Lack of Selection and Imperfect Managerial Contracts: Firm Dynamics in Developing Countries*

Ufuk Akcigit
University of Pennsylvania and NBER

Harun Alp
University of Pennsylvania

Michael Peters
London School of Economics

June 6, 2014

Abstract

Firm dynamics in poor countries show striking differences to those of rich countries. While some firms indeed experience growth as they age, many firms are simply stagnant in that they neither exit nor expand. We interpret this fact as a lack of selection, whereby producers with little growth potential survive because innovating firms do not expand enough to force them out of the market. Our theory stresses the role of imperfect managerial contracts. If managerial effort provision is non-contractible, firms will endogenously limit managerial authority to reduce the extent of hold-up. As large producers will have a higher incentive to put such inefficient monitoring policies in place, the returns to innovation decline rapidly. Improvements in the degree of contract enforcement will therefore raise the returns of growing large and increase the degree of creative destruction; innovative firms will replace inefficient producers quickly. To discipline the quantitative importance of this mechanism, we incorporate such incomplete managerial contracts into an endogenous growth model and calibrate it to firm level data from India. Improvements in the contractual environment can explain a sizable fraction of the difference between US and Indian life-cycles of plants. The model also suggests that policies targeted toward small firms could indeed be detrimental to welfare as they slow down the process of selection.

Keywords: Development, growth, selection, competition, firm dynamics, contracts, management, entrepreneurship.

JEL classification: O31, O38, O40

PRELIMINARY, COMMENTS WELCOME

*We thank the seminar and conference participants at the LSE, the London Macroeconomics Conference, the Cologne Workshop on Macroeconomics and UPenn Growth Reading Group for helpful comments, and Nick Bloom, Francesco Caselli, Matthew Cook, Jeremy Greenwood, John Haltiwanger, Chang-Tai Hsieh, Daniel Keniston, Pete Klenow, Sam Kortum, Aart Kraay, Giuseppe Moscarini, Luis Serven, and Silvana Tenreyro for the very valuable feedback. Chang-Tai Hsieh and Pete Klenow have kindly shared their data with us to use in this project. Akcigit gratefully acknowledges financial support from the World Bank.

1 Introduction

Firms in poor countries are much smaller than those in rich country countries. This is due to differences in life-cycle growth. More specifically, it is not the case that rich country firms enter at a much bigger size, but rather that they grow as they age (Hsieh and Klenow (2011)). In this paper, we argue that this difference is due to a lack of selection in poor countries. While firms in poor countries indeed do not grow on average, this average hides an important regularity: Although there are producers that grow over their lifetime, the vast majority of firms are simply stagnant in that they neither exit nor expand. In fact, this dichotomy between innovators and stagnant firms is not limited to developing countries. As shown by Hurst and Pugsley (2012) there are also many firms in the US that do not expand. The striking difference between poor and rich countries however, is their aggregate importance. While such firms in the US account for relatively little aggregate employment and shrink in importance as they age, in poor countries essentially the entirety of employment is allocated toward these producers and their aggregate importance remains stubbornly high. The problem in developing countries therefore seems not to be a failure of most firms to grow. The problem is rather that firms that do have innovative potential do not grow quickly enough to push stagnant producers out of the market. This paper provides both a theory and empirical evidence from the Indian manufacturing sector for this lack of selection.

Why is the degree of creative destruction, whereby innovative firms replace stagnant firms, so low in India? We focus on one particular mechanism, namely frictions in the market for managers. If managers add value to the firm by increasing the scale of operation, inefficiencies in how managerial services can be provided will lower the return to growth and thereby reduce the competitive pressure on stagnant firms. The idea that managerial inputs are crucial for the process of firm dynamics has a long tradition in development economics. Of particular importance is the seminal work of Penrose (1959), which not only argues that managerial resources "create a fundamental and inescapable limit to the amount of expansion a firm can undertake at any time" but also that it is precisely this scarcity of managerial inputs which prevents the weeding out of small firms as "the bigger firms have not got around to mopping them up" (Penrose (1959, p. 221)). Recently, Nick Bloom and John Van Reenen have provided empirical support to this view. First, they show that managerial practices differ across countries (Bloom and Reenen (2007, 2010)). Second, they suggest that it is not merely differences in managerial technology (or human capital) that determine managerial efficiency, but that contractual imperfections are likely to be at the heart of why firms in poor countries might be "management constrained". In their empirical study on Indian textile firms, they find that "managerial time was constrained by the number of male family members. Non–family members were not trusted by firm owners with any decision-making power, and as a result firms did not expand beyond the size that could be managed by close (almost always male) family members." (Bloom et al. (2010))

We embed these features into an otherwise standard endogenous growth model in the tradition of Klette and Kortum (2004). We model firm dynamics as the outcome of creative destruction, whereby firms expand into new product lines by investing in productivity-enhancing investment activities. To study the importance of selection, we allow for two types of firms. While innovators have the potential to grow by investing in technological improvements, stagnant firms are endowed with an inefficient innovation technology, which makes them choose to remain small. To analyze the consequences of imperfect managerial contracts, we model the strategic interaction between managers and firm owners as an incomplete contracting game as in Grossman and Hart (1986) and Acemoglu et al. (2007).

In particular we assume that the provision of managerial effort is both relationship-specific and non-contractable, and that the manager and the firm bargain over the joint surplus ex-post. To limit managerial hold-up, the firm can decide to monitor some actions of the manager. Doing so allows the firm to enforce the provision of effort in those tasks. We will loosely refer to this choice of monitoring as the firm's allocation of authority. While monitoring is valuable ex-post, as it increases the firm's bargaining share, it is detrimental to efficiency in that it lowers the manager's incentive to provide relationship-specific investments ex-ante. The crucial prediction of the theory is that the incentives for monitoring are higher for larger firms than for smaller firms. Intuitively: As firms with larger revenue face a more severe hold-up problem in the bargaining stage, their incentives to distort managerial effort provision on the margin increase. Contractual imperfections lead to a schedule of marginal costs that are endogenously *increasing* in firm size. From a dynamic point of view, firms anticipate that the marginal costs of production increase as they expand. The value of growing large is low when contractual imperfections are severe. This in turn lowers innovation incentives for innovators, and with it, the degree of creative destruction. Contractual frictions therefore limit the process by which innovators "mop up" stagnant producers.

Our preliminary analysis shows that these contractual frictions are quantitatively important. After characterizing the dynamic equilibrium of our model, we take it to the data and calibrate its structural parameters to the Indian establishment level data. Our model is able to the targeted moments well. Then we conduct the following counterfactual exercise: In order to understand the potential quantitative importance of our mechanism, we replace the contractual environment in India with the US counterpart while keeping all other parameters fixed. Improving the contractual environment makes the firm dynamics in India closely resemble those in the US. In particular, 60% of the observed firm size difference between old firms in India and US is explained by this channel. Similarly, this leads to 50% reduction in the number of low-type firms in the economy within the first 15 years of their lifetime. Finally, the share of total employment by 10-year and older firms increases from 6% to 90%. Overall, these results show that the contractual frictions can potentially go a long way to explain the differences in firm selection and creative destruction.

