is log-convex, which is a significantly weaker condition.

Remark. For the $|\varepsilon_{D,p}| > 1$ case, it can be seen that this result will hold for distributions that are not "too log-concave" in the sense that all that is required is that

$$\left[\left.\frac{d\left(\varphi\mu\left(\varphi,p^{\delta}\right)^{\theta}\right)}{d\varphi}\right|_{\varphi^{*}}\right]^{2} Var\left(\varphi\middle|\,\varphi \geq \varphi_L^{\delta}\right) > Var\left(\varphi|\,\varphi \geq \varphi_S^{1}\right),$$

where φ^{*} is the approximation point for a variance approximation. Similarly, if $|\varepsilon_{D,p}| < 1$, then this result will hold for distributions that are not "too log-convex."

References

[1] Acemoglu, Daron, Pol Antras, and Elhanan Helpman. 2007. Contracts and Technology Adoption. *American Economic Review*, 97(3): 916-943.

[2] Aghion, Philippe, Nicholas Bloom, Brian Lucking, Raffaella Sadun, and John Van Reenen. 2015. Growth and Decentralization in Bad Times. *Working Paper*.

[3] Alfaro, Laura, Paola Conconi, Harald Fadinger, and Andrew Newman. Forthcoming. Do Prices Determine Vertical Integration? *Review of Economic Studies*.

[4] Asker, John, Allan Collard-Wexler, and Jan De Loecker. 2014. Dynamic Inputs and Resource (Mis)Allocation. *Journal of Political Economy*, 122(5): 1013-1063.

[5] Baily, Martin, Eric Bartelsman, and John Haltiwanger. 2001. Labor Productivity: Structural Change and Cyclical Dynamics. *Review of Economics and Statistics*, 83(3): 420-433.

[6] Banerjee, Abhijit and Esther Duflo. 2005. Growth Theory Through the Lens of Development Economics. In *Handbook of Economic Growth*, Vol. 1, eds. Philippe Aghion and Steven Durlauf. Amsterdam: Elsevier.

[7] Banerjee, Abhijit and Benjamin Moll. 2010. Why Does Misallocation Persist? *American Economic Journal: Macroeconomics*, 2(1): 189-206.

[8] Barron, Daniel and Jin Li. 2016. Financing Constraints and Relational Contracts. *Working Paper*.

[9] Barron, Daniel and Michael Powell. 2016. Policies in Relational Contracts. *Working Paper*.

[10] Bartelsman, Eric, John Haltiwanger, and Stefano Scarpetta. 2013. Cross-Country Differences in Productivity: The Role of Allocation and Selection. *American Economic Review*, 103(1): 305-334.

[11] Bartelsman, Eric and Mark Doms. 2000. Understanding Productivity: Lessons from Longitudinal Microdata. *Journal of Economic Literature*, 38(3): 569-594.

[12] Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2009. The Organization of Firms Across Countries. *NBER Working Paper No. 15129*.

[13] Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. The Organization of Firms Across Countries. *Quarterly Journal of Economics*, 127(4): 1663-1705.

[14] Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2015. Management as a Technology. *Working Paper*.

[15] Bloom, Nicholas and John Van Reenen. 2007. Measuring and Explaining Management Practices Across Firms and Countries. *Quarterly Journal of Economics*, 122(4): 1351-1408.

[16] Board, Simon and Moritz Meyer-ter-Vehn. 2015. Relational Contracts in Competitive Labour Markets. *Review of Economic Studies*, 82(2): 490-534.

[17] Buera, Francisco, Joseph Kaboski, and Yongseok Shin. 2011. Finance and Development: A Tale of Two Sectors. *American Economic Review*, 101(5): 1964-2002.

[18] Bull, Clive. 1987. The Existence of Self-Enforcing Implicit Contracts. *Quarterly Journal of Economics*, 102(1): 147-159.

[19] Burdett, Kenneth. 1996. Truncated Means and Variances. *Economics Letters*, 52(3): 263-267.

[20] Caballero, Ricardo and Mohamad Hammour. 1998. The Macroeconomics of Specificity. *Journal of Political Economy*, 106(4): 724-767.

[21] Chassang, Sylvain. 2010. Building Routines: Learning, Cooperation, and the Dynamics of Incomplete Relational Contracts. *American Economic Review*, 100(1): 448-465.

[22] Cooley, Thomas, Ramon Marimon, and Vincenzo Quadrini. 2004. Aggregate Consequences of Limited Contract Enforceability. *Journal of Political Economy*, 112(4): 817-847.

[23] Diamond, Peter and James Mirrlees. 1971. Optimal Taxation and Public Production I: Production Efficiency. *American Economic Review*, 61(1): 8-27.

[24] Djankov, Simeon, Rafael La Porta, Florencio Lopez-de-Silanes, and Andrei Shleifer. 2003. Courts. *Quarterly Journal of Economics*, 118(2): 453-517.

[25] Ellison, Glenn and Richard Holden. 2014. A Theory of Rule Development. *Journal of Law, Economics, and Organization*, 30(4): 649-682.

[26] Francois, Patrick and Joanne Roberts. 2003. Contracting Productivity Growth. *Review of Economic Studies*, 70(1): 59-85.

[27] Gali, Jordi and Thijs Van Rens. 2014. The Vanishing Procyclicality of Labor Productivity. *Working Paper*.

[28] Garicano, Luis, Claire LeLarge and John Van Reenen. Forthcoming. Firm Size Distortions and the Productivity Distribution: Evidence from France. *American Economic Review*.

