Productivity and Credibility in Industry Equilibrium

Michael Powell*

June 26, 2012

Abstract

Productivity dispersion among seemingly similar firms has been widely documented and often viewed as symptomatic of an underlying misallocation of resources. Why does a firm that is marginally more productive than others not expand? Following Chandler (1962), I argue that in order to produce efficiently, a large firm must decentralize operating decisions to managers. In order to decentralize, a firm's owner must make credible promises to reward judicious use of the firm's resources, and such credibility may be in short supply. I therefore develop a model of relational contracts in a competitive environment with heterogeneous firms. Credibility requires collateral, which takes the form of future competitive rents. In equilibrium, competitive rents are allocated inefficiently: high-ability firms are better able to solve their credibility problem than low-ability firms and therefore, the marginal collateral value of competitive rents is not equalized across firms. Improvements in formal contracting institutions reduce the importance of credibility and therefore disproportionately benefit low-ability firms. Cross-country differences in contracting institutions can therefore partially explain the observed pattern that productivity dispersion is greater in developing countries. (JEL D21, 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 Abhijit Banerjee, Daniel Barron, Nick Bloom, Dan Cao, Ricardo Caballero, Sebastian di Tella, Glenn Ellison, Luis Garicano, Jean-Paul L'Huillier, Hongyi Li, Iuliana Pascu, Michael Peters, Andrea Prat, Alp Simsek, Erik Snowberg, Chad Syverson, Steven Tadelis, Heiwai Tang, Eric Van den Steen, Xiao Yu Wang, and seminar participants at Caltech, Columbia, Duke, LSE, MIT, Northwestern, Rochester, Stanford, and UPenn. I thank MIT Sloan's Program on Innovation in Markets and Organizations and the National Science Foundation for financial support.

1 Introduction

In order for a large firm to produce efficiently, the owner must decentralize daily operating decisions to a team of managers (Chandler (1962)). In the absence of perfect formal contracts, decentralization requires trust: the owner must trust that the managers will not squander the firm's resources, and the managers must trust that judicious use of resources will be rewarded appropriately. Recent empirical work by Bloom, Sadun, and Van Reenen (Forthcoming) demonstrates that lack of trust constrains firm size by limiting decentralization.[1] As a result, firms that are highly productive at the margin may like to expand but be unable to do so. In this paper, I develop a simple model of relational contracts (informal, self-enforcing agreements) in a competitive environment to analyze the consequences of limited trust on the steady-state distribution of firm size and productivity.

I model trust as credibility 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 some resources to each manager. He would like each manager to utilize those resources appropriately, but formal contracts are unavailable. He can promise to pay a pre-specified reward if the manager appropriately utilizes the resources he has been allocated.[2] The owner lacks commitment, so in a one-shot game, after the manager's utilization choice has been made, the owner would always prefer not to pay the reward (i.e., to renege) and will do so; forward-looking managers working for such a "fly-by-night" firm will squander their firm's resources and therefore will not be given any to begin with. A long-lived firm, however, can make credible promises of future rewards, since failure to uphold promises may put the future of the firm at stake: the future competitive rents the firm generates can thus be used as collateral in the firm's promises.

Competitive rents are endogenous. Output generated by the owner-manager problem is sold into a competitive product market. The market consists of many firm owners of heterogeneous ability,[3] and production exhibits decreasing returns to scale. As in Lucas (1978), this implies that firms of different total factor productivity levels will coexist in equilibrium. The novel element of this model is that firms of different *marginal* productivities will coexist in equilibrium, even though all firms face the same factor prices: heterogeneous firms will be

[1] Firm owners may lock up spare parts for machines, depriving local managers of the ability to perform repairs when a machine breaks down, for if they did not, the managers might steal the parts, sell them, and replace them with low-quality parts. Owners require managers to obtain approval to make capital investments greater than \$500, hire non-temporary personnel, or make sales and marketing decisions.

[2] Discretionary payments take the form of monetary bonuses in the literature on relational incentive contracts. These payments can be interpreted more broadly as raises, promotions, additional freedom, policy commitments, and improved working conditions that can be awarded to an agent in a contingent way.

[3] Ability can be thought of as anything that is valuable for production, scarce, and non-transferable.

heterogeneously constrained, and therefore there will be misallocation of production. The credibility 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.

Because competitive rents serve as collateral, their allocation matters for efficiency. Initial advantage begets further advantage: in equilibrium, high-ability owners achieve high levels of rents and hence collateral, which in turn gives rise to even greater rents. This "Matthew Effect" (Merton (1968)) 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.

Low-ability firms face relatively tighter credibility constraints. Their productivity is therefore relatively more sensitive to changes in factors that determine competitive rents, such as shifts in aggregate demand. Improvements in formal contracting institutions reduce the importance of credibility in sustaining decentralization and therefore disproportionately benefit low-ability firms, leading to a greater dispersion of total factor productivity in weaker contracting environments. Cross-country differences in contracting institutions can thus partially explain the observed pattern that productivity dispersion is more pronounced in developing countries (Bartelsman, Haltiwanger, and Scarpetta (Forthcoming)). Further, differences in formal contracting institutions also lead to differences in the price level. I show that when one takes price effects into account, improvements in formal contracting institutions also leads to compression in firm size: small firms produce more and large firms produce less.

This paper is related to the recent 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 ability to allocate resources efficiently across firms can explain a substantial portion of the differences in per-capita GDP. These papers argue that misallocation of productive resources is ubiquitous, but it is more pronounced in developing countries than in developed countries. Hsieh and Klenow show that improving the allocation of capital and labor in China and India to U.S. levels would result in a one-off increase in per-capita GDP by 30-50% and 40-60% respectively.

To design effective policy for improving the allocation of resources, we need to understand why resources were not allocated efficiently to begin with. Several recent papers in the macro tradition (Banerjee and Moll (2010), Moll (2011), Buera, Kaboski, and Shin (2011), and Midrigan and Xu (2010)) have focused on the role of underdeveloped financial markets.[4]

[4]In a static model, mismatch between the quality of ideas of an entrepreneur and the funding necessary to take the idea to fruition can lead to misallocation. However, in the long run, entrepreneurs with the

Others include Peters (2011b), who argues that in a monopolistic-competition framework, heterogeneity in entry rates leads to heterogeneity in markups, which in turn leads to a distortion in relative output prices and thus misallocation. Collard-Wexler, Asker, and De Loecker (2011) argue that much of the misallocation is driven by adjustment costs. Guner, Ventura, and Xu (2008) and Garicano, Lelarge and Van Reenen (2011) highlight the importance of existing size-dependent policies on whether or not firms operate at their efficient scale.

The normative implications of each of these explanations differ. For example, if misallocation is driven solely by adjustment costs, then there is little scope for beneficial policy. If, on the other hand, heterogeneity in markups is the driving factor, then we want to understand why there is heterogeneity in entry rates and perhaps remedy this by selectively reducing entry barriers in certain industries. If underdeveloped financial markets are the problem, then top-down improvements in financial markets could reduce misallocation. My model generates persistent misallocation in a perfectly competitive environment with no adjustment costs or credit rationing and is therefore complementary to existing views. It suggests that policy should focus on improving the quality of formal contracting institutions. Absent such policy instruments, productive efficiency can be improved through policies that reallocate profits among heterogeneous firms.

This paper is also related to the literature on the large and persistent differences in productivity levels across producers (for a recent empirically oriented survey, see Syverson (2011)), and it is methodologically similar to Board and Meyer-ter-Vehn (2011), who augment Shapiro and Stiglitz (1984)'s model of efficiency wages with on-the-job search and show that wage, and hence productivity, dispersion emerges in a stationary industry equilibrium, even with ex ante identical firms. In their model, credible incentives are derived from endogenous quasi-rents: workers are motivated by the prospect of obtaining or losing high-paying jobs. In my model, credibility is derived from competitive rents: a firm upholds its promises out of fear of losing future profits. Both quasi-rents and competitive rents are important in determining the strength of ongoing relationships, and since their determinants differ, these approaches are complementary. Also closely related are Chassang (2010) and Gibbons and Henderson (2011), who argue that firm-level heterogeneity in productivity is due to differences in (ex ante identical) firms' success in developing relational contracts that put them on the production possibilities frontier.[5] I assume that all firms succeed in implementing optimal relational contracts. Small differences in the ability of firm owners translate into

best ideas should be able to save their way out of capital constraints. This argument underpins the title of Banerjee and Moll (2010)'s paper, "Why Does Misallocation Persist?"

[5]Also related is Ellison and Holden (Forthcoming), who show the potential for path-dependence in the efficiency of organizational rules.

differences in continuation values and potentially large differences in decentralization and productivity. Relational incentive contracts can therefore amplify existing differences. The analysis in this paper is silent on firm dynamics, unlike Chassang (2010) and Ellison and Holden (Forthcoming). It provides a theory of steady-state misallocation, not a theory of the process that leads to it.

Finally, this paper contributes to the literatures on firm governance in industry equilibrium (Grossman and Helpman (2002), Legros and Newman (2012), and Gibbons, Holden, and Powell (Forthcoming)) and on the aggregate implications of contractual incompleteness (Caballero and Hammour (1998), Francois and Roberts (2003), Cooley, Marimon, and Quadrini (2004), and Acemoglu, Antras, and Helpman (2007)). My analysis is most similar to Acemoglu, Antras, and Helpman (2007), who examine the role of incomplete contracts and unresolved hold-up on technology adoption. In contrast, I explore how the success of attempts to resolve contractual incompleteness using relational contracts varies with underlying firm characteristics and with the competitive environment in which the firm operates.

Section 2 sets up the basic model and defines terminology. Section 3 characterizes the solution in the complete-contracts case. Section 4 analyzes optimal relational incentive contracts in the absence of formal contracts, and section 5 explores the efficiency of the resulting industry equilibrium. Section 6 explores empirical implications of the model and extends the model to incorporate the possibility of formal contracting. Section 7 concludes.

2 Setup and Technology

There is a unit mass of firms, indexed by $i \in [0,1]$. Each firm is run by a risk-neutral owner who is the residual claimant. Output requires capital and managers, who must be given resources in order to be productive. Contracting institutions are weak, and therefore judicious use of resources by managers cannot be directly contracted upon. Throughout, we will assume that there is a large enough mass of risk-neutral managers so that in equilibrium, they are indifferent between working and not. Play is infinitely repeated, and we denote by $t = 1, 2, 3, \dots$ the period. All players share a common discount factor, which we will express in terms of a discount rate $\frac{1}{1+r}$ with $r < 1$. The product of the owner-manager problem is output, which is homogeneous across firms and sold into a competitive product market. Aggregate demand is assumed to be stationary, $D_t(p_t) = D(p_t)$, where p_t is the output price in period t, downward-sloping ($D' < 0$), and generated by consumers who have quasilinear preferences.

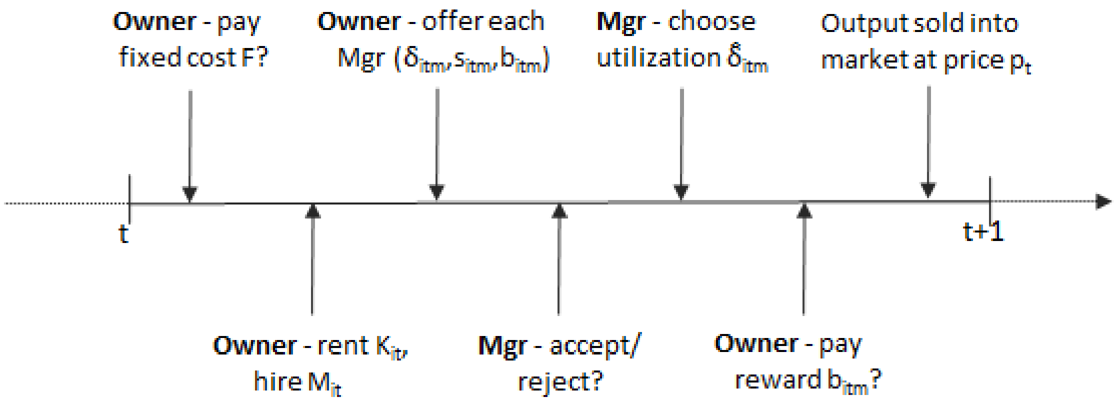

Figure 1: Timing

Each period consists of seven stages. In the first stage, owner i decides whether or not to pay the fixed cost of production, F. If he chooses to, in the second stage, he decides how much capital K_{it} to rent at constant rental rate R and the mass M_{it} of managers to whom he would like to make an offer. In stage 3, the owner offers each manager $m \in [0, M_{it}]$ a triple $(\delta_{itm}, s_{itm}(\rho_{itm}), b_{itm})$, where δ_{itm} is a level of discretionary resources entrusted to manager m,[6] $s_{itm}(\rho_{itm})$ is a payment that potentially depends on a contractible measure ρ_{itm} of the level of resources manager m uses appropriately,[7] and b_{itm} is a reward that the owner intends to pay manager m if and only if he utilizes all the resources he has been entrusted with. In the fourth period, each manager m decides whether or not to accept this proposed contract or reject it in favor of an outside opportunity that yields utility $W > 0$. If manager m accepts the contract, in stage 5, he chooses a resource utilization level $\hat{\delta}_{itm} \geq 0$ and keeps $\delta_{itm} - \hat{\delta}_{itm}$. This utilization level is commonly observed, and in stage 6, the owner decides whether or not to pay manager m a reward of b_{itm}. Output for firm i is then realized and sold into the market at price p_t in stage 7.