Related Literature This paper provides a theory of firm dynamics in developing countries.[1] While many recent papers have aimed to measure and explain the static differences in allocative efficiency across firms[2], there has been little theoretical work explaining why firm dynamics differ so much across countries. A notable exception is the work by Cole et al. (2012), which argues that cross-country differences in the financial system will affect the type of technologies that can be implemented. Like them, we let the productivity process take center stage. However, we turn toward the recent generation of micro-founded models of growth, in particular Klette and Kortum (2004). While such models have been built to study firm dynamics in developed economies (Lentz and Mortensen (2008), Acemoglu et al. (2012), Akcigit and Kerr (2010)), this is not the case for developing countries.[3] We believe endogenous technical change models are a natural environment to study this question, as they focus on firms' productivity-enhancing investment decisions. We believe that models of endogenous growth have been under-utilized in the development literature, partly due to lack of data to discipline these models, and partly due to the fact that early models of endogenous growth have been mainly constructed to model innovation decisions of firms in developed countries.[4] Hence, these early models have been harmonized with terminologies such as innovation, R&D, patent protection, and innovation policy, which do not seem to properly capture the reality of firms in developing countries. For the remainder of this paper, we therefore refer to innovation in a broad sense,

[1] An overview over some regularities of the firm size distribution in India, Indonesia and Mexico is contained in Hsieh and Olken (2014).

[2] The seminal papers for the recent literature on misallocation are Restuccia and Rogerson (2008) and Hsieh and Klenow (2009). As far as theories are concerned, there is now a sizable literature on credit market frictions (Buera et al., 2011; Moll, 2010; Midrigan and Xu, 2010), size-dependent policies Guner et al. (2011), monopolistic market power (Peters, 2013) and adjustment costs (Collard-Wexler et al., 2011). A synthesis of the literature is also contained in Hopenhayn (2012) and Jones (2013).

[3] An exception is Peters (2013), who applies a dynamic Schumpeterian model to firm level data in Indonesia.

[4] A major impediment of bringing the first generation models of endogenous growth to the data is that these were aggregate models, which do not have direct implications at the firm-level (Romer, 1990; Aghion and Howitt, 1992; Grossman and Helpman, 1991).

capturing not only the implementation of new ideas but also a variety of costly productivity enhancing activities, encompassing also training, reorganization or the acquisition of high quality complementary factors.

We focus on inefficiencies in the interaction between managers and owner of firms to explain the differences in firms' demand for expansion. Hence, particularly relevant contributions are Caselli and Gennaioli (2012) and Powell (2012). Caselli and Gennaioli (2012) also stress the negative consequences of inefficient management. Their focus is on the efficiency of the "market for control", i.e. the market where (untalented) firm-owners are able to sell their firms to (talented) outsiders. With imperfect financial markets, such transactions might not take place as outsiders might be unable secure the required funds.[5] Our economy does not have any exogenous heterogeneity in productivity so that there is not any notion of *static* misallocation. In contrast, we argue that managerial frictions *within* the firm reduces growth incentives and hence prevent competition from taking place sufficiently quickly on product markets. Such within-firm considerations are also central in Powell (2012), who studies an economy, where firms ("owners") need to hire managers as inputs to production but contractual frictions prevent owners to commit to pay the promised managerial compensation after managerial effort has been exerted. He studies the properties of the optimal long-term relational contract in a stationary equilibrium, whereby owners are disciplined to pay their promises through reputation concerns. There are two important differences to our paper. First of all, Powell (2012) studies an economy where firm productivity is constant, i.e. there is no interaction between contractual frictions in the market for managers and firms' innovation incentives. Secondly, while he studies the implications of *owners* not being able to write contracts on their wage promises, we focus on *managers* not being able to contractually commit themselves to their choice of effort. This difference is important in that it determines the distribution of costs of imperfect legal systems. While in our model, contractual frictions will especially hurt *large* firms, for which hold-up is costly, Micael Powell's model implies that it will be *small* producers, who will be particularly affected, as they have little reputational capital to pledge.

The remainder of the paper is organized as follows. The next section presents evidence on the two main ingredients of our theory. In particular we present three regularities of managerial employment across countries and use Indian microdata to show the importance of our assumption on innovating and stagnant firms. In Section 3 we describe the theoretical model. Section 4 contains the quantitative analysis. We first calibrate the model to the Indian micro data and then consider the two policy exercises discussed above. Section 5 concludes.

2 Motivating Evidence

This paper proposes a theory of firm dynamics in developing economies. The theory has two main ingredients. First of all, we argue that it is important to think of the economy as being populated by different *types of firms*. Some innovate, and some remain in the market without expecting to grow. Secondly, we link the speed at which the market is able to drive stagnant firms out of the market to contractual frictions between managers and entrepreneurs. In this section, we present some evidence on both of these ingredients. This not only aims to motivate the environment we have in mind, but we will also use some of these regularities as explicit calibration targets in our quantitative exercise.

[5] Another reason for untalented owners to *not* sell their firm is that individual wealth can substitute for managerial incompetence if financial markets are imperfect. Hence, financial frictions will also reduce the supply of firms and not only the demand from credit-constrained outsiders.

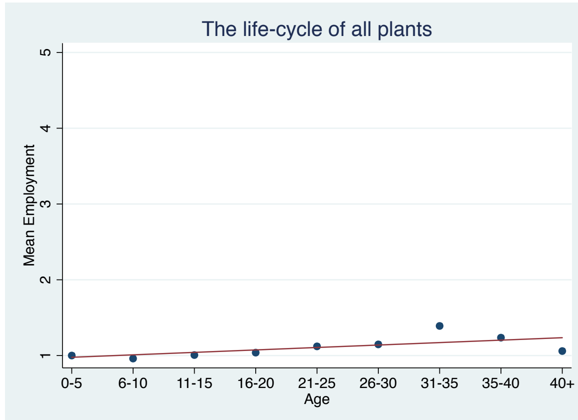


Figure 1: The life-cycle of manufacturing plants in India

2.1 Innovators and Stagnant Firms in India: Empirical Evidence

In this section, we present micro evidence on the pervasiveness of stagnating firms in the Indian economy. To get a picture of the population of Indian firms, we follow Hsieh and Klenow (2011) and Hsieh and Olken (2014) to construct a firm-level dataset by merging the Annual Survey of Industries (ASI) and the National Sample Survey, Schedule 2.2 (NSS). Broadly speaking, the ASI contains the universe of establishments with more than 100 employees and a random sample of establishments with 20 to 100 employees. The NSS contains is a survey of informal establishments. Using this information we extrapolate to the whole economy using the sampling weights provided in the data. A more detailed description of the data is given in Section 4.