[29] Gibbons, Robert and Rebecca Henderson. 2012. Relational Contracts and Organizational Capabilities. *Organization Science*, 23(5): 1350-1364.

[30] Gibbons, Robert and Rebecca Henderson. 2013. Relational Contracts, Managerial Practices, and Organizational Capabilities. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts. Princeton, NJ: Princeton University Press.

[31] Gibbons, Robert, Richard Holden, and Michael Powell. 2012. Organization and Information: Firms' Governance Choices in Rational-Expectations Equilibrium. *Quarterly Journal of Economics*, 127(4): 1813-1841.

[32] Grossman, Gene and Elhanan Helpman. 2002. Integration Versus Outsourcing in Industry Equilibrium. *Quarterly Journal of Economics*, 117(1): 85-120.

[33] Guner, Nezih, Gustavo Ventura, and Yi Xu. 2008. Macroeconomic Implications of Size-Dependent Policies. *Review of Economic Dynamics*, 11(4): 721-744.

[34] Hopenhayn, Hugo. 1992. Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica*, 60(5): 1127-1150.

[35] Hsieh, Chang-Tai and Peter Klenow. 2009. Misallocation and Manufacturing TFP in China and India. *Quarterly Journal of Economics*, 124(4): 1403-1448.

[36] Hsieh, Chang-Tai and Peter Klenow. 2014. The Life Cycle of Plants in India and Mexico. *Quarterly Journal of Economics*, 129(3): 1035-1084.

[37] Hultgren, Thor. 1960. *Changes in Labor Cost During Cycles in Production in Business*. Cambridge, MA: National Bureau of Economic Research.

[38] Jeong, Hyeok and Robert Townsend. 2007. Sources of TFP Growth: Occupational Choice and Financial Deepening. *Economic Theory*, 32(1): 179-221.

[39] Johnson, Simon, John McMillan and Christopher Woodruff. 2002. Courts and Relational Contracts. *Journal of Law, Economics, and Organization*, 18(1): 221-227.

[40] Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi. 2006. Governance Matters V: Aggregate and Individual Governance Indicators for 1996-2005. *World Bank Policy Research Working Paper*.

[41] Kehrig, Matthias. 2015. The Cyclical Nature of the Productivity Distribution. *Working Paper*.

[42] La Porta, Rafael, Florencio Lopez-de-Silanes, and Andrei Shleifer. 1999. Corporate Ownership Around the World. *Journal of Finance*, 54(2): 471-517.

[43] Laeven, Luc and Christopher Woodruff. 2007. The Quality of the Legal System, Firm Ownership, and Firm Size. *Review of Economics and Statistics*, 89(4): 601-614.

[44] Legros, Patrick and Andrew Newman. 2013. A Price Theory of Vertical and Lateral Integration. *Quarterly Journal of Economics*, 128(2): 725-770.

[45] Levin, Jonathan. 2003. Relational Incentive Contracts. *American Economic Review*, 93(3): 835-857.

[46] Li, Jin, Niko Matouschek, and Michael Powell. Forthcoming. Power Dynamics in Organizations. *American Economic Journal: Microeconomics*.

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

[48] Macchiavello, Rocco and Ameet Morjaria. 2015. The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports. *American Economic Review*, 105(9): 2911-2945.

[49] Macchiavello, Rocco and Ameet Morjaria. 2016. Competition and Relational Contracts: Evidence from Rwanda's Coffee Mills. *Working Paper*.

[50] MacLeod, Bentley and James Malcomson. 1989. Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment. *Econometrica*, 57(2): 447-480.

[51] Malcomson, James. 2013. Relational Incentive Contracts. In *Handbook of Organizational Economics*, eds. Robert Gibbons and John Roberts. Princeton, NJ: Princeton University Press.

[52] Martimort, David and Thierry Verdier. 2004. The Agency Cost of Internal Collusion and Schumpeterian Growth. *Review of Economic Studies*, 71(4): 1119-1141.

[53] Manova, Kalina. 2013. Credit Constraints, Heterogeneous Firms, and International Trade. *Review of Economic Studies*, 80(2): 711-744.

[54] McMillan, John and Christopher Woodruff. 1999. Interfirm Relationships and Informal Credit in Vietnam. *Quarterly Journal of Economics*, 114(4): 1285-1320.

[55] Midrigan, Virgiliu and Daniel Xu. 2014. Finance and Misallocation: Evidence from Plant-Level Data. *American Economic Review*, 104(2): 422-458.

[56] Moll, Benjamin. 2014. Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation? *American Economic Review*, 104(10): 3186-3221.

[57] Peters, Michael. 2013. Heterogeneous Markups and Endogenous Misallocation. *Working Paper*.

[58] Restuccia, Diego and Richard Rogerson. 2008. Policy Distortions and Aggregate Productivity with Heterogeneous Establishments. *Review of Economic Dynamics*, 11(4): 707-720.

[59] Schroth, Josef. 2016. Optimal Intermediary Rents. *American Economic Journal: Macroeconomics*, 8(1): 98-118.

[60] Syverson, Chad. 2011. What Determines Productivity? *Journal of Economic Literature*, 49(2): 326-365.

[61] Tang, Hsiu Khuern and Chuen Teck See. 2009. Variance Inequalities Using First Derivatives. *Statistics and Probability Letters*, 79(9): 1277-1281.

[62] Wu, Yanhui. Forthcoming. Incentive Contracts and the Allocation of Talent. *Economic Journal*.