Owners have heterogeneous ability.[8] Let φ_i denote the ability of owner i. Assume the realized distribution of ability is $\varphi_i \sim \Phi$, where Φ is a distribution function. Given capital

[6] This model is formally equivalent to one in which the owner asks managers to exert non-contractible effort. Interpreting the non-contractible variable as resource utilization allows for a tighter connection with the stylized facts on decentralization.

[7] Throughout, we assume that perfectly enforceable contracts can be written on ρ_{itm} but that no contracts can be written directly on the manager's utilization choice.

[8] Alternatively, the ideas that firms are based on are of heterogeneous quality, or some owners possess more appropriate skills for the environment they operate in or are more successful in adopting good management practices.

K_{it} and a mass M_{it} of managers who choose utilization levels $\hat{\delta}_{it} \equiv \left\{\hat{\delta}_{itm}\right\}_{m \in M_{it}}$, firm i's production in period t is given by

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

Throughout, make the following assumption.

Assumption 1. $\theta < 1 - \alpha - \theta$.

Assumption 1 ensures that utilization levels across managers are substitutes. Further, it is a sufficient condition for the first-order conditions for the unconstrained problem to be sufficient. In period t, if owner i pays all rewards, his profits are

$$\pi_i\left(\hat{\delta}_{it}, K_{it}, M_{it}, p_t\right) = p_t y_i\left(\hat{\delta}_{it}, K_{it}, M_{it}\right) - RK_{it} - \int_0^{M_{it}}\left(\delta_{itm} + s_{itm}\left(\rho_{itm}\right) + b_{itm}\right) dm - F.$$

We will analyze the owner's optimal solution to this problem when different performance measures are available. The next section analyzes the case where $\rho_{itm}\left(\hat{\delta}_{itm}\right) = \hat{\delta}_{itm}$, so that formal contracts can be written directly on utilization levels (obviating the need to use relational incentives), and the section that follows examines the pure relational incentives case, where $\rho_{itm}\left(\hat{\delta}_{itm}\right)$ is constant. Intermediate cases are considered in Section 5.

Throughout, I assume that the rental rate of capital is exogenously given and constant at R.[9] Additionally, I will maintain the assumption of perfect competition in the product market. Alternatively, as I show in Appendix B, this model is equivalent to a monopolistic competition model, where φ_i is a function of the size of the market for the variety that firm i produces. Finally, the mass of firms in the economy is fixed at 1. In Appendix C, I allow for endogenous firm entry. As in Hopenhayn (1992), a firm can pay a sunk cost F^e to enter the market and draw a value $\varphi_i \sim \Phi$. The resulting mass of entrants is determined by an indifference condition.

3 Complete Contracts

As a benchmark, consider the case where $\rho_{itm}\left(\hat{\delta}_{itm}\right) = \hat{\delta}_{itm}$, so that the owner can use the contractible portion of the payment, s_{itm}, to both pin each manager to his (IR) constraint and directly choose his utilization level (say, by setting $s_{itm}\left(\hat{\delta}_{itm} \neq \delta_{itm}\right) = -\infty$). Because in this case, there are no intertemporal linkages in the problem, each firm can solve its

[9] I will be focusing on the steady state of this economy. Consequently, it is possible to microfound the stationary aggregate demand function and constant rental rate by specifying an explicit consumer choice model, but for simplicity, I do not do this.

profit-maximization problem period-by-period. Given a price level p_t, owner i wants to choose M_{it}, $\{\delta_{itm}\}_{m\in[0,M_{it}]}$, and K_{it} to solve the following problem.

$$\max_{K_{it},M_{it},\{\delta_{itm},s_{itm}\}_{m\in[0,M_{it}]}} p_t\varphi_i K_{it}^{\alpha}\left(\int_0^{M_{it}}(e_{itm})^{\frac{\theta}{1-\alpha-\theta}}dm\right)^{1-\alpha-\theta}-RK_{it}-\int_0^{M_{it}}(\delta_{itm}+s_{itm})\,dm-F \tag{1}$$

subject to each manager's individual rationality constraint, which will hold with equality: $s_{itm}(\delta_{itm})=W$.

By Assumption 1, the firm's problem is concave in $\{\delta_{itm}\}_{m\in[0,M_{it}]}$, and managers are symmetric, so any optimal solution must satisfy $\delta_{itm}=\delta_{it}$ for all m. Recognizing this and substituting the (IR) constraint into (1), the problem becomes

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

There will be some shutdown value of ability, φ_S, for which $\varphi_i<\varphi_S$ implies that a firm with potential φ_i should optimally not produce. The solution to this problem is captured in the following proposition.

Proposition 1 Let

$$\begin{aligned}\varphi_S(p_t) &= \frac{F^{\theta}}{p_t}\left(\frac{1}{\theta^2}\right)^{\theta}\left(\frac{R}{\alpha}\right)^{\alpha}\left(\frac{W}{1-\alpha-2\theta}\right)^{1-\alpha-2\theta}\\ H(\varphi_i,p_t) &= (p_t\varphi_i)^{\frac{1}{\theta}}\left(\frac{\alpha}{R}\right)^{\frac{\alpha}{\theta}}\left(\frac{1-\alpha-2\theta}{W}\right)^{\frac{1-\alpha-\theta}{\theta}}.\end{aligned}$$

The unconstrained solution to firm i's problem is

$$\begin{aligned}\delta^{FB} &= \frac{W}{1-\alpha-2\theta}\theta;\\ M^{FB}(\varphi_i,p_t) &= \frac{1-\alpha-2\theta}{W}H(\varphi_i,p_t)\,\delta^{FB}\\ K^{FB}(\varphi_i,p_t) &= \frac{\alpha}{R}H(\varphi_i,p_t)\,\delta^{FB}\end{aligned}$$

if $\varphi_i\geq\varphi_S$. If $\varphi_i<\varphi_S$, firm i optimally does not produce. Equilibrium total factor productivity for a firm with ability φ_i is given by

$$A_i^{FB}(\varphi_i,p_t)=\frac{y_i}{K_i^{\alpha}M_i^{1-\alpha-\theta}}=\varphi_i\left(\delta^{FB}\right)^{\theta}$$

Since the solution to the period t problem does not depend on variables from any other period, and demand is stationary, output prices will be constant, $p_t = p$ for all t. A *competitive equilibrium* is then a price level p and a vector of firm-level choices $\{K_i, M_i, \delta_i\}_{i\in[0,1]}$ such that these choices are optimal given the price level, and the price level clears the market in each period. It is straightforward to verify that a competitive equilibrium exists and is unique.

It is 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, δ^{FB}. It does not depend on the interest rate, r, or the equilibrium prices, p. This will stand in contrast to the results of the following section, where managers' utilization choices are not directly contractible.

4 Relational Incentive Contracts

We now turn to the heart of the model and assume that $\rho_{itm} = \emptyset$ for all $\hat{\delta}_{itm}$. Resource utilization is non-contractible, and therefore s_{itm} is constant. The owner would like to incentivize his agents to utilize resources, but he can only do so by making a promise that he will pay a pre-specified reward if the manager chooses a particular utilization level. The owner cannot commit to doing so, so in a one-shot game, after the manager's utilization choice has been sunk, the owner would always prefer not to pay the reward, and thus, a forward-looking manager will not choose a positive utilization level (and therefore the owner will not entrust any resources to the manager). However, the owner may use future competitive rents as a partial commitment device.

His ability to do so depends on the clarity with which his failure to pay promised rewards gets communicated to his current and potential future managers. Throughout, I make the following assumption of perfect observability.

Assumption 2. A firm's potential future managers commonly observe the entrusted resources and utilization choices of individual managers and whether or not they were paid their promised rewards.

This assumption ensures that the totality of a firm's future competitive rents can be used as collateral in its promises. Relaxing Assumption 2 to the case of all-or-nothing public monitoring makes the goal of dynamic enforcement more difficult to achieve but does not qualitatively change any of the results (see Appendix A).[10][11]

In addition, I assume that managers can be rematched to another firm at no cost, which

[10] Bull (1987) justifies this form of monitoring in an overlapping-generations model in which young workers, who are eventually promoted to management, observe whether the promises made to the previous generation of managers were upheld.

[11] Ghosh and Ray (1996) and Kranton (1996) show that completely shutting down observability with respect to future potential managers can lead to interesting relationship dynamics.

prevents the firm from leveraging quasi-rents derived from labor market frictions to aid in dynamic enforcement–unlike in Shapiro and Stiglitz (1984), unemployment can not serve as a worker discipline device. I similarly assume that capital is not firm-specific.

Assumption 3. Managers can be rematched with a different firm at no cost. Capital is not firm-specific.

4.1 Dynamic Enforcement

Under what conditions does the owner have the credibility necessary to ensure that managers will utilize all the resources they have been entrusted with? To examine this, we will look for an equilibrium in which a firm's managers begin by fully utilizing the resources they have been entrusted with, and owners respond to this by rewarding them as promised. In any given period, if in the past, the owner failed to pay any number of managers their promised reward following full utilization, players revert to the unique SPNE of the state game: each manager utilizes zero resources, the owner does not entrust any resources to any manager, managers reject all offers, and the owner does not pay the fixed cost of production. Such a trigger strategy constitutes an optimal penal code (Abreu (1988)).[12][13] In this section, we will derive conditions under which utilization levels $[\delta_{it}]_{m\in M_{it}}$ are sustainable as part of a relational contract.

Suppose manager m believes the owner will pay reward b_{itm} if and only if he chooses utilization level $\hat{\delta}_{itm} = \delta_{itm}$. Then he will choose $\hat{\delta}_{itm} = \delta_{itm}$ (instead of $\hat{\delta}_{itm} = 0$, in which case he walks away with δ_{itm}) if

$$b_{itm} + \frac{1}{1+r}\left(U_{i,t+1,m} - \tilde{U}_{i,t+1,m}\right) \geq \delta_{itm}, \tag{2}$$

where $U_{i,t+1,m}$ is the continuation utility manager m receives from $t+1$ on if the relationship is not terminated, and $\tilde{U}_{i,t+1,m}$ is the continuation utility he receives if separation occurs. Thus, he will choose full utilization if and only if the sum of the reward and the change in the continuation value exceeds the value of the resources.

If the manager chooses any utilization level other than full utilization, the owner has no incentive to pay the reward and therefore will not. If the manager fully utilizes resources

[12] Such a trigger strategy is not renegotiation-proof. However, since monetary transfers are possible, it is outcome equivalent to a renegotiation-proof strategy in which, following a deviation by the principal, all the competitive rents are given to the managers.

[13] The possibility of small perturbations could potentially render this trigger strategy suboptimal. However, a richer model involving such perturbations would exhibit a richer equilibrium, which would likely still have the desired property that the set of enforceable decentralization levels is increasing in future profits.

δ_{itm}, the owner will pay the promised reward b_{itm} if

$$\frac{1}{1+r}\left(\Pi_{i,t+1,m}-\tilde{\Pi}_{i,t+1,m}\right)\geq b_{itm}, \tag{3}$$

where $\Pi_{i,t+1,m}$ and $\tilde{\Pi}_{i,t+1,m}$ are, respectively, owner i's continuation value if he pays the promised reward and if he does not pay the promised reward. Thus, the change in continuation value of the firm must exceed the size of the promised bonus.

We know from MacLeod and Malcomson (1989) and Levin (2003) that (2) and (3) can be pooled together to provide necessary and sufficient conditions for the manager to choose full utilization and the principal to pay the promised reward. That is, if we let $S_{i,t+1.m} = U_{i,t+1,m} + \Pi_{i,t+1,m}$ and $\tilde{S}_{i,t+1,m} = \tilde{U}_{i,t+1,m} + \tilde{\Pi}_{i,t+1,m}$, we need that

$$\frac{1}{1+r}\left(S_{i,t+1,m}-\tilde{S}_{i,t+1,m}\right)\geq \delta_{itm} \tag{4}$$

is satisfied. $\tilde{S}_{i,t+1,m}$ is in principle not a straightforward object, since other relationships within the firm may be altered, and the owner may choose to renege on multiple managers simultaneously. However, the candidate equilibrium described above involves multilateral punishment: an owner's choice to renege on a single manager leads all current and potential future managers to stop cooperating[14]. The only potentially appealing reneging temptation on the part of the owner, then, is one in which he pays no bonuses whatsoever. This allows us to focus only on the aggregate reneging temptation, which can be expressed as

$$\frac{1}{1+r}\left(S_{it+1}-\tilde{S}_{it+1}\right)\geq \int_0^{M_{it}} \delta_{itm}dm, \tag{5}$$

where $S_{i,t+1}$ represents the total profits generated by the owner and the managers she hires net of their outside opportunities, and $\tilde{S}_{i,t+1} = 0$ is the firm's outside option. Finally, note that $S_{i,t+1}$ depends on the whole future stream of prices and future promises. Given a conjecture $\{p_\tau\}_{\tau=t}^{\infty}$ that is shared by the owner and the managers, S_{it+1} is given by

$$\sum_{\tau=t+1}^{\infty}\left(\frac{1}{1+r}\right)^{\tau-t-1}\left[p_\tau\varphi_i K_{i\tau}\left(\int_0^{M_{it}}\left(\delta_{itm}\right)^{\frac{\theta}{1-\alpha-\theta}}dm\right)^{1-\alpha-\theta} - RK_{i\tau} - \int_0^{M_{it}}\left(W+\delta_{i\tau m}\right)dm - F\right]. \tag{6}$$

$\{p_\tau\}_{\tau=t}^{\infty}$ are determined jointly by the production capabilities and relational contracts of all the firms in the economy as well as demand conditions.