As in Hsieh and Klenow (2011) we focus mainly on the cross-sectional size-age relationship and interpret this schedule as the life-cycle of a representative cohort. This will be exactly true in our theory. In general, the cross-sectional pattern could be driven by cohort effects and not be informative about the life-cycle. While we could look at the firm dynamics more directly using the panel version of the ASI data, the NSS data is only available in repeated cross-sections every five years; we focus on the cross-section for now but refer to it as the "life-cycle".[6]

For comparison with the literature, we first want to ensure that we are able to replicate other findings in the literature. First, we focus on the firm-size distribution and replicate the findings of Hsieh and Olken (2014). We report these results in Section 6.1 in the Appendix. Second, we focus on the life-cycle of manufacturing plants in India. In Figure 1 below we first replicate the findings of Hsieh and Klenow (2011) using our data. In particular we calculate mean employment for different age bins and plot average firm size by age relative to the size of the youngest cohort, which we will sometimes refer to as entrants. As in Hsieh and Klenow (2011) we see little growth along the life-cycle. For ease of comparison, we chose the axis to approximately represent the growth of typical firms in a developed economy, which is roughly equal to five.

Figure 1 can be driven by two types of theories. It could either be the case that the representative Indian firm grows less than its US counterpart. Or it could be the case that the flat average profile in India is driven by a plethora of firms, which do not grow at all, while some innovative firms actually do grow — there are just too few of them in the aggregate to affect the average age-size relationship in a meaningful way. As explained above, we opt for the second explanation: the *selection hypothesis.* In what follows, we are going to decompose Figure 1 in various ways to show that this is indeed a useful look at the data:

[6] We could of course follow Hsieh and Klenow (2011) and construct the synthetic life-cycle from the comparison of repeated cohorts in different years of the sample. We have not explored this in detail yet.

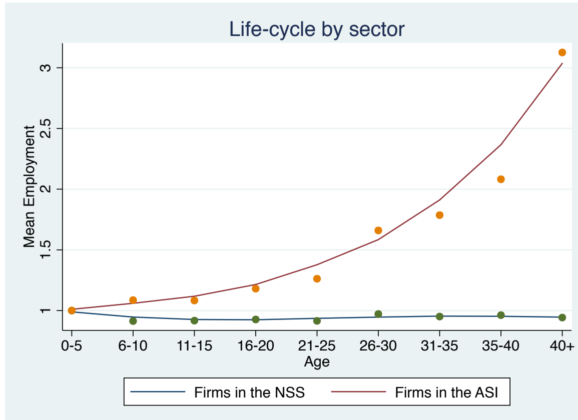


Figure 2: The life-cycle of manufacturing plants in India: ASI versus NSS

There are growing firms in India, but they do not grow sufficiently quickly to force the stagnant firms to exit the market.

As a first cut of the data, we analyze the ASI and NSS data separately. This is problematic because being in one of the samples is not a fixed firm characteristic, but to a large degree a function of size. The split is nevertheless useful because it is a transparent decomposition of the data. In Figure 2 we plot the life-cycle for the two different samples. The differences are apparent: While there is growth along the life-cycle for plants in the ASI, there is no growth for firms in the NSS.

Figure 2 is not straightforward to interpret due to selection into the ASI. Specifically: Growing firms will leave the NSS and migrate into the ASI as they formalize and cross the size threshold of 20 employees. However, given that the vast majority of firms in the NSS is far from the threshold (the median firm has 2 employees, and the 99% quantile is 13 employees) we think it is unlikely these transitions account for the majority of the differences in the age-size relationship.[7] However, we will nevertheless look at cuts of the data, that are less subject to these concerns.

If the Indian economy is characterized by a small number of innovative firms, we expect the firms to be bigger at point in time and, more importantly, to actually increase their importance as they age. The intuition is that if only the best firms grow, the distribution of firm size should "fan out" in the upper tail relative to the rest of the economy. The data indeed reflect this.

Figure 3 plots the life-cycle of the top 5% of firm observations in each age group. More specifically, we take the top 5% of observations — from entire sample consisting of ASI and NSS firms — and track their average employment as they age. The results mirror our findings in Figure 2: The top firms in India actually do grow quite substantially. The average employment of the top 5% increases by a factor of 6. Hence, it is not the case that the Indian manufacturing sector is stagnant; the majority of firms is, but there are vibrant pockets that do expand with age. One comment about Figure 3 is in order. To construct the figure we chose the top 5% of observations in the data. These 5% of observations account for less than 5% of firms because small firms (primarily in the NSS) get a higher sampling weight than do bigger firms (primarily in the ASI). If bigger firms have lower sampling weights as they age, we might be selecting on fewer and fewer firms by focusing on the top 5% of observations. In Section 6.2 in the Appendix we perform various robustness checks to Figure 3, which give the same answer qualitatively.

Finally, we look at one particular firm characteristic, which is easily observable and argued to be an import dimension of heterogeneity: family firms. Both the NSS and ASI identify whether or not firms are

[7] To reconcile Figures 1 and 2, note that only roughly 1% of firms are part of the ASI. Hence, the aggregate picture (Figure 1) is dominated by the behavior of NSS firms, which do not grow. See Figure 19 in the Appendix, which plots the share of firms in the ASI as a function of age.

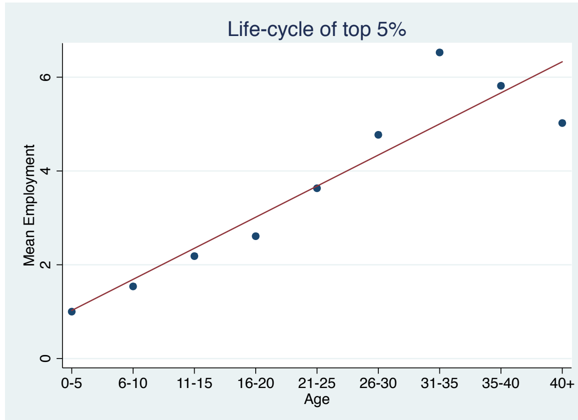


Figure 3: The life-cycle of the top 5% of firms in India

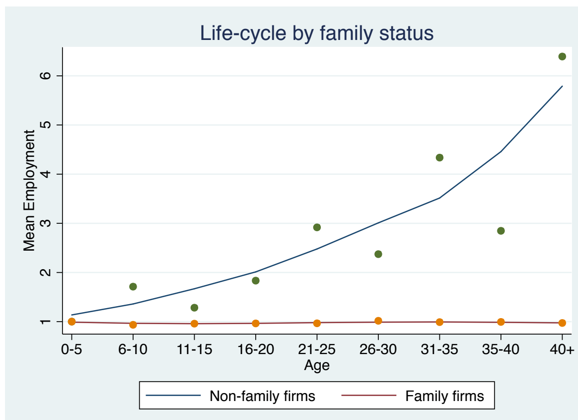


Figure 4: The life-cycle of family firms versus non-family firms

organized within the family, i.e. if family members hold the property rights to the firm. Figure 4 below performs the life-cycle exercise for the subsample of family and non-family firms. Again, the importance of that characteristic is striking: While family firms do not grow as they age, firms outside the scope of the family increase employment by a factor 6.