[14] Levin (2002) shows that any level of cooperation that can be sustained in a sequence of bilateral relational contracts can also be sustained in a multilateral relational contract, and in fact the latter can sometimes sustain strictly higher levels of cooperation.

4.2 Rational Expectations Equilibrium

Throughout, we will focus on rational expectations equilibria in which all firms conjectured the same price sequence, and this price sequence in fact clears the market in each period. Because it is not essential for the model, assume the capital rental rate is exogenously fixed at R.

Definition 1 A **rational-expectations equilibrium (REE)** is a sequence of prices $\{p_t\}_{t=1}^{\infty}$, a sequence of capital and management choices $\{K_{it}, M_{it}\}_{it}$, a sequence of relational contracts $\{\delta_{itm}, s_{itm}, b_{itm}\}_{itm}$, and a sequence of utilization choices $\left\{\hat{\delta}_{itm}\right\}_{itm}$ such that at each time t

1. Given promised reward b_{itm} and resources δ_{itm}, manager m for firm i optimally chooses full utilization $\hat{\delta}_{itm} = \delta_{itm}$
2. Given the conjectured price sequence $\{p_t\}_{t=1}^{\infty}$, owner i optimally offers relational contract $\{\delta_{tm}, s_{tm}, b_{tm}\}_{tm}$ and chooses capital and management levels $\{K_{it}, M_{it}\}_t$
3. $\{p_t\}_{t=1}^{\infty}$ clears the output market for all t

Throughout, I will focus on **stationary REEs** with constant prices $p_t = p$, since they are the direct analogue of the (unique) stationary competitive equilibrium in the complete-contracts case. The following proposition establishes existence and uniqueness of a stationary REE.[15] The proof is constructive.

Theorem 1 Suppose D is smooth, decreasing, and satisfies $\lim_{p\to 0} D(p) = \infty$, $\lim_{p\to\infty} D(p) = 0$, and suppose Φ is absolutely continuous. There exists a unique stationary REE.

Proof. 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 use a stationary relational contract $\{\delta_{jtm}, s_{jtm}, b_{jtm}\} = \{\delta_{jm}, s_{jm}, b_{jm}\}$ and choose constant capital and management levels $\{K_{jt}, M_{jt}\} = \{K_j, M_j\}$. Further, suppose firm i chooses constant capital and management levels $\{K_{it}, M_{it}\} = \{K_i, M_i\}$. From firm i's perspective, the environment is stationary. By Levin (2003), firm i can replicate any optimal relational

[15] In a related model, I have shown that there may exist a nonstationary REE with price cycles. The basic intuition is the following. Suppose all firms believe that output prices will be high and then low and then high and so forth. Then from the perspective of a period in which prices are high, the future looks relatively grim, as prices will be low in the future. This constrains the level of decentralization firms can sustain as part of an optimal relational contract today, which leads to a restriction in quantity and hence a high price today. From tomorrow's perspective, future prices will be high, and thus the firm's competitive rents are sufficient for sustaining high levels of decentralization. Quantity is high and therefore prices are low. Thus, this two-point alternating price sequence is consistent with equilibrium.

contract with a stationary relational contract. Thus, $\{\delta_{itm}, s_{itm}, b_{itm}\} = \{\delta_{im}, s_{im}, b_{im}\}$, which rationalizes the firm's choice of a constant capital and management sequence. This implies a constant aggregate production sequence, which yields aggregate supply $S(p)$.

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 smooth, 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. ■

4.3 Equilibrium Optimal Relational Contracts

The remaining sections characterize optimal relational contracts in the stationary REE and examine the aggregate implications of dynamic enforcement constraints. The proof of Theorem 1 charts a roadmap for how to construct the stationary REE: (1) fix output prices $p_t = p$ and solve for each firm's optimal stationary relational contract, (2) aggregate up the production of individual firms to generate the industry supply curve $S(p)$, and (3) solve for the equilibrium price p^* that satisfies $S(p^*) = D(p^*)$.

By Assumption 1, production is concave in individual utilization levels. Since managers are symmetric, any optimal relational contract will involve $\delta_{im} = \delta_i$ for all m. At the steady state, per period profits for firm i are given by

$$\pi_i(K_i, M_i, \delta_i) = p\varphi_i \delta_i^\theta K_i^\alpha M_i^{1-\alpha-\theta} - RK_i - (W + \delta_i) M_i - F.$$

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

$$\max_{K_i, M_i, \delta_i} \pi_i(K_i, M_i, \delta_i) \tag{7}$$

subject to

$$\pi_i(K_i, M_i, \delta_i) \geq r M_i \delta_i. \tag{8}$$

In the formulation of the production function, I have assumed that if all managers choose the same utilization levels, production exhibits decreasing returns to scale in K and M. This is a standard assumption in models in which firms of different productivities coexist in equilibrium (e.g. Lucas (1978)).[16] Alternatively, the entire model can be reformulated as a

[16] Identifying the nature of organizational decreasing returns to scale is an unresolved question and is beyond the scope of the current paper. Such decreasing returns to scale are, however, partly offset by the credibility channel, and this might lead to inefficiently large (or in a richer model with multiproduct firms,

monopolistic competition model in which each firm's production exhibits constant returns to scale in K and M, but each firm faces a downward-sloping demand curve (see Appendix B). Proposition 7 in the appendix shows that if revenues exhibit constant returns in K and M, there does not exist an REE. With constant returns to scale, equilibrium prices will be such that firms that produce make zero profits, which in turn precludes such firms from producing at all.

We can think of the interest rate the firm faces as an effective interest rate that combines firm turnover (i.e. an exogenous probability of firm destruction), pure time preferences, monitoring technology on the part of the firm (i.e. can the firm see whether or not a manager has chosen the correct utilization level?), social connections on the part of the population of managers (i.e. can future managers see if the owner has paid the promised rewards?).[17] In other words, think of r as fairly large.

The next proposition characterizes the solution to the constrained problem (7) subject to (8).

Proposition 2 In this model, the solution to the constrained problem satisfies

$$\frac{\delta^*(\varphi_i, p)}{\delta^{FB}} = \frac{M^*(\varphi_i, p)}{M^{FB}(\varphi_i, p)} = \frac{K^*(\varphi_i, p)}{K^{FB}(\varphi_i, p)} = \mu^*(\varphi_i, p),$$

where $0 \leq \mu^*(\varphi_i, p) \leq 1$ is (weakly) increasing in output prices and (weakly) decreasing in the capital rental rate, the outside options of managers, and the firm's effective discount rate. Further, if we define

$$\varphi_L(p) \equiv (1+r)^\theta \varphi_S(p); \; \varphi_H(p) = (1-r)^{-\theta} \varphi_S(p),$$

then

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

Equilibrium total factor productivity is given by

$$A_i(\varphi_i, p) = \frac{y_i}{K_i^\alpha M_i^{1-\alpha-\theta}} = \varphi_i \mu^*(\varphi_i, p)^\theta \left(\delta^{FB}\right)^\theta.$$

inefficiently broad) firms.

[17] Proposition 8 in the appendix shows that if with probability q_O and q_M, deviations by the owner and by each manager, respectively, are publicly detected, and if with probability $1 - q_X$, the firm exogenously exits the industry, then the effective interest rate is $\tilde{r} = \frac{r}{q_O q_M q_X}$.

Proof. See Appendix. ∎

Figure 2 characterizes this solution as a function of φ. In the complete contracts model, $\delta^*(\varphi_i, p)$ equals zero if φ_i is not large enough for the firm to cover its fixed costs of production and $\delta^*(\varphi_i, p) = \delta^{FB}$ otherwise. When formal contracts are unavailable, there are three additional regions. For $\varphi_S \leq \varphi_i < \varphi_L$, the firm should produce but is unable to. For $\varphi_L \leq \varphi_i < \varphi_H$, the dynamic enforcement constraint is binding, and the firm is unable to produce efficiently. For $\varphi_i \geq \varphi_H$, the firm is unconstrained and thus produces according to first-best.

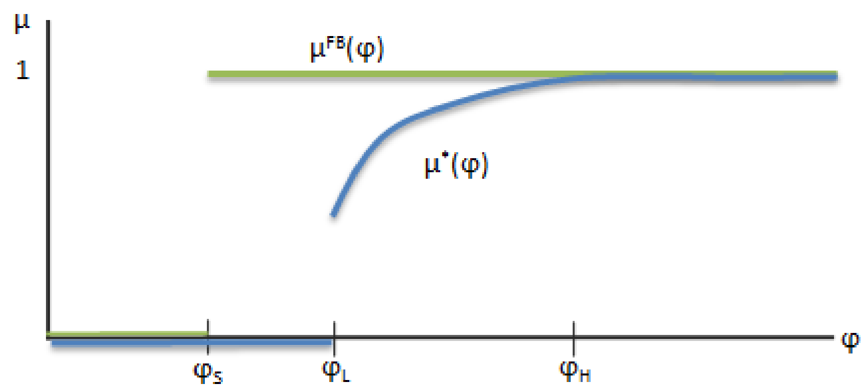

Figure 2: Equilibrium and FB solution for given prices

Equilibrium total factor productivity for a firm with ability φ_i is proportional to the first-best total factor productivity for a firm of ability φ_i: $A_i(\varphi_i, p) = \mu^*(\varphi_i, p)^{\theta} A_i^{FB}(\varphi_i, p)$. The relationship between equilibrium TFP and first-best TFP, holding prices constant is shown in Figure 3.

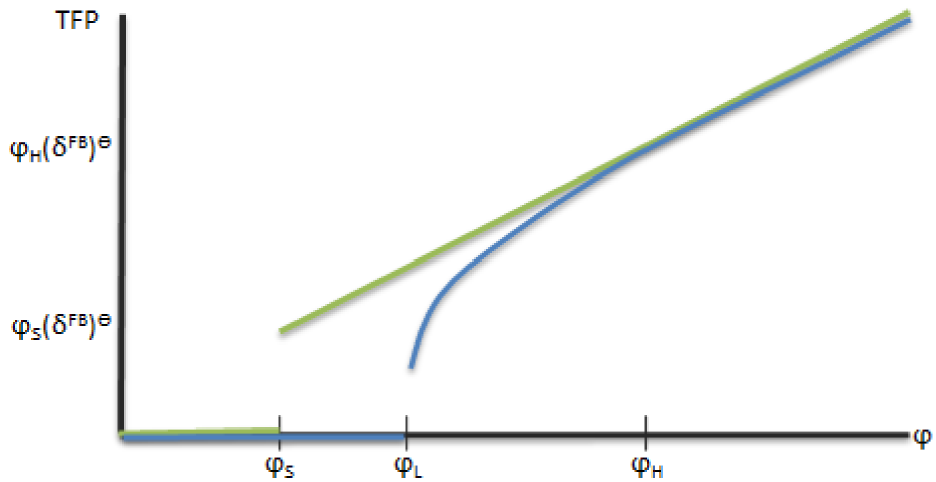

Figure 3: Equilibrium and FB TFP levels for given prices

A firm's total factor productivity depends on the effective discount rate a firm faces and is therefore decreasing in firm turnover and increasing in the clarity with which deviations are communicated. The quality of communication technology and the strength of social connections may therefore play a role in determining a firm's total factor productivity. 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 growth model.

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, in the absence of formal contracts, firms of heterogeneous marginal productivity coexist in equilibrium: high-ability firms are less constrained, implying a lower marginal productivity. Further, high-ability firms are able to sustain higher levels of decentralization and therefore have higher total factor productivity. This implies a negative relationship between total factor productivity and marginal productivity in equilibrium.

So far, I have derived that, for a constant 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

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

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 have enough credibility to sustain any positive level of decentralization if prices are constant at p. It is worth noting that $\varphi_L(p)$ is continuous and decreasing in p: if prices are higher, then future competitive rents are higher, and therefore firms with lower ability will be able to sustain positive levels of decentralization. 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 greater. Therefore, $S(p)$ is strictly increasing in p. Equilibrium prices, p^* therefore solve $D(p^*) = S(p^*)$ in each period.