None of the exercises displayed in Figures 2, 3 and 4 is perfect, and these are not mutually exclusive, as these samples are correlated: The top firms are likely to come from the ASI, and are in turn less likely to be family run. Also, a firm's family status is obviously not a "stamp on the head" of the firm, but an endogenous choice of the owner. This endogeneity is, however, at the heart of our argument that the substantial problem of the Indian economy is one of selection. While some firms manage to expand in the Indian business environment, these firms are too few in the aggregate to draw resources from the stagnant firms of the economy sufficiently quickly to force these firms to exit.

To see this lack of selection in the micro-data, finally consider Figure 5, which shows the share of firms with at most 2 workers by age. Roughly 70% of firms fall in this category. More notably, this share is almost constant by age. Hence, these firms - which are probably run by an owner and another family member - neither exit the economy nor grow out of these family boundaries. From a literal life-cycle interpretation, Figure 5 suggests that the majority of entrepreneurs in India start a firm with two employees and remain at this size for their entire life. While we do not have access to the US microdata to redo the same exercise, Table 1 reports the data about small firms from Hurst and Pugsley (2012) and

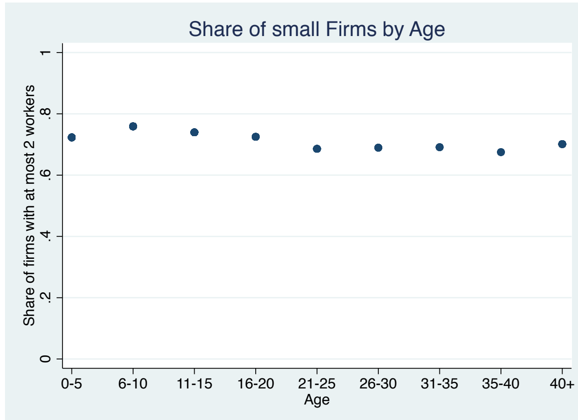


Figure 5: The share of small firms in the Indian economy

	Share of firms with less then 20 employees		
	0-10 Years Old	10-25 Years Old	Full Sample
U.S (2005)	85.5	71.5	72.4
India (1995)	99.3	99.5	99.6
	Share of aggregate employment in firms with less than 20 employees		
	0-10 Years Old	10-25 Years Old	Full Sample
U.S (2005)	34.6	16	8.5
India (1995)	81.6	76.2	74.5

Table 1: Importance of small firms across age: US versus India

compares them with our India data.

Small firms are defined as firms with fewer than 20 employees. What we find notable is the difference in selection. While the aggregate importance of small firms in the US drops by more than 50% as the cohort ages (i.e. it drops from 34.6% to 16%), the corresponding number in India only drops by 5% (from 81.6% to 76.2%). This is due to the fact that large firms are larger in the US, and to the fact that small firms do not exit in India.

To understand the aggregate evolution of the manufacturing economy, one has to understand (a) why innovative firms do not grow enough to force inefficient firms out of the market, and (b) why there is such prevalent entry of firms that neither grow nor exit. The first aspect concerns the innovation incentives of potential innovators, and second aspect concerns the apparently low opportunity costs of stagnant producers. In this paper, we will provide a model, which formalizes these two margins by introducing imperfect contracts in the relationship between owners and managers.

2.2 Imperfect Contracts and Managerial Employment

Our focus on the linkage between the cross-country variation in the state of the contractual environment and the interaction between owners and managers is partly determined by three broad macro facts, which are depicted in Figure 6 below. In the left panel, we depict the cross-sectional relationship between the country-wide employment share of managerial personnel and the "Rule of Law"-Index of the World Bank in 2010. It is clearly seen that there is robust positive correlation in that better governance leads to an increase in the provision of managerial positions. In the right panel, we show the cross-sectional correlation of the rule of law index and the importance of self-employment. As expected (and consistent with Gollin

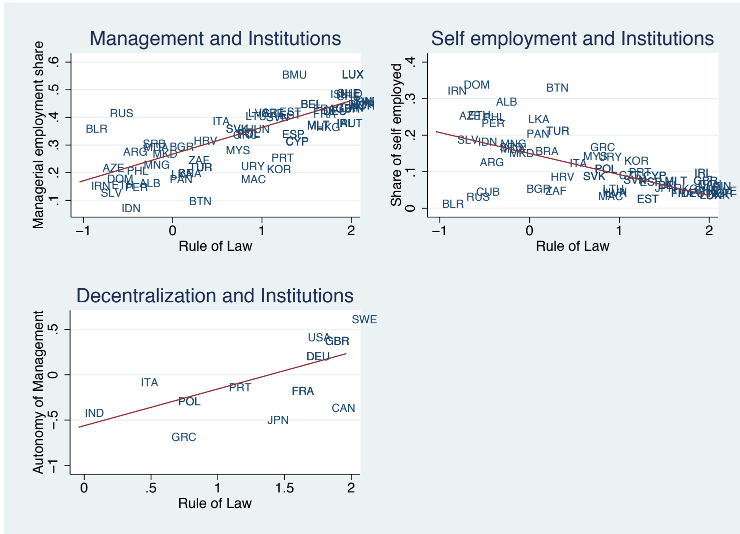


Figure 6: Management and Contracts

(2008)), there is a strong negative correlation in that petty entrepreneurship seems to flourish in bad legal systems. Finally, the last panel uses the data from Bloom et al. (2009) on within-firm decentralization and shows that countries with better legal systems see more decentralization in that more decision power is granted to plant managers.

Our theory will (a) connect these three facts and (b) show why (and how) these regularities are predictive of sclerotic selection as shown in Section 2.1. Our basic narrative is the following: In our theory, owners and managers interact in a process of joint production. If contracts are imperfect, owners are subject to managerial hold-up. As a response, owners will endogenously limit managerial authority (Fact 3). Such limits to authority however are costly as they reduce managerial effort and with it firm profitability. Hence, they lower firms' demand for managers and consequently the equilibrium level of managerial personnel (Fact 1). Moreover, as far as innovation incentives are concerned, reduced profitability is akin to a scale-effect from the point of view of the firm: Contractual frictions reduce innovation incentives for high types so that low types will survive longer as there is little threat of creative destruction. As small firms will be predominantly low types and smaller firms are (in our theory) less likely to hire managers from outside and hence more likely to be self-employed entrepreneurs, contractual frictions will drive up the rate of self-employment (Fact 2).