5 Efficiency of REE

We now examine some of the properties of the stationary rational expectations equilibrium derived in the previous section. To build intuition for the nature of the inefficiencies in this economy, let

$$\pi^*(\varphi, p^*, F) = \max_{K,M,\delta} \{\pi(K, M, \delta; \varphi, p^*, F) : \pi(K, M, \delta; \varphi, p^*, F) \geq rM\delta\}$$

denote the optimal per-period profits of a firm with ability φ when equilibrium prices are p^*, and let $\lambda^*(\varphi, p^*)$ denote the shadow cost of the dynamic enforcement constraint at the optimum. By the envelope theorem,

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

In addition to the static effect on per-period profits, a reduction in fixed costs increases future profits–this in turn increases the firm's credibility and allows it to increase decentralization. The dynamic effect is greater the more constrained the firm is. From the previous section, we know that $\mu(\varphi, p^*)$ is increasing in φ. It can be shown (by evaluating the first-order conditions of the Lagrangian at the constrained optimum) that

$$\lambda^*(\varphi, p^*, F) = \frac{1}{1+r} \frac{1 - \mu(\varphi, p^*, F)}{\mu(\varphi, p^*, F) - \frac{1}{1+r}}.$$

From this, we see that $\lambda^*(\varphi, p^*, F)$ is decreasing in φ (indeed, $\lambda^*(\varphi_L(p^*), p^*, F) = \infty$ and $\lambda^*(\varphi, p^*, F) = 0$ for $\varphi \geq \varphi_H(p^*)$). Higher ability firms are less constrained in equilibrium, and therefore benefit less from an increase in future profits.

Ignoring potential price effects, suppose a social planner can permanently redistribute fixed costs of production among existing firms, allocating fixed costs $F(\varphi)$ to a firm of ability φ. The planner's problem is

$$\max_{F(\varphi)} \int_{\varphi_L(p^*)}^{\infty} \pi^*\left(\varphi, p^*, F(\varphi)\right) dG(\varphi) \tag{9}$$

subject to

$$\int_{\varphi_L(p^*)}^{\infty} F(\varphi)\, dG(\varphi) = F. \tag{10}$$

Proposition 3 A social planner who maximizes (9) subject to (10) will choose

$$F^*(\varphi) = \frac{\varphi^{1/\theta} \cdot \left(1 - \Phi\left(\varphi_L(p^*)\right)\right)}{E\left[\varphi^{1/\theta} \middle| \varphi \geq \varphi_L(p^*)\right]} F.$$

Proof. See Appendix. ■

The solution to this problem satisfies $F^*(\varphi) > F$ if $\varphi > \left(\frac{E\left[\varphi^{1/\theta} \middle| \varphi \geq \varphi_L(p^*)\right]}{1-\Phi(\varphi_L(p^*))}\right)^{\theta}$ and $F^*(\varphi) < F$ otherwise. That is, in the optimal reallocation, high-ability firms are taxed and low-ability firms are subsidized: rents are allocated too progressively. Of course, this is a partial equilibrium statement, since reallocating fixed costs of production would lead to an increase in output and hence a reduction in prices. Even accounting for equilibrium changes in the price level, a social planner could improve upon the allocation. To see this, suppose Φ is unbounded from above, so that for any price level p, there will be a positive mass of firms with $\varphi_i > \varphi_H(p) + \zeta$ for some small but positive ζ, so they are unconstrained. Consider a persistent proportional output tax τ on such firms. Let $T(\tau)$ be tax revenues generated by this tax scheme, and define p^τ to solve

$$D(p^\tau) = S(p^\tau; \tau)$$

Total per-period welfare is given by

$$\begin{aligned} W(\tau) \quad = \quad & \underbrace{\int_{\varphi_L(p^\tau)}^{\varphi_H(p^\tau)+\zeta} \pi^*(p^\tau, \varphi; 0)\, dG(\varphi)}_{\text{Producer surplus of untaxed firms}} + \underbrace{\int_{\varphi_H(p^\tau)+\zeta}^{\infty} \pi^*(p^\tau, \varphi; \tau)\, dG(\varphi)}_{\text{Producer surplus of taxed firms}} \\ & + \underbrace{\int_{p^\tau}^{\infty} D(p)\, dp}_{\text{Consumer surplus}} + \underbrace{T(\tau)}_{\text{Revenues}}, \end{aligned} \tag{11}$$

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

Theorem 2 $W'(0) > 0$.

Proof. See appendix. ■

The basic idea of the proof is that a small increase in taxes on unconstrained firms leads to a price increase, which induces a transfer from consumers to the constrained firms. Statically, this is merely a transfer. However, the dynamic effects of this transfer in each period result in a relaxation of the dynamic enforcement constraint and hence an increase in efficiency of the constrained firms. This proposition highlights the nature of the inefficiencies: high-ability firms induce a first-order negative pecuniary externality on low-ability firms. The result of this is that competitive rents are allocated too progressively.

A full treatment of optimal corporate taxation in the presence of credibility constraints is beyond the scope of this paper, but it is interesting to note that, in contrast to classical results on optimal tax theory (Diamond and Mirrlees (1971)), 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. That is, there is no misallocation of productive resources across independent production units.

6 Empirical Implications Within and Across Countries

This section explores two sets of empirical implications of this model. First, I examine the model's within-country implications for firm-level productivity changes in response to persistent changes in aggregate demand. Second, I consider its implications for cross-country differences in the distribution of firm-level productivity. Both sets of implications will build upon the idea that low-ability firms are more sensitive than high-ability firms to changes in competitive rents. To formalize this, recall that, in equilibrium, the total factor productivity for a firm with ability φ is given by

$$A\left(\varphi, p^*, F\right) = \varphi \mu^*\left(\varphi, p^*, F\right)^\theta \left(\delta^{FB}\right)^\theta .$$

The following proposition shows that the sensitivity of total factor productivity to future competitive rents is greater for low-ability firms than for high-ability firms.

Proposition 4 $\log A^*\left(\varphi, p, F\right)$ is increasing and exhibits decreasing differences in $(\varphi, -F, p)$.

Proof. See appendix. ■

In response to an unexpected, persistent increase (decrease) in aggregate demand, this proposition suggests that more constrained firms will see a proportionally larger increase (decrease) in their productivity than less constrained firms. This is consistent with the micro data from Baily, Bartelsman, and Haltiwanger (2001) on firm-level responses to the business cycle from 1979 to 1989. Section 6.1 explores this implication.

In Section 6.2, I compare productivity distributions across countries that differ in the strength of formal contracting institutions. An increase in the strength of formal contracting institutions reduces the importance of credibility in sustaining decentralization, and therefore disproportionately benefits low-ability firms. This leads to a compression of the left tail and a reduction in productivity dispersion. These effects hold even after accounting for the general equilibrium effects that result (provided that the price effects are not too large): if all firms produce more at a given price, equilibrium prices must fall. This reduction in prices in equilibrium reduces the output of high ability, unconstrained firms, which leads to a compression in the distribution of output across firms.

6.1 Responses to Sustained Changes in Aggregate Demand

This section analyzes firm-level responses to unexpected, sustained changes in aggregate demand. Suppose initially, aggregate demand is given by $D(p;\psi_M)$. We will compare the steady state equilibrium in this economy to that of two other economies: one in which aggregate demand is given by $D(p;\psi_L)$ (i.e. a low-demand economy) and one in which aggregate demand is $D(p;\psi_H)$. Throughout, assume D is increasing in ψ and $\psi_H > \psi_M > \psi_L$. In equilibrium, it will be the case that $p_H > p_M > p_L$.

Proposition 7 implies that all surviving firms will experience an increase (weakly) in productivity in ψ_H relative to ψ_M, and a decrease (weakly) in productivity in ψ_L relative to ψ_M. If we view changes in aggregate demand as the dawning of a boom or a bust, then the model predicts pro-cyclical within-firm productivity changes. This contrasts with an efficiency wage story: downward-rigid wages and increased unemployment during recessions should enable firms to implement higher levels of effort, and thus we would expect to see within-firm productivity increase during a recession. Further, since A exhibits decreasing differences in (φ, p), the incidence of these changes will be primarily centered around low-ability firms: firms whose futures are not necessarily bright.

Macroeconomic evidence, dating back to at least Hultgren (1960), strongly suggests that aggregate productivity is pro-cyclical.[18] 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 this procyclicality was driven by within-firm productivity declines during the slump that occurred between 1977 and 1982 and within-firm productivity increases during the boom that occurred between 1982 and 1987. Within-firm productivity changes

[18] At least until the most recent recessions. See Gali (2010) and Berger (2011).

were found to be procyclical.

Baily, Bartelsman, and Haltiwanger (2001) decompose these productivity changes further. They examine firm-level changes in productivity over the period between 1979 and 1988. They find that the productivity of firms with bright long-run prospects (as predicted by variables that are observable in 1979) was procyclical, but not very much so. Firms with poor long-run predicted prospects, on the other hand, exhibited much greater degrees of procyclicality. This is consistent with the predictions of Proposition 4, as shown in Figure 4: firms with observably dim prospects (low-ability firms) are expected to be much more sensitive to changes in aggregate demand.

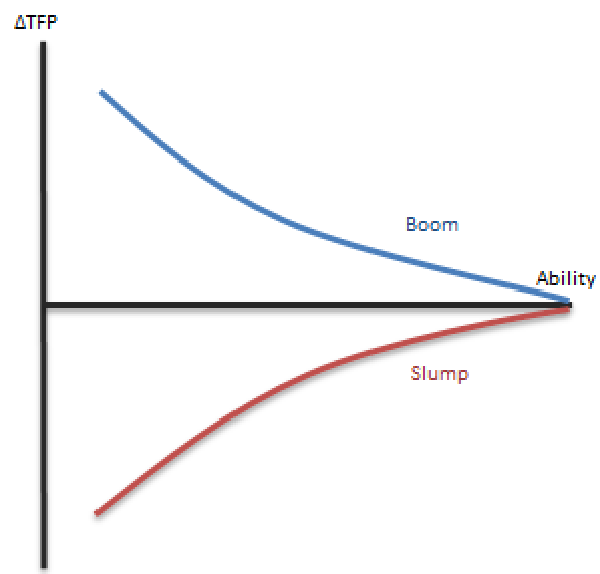

Figure 4: Patterns of procylical TFP

6.2 Differences in Formal Contracting Institutions

Bartelsman, Haltiwanger, and Scarpetta (Forthcoming) and Hsieh and Klenow (2009) document both substantial dispersion in within-country productivity, controlling for industry composition, and heterogeneity in productivity dispersion across countries. Loosely speaking, there is more productivity dispersion in less-developed countries. Other authors have similarly documented "... huge variation among countries in the speed and quality of courts." (Djankov, La Porta, Lopez-de-Silanes, and Shleifer (2003)). The objective of this section is to connect these two sets of facts.

To do so, I begin by extending the model to allow for imperfect formal contracts that can supplement relational contracts. This is in line with Johnson, McMillan, and Woodruff

(2002)'s finding that "... entrepreneurs who say the courts are effective have measurably more trust in their trading partners... ." With stronger formal contracting institutions, credibility becomes relatively less important for sustaining decentralization, consistent with the positive correlation between Kaufmann, Kraay, and Mastruzzi (2006)'s country-level measure of "rule of law" and Bloom, Sadun, and Van Reenen (Forthcoming)'s measure of decentralized decision making in organizations. Since in equilibrium, low-ability firms are more constrained by lack of credibility, an improvement in formal contracting institutions will disproportionately benefit such firms.

6.2.1 Imperfect Formal Contracts

Suppose a third-party enforcer observes δ_{itm} and $\hat{\delta}_{itm}$. However, the third-party enforcer will only enforce deviations that are at least $(1-\omega)$-egregious, for $\omega \in [0,1]$, which implies that the formal portion of the contract can be contingent on whether or not $\rho_{itm}\left(\hat{\delta}_{itm}, \delta_{itm}\right) = 1\left\{\hat{\delta}_{itm} \geq \omega\delta_{itm}\right\}$, and hence $s_{itm}\left(\rho_{itm} = 0\right)$ can be set to $-\infty$. Enforcement is otherwise costless. Throughout, I will restrict attention to "full-utilization" relational contracts, in which any choice $\hat{\delta}_{itm} < \delta_{itm}$ is viewed as a deviation, which results in punishment.[19]

We refer to ω as the quality of formal contracting institutions. Given quality ω, suppose the principal entrusts manager m with δ_{itm} resources and promises to pay a reward b_{itm} if and only if he fully utilizes them. Manager m will choose full utilization over $\hat{\delta}_{itm} = \omega\delta_{itm}$ if and only if

$$b_{itm} + \frac{1}{1+r}\left(U_{i,t+1,m} - \tilde{U}_{i,t+1,m}\right) \geq (1-\omega)\,\delta_{itm},$$

and following full utilization, the owner will pay the promised reward if and only if

$$\frac{1}{1+r}\left(\Pi_{i,t+1,\ell} - \tilde{\Pi}_{i,t+1,\ell}\left(\omega\right)\right) \geq b_{itm}$$

As always, these constraints can be pooled within the pair and across all managers in the firm to give

$$\frac{1}{1+r}\left(S_{it+1} - \tilde{S}_{it+1}\left(\omega\right)\right) \geq (1-\omega)\,M_{it}\delta_{it\ell}.$$

Stationarity and manager symmetry implies that $S_{it+1} = \frac{1+r}{r}\pi_i$, $\tilde{S}_{it+1}\left(\omega\right) = \frac{1+r}{r}\max\left\{\tilde{\pi}_i\left(\omega\right), 0\right\}$, $M_{it} = M_i$, and $\delta_{it\ell} = \delta_i$. Here, $\tilde{\pi}_i\left(\omega\right)$ represents the per-period profits that firm i can earn

[19] This is consequential: relaxing this assumption enables each firm to achieve first-best utilization levels by setting $\delta_i = \frac{\delta^{FB}}{\omega}$, allowing each manager to choose $\hat{\delta}_i = \omega\delta_i = \delta^{FB}$ and keep the remaining $\frac{1-\omega}{\omega}\delta^{FB}$. The salary component of the contract then extracts this ex post "reward." This undesirable feature can be sidestepped if I allow instead $\rho_{itm} = 1\left\{\hat{\delta}_{itm} \leq \omega\delta^{FB}\right\}$. Such a model delivers qualitatively similar results, but its solution can only be computed numerically.

using only formal contracts.[20] The (DE) then becomes

$$\pi_i \geq (1-\omega)\, r M_i \delta_i = \tilde{r} M_i \delta_i.$$

In other words, the strength of the formal contracting institutions enters the (DE) constraint as an effective decrease in the interest rate. Under the current formulation, an increase in ω affects all firms equally, but since some firms are more constrained than others, this increase in ω disproportionately benefits such firms. Under this formulation, the following proposition is true.

Proposition 5 $\log y^* (\varphi, p, F, \omega)$ and $\log A^* (\varphi, p, F, \omega)$ are increasing in ω and exhibit decreasing differences in (φ, ω).

Proof. See appendix. ■

Holding prices constant, an increase in ω leads to an increase in total factor productivity and output. Further, low-ability firms disproportionately benefit from this increase in ω. This leads to a convergence in the productivity distribution among existing firms. However, it will also potential lead to the entry of low-ability firms. The new entrants and all existing firms produce more the greater is ω, so supply increases and therefore prices must fall. Let p^ω solve $D(p^\omega) = S(p^\omega)$. Then p^ω is decreasing in ω. This decrease in prices leads to a net reduction in production of unconstrained firms, since the increase in ω does not allow them to produce more.

Proposition 6 Suppose that 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 $Var\left(A^*(\varphi, p^\omega, \omega) \middle| \varphi \geq \varphi_L^\omega\right)$ is greater for $\omega = 0$ than for $\omega = 1$. Additionally, there exists some $\hat{\varphi}^\omega$ such that $y^*(\varphi, p^\omega, \omega)$ is increasing in ω for $\varphi < \hat{\varphi}^\omega$ and decreasing in ω for $\varphi > \hat{\varphi}^\omega$.

Proof. See appendix. ■

6.2.2 Evidence

The three main predictions from the previous section are: (1) productivity dispersion is lower in high-ω countries, (2) the distribution of productivity in high-ω countries will have a thinner left tail, and (3) output (firm-size) dispersion is lower in high ω countries. In

[20] For what follows, I will assume $\tilde{\pi}_i(\omega) < 0$ for all ω and for all i, though considering the case where $\tilde{\pi}_i(\omega) > 0$ could be an interesting extension. One way of rationalizing $\tilde{\pi}_i(\omega) < 0$ is that no matter how strong a formal contract is, certain non-contractible actions must be taken for any production to take place. A firm is unlikely to survive if its management team works "to rule."

order to examine (1), I gathered country-level measures of (a) labor productivity dispersion from Bartelsman, Haltiwanger, and Scarpetta (Forthcoming),[21][22] (b) the quality of formal contracting institutions ("Rule of Law") from Kaufmann, Kraay, and Mastruzzi (2006),[23][24] and (c) "Private Credit" from Manova (2011),[25] which proxies for the quality of capital markets. Figure 5 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. Of course, "Rule of Law" is not the only factor that varies across countries. If, as we would expect, "Rule of Law" is highly correlated with the quality of capital markets, which in turn largely determine the level of productivity dispersion in a country, then Figure 5 may simply be capturing this relationship. I show in figure 6 that the relationship between "Rule of Law" and productivity dispersion is robust to controlling for Manova (2011)'s measure of the quality of capital markets.

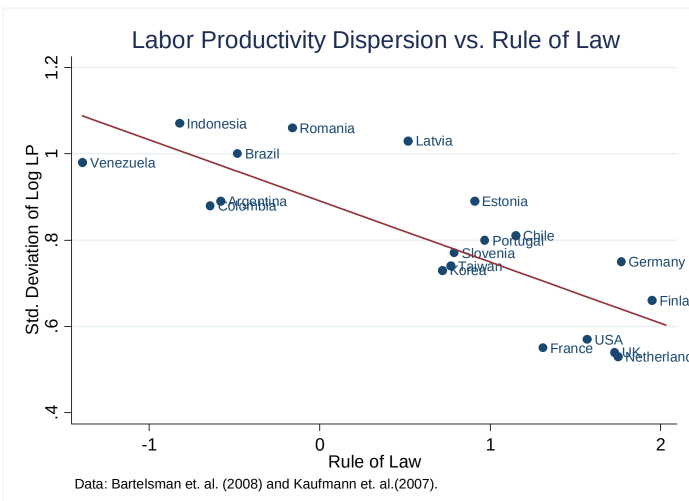

Figure 5: LP Dispersion vs Rule of Law

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

[22] Productivity measures in the empirical literature are revenue-based, which conflates technological and market-power considerations (see Foster, Haltiwanger, and Syverson (2008)). In this model, firms produce homogeneous goods and do not have market power, allowing me to sidestep these issues.

[23] 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.

[24] The results are qualitatively similar using Gwartney and Lawson (2001)'s measure of economic freedom or either of Djankov, et. al. (2003)'s measures of the quality of court enforcement.

[25] "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.

For the second prediction, Figure 7 shows Hsieh and Klenow (2009)'s plots of the log (TFP) distributions in India and the U.S., controlling for industry composition. A striking feature is the thickness of the left tail of productivity in India relative to the U.S.. In contrast, a model of Cobb-Douglas production with capital constraints would suggest that low-ω environments should see a thick *right* tail of financially constrained, and therefore excessively high average-productivity, firms.

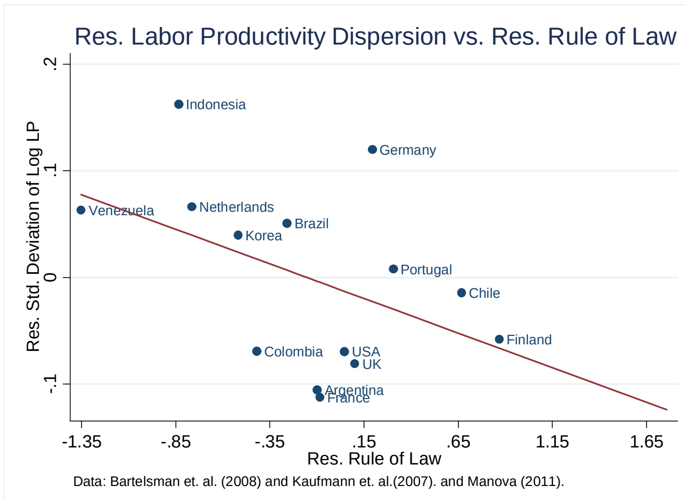

Figure 6: Controlling for "private credit"

With regards to prediction (3), the evidence is limited. Alfaro, Charlton, and Kanczuk (2008) show that *establishment* size is less variable in countries with higher GDP per capita (which is correlated with rule of law).[26] Others (Tybout (2000), Ayyagari, Demirguc-Kunt, and Beck (2003)) describe the phenomenon of the "missing middle" in developing countries: in high-income countries, medium-sized firms are responsible for a much larger share of GDP ($\approx 50\%$) than in low-income countries ($\approx 17\%$). Low-income countries tend to be dominated by firms that are either very small, often informal, or very large.

Though this model does not literally generate a "missing middle," an extension, along the lines of Peters (2011a) potentially could. In such an extension, firms can potentially choose between two technologies: one is a low productivity (traditional) technology that does not require the owner to decentralize decision-making to managers, and the other (modern) technology is given by the current model. For sufficiently weak contracting institutions, both types of firms could coexist in equilibrium–low-ability owners will choose the traditional

[26] Of course, it may be the case that, in countries with higher GDP per capita, firms expand by adding establishments, not by increasing the size of existing establishments.

technologies, and owners with sufficiently high ability to sustain decentralization will choose the modern technology. Improved formal contracting will cause some marginal traditional producers to switch to modern technologies. This will lead to increased output and hence decreased prices, which could in turn drive out some of the less productive traditional producers.

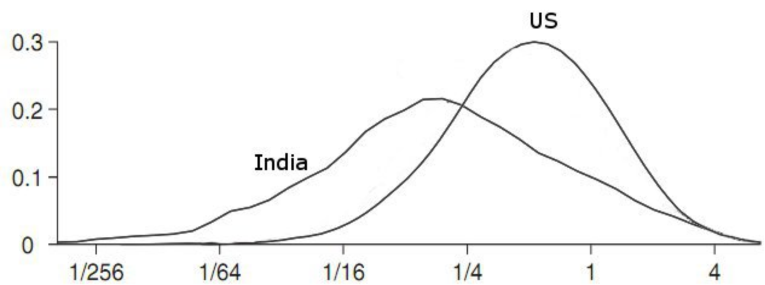

Figure 7: $\log(TFP)$ distribution (Hsieh and Klenow)

6.2.3 Alternative Solutions to the Credibility Problem

An entrepreneur who is constrained by the limited credibility of his promises may pursue policies aimed at relaxing this constraint. For example, he may purchase capital that is specific to the firm and therefore loses value if the firm is dissolved. In a multiproduct firm, he may leverage the profits earned in one product line as collateral for promises made to managers responsible for other product lines. Finally, he may hire managers with whom he interacts more frequently (perhaps relatives). Such policies are likely to be more prevalent in countries with poor formal contracting institutions.

Though privately (and potentially socially) beneficial, these alternative firm-level policies do not eliminate the inefficiency of the competitive equilibrium, however. A firm investing in capital that is otherwise suboptimally firm-specific will have inefficiently low capital productivity. A conglomerate pursuing breadth for the sake of leveraging its profits may crowd out more efficient (but narrow) producers of other goods. Firms may overemploy trustworthy family members, even if they are not a good fit for the job; further, skilled entrepreneurs may lack the familial connections necessary to profitably expand his enterprise to its optimal size. To the extent that a firm's size and scope is determined by factors orthogonal to its marginal profitability, the allocation of profits will be inefficient: some firms will be too small and others will be too large.

7 Discussion and Conclusion

Following Chandler (1962), I have argued that in order for a large firm to produce efficiently, the owner of the firm must decentralize daily operating decisions to a team of managers. Absent perfect formal contracts, decentralization requires trust: the owner must trust that the managers will not make reckless decisions for their own private gains, and the managers must trust that the owner will reward them appropriately for judiciously using the firm's resources. This paper views trust as credibility in a relational contract–the credibility of the owner's promises is derived from the value of the owner's reputation in the labor market. This value is, in turn, limited by the firm's potential future competitive rents. Competitive rents, credibility, and therefore firms' decentralization levels and hence productivity are jointly determined in industry equilibrium.

The theory of relational contracts generates a mechanism through which future profits can be an important determinant of current productivity, resulting in firm-level income effects that have efficiency consequences. The model presented in this paper argues that these firm-level income effects are decreasing–the marginal returns to a dollar-a-day increase in profits is higher for less productive firms–and therefore, profits are inefficiently concentrated at the top in a competitive equilibrium. On the normative side, this view suggests that in weak formal-contracting environments, a progressive corporate tax may improve aggregate productivity by distorting production away from high average- but low marginal-productivity firms to low average- but high marginal-productivity firms.

Other policies that reduce the concentration of profits can have similar effects. Consider two countries opening up to trade. In a Melitz (2003) model of heterogeneous firms, high-ability–and hence high profitability–firms will export. In so doing, they will drive up domestic factor prices, reducing the profitability of those firms that do not export (as well as the marginal exporting firms). Trade liberalization therefore leads to a further concentration of profits at the top. If the countries involved have poor formal contracting institutions, this concentration of profits could result in a reduction in aggregate productivity among the smaller firms, and these losses may exceed the Melitz (2003) reallocation benefits. Trade liberalization, therefore, may harm aggregate productivity in countries with weak formal contracting institutions.

On the positive side, low-ability firms face tighter credibility constraints, making their productivity more sensitive to the environment in which they operate. This effect forms the basis for two sets of empirical implications: (1) within-country, over time, and (2) across-country. Low-ability firms are uniformly more responsive to persistent changes in aggregate demand, which is consistent with micro evidence on firm-level productivity responses to business cycles (Baily, Bartelsman, and Haltiwanger (2001)). Improvements in the strength of formal contracting institutions reduce the importance of credibility in sustaining decentralization and therefore disproportionately improve the productivity of low-ability firms, leading to a reduction in the dispersion of productivity. Cross-country evidence supports the predictions of an upward compression of the left tail of the productivity distribution in high rule-of-law countries and a negative relationship between the strength of formal contracting institutions and productivity dispersion.