An important implication for micro-data is that some firms in India will be *managerially constrained,* in that contractual frictions between owners and managers cause an effective undersupply of managerial resources. The selection hypothesis implies that the problem in India is not so much that small firms do *not* grow but that big firms do not grow *even more*. Hence, the marginal product of managerial resources should be especially high for large firms. In Figure 7 we plot the non-parametric regression of the *average* product of managers, that is, the log of the value added per manager as a function of firm size.[8] Hence, as long as the average product carries information about the marginal product (which it will in our theory), Figure 7 suggests that the marginal value of managerial efficiency units is particularly high in large firms, i.e. it is precisely large firms that seem to be constrained on the managerial margin.[9] This

[8]Figure 7 simply plots the raw non-parametric regression without any covariates. In the Appendix we show that the positive correlation between the average product of managerial inputs and firm size is robust to a host of controls.

[9]This is consistent with the findings of Hsieh and Olken (2014), who show that the average product of capital and the average product of labor also seems to be increasing in firm size.

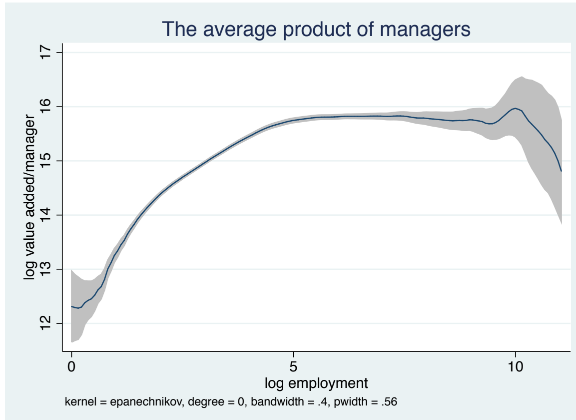


Figure 7: The average product of managers in the cross-section of firms

cross-sectional relationship between size the marginal products will be informative about the degree of contractual frictions. Hence, we will use it as an explicit micro-moment, we calibrate our theory against.

3 The Model

To model these issues, we consider a firm-based model of endogenous growth in the spirit of Klette and Kortum (2004). We augment this framework with three ingredients:

1. We assume that firms are heterogenous in their innovation potential.
2. We allow for a margin of occupation choice, whereby agents can either enter the economy as an entrepreneur or work as managers for existing firms.
3. We explicitly introduce contractual frictions in the interaction between firm owners and managers.

The the former two allows us to meaningfully speak about a process of selection, it is the last ingredient that will determine how quickly this process will take place. As with our evidence presented in Figure 6, we will be using the degree of contractual frictions in 3. as our source of variation across countries.

More precisely, we think of an economy, which is populated by two type of agents: workers and potential entrepreneurs. We call them potential entrepreneurs as some of them might not chose to pursue an entrepreneurial activity but rather work as a manager for another firm. Entrepreneurs are endowed with production possibilities ("firms") and have the capacity to grow their firms through innovation. Entrepreneurs come in two types, which differ in their innovation costs: while high types can perform innovation activities and hence generate sustained productivity growth through creative destruction, low types are not capable of starting a thriving business in that they have no talent for innovation. Hence, our model is a heterogeneous firm model, where firms do not differ in their exogenous TFP as in Lucas (1978), but where firms differ in the efficiency of innovation. The process of creative destruction, which is ignited by the high types, determines how long low types can remain in business. Hence, at the heart of our selection process is the demand for growth of high types.

Entrepreneurs combine their technology with two inputs of production: workers and managerial effort. By increasing the amount of managerial effort, firms can increase the efficiency of their physical production factors. Hence, well managed firms have high "x-efficiency" in that they combine their technology and their production workers more effectively. While workers are simply hired on a frictionless spot market, the provision of managerial effort is more involved. In particular, managerial services can either be provided by the entrepreneur himself or can be outsourced to a specialized manager, i.e. a potential entrepreneur who decided to forgo his entrepreneurial opportunities to become a manager himself. Specialized managers are useful in that they can provide managerial services more efficiently (for example because the owner needs to split his available time between managerial tasks and additional strategic decisions). However, the interaction between entrepreneurs and managers however is subject to contractual frictions, which will taint the efficiency at which managers can be employed. This has important dynamic ramifications: because large firms will - endogenously - be harmed more by contractual imperfections, the incentives to grow large are low when contracts are hard to enforce. Hence, the demand for creative destruction will - endogenously - be low and the economy will be sclerotic for two reasons. First of all, low types, i.e. firms without any growth potential, will survive for a long time conditional on entry. Secondly, contractual frictions reduce the demand for outside managers, which implies that the share of low-type agents entering as an entrepreneur will be bigger.

For simplicity we assume that both managers and entrepreneurs are short-lived. More precisely: entrepreneurs live for one period and then hand over the firm to their off-spring, who also lives for one period. This is isomorphic to an environment where entrepreneurs are infinitely lived but have a planning horizon of only one period. This is useful for analytical tractability and captures all the economic intuition.

3.1 Preferences and technology

On the demand side, we model workers as a representative household, with standard preferences

$$U_0 = \int_0^\infty \exp(-\rho t) \ln C_t dt, \tag{1}$$

where, as usual, $\rho > 0$ is the discount factor. Given the unitary intertemporal elasticity of substitution, the Euler equation along the balanced growth path is simply given by

$$g = r - \rho,$$

where g is the growth rate of the economy and r is the interest rate. Assuming workers to be long-lived is useful in that it determines the equilibrium interest rate. However, it is not essential for the main points of this paper.

The final good, which we take as the numeraire of the economy, is a composite of a continuum of products, which for simplicity takes the Cobb-Douglas form

$$\ln Y_t = \int_0^1 \ln y_{jt} dj, \tag{2}$$

and y_{jt} is the amount of product j produced at time t. Production takes place by heterogeneous firms and uses both production workers and managers. In particular, the production function for good j at time t is given by

$$y_{jf} = q_{jf} m(e_f) l_{jf}, \tag{3}$$

where q_{jf} is the firm-product specific production technology, $m(e_f)$ denotes the amount of managerial efficiency units employed by firm j and l_{jf} is the number of workers employed for producing intermediate good j. Naturally, $m(e)$, which we will specify below, is strictly increasing. Note that managers are employed at the firm level, so that $m(e_f)$ has no j index. Anticipating our results slightly: With incomplete managerial contracts, it will be hard to elicit the efficient level of managerial effort e_j. Hence, e_j will be derived endogenously from the principal agent relationship between the firm owner and the manager (in case the firm decides to outsource managerial effort provision). The distribution of efficiencies $q_{j,t}$ will evolve endogenously through firms' choices of innovation spending and will determine which firm produces which product.[10] As production workers are in fixed supply, the labor market clearing condition is given by