These patterns, 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 view proposed in this paper is establishing direct evidence of these mechanisms.[27]

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 as being made possible only by non-contractible 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 weak formal contracting institutions. Such a model may be able to generate results consistent with the recent Hsieh and Klenow (2012) facts on firm growth.

[27]There are at least two imperfect, but more direct, pieces of evidence. Olley and Pakes (1996) find that AT&T's productivity fell in response to the 1982 announcement of its 1984 divestiture. Bertrand and Mullainathan (2001) find that shocks to oil prices and exchange rates affect current performance pay even though their effects on current performance are out of control of executives. Since there is persistence in oil prices (Hamilton (2008)) and exchange rates (Rogoff (1996)), such shocks are informative about future profits and therefore may affect the credibility of firms' promises with regards to discretionary payment schemes.

Appendix A: Proofs and Derivations

Solution to the Model

Proposition 7 If production exhibits constant returns to scale in labor and management, there does not exist an REE.

Proof. Suppose production is $y_{it}(\delta_{it}, K_{it}, M_{it}) = \varphi_i \delta_{it}^{\theta} K_{it}^{\alpha} M_{it}^{1-\alpha}$. Then, in period t, the firm with the highest value of φ_i will continue to produce as long as $p_t y_{it} - RK_{it} - (W + \delta_{it}) M_{it} \geq 0$. Market clearing with finite demand thus implies that $p_t y_{it} - RK_{it} - (W + \delta_{it}) M_{it} = 0$ for all t. This in turn implies that the left-hand side of the dynamic enforcement constraint is equal to $-F$, which implies that no production can be sustained. ■

Proposition 8 Suppose with probability q_O, deviations by the owner are publicly detected, and with probability q_M, deviations by a manager are publicly detected. Finally, suppose with probability $1 - q_X$, the firm exogenously is forced to exit the industry. Then the effective interest rate in (8) is $\tilde{r} = \frac{r}{q_O q_M q_X}$.

Proof. If we rewrite (2) and (3) recognizing that (a) the owner will choose s_t to pin each manager to his (IR) constraint and that (b) the optimal relational contract will be stationary, and we introduce $q_O, q_M, q_X > 0$, these become

$$b_{im} \geq \frac{1}{q_M}\delta_{im} \text{ and } q_O q_X \frac{\pi_{im}}{r} \geq b_{im}.$$

If we pool these across agents, this becomes

$$q_O q_X \frac{\pi_{im}}{r} \geq \int b_{im} \geq \frac{1}{q_M} M_i \delta_i,$$

and therefore a reward scheme supporting δ_i exists if and only if $\pi_i \geq \frac{r}{q_O q_M q_X} M_i \delta_i \equiv \tilde{r} M_i \delta_i$, which is the desired result. ■

Proposition 2 In this model, the solution to the constrained problem satisfies

$$\frac{\delta^*(\varphi)}{\delta^{FB}} = \frac{M^*(\varphi)}{M^{FB}(\varphi)} = \frac{K^*(\varphi)}{K^{FB}(\varphi)} = \mu^*(\varphi),$$

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

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

where

$$\begin{aligned}
\varphi_L &= \frac{F^\theta}{p}\left((1+r)\frac{1}{\theta^2}\right)^\theta\left(\frac{R}{\alpha}\right)^\alpha\left(\frac{W}{1-\alpha-2\theta}\right)^{1-\alpha-2\theta} \\
\varphi_H &= \frac{F^\theta}{p}\left(\frac{1}{1-r}\frac{1}{\theta^2}\right)^\theta\left(\frac{R}{\alpha}\right)^\alpha\left(\frac{W}{1-\alpha-2\theta}\right)^{1-\alpha-2\theta}.
\end{aligned}$$

Proof. 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 $\delta_m = \delta$ for all $m \in [0, M]$. The firm's problem is then

$$\max_{K,M,\delta} p\varphi\delta^\theta K^\alpha M^{1-\alpha-\theta} - RK - (W+\delta)M - F$$

subject to

$$p\varphi\delta^\theta K^\alpha M^{1-\alpha-\theta} - RK - (W+\delta)M - F \geq rM\delta.$$

Since an increase in K increases the objective function as well as the left-hand side of the constraint, capital will be chosen efficiently, given M and δ. Define

$$\pi\left(K^*(M,\delta), M, \delta\right) = py\left(K^*(M,\delta), M, \delta\right) - RK^*(M,\delta) - (W+\delta)M - F.$$

The firm's problem is then to $\max \pi\left(K^*(M,\delta), M, \delta\right)$ subject to $\pi\left(K^*(M,\delta), M, \delta\right) \geq rM\delta$. Suppose the firm is constrained at the optimum. Define $M(\delta)$ such that the constraint holds with equality. The unconstrained problem is then

$$\max_\delta rM(\delta)\delta.$$

Taking first-order conditions, the firm chooses δ such that $\frac{M'(\delta)}{M(\delta)}\delta = -1$. Implicitly differentiating the constraint with respect to δ and substituting this into the first-order condition yields

$$\frac{py(K^*, M^*, \delta^*) - RK^*}{M^*} = \frac{1-\alpha}{1-\alpha-2\theta}W \tag{12}$$

and we know from the constraint that

$$\frac{py(K^*, M^*, \delta^*) - RK^*}{M^*} = (W + (1+r)\delta^*) + \frac{F}{M^*}. \tag{13}$$

(12) implies

$$M^*(\delta^*) = \left(\frac{1-\alpha-2\theta}{W}\right)^{\frac{1-\alpha}{\theta}}(pA)^{\frac{1}{\theta}}\left(\frac{\alpha}{R}\right)^{\frac{\alpha}{\theta}}\delta^*,$$

and substituting this into (13), we have that δ^* solves a quadratic equation. The linearity of $M^*(\delta^*)$ results from the assumption that production is constant returns to scale in (K, M, δ). Without this assumption, δ^* 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

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

Optimal capital and management are linear in δ^*. 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. ■

Structure of Inefficiencies

Proposition 3 Suppose a social planner wants to

$$\max_{F(\varphi)} \int_{\varphi_L(p^*)}^{\infty} \pi^*\left(\varphi, p^*; F(\varphi)\right) dG(\varphi)$$

subject to $\int_{\varphi_L(p^*)}^{\infty} F(\varphi)\, dG(\varphi) \geq F$ (with Lagrange multiplier $1+\Lambda$). The optimal solution satisfies

$$F^*(\varphi) = \frac{\varphi^{1/\theta}}{E\left[\varphi^{1/\theta} \middle| \varphi \geq \varphi_L(p^*)\right]} F.$$

Proof. The social planner's Lagrangian is

$$\mathcal{L} = \int_{\varphi_L(p^*)}^{\infty} \pi^*\left(\varphi, p^*; F(\varphi)\right) dG(\varphi) + (1+\Lambda)\left(\int_{\varphi_L(p^*)}^{\infty} F(\varphi)\, dG(\varphi) - F\right).$$

Any solution will involve $\lambda^*\left(\varphi, p^*; F^*(\varphi)\right) = \Lambda$. In order for $\mu\left(\varphi, p^*; F^*(\varphi)\right) = \mu\left(\varphi', p^*; F^*(\varphi')\right)$ for all φ, φ', it must be the case that $F^*(\varphi) = \left(\frac{\varphi}{\varphi'}\right)^{1/\theta} F^*(\varphi')$. Set $\varphi' = \varphi_L(p^*)$, so that we have $F^*(\varphi) = \left(\frac{\varphi}{\varphi_L(p^*)}\right)^{1/\theta} F^*\left(\varphi_L(p^*)\right)$. We then have

$$\int_{\varphi_L(p^*)}^{\infty} F^*(\varphi)\, dG(\varphi) = \frac{F^*\left(\varphi_L(p^*)\right)}{\varphi_L(p^*)^{1/\theta}} \frac{E\left[\varphi^{1/\theta} \middle| \varphi \geq \varphi_L(p^*)\right]}{1-\Phi\left(\varphi_L(p^*)\right)} = F,$$

which implies $F^*\left(\varphi_L(p^*)\right) = \frac{\varphi_L(p^*)^{1/\theta}\left(1-\Phi\left(\varphi_L(p^*)\right)\right)}{E\left[\varphi^{1/\theta} \middle| \varphi \geq \varphi_L(p^*)\right]} F$, and therefore $F^*(\varphi) = \frac{\varphi^{1/\theta}\left(1-\Phi\left(\varphi_L(p^*)\right)\right)}{E\left[\varphi^{1/\theta} \middle| \varphi \geq \varphi_L(p^*)\right]} F$.
■

Theorem 2 Let $W(\tau)$ be as defined in (11). Then $W'(0) > 0$.

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

$$\int_{\varphi_L(p^\tau)}^{\varphi_H(p^\tau)+\zeta} \pi^*(p^\tau,\varphi;0)\,dG(\varphi) + \int_{\varphi_H(p^\tau)+\zeta}^{\infty} \pi^*(p^\tau,\varphi;\tau)\,dG(\varphi) + T(\tau) = \int_{\varphi_L(p^\tau)}^{\infty} \pi^*(p^\tau,\varphi;0)\,dG(\varphi).$$

Next, using Leibniz's rule,

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

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

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

which establishes the claim. ■

Partial Equilibrium Comparative Statics

Notation 1 *Let $\underline{\mu}(\omega) = \frac{1}{1+(1-\omega)r}$. For applications with $\omega = 0$, $\underline{\mu} = \frac{1}{1+r}$. Let $\xi(\varphi,p,F,\omega) = \frac{\underline{\mu}(\omega)-\frac{1}{2}\mu(\varphi,p,\omega)}{\mu(\varphi,p,\omega)-\underline{\mu}(\omega)}$. For applications with $\omega = 0$, denote $\xi(\varphi,p,F) = \xi(\varphi,p,F,0)$. Finally, the following definitions will be useful in what follows*

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

Remark 1 *For $X \in \{\varphi,p,-F\}$,*

$$\frac{\partial\mu}{\partial X} = \frac{\mu\xi}{Z_X};\ \frac{\partial\mu}{\partial\omega} = \frac{\mu\xi}{Z_\omega}\frac{1+\xi}{\xi},$$

and for $X \in \{\varphi,p,-F,\omega\}$, $\frac{\partial\xi}{\partial X} = -\frac{\Xi}{Z_X}$. 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 1 $\log\mu(\varphi,p,F,\omega)$ is increasing in and exhibits decreasing differences in $(\varphi,p,-F,\omega)$.

Proof. That $\log\mu$ is increasing in $(\varphi, p, -F, \omega)$ follows from their characterizations in the remark above. To examine decreasing differences, we simply must check the cross-partials. For $X \in \{\varphi, p, -F\}$, the first derivatives are

$$\frac{\partial \log\mu}{\partial X} = \frac{1}{\mu}\frac{\partial \mu}{\partial X}.$$

With some effort, it can be shown that for $X, Y \in \{\varphi, p, -F, \omega\}$, $X \neq Y$,

$$\frac{\partial^2 \log\mu}{\partial X \partial Y} = -\frac{\Xi}{Z_X Z_Y}.$$

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

Proposition 4 $\log A^*(\varphi, p, F)$ is increasing and exhibits decreasing differences in $(\varphi, -F, p)$.

Proof. Since $\log A^* = \log\varphi + \theta\log\mu + \theta\log\delta^{FB}$, $\log A^*$ is increasing in $(\varphi, -F, p)$ since $\log\varphi$ is increasing in φ and $\log\mu$ is increasing in $(\varphi, -F, p)$ from the previous lemma. Since μ is the only term that contains interactions, $\log A^*$ exhibits decreasing differences in $(\varphi, -F, p)$ if $\log\mu$ exhibits decreasing differences in $(\varphi, -F, p)$, which it does by the previous lemma.which is negative. ■

Proposition 5 $\log y^*(\varphi, p, F, \omega)$ and $\log A^*(\varphi, p, F, \omega)$ are increasing in ω and exhibit decreasing differences in (φ, ω).

Proof. Note that

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

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

Industry Equilibrium Comparative Statics

This section provides a proof of proposition 9. It proceeds first by establishing three lemmas. The first lemma connects the equilibrium price response to properties of the industry supply and demand curves. The second lemma shows that when price effects are small (large), the equilibrium ability cutoff is decreasing (increasing) in the strength of formal contracts. The third lemma shows that the lowest observed productivity level will be lower when formal contracts are weaker. Lemma 4 shows that the slope of TFP with respect to ability is higher when formal contracts are weaker. These lemmas are used in the proof of proposition 9.