$$L = L_t^P. \tag{4}$$

3.2 Static equilibrium

Now consider the equilibrium in the product market. At each point in time, each product line j is populated by a set of firms that can produce this good with productivity $\left[q_{jt}^f\right]_f$, where f identifies the firm. We will make sufficient assumptions on $m(.)$, that the most productive firm (which we will sometimes refer to as the (quality) leader) will be the sole producer of product j. Intuitively: while managerial slack can and will be a drag on efficiency, it can never reverse comparative advantage based on physical efficiency q. This assumption will make the structure of the optimal contract between entrepreneurs and managers slightly easier but is not essential for our results. Additionally we assume that fringe firms (i.e. the followers) can produce the good at a technological disadvantage. Specifically, we assume that there is imperfect diffusion of technology, i.e., if the leader in product j can produce the product with efficiency q_{jt}, the remaining firms can produce it with efficiency $\frac{q_{jt}}{\gamma}$ for some $\gamma > 1$. This assumption allows us to sidestep some issues of mark-up heterogeneity, which we do not think to be of first-order importance to understand differences in the life-cycle of firms across countries.[11]

The Cobb-Douglas structure in (2) implies that the demand for an individual product will have unitary demand elasticity. Hence, the leader will always be forced to engage in limit pricing. Given this assumption, the equilibrium price for product j is given by

$$p_{ij} = \frac{\gamma w_t}{q_{jt}}, \tag{5}$$

as $\frac{\gamma w_t}{q_{jt}}$ are exactly the competitive fringe's marginal costs of producing product j. Equation (2) then implies that the demand for product j is given by

$$y_{jt} = \frac{Y_t}{p_{ij}} = \frac{q_{jt} Y_t}{\gamma w_t}, \tag{6}$$

so that total sales are simply $S_{jt} = p_{jt} y_{jt} = Y_t$, i.e. equalized per product. This of course does not imply that the distribution of sales is also equalized across firms - as some firms will (endogenously) have more products than other firms, the distribution of sales is fully driven by the distribution of products. This tight link between firm-level sales and firms' product portfolio is not only analytically attractive but also conceptually useful in that it clarifies that our model attributes firm dynamics to a single mechanism: why do countries differ in the speed at which firms accumulate (and lose) products along their life-cycle.

Similarly, the allocation of labor demand is simply

$$l_{jft} = \frac{y_{jft}}{q_{jft} m(e_{ft})} = \frac{Y_t}{m(e_{ft}) \gamma w_t}, \tag{7}$$

i.e., the allocation of labor across products depends on firm's managerial choices. While all products within a firm have the same number of workers, managerial efficiency is *labor-saving.* Intuitively: an increase in managerial effort *increases* profitability as it increases the firms' sustainable mark-up. To see this, note that equilibrium mark-ups are given by

$$\zeta_{jft} = \frac{p_{jft}}{MC_{jft}} = \frac{\frac{\gamma w_t}{q_{jt}}}{\frac{w_t}{q_{jt} m(e_{ft})}} = \gamma m(e_{ft}), \tag{8}$$

[10] We will be using the terms efficiency and productivity interchangeably when referring to q.

[11] See Peters (2013) for a related model that focuses on heterogeneous mark-ups.

i.e. well managed firms can keep their competitors at bay, sustain high prices and hence move up on their product demand curve. The resulting profit (before paying the managers) for producer f of variety j is then simply

$$\tilde{\pi}_{jft} = \left[\frac{\gamma w_t}{q_{jt}} - \frac{w_t}{q_{jt} m\left(e_{ft}\right)}\right] \frac{q_{jt} Y_t}{\gamma w_t} = \frac{\gamma m\left(e_{ft}\right) - 1}{m\left(e_{ft}\right)\gamma} Y_t, \tag{9}$$

i.e. profits *only* depend on how well the respective firm can incentive their managers. In particular, $\tilde{\pi}_{jft}$ is increasing in e_{ft}: better managerial practices increase mark-ups and hence profit per product. (9) contains the main intuition about the interaction between contractual frictions and innovation incentives: As contractual frictions will be detrimental to managerial effort provision, firms will be unable to sustain high mark-ups as they grow. The marginal product will therefore be less profitable than the average product and incentives to break into new products will be low.

Substituting (6) into (2) we get that equilibrium wages are given by

$$w_t = \frac{1}{\gamma} Q_t,$$

where Q_t is the Cobb-Douglas composite of individual efficiencies

$$\ln Q_t \equiv \int_0^1 \ln q_{jt} dj.$$

Using (7), we get that $l_{jft} = \frac{1}{Q(t) m\left(e_{ft}\right)} Y_t$, so that labor market clearing implies that total output is given by

$$Y_t = Q_t M_t L, \tag{10}$$

where

$$M_t = \left[\int m\left(e_{ft}\right)^{-1} df\right]^{-1}. \tag{11}$$

Here: M_t is the endogenous TFP term based on managerial effort. In particular, increases in x-efficiency, i.e. managerial effort, will increase aggregate TFP.[12]

3.3 Innovation and Entry

As usual in firm-based models of endogenous growth, growth stems from two margins: entry and innovation by incumbent firms. In order to focus on the process of selection (or lack thereof), we assume that each period there is a measure N of potential entrepreneurs entering the economy at each point in time. This can be thought as an exogenous flow of business ideas to outsiders, which enter the economy as new entrepreneurs. Importantly, entrants are heterogeneous and are either of high or of low types as discussed above. Formally, upon entry, each new entrant draws a firm type $\theta \in \{\theta_H, \theta_L\}$ from a Bernoulli distribution, where

$$\theta = \begin{cases} \theta^H & \text{with probability } \alpha \\ \theta^L & \text{with probability } 1-\alpha \end{cases}.$$

The type of the firm determines its innovation productivity or growth potential. In particular, each firm is endowed in with an innovation technology. If a firm of type θ with n products in its portfolio invests R units of the final good in R&D, it generates a flow rate of innovation of

$$X\left(R; \theta, n, Q\left(t\right)\right) = \theta \left(\frac{R}{Q\left(t\right)}\right)^{\zeta} n^{1-\zeta}. \tag{12}$$

[12] Note that the integral in (11) integrates over *firms* and not products.

Hence, θ parametrizes the efficiency of innovation resources. For simplicity we assume that $\theta_L = 0$, i.e. low types will never be able to grow and we can focus on the high types' decisions. The other terms in the innovation technology are the usual scaling variables in many models of growth. Because we denote innovation costs in terms of the final good, the scalar $Q(t)$ is required to keep the model stationary and the presence of n implies that the cost of innovation do not scale in firm size. To see this, note that the cost function of an innovation rate per product $x = \frac{X}{n}$ is given by

$$C(x \mid n, Q(t), \theta) = nQ(t)\left[\frac{x}{\theta}\right]^{\frac{1}{\zeta}}. \tag{13}$$

These potential entrepreneurs face an occupational choice decision after observing their type. They can either try to become entrepreneurs or they can become managers and work for existing firms. All agents who decide to become entrepreneurs are successful in doing so with flow rate z, i.e. with flow rate z they enter the economy as a single product firm. Unsuccessful entrepreneurs exit the economy. Hence, letting ω_i be the share of type i agents that decide to try the entrepreneurial option, the total amount of entrants is given by

$$\text{Entry} = z \times N \times (\omega_H \alpha + \omega_L (1-\alpha)).$$

The total supply of managers is given by

$$M^S = N \times ((1-\omega_H)\alpha + (1-\omega_L)(1-\alpha)).$$

As the value of entrepreneurship will be higher for the high types (see below), in equilibrium we will have $\omega_H = 1$ and ω_L be determined by market clearing for managers. Hence, the share of low types upon entry is given by $\frac{\omega_L(1-\alpha)}{\alpha+\omega_L(1-\alpha)}$ - the more managers other firms will hire, the more low types are already "weeded" out before the even the economy as self-employed entrepreneurs.

After entry decisions have taken place, firms can try to innovate. Letting $V(n,t)$ be the value of having a firm with n products (to be defined below), the value of an incumbent of type i with n products prior to the innovation stage is given by

$$W_i^{INC}(n,t) = V(n,t) + \max_x \{xn[V(n+1,t) - V(n,t)] - C(x \mid n, Q(t), \theta_i)\}, \tag{14}$$

where $C(x \mid n, Q(t), \theta)$ is given in (13). Hence the profit-maximizing innovation rate is given by

$$\begin{aligned} x^*(n,t) &= \arg\max_x \left\{x[V(n+1,t) - V(n,t)] - Q(t)\left[\frac{x}{\theta}\right]^{\frac{1}{\zeta}}\right\} \\ &= [\zeta\Delta(n,t)]^{\frac{\zeta}{1-\zeta}} \theta^{\frac{1}{1-\zeta}}, \end{aligned} \tag{15}$$

where $\Delta(n,t)$ denotes the marginal return to innovation

$$\Delta(n,t) \equiv \frac{V(n+1,t) - V(n,t)}{Q(t)}. \tag{16}$$

As $x_L(n,t) = 0 < x_H(n,t)$, (14) implies that $W_H^{INC}(n) > W_L^{INC}(n)$. Hence, the value of entry into entrepreneurship for type i is given by

$$W_i^E = zW_i^{INC}(1), \tag{17}$$

so that $W_H^E > W_L^E$. (16)

To study the aggregate consequences of selection, we need to keep track of the share of product lines belonging to high and low types respectively. Let us denote the share of the product lines that belong an

$n-$product high-type firm by μ_n^H and the share of the product lines that belong to all low type firms by μ^L . Then

$$\mu^L + \sum_{n=1}^{\infty} \mu_n^H = 1. \tag{18}$$

Firms loose products if they are replaced by either new entrants or successful incumbents. Let us denote the aggregate creative destruction, i.e. the rate at which the producer of given product is replaced, by τ, where

$$\tau \equiv \underbrace{\sum_{n=1}^{\infty} x_n \mu_n^H}_{\text{Incumbent high types}} + \underbrace{z\alpha\bar{n}_e^H}_{\text{Entry of high types}} + \underbrace{z(1-\alpha)\bar{n}_e^L}_{\text{Entry of low types}},$$

and $\overline{n}_e^H$ and $\overline{n}_e^L$ denote the average number of products of high and low types upon entry, i.e. $\bar{n}_e^j = \sum_{n=1}^{\infty} n f_n^j$, where we denote the initial entry distribution $f(n_0|\theta_j)$ by f_n^j.

In steady-state, the amount of entry of low types must be equal to the rate at which low types are bing replaced. Hence,

$$\mu^L \tau = z(1-\alpha)\bar{n}_e^L,$$

where the LHS denotes the aggregate number of low-type products, which are being replaced and the RHS is the gross number of products, which are entered in by new low type firms. Similarly, the amount of entry by high-types must be equal to the amount of exit of high-type producers. As firms exit whenever they lose their last product, it has to be that

$$z\alpha\bar{n}_e^H = \tau\mu_1^H.$$

In general, the flow equations for the set of high type producers are

$$\mu_n^H n[\tau + x_n] = z\alpha n f_n^H + \mu_{n-1}^H [n-1] x_{n-1} + \mu_{n+1}^H \tau [n+1]. \tag{19}$$

Here the LHS of (19) is the number of high-type firms that exit state n and the RHS gathers the number of high types that enter state n. These can come from three sources: Either they enter with n products, they grow from being $n-1$ to being n firms or they used to have $n+1$ products but lose one product against another competitor. For given innovation schedules $\{x_n\}$, (18)-(19) fully characterize the stationary distribution of the economy. As x_n is only dependent on Δ_n (see (15)), (18)-(19) are also sufficient to solve for the dynamic evolution of the economy given a schedule of *marginal returns* $\{\Delta_n\}_n$. It is precisely this marginal return schedule, which we will construct from the incomplete contracts game between owners and managers.

3.4 The value of the firm

After innovation outcomes have been realized and observed, firms hire production workers, set prices and decide whether or not to hire an outside manager. We assume that matching between firms and managers is frictionless and that the market clears in equilibrium. The important aspect of the model is that we introduce an explicit imperfect contract into the theory. While the entrepreneur makes the innovation decision (i.e. solves the problem (15)), the day-to-day affairs can be run by outside managerial personnel. The contractual incompleteness arises in that managerial effort is not directly contractible. In particular we follow Grossman and Hart (1986) to assume that the manager and the entrepreneur engage in Nash-Bargaining about the joint surplus ex-post. As an imperfect substitute of a contract to limit the firms' hold-up, we assume that the firm can monitor the manager and thereby ensure a fraction of the managerial effort to be exerted ((Bloom et al., 2009, 2010)). This monitoring technology allows the owner to limit the exposure of managerial hold-up ex-post, but it also reduces managerial incentives ex-ante. As will be clear in the theory, intense monitoring is as if firms were granting little authority to the manager. We will see that in equilibrium firms will be heterogeneous in their demand for managerial monitoring. Moreover, it is precisely the pattern of optimal monitoring, which will determine the marginal return schedule $\{\Delta_n\}_n$. Clearly, not all firms hire managers from outside. This is particularly true in India, where the vast majority of small firms are family firms. To capture this important aspect in our theory, entrepreneurs do not have to delegate authority to outside managers. In contrast, they can also decide to manage their own firm. The advantage of doing so is that contractual frictions do not apply and that the firm can save on managerial wages. The downside is similar to Caselli and Gennaioli (2012): owners might not be the best managers.

We assume that managerial effort enters the production function according to

$$m(e) = \frac{1}{\gamma - (\gamma - 1) e^{\sigma}}.$$

This functional form is convenient, because it implies (see (9)) that the the flow profits of a firm (before paying for any managers in case the firm decides to hire some) are given by

$$\pi(n, t) = \frac{\gamma - 1}{\gamma} e^{\sigma} Y(t) n.$$

The managerial effort bundle in turn is an aggregate of many managerial *tasks*. In particular, we assume that a unit continuum of tasks to be performed and that

$$e = exp\left(\int_0^1 ln(e(i))\, di\right),$$

where $e(i)$ denotes managerial effort for task i. This structure is convenient once we want to parametrize the quality of the contractual system (see below).

Given this structure, the firms now have the choice to "buy" e from managers or to provide them on their own. The benefit of hiring managers is that specialized managers can provide managerial efficiency more effectually. However, hiring managers is also costly. Not only do managers have to get remunerated for their services and their opportunity costs (as they could have been entrepreneurs themselves) but the interaction between owners and outside managers is plagued by contractual frictions.

3.4.1 The case of no managerial delegation

Consider first a firm that decides to not hire a manager. Effort provision is costly. In particular, the owner suffers utility costs of

$$C([e(i)]_i, Q(t)) = v_H n Q(t) \left[\int_0^1 e(i)\, di\right], \tag{20}$$

where ν_H is a cost-shifter. The value of the firm is hence given by

$$V^{NM}(n) = \max_{[e(i)]} \left\{\frac{\gamma - 1}{\gamma} e^{\sigma} Y(t) n - v_H n Q(t) \left[\int_0^1 e(i)\, di\right]\right\}.$$

The solution to this problem is given by

$$V^{NM}(n) = (1 - \sigma) \frac{\gamma - 1}{\gamma} (e_{NM})^{\sigma} Y(t) n, \tag{21}$$

where

$$e_{NM} = \left(\frac{\gamma - 1}{\gamma} \frac{\sigma}{v_H} \frac{Y_t}{Q_t} \right)^{\frac{1}{1-\sigma}}, \tag{22}$$

and $\frac{Y_t}{Q_t}$ is constant in a stationary equilibrium. In particular, we can also express V^{NM} as

$$V^{NM}(n) = (1-\sigma) \pi^{NM} Y(t) n, \tag{23}$$

where

$$\pi^{NM} \equiv \frac{\gamma - 1}{\gamma} \left(\frac{\gamma - 1}{\gamma} \frac{\sigma}{v_H} \frac{Y_t}{Q_t} \right)^{\frac{\sigma}{1-\sigma}} \tag{24}$$

is a constant.

3.4.2 The case of managerial delegation

Now suppose the firms were to hire an outside manager. We assume that firms are matched with managers and make take-it-or-leave-it offers. Once being matched, managers have the outside option of starting their own firm, i.e. they can again decide to enter the market as a single product firm without hiring a manager on their own. Intuitively, training an outside manager takes some time. Hence, the outside option of manager is given by

$$U^{Out} = z \times V^{NM}(1), \tag{25}$$

where z is again the flow rate of entry. Note from (14) that

$$W_H^E = zW_1^{INC}(1) > zV^{NM}(1) = W_L^E = U^{Out}.$$

While low types are indifferent between entrepreneurship and working as a manager, high types are strictly better off entering the entrepreneurial market - as a manager their outside option is to be a one-product firm forgoing the option value of innovation (which is inexistent for low types). Taking the outside option (25) as given, the firm and the manager enter a contracting game, which is similar to Acemoglu et al. (2007). The ease at which contracts can be written depends on the *contractual environment,* which is defined be the subset of tasks, which are contractible. In particular, we assume that of the unit mass of tasks only a the measure $[0, \mu]$ can directly be contracted on in that the required managerial effort levels $e(i)$ are enforceable in court. We take μ as a country characteristic, with its empirical analog being the rule of law index, which we used in our empirical section above.

To model the game between owners and managers, we follow the standard incomplete contract literature that managerial effort provision requires an ex-ante investment and that the manager and the entrepreneur engage in Nash-Bargaining about the joint surplus ex-post. As an imperfect substitute of a contract to limit the firms' hold-up, we assume that the firm can monitor the manager and thereby ensure a fraction of the managerial effort to be exerted. This monitoring technology (which we think of as an endogenous limit to managerial authority as in the work of Nick Bloom and John van Reenen) allows the owner to limit the exposure of managerial hold-up ex-post, but it also reduces managerial incentives ex-ante. The precise timeline of the game is as follows:

1. The entrepreneur is matched to a manager, who has an outside option given in (25). The entrepreneur offers a contract, which specifies an ex-ante transfer τ, managerial efforts in contractable tasks $[e_C(i)]_{i=0}^{\mu}$ and a degree of monitoring $\delta \in [0, 1]$. The contract will not be able to enforce effort in non-contractable tasks $[e_{NC}(i)]_{i=\mu}^{1}$ and managers are subject to ex-ante liquidity constraints $\tau \geq -\kappa Y(t)$. Such liquidity constraints limit the degree to which payoffs are directly transferable.

2. Given the contract, the manager choses the effort levels of *all* tasks, i.e. $[e_C(i)]_{i=0}^{\mu}$ and $[\hat{e}_{NC}(i)]_{i=\mu}^{1}$. Doing so is subject to utility costs

$$C^M\left([e(i)]_i, Q(t)\right) = v_L n Q(t)\left[\int_0^1 e(i)\,di\right], \tag{26}$$

where $\nu_L < \nu_H$. In contrast to (20), managers are more productive in providing managerial tasks ($\nu_L < \nu_H$) and the span of control is larger in that costs scale with n rather than n^{ε}. This reflects the efficiency gains of delegation and division of labor.

3. The firm hires production workers and is committed to pay their wage $w(t)$.

4. At the end of the period, the manager can threat to withhold his services in the non-contractable activities $[\hat{e}_{NC}(i)]_{i=\mu}^{1}$, i.e. he can threat to set $\tilde{e}(i) = \delta\hat{e}(i)$ instead of $\hat{e}(i)$. Hence, through monitoring the firm can make[13]

5. He demands a payment P in return of actually performing an effort level of $[\hat{e}_{NC}(i)]_{i=\mu}^{1}$ and not to only set $[\delta\hat{e}_{NC}(i)]_{i=\mu}^{1}$. P is decided about by Nash-Bargaining between the firm and the manager. Note that the ex-ante transfer τ is already paid, i.e. is not subject to the bargaining. As there are gains from trade, they will bargain to the efficient ex-post outcome and the manager will put in $\hat{e}$ units of effort.

6. Production takes place and payments are settled.

While our interpretation is different, our game is a standard incomplete contract game, where a "supplier" (the manager) has to do a relationship-specific investment but can threat to withhold his services in the process of joint production. Note that the firm *also* engages in a relationship-specific investment - the innovation. However, because the parties can transfer utility (via the ex-ante payment τ) *after