Lemma 2 Let p^ω solve $D(p^\omega) = S(p^\omega, \omega)$. Then

$$\left.\frac{dp}{d\omega}\frac{\omega}{p}\right|_{p^\omega} = -r\omega\theta\underline{\mu}\frac{\frac{\partial S}{\partial p}p + D}{\frac{\partial S}{\partial p}p + |\varepsilon_{D,p}|\, D}.$$

Proof. If we totally differentiate the market-clearing condition and rearrange, we get the following

$$\frac{dp^\omega}{d\omega}\frac{\omega}{p} = -\frac{\frac{\partial S}{\partial \omega}\omega}{\frac{\partial S}{\partial p}p + \left|\frac{dD}{dp}\right| p}$$

We now seek to derive a relationship between $\frac{\partial S}{\partial \omega}\omega$ and $\frac{\partial S}{\partial p}p$. Supply is

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

and therefore, using Leibniz's rule,

$$\begin{aligned}
\frac{\partial S}{\partial p} &= -\frac{\partial \varphi_L}{\partial p} y^*(\varphi_L, p, \omega)\, g(\varphi_L) + \int_{\varphi_L}^{\infty} \frac{\partial}{\partial p}\left(\mu y^{FB}\right) g(\varphi)\, d\varphi \\
\frac{\partial S}{\partial \omega} &= -\frac{\partial \varphi_L}{\partial \omega} y^*(\varphi_L, p, \omega)\, g(\varphi_L) + \int_{\varphi_L}^{\infty} \frac{\partial}{\partial \omega}\left(\mu y^{FB}\right) g(\varphi)\, d\varphi
\end{aligned}$$

Recall that $\varphi_L(p,\omega) = (1 + (1-\omega) r)^\theta \varphi_S(p)$, so that

$$\begin{aligned}
\frac{\partial \varphi_L}{\partial p}\frac{p}{\varphi_L} &= -1;\ \frac{\partial \varphi_L}{\partial \omega}\frac{\omega}{\varphi_L} = -\theta\omega r\underline{\mu} \\
\frac{\partial}{\partial p}\left(\mu y^{FB}\right) &= \frac{y^*}{p}\left[\frac{1+\xi}{\theta} - 1\right],
\end{aligned}$$

we get

$$\begin{aligned}
\frac{\partial S}{\partial p}p &= \varphi_L y^*(\varphi_L, p, \omega)\, g(\varphi_L) + \int_{\varphi_L}^{\infty}\left[\frac{1+\xi}{\theta} - 1\right] y^* g(\varphi)\, d\varphi \\
\frac{\partial S}{\partial \omega}\frac{\omega}{r\omega\theta\underline{\mu}} &= \varphi_L y^*(\varphi_L, p, \omega)\, g(\varphi_L) + \int_{\varphi_L}^{\infty}\frac{1+\xi}{\theta} y^* g(\varphi)\, d\varphi
\end{aligned}$$

Finally, note that

$$\frac{\partial S}{\partial \omega}\frac{\omega}{r\omega\theta\underline{\mu}} - \frac{\partial S}{\partial p}p = S(p,\omega).$$

At $p = p^\omega$, $S(p^\omega, \omega) = D(p^\omega)$, so the result follows. ■

Lemma 3 If $|\varepsilon_{D,p}| > 1$, then $\varphi_L^0 > \varphi_S^1$. If $|\varepsilon_{D,p}| < 1$, then $\varphi_L^0 < \varphi_S^1$.

Proof. We know that $\varphi_L(p^\omega,\omega) = (1+(1-\omega)r)^\theta \varphi_S(p^\omega)$ and

$$\frac{d\varphi_S(p^\omega)}{d\omega} = -\frac{\varphi_S(p^\omega)}{p^\omega}\frac{dp^\omega}{d\omega}.$$

Therefore,

$$\begin{aligned}\frac{d\varphi_L}{d\omega} &= -r\theta(1+(1-\omega)r)^{\theta-1}\varphi_S(p^\omega) + (1+(1-\omega)r)^\theta \frac{d\varphi_S}{d\omega} \\ &= -\frac{\varphi_L(p^\omega,\omega)}{\omega}\left(r\omega\theta\underline{\mu} + \frac{\omega}{p^\omega}\frac{dp^\omega}{d\omega}\right) = \frac{\varphi_L(p^\omega,\omega)}{\omega} r\omega\theta\underline{\mu}\left(\frac{D(1-|\varepsilon_{D,p}|)}{\frac{\partial S}{\partial p}p + |\varepsilon_{D,p}|D}\right).\end{aligned}$$

By definition, $\varphi_S^1 = \varphi_L^1$, and by the fundamental theorem of calculus,

$$\varphi_L^1 = \varphi_L^0 + \int_0^1 \frac{d\varphi_L}{d\omega}d\omega.$$

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

Lemma 4 $\varphi_L^0\mu(\varphi_L^0)^\theta(\delta^{FB})^\theta < \varphi_S^1(\delta_{FB})^\theta$.

Proof. We know that $\mu(\varphi_L^0) = \frac{1}{1+r}$ and $\varphi_L = (1+r)^\theta\varphi_S$. This implies that

$$\varphi_L^0\mu(\varphi_L^0)^\theta(\delta^{FB})^\theta = \varphi_S^0(\delta^{FB})^\theta < \varphi_S^1(\delta^{FB})^\theta,$$

where in the last inequality I used the facts that $p^1 < p^0$ and therefore $\varphi_S^1 > \varphi_S^0$. ■

Lemma 5 $\varphi\mu(\varphi,p,F,0)$ increases faster than φ for $\varphi \geq \varphi_L^0$.

Proof. We know that

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

If this expression is negative, then $\frac{d}{d\varphi}\varphi < \frac{d}{d\varphi}\varphi\mu(\varphi,p,F,0)$. Note that

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

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

Proposition 6 Suppose that 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 $Var\left(A^*(\varphi,p^\omega,\omega)\middle|\,\varphi \geq \varphi_L^\omega\right)$ is greater for $\omega = 0$ than for $\omega = 1$. Additionally, there exists some $\hat{\varphi}^\omega$ such that $y^*(\varphi,p^\omega,\omega)$ is increasing in ω for $\varphi < \hat{\varphi}^\omega$ and decreasing in ω for $\varphi > \hat{\varphi}^\omega$.

Proof. From the previous lemma, we know that $\left|\frac{dA^0}{\partial\varphi}\right| < \left|\frac{dA^1}{\partial\varphi}\right|$ for all $\varphi \geq \varphi_L^0$. Tang and See (2009) show that if f and g are functions of a random variable and $|f| < |g|$ almost everywhere, then $Var(f) < Var(g)$. This implies that

$$Var\left(A^0(\varphi)\middle|\,\varphi \geq \varphi_L^0\right) > Var\left(A^1(\varphi)\middle|\,\varphi \geq \varphi_L^0\right) = \left(\delta^{FB}\right)^{2\theta} Var\left(\varphi\middle|\,\varphi \geq \varphi_L^0\right).$$

If $|\varepsilon_{D,p}| > 1$ and φ is log-convex, then we have that $Var(\varphi|\,\varphi \geq k)$ is increasing in k (see Burdett 1996), and the result follows, since $\varphi_L^0 > \varphi_S^1$. If $|\varepsilon_{D,p}| < 1$ and φ is log-concave, then $Var(\varphi|\,\varphi \geq k)$ is decreasing in k and the result follows since $\varphi_L^0 < \varphi_S^1$.

For the second result, let $y(\varphi, p^\omega, \omega) = \mu(\varphi, p^\omega, \omega)\, y^{FB}(\varphi, p^\omega)$ for $\varphi < \varphi_H(p^\omega, \omega)$ and $y(\varphi, p^\omega, \omega) = y^{FB}(\varphi, p^\omega)$ for $\varphi > \varphi_H(p^\omega, \omega)$. For $\varphi < \varphi_H$,

$$\begin{aligned}\frac{dy}{d\omega} &= \underbrace{\frac{\partial\mu}{\partial\omega}y^{FB}}_{>0} + \underbrace{\left(\frac{\partial\mu}{\partial p}y^{FB} + \mu\frac{\partial y^{FB}}{\partial p}\right)\frac{dp}{d\omega}}_{<0} \\ &= r\underline{\mu}\left(\frac{\left(|\varepsilon_{D,p}| - 1\right)D}{\frac{\partial S}{\partial p}p + |\varepsilon_{D,p}|\,D}(1+\xi) + \frac{\frac{\partial S}{\partial p}p + D}{\frac{\partial S}{\partial p}p + |\varepsilon_{D,p}|\,D}\theta\right)\mu y^{FB}\end{aligned}$$

and for $\varphi \geq \hat{\varphi}^\omega = \varphi_H$, $\frac{dy}{d\omega} = \left(\frac{\partial\mu}{\partial p}y^{FB} + \mu\frac{\partial y^{FB}}{\partial p}\right)\frac{dp}{d\omega}$. When $|\varepsilon_{D,p}| > 1$, all firms with $\varphi < \hat{\varphi}^\omega$ expand production, and all firms with $\varphi > \hat{\varphi}^\omega$ reduce production. When $|\varepsilon_{D,p}| < 1$, there is a cutoff value $\hat{\varphi}^\omega < \varphi_H$ such that all firms with $\varphi < \hat{\varphi}^\omega$ expand production and all firms with $\varphi > \hat{\varphi}^\omega$ reduce production. ■

Remark 2 *In fact the Burdett 1996 result shows that a sufficient condition for* $Var(\varphi|\,\varphi \geq k)$ *to be increasing in* k *is that the triple cumulative integration of* φ *is log-convex, which is a significantly weaker condition.*

Remark 3 *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*

$$\left[\left.\frac{d\left(\varphi\mu\left(\varphi, p^0\right)^\theta\right)}{d\varphi}\right|_{\varphi^*}\right]^2 Var\left(\varphi\middle|\,\varphi \geq \varphi_L^0\right) > Var\left(\varphi\middle|\,\varphi \geq \varphi_S^1\right),$$

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."*

Appendix B: Monopolistic Competition Version

In this appendix, I show that this model can be equivalently formulated as a monopolistic competition model in which production exhibits constant returns to scale in capital and labor. Under a restriction on the effort cost function, total factor revenue product, $(TFPR \equiv p \cdot A)$, is equated across firms, as in Hsieh and Klenow. Higher ability firms produce more and therefore have lower output prices. Distortions in the allocation of productive resources across firms can then be inferred from heterogeneity in $TFPR$. Suppose each firm faces a demand curve of the form $y = \tilde{L}p^{-\frac{1}{1-\sigma}}$, where $\sigma \in [0,1]$ is a measure of the elasticity of demand. The inverse demand is then $p = \tilde{L}^{1-\sigma}y^{-(1-\sigma)}$ and thus revenues are given by $py = \tilde{L}^{1-\sigma}y^{\sigma}$, where $\tilde{L}$ is a function of market size and the price aggregate. Per-period profits are then

$$\pi = \tilde{L}^{1-\sigma}y^{\sigma} - RK - (W+\delta) M$$

where $y = \tilde{\varphi}\delta^{\tilde{\theta}}K^{\tilde{\alpha}}M^{\tilde{\eta}}$. Define $\theta = \tilde{\theta}\sigma, \alpha = \tilde{\alpha}\sigma, \eta = \tilde{\eta}\sigma$, and normalize discretionary resources so that $\theta + \alpha + \eta = 1 = \sigma\left(\tilde{\alpha} + \tilde{\theta} + \tilde{\eta}\right)$. Per-period profits become

$$\begin{aligned} \pi &= \tilde{L}^{1-\sigma}\left(\tilde{\varphi}\delta^{\tilde{\theta}}K^{\tilde{\alpha}}M^{\tilde{\eta}}\right)^{\sigma} - RK - (W+\delta) M \\ &= L\delta^{\theta}K^{\alpha}M^{1-\alpha-\theta} - RK - (W+\delta) M \end{aligned}$$

where $L = \tilde{L}^{1-\sigma}\tilde{\varphi}^{\sigma}$. We know from the previous analysis that the unconstrained solution to this problem is

$$\begin{aligned} \delta^{FB} &= \frac{W}{1-\alpha-2\theta}\theta \\ M^{FB}(L) &= L^{\frac{1}{\theta}}\left(\frac{\alpha}{R}\right)^{\frac{\alpha}{\theta}}\left(\frac{1-\alpha-2\theta}{W}\right)^{\frac{1-\alpha}{\theta}}\delta^{FB} \\ K^{FB}(L) &= L^{\frac{1}{\theta}}\left(\frac{\alpha}{R}\right)^{\frac{\alpha+\theta}{\theta}}\left(\frac{1-\alpha-2\theta}{W}\right)^{\frac{1-\alpha-\theta}{\theta}}\delta^{FB} \end{aligned}$$

TFPR is then

$$TFPR = p\tilde{\varphi} = \tilde{\varphi}\tilde{L}^{1-\sigma}\frac{y^{\sigma}}{y} = K\tilde{L}^{1-\sigma}\tilde{L}^{(1-\sigma)\left(1-\frac{1}{\sigma}\right)\left(\frac{\tilde{\alpha}+\tilde{\eta}}{\tilde{\theta}}\right)}\tilde{\varphi}^{\frac{1-(\tilde{\alpha}+\tilde{\eta})}{\tilde{\theta}}}.$$

TFPR is thus independent of $\tilde{\varphi}$ if and only if $\tilde{\alpha} + \tilde{\eta} = 1$, that is, if production exhibits constant returns to scale in capital and labor, then TFPR is independent of $\tilde{\varphi}$. This restriction amounts to

$$\frac{1}{\sigma} = \left(\tilde{\alpha} + \tilde{\eta} + \tilde{\theta}\right) = \left(1 + \tilde{\theta}\right)$$

or $\tilde{\theta} = \frac{1-\sigma}{\sigma} > 0$.

Appendix C: Free Entry

In the main text of the paper, I have assumed that there is a mass $N = 1$ of firms. This appendix follows Hopenhayn (1992) in endogenizing the mass of firms in the economy. Suppose there is a period 0 at which a firm can pay a cost F^e to enter the economy and take a productivity potential draw $\varphi \sim \Phi(\varphi)$. Throughout, I will focus on the stationary REE. Suppose a mass N of firms has entered. Aggregate supply in each period as a function of the steady state price level p is given by

$$\begin{aligned} S(p;N) &= Np^{\frac{1-\theta}{\theta}}\left(\frac{1-\alpha-2\theta}{W}\right)^{\frac{1-\alpha-\theta}{\theta}}\left(\frac{\alpha}{R}\right)^{\frac{\alpha}{\theta}}\delta^{FB}\cdot \\ &\quad \int_{\varphi_L(p)}^{\infty}\varphi^{1/\theta}\mu^*(\varphi,p)\,d\Phi(\varphi), \end{aligned}$$

where I have made the dependence of μ on p explicit. Note that $\varphi_L(p)$ is decreasing in p and μ is increasing in p, so $S(p;N)$ is increasing in p. Demand is stationary and given by $D(p)$, which is smooth, downward-sloping, and satisfies $\lim_{p\to 0} D(p)=\infty$ and $\lim_{p\to\infty} D(p)=0$. The unique equilibrium price $p^*(N)$ thus solves

$$S(p^*(N);N)=D(p^*(N)).$$

Define first-best gross profits for a firm with ability φ as

$$\pi^{gross,FB}(\varphi,p)=\theta(p\varphi)^{\frac{1}{\theta}}\delta^{FB}\left(\frac{\alpha}{R}\right)^{\frac{\alpha}{\theta}}\left(\frac{1-\alpha-2\theta}{W}\right)^{\frac{1-\alpha-\theta}{\theta}}.$$

Net profits can be shown to be

$$\pi(\varphi,p)=\max\left\{\mu^*(\varphi,p)(2-\mu^*(\varphi,p))\,\pi^{gross,FB}(\varphi,p)-F,0\right\}.$$

Expected gross profits are given by

$$\bar{\pi}(p)=\int\pi(\varphi,p)\,d\Phi(\varphi)$$

The free entry condition is then given by

$$F_e=\frac{\bar{\pi}(p)}{r}$$

Since D is downward-sloping and S is increasing in p and N, $p^*(N)$ is decreasing in N. The free entry condition pins down N by $\bar{\pi}(p^*(N))-rF_e=0$. Since $\bar{\pi}$ is increasing in p, this is monotonically decreasing in N. Since the choke price is infinity, we will have that $\bar{\pi}(p^*(0))>rF_e$. As long as $\bar{\pi}\left(p^*\left(\hat{N}\right)\right)<rF_e$ for some $\hat{N}$ large, there will be a unique value N^* that satisfies the free entry condition.

References

Abreu, D., 'On the Theory of Infinitely Repeated Games with Discounting', *Econometrica*, pp. 383–396 (1988).

Acemoglu, D., Antras, P. and Helpman, E., 'Contracts and Technology Adoption', *American Economic Review*, vol. 97, no. 3, 916–943 (2007).

Alfaro, L., Charlton, A.H.G. and Kanczuk, F., 'Firm-size distribution and cross-country income differences', *Working Paper* (2008).

Ayyagari, M., Demirguc-Kunt, A. and Beck, T., 'Small and medium enterprises across the globe: A new database', *World Bank Policy Research Working Paper No. 3127* (2003).

Baily, M.N., Bartelsman, E.J. and Haltiwanger, J., 'Labor Productivity: Structural Change and Cyclical Dynamics', *Review of Economics and Statistics*, vol. 83, no. 3, 420–433 (2001).

Banerjee, A. V and Duflo, E., 'Growth Theory Through the Lens of Development Economics', *Handbook of Economic growth*, vol. 1, 473–552 (2005).

Banerjee, A. V and Moll, B., 'Why Does Misallocation Persist?' *American Economic Journal: Macroeconomics*, vol. 2, no. 1, 189–206 (2010).

Bartelsman, E., Haltiwanger, J. and Scarpetta, S., 'Cross Country Differences in Productivity: The Role of Allocation and Selection', *American Economic Review* (2013).

Bartelsman, E.J. and Doms, M., 'Understanding productivity: Lessons from longitudinal microdata', *Journal of Economic literature*, vol. 38, no. 3, 569–594 (2000).

Berger, D., 'Countercyclical Restructuring and Jobless Recoveries', *Working Paper* (2011).

Bertrand, M. and Mullainathan, S., 'Are CEOs Rewarded for Luck? The Ones Without Principals are', *The Quarterly Journal of Economics*, vol. 116, no. 3, 901–932 (2001).

Bloom, N., Sadun, R. and Van Reenen, J., 'The Organization of Firms Across Countries', *Quarterly Journal of Economics* (2012).

Board, S. and Meyer-ter-Vehn, M., 'Relational Contracts in Competitive Labor Markets', *Working Paper* (2011).

Buera, F. J, Kaboski, J. P and Shin, Y., 'Finance and Development: A Tale of Two Sectors', *American Economic Review*, vol. 101, 1964–2002 (2011).

Bull, C., 'The Existence of Self-Enforcing Implicit Contracts', *Quarterly Journal of Economics*, vol. 102, no. 1, 147–159 (1987).

Burdett, K., 'Truncated Means and Variances', *Economics Letters*, vol. 52, no. 3, 263–267 (1996).

Caballero, R and Hammour, M. L, 'The Macroeconomics of Specificity', *Journal of Political Economy*, vol. 106, no. 4, 724–767 (1998).

Chandler, A. D, *Strategy and Structure: Chapters in the History of the Industrial Enterprise* (MIT Press, 1962).

Chassang, S., 'Building Routines: Learning, Cooperation, and the Dynamics of Incomplete Relational Contracts', *American Economic Review*, vol. 100, no. 1, 448–465 (2010).

Collard-Wexler, A., Asker, J. and De Loecker, J., 'Productivity Volatility and the Misallocation of Resources in Developing Economies', *Working paper* (2011).

Cooley, T., Marimon, R. and Quadrini, V., 'Aggregate Consequences of Limited Contract Enforceability', *Journal of Political Economy*, vol. 112, no. 4, 817–847 (2004).

Diamond, P.A. and Mirrlees, J.A., 'Optimal Taxation and Public Production I: Production Efficiency', *The American Economic Review*, vol. 61, no. 1, 8–27 (1971).

Djankov, S., La Porta, R., Lopez-de-Silanes, F. and Shleifer, A., 'Courts', *Quarterly Journal of Economics*, vol. 118, no. 2, 453–517 (2003).

Ellison, G. and Holden, R., 'A Theory of Rule Development', *Journal of Law Economics and Organization* (2013).

Foster, L., Haltiwanger, J. and Syverson, C., 'Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?' *American Economic Review*, vol. 98, no. 1, 394–425 (2008).

Francois, P. and Roberts, J., 'Contracting Productivity Growth', *Review of Economic Studies*, vol. 70, no. 1, 59–85 (2003).
URL: *http://onlinelibrary.wiley.com/doi/10.1111/1467-937X.00237/full*

Gali, J. and van Rens, T., 'The Vanishing Procyclicality of Labor Productivity', *Working Paper* (2010).

Garicano, Luis, Lelarge, Claire and Van Reenen, John, 'Firm Size Distortions and the Productivity Distribution: Evidence from France', *Working paper* (2011).

Ghosh, P. and Ray, D., 'Cooperation in Community Interaction Without Information Flows', *The Review of Economic Studies*, vol. 63, no. 3, 491–519 (1996).

Gibbons, R., Holden, R. and Powell, M., 'Organization and Information: Firms' Governance Choices in Rational-Expectations Equilibrium', *Quarterly Journal of Economics* (2012).

Gibbons, Robert and Henderson, Rebecca, 'Relational Contracts, Managerial Practices, and Organizational Capabilities', *Handbook of Organizational Economics, eds. R. Gibbons and J. Roberts* (2011).

Grossman, G. M and Helpman, E., 'Integration Versus Outsourcing in Industry Equilibrium', *Quarterly Journal of Economics*, vol. 117, no. 1, 85–120 (2002).

Guner, N., Ventura, G. and Xu, Y., 'Macroeconomic Implications of Size-Dependent Policies', *Review of Economic Dynamics*, vol. 11, no. 4, 721–744 (2008).

Gwartney, J. and Lawson, R., *Economic freedom of the world 2001 annual report* (The Fraser Institute, 2001).

Hamilton, J.D., 'Understanding Crude Oil Prices', *Working Paper* (2008).

Hopenhayn, H. A, 'Entry, Exit, and Firm Dynamics in Long Run Equilibrium', *Econometrica*, vol. 60, no. 5, 1127–1150 (1992).

Hsieh, C. T and Klenow, P. J, 'Misallocation and Manufacturing TFP in China and India', *Quarterly Journal of Economics*, vol. 124, no. 4, 1403–1448 (2009).

Hsieh, C. T and Klenow, P. J, 'The Life Cycle of Plants in India and Mexico', *Working paper* (2012).

Hultgren, T., 'Changes in Labor Cost During Cycles in Production and Business', *NBER Books* (1960).

Jeong, H. and Townsend, R. M, 'Sources of TFP growth: Occupational Choice and Financial Deepening', *Economic Theory*, vol. 32, no. 1, 179–221 (2007).

Johnson, S., McMillan, J. and Woodruff, C., 'Courts and relational contracts', *Journal of Law, Economics, and Organization*, vol. 18, no. 1, 221–227 (2002).

Kaufmann, D., Kraay, A. and Mastruzzi, M., 'Governance matters V: aggregate and individual governance indicators for 1996-2005', *World Bank Policy Research Working Paper No. 4012* (2006).

Kranton, R. E, 'The Formation of Cooperative Relationships', *Journal of Law, Economics, and Organization*, vol. 12, no. 1, 214–233 (1996).

Legros, P. and Newman, A. F, 'A Price Theory of Vertical and Lateral Integration', *Working paper* (2012).

Levin, J., 'Multilateral Contracting and the Employment Relationship', *Quarterly Journal of Economics*, vol. 117, no. 3, 1075–1103 (2002).

Levin, J., 'Relational Incentive Contracts', *American Economic Review*, vol. 93, no. 3, 835–857 (2003).

Lucas, R. E, 'On the Size Distribution of Business Firms', *Bell Journal of Economics*, vol. 9, no. 2, 508–523 (1978).

MacLeod, W. B and Malcomson, J. M, 'Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment', *Econometrica*, pp. 447–480 (1989).

Manova, K., 'Credit constraints, heterogeneous firms, and international trade', *Working Paper* (2011).

Melitz, M. J, 'The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity', *Econometrica*, vol. 71, no. 6, 1695–1725 (2003).

Merton, R.K., 'The Matthew effect in science', *Science*, vol. 159, no. 3810, 56 (1968).

Midrigan, V. and Xu, D., 'Finance and Misallocation: Evidence from Plant-Level Data', *Working paper* (2010).

Moll, B., 'Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?' *Working paper* (2011).

Olley, G.S. and Pakes, A., 'The Dynamics of Productivity in the Telecommunications Equipment Industry', *Econometrica*, vol. 64, no. 6, 1263–1297 (1996).

Peters, M, 'Why Do Inefficient Firms Survive? Management and Economic Development', *Working Paper* (2011).

Peters, Michael, 'Heterogeneous Markups and Endogenous Misallocation', *Working paper* (2011).

Restuccia, D. and Rogerson, R., 'Policy Distortions and Aggregate Productivity with Heterogeneous Establishments', *Review of Economic Dynamics*, vol. 11, no. 4, 707–720 (2008).

Rogoff, K., 'The Purchasing Power Parity Puzzle', *Journal of Economic literature*, vol. 34, no. 2, 647–668 (1996).

Shapiro, C. and Stiglitz, J. E, 'Equilibrium Unemployment as a Worker Discipline Device', *American Economic Review*, vol. 74, no. 3, 433–444 (1984).

Syverson, C., 'What Determines Productivity?' *Journal of Economic Literature*, vol. 49, no. 2, 326–365 (2011).

Tang, H.K. and See, C.T., 'Variance Inequalities Using First Derivatives', *Statistics & Probability Letters*, vol. 79, no. 9, 1277–1281 (2009).

Tybout, J.R., 'Manufacturing firms in developing countries: How well do they do, and why?' *Journal of Economic literature*, vol. 38, no. 1, 11–44 (2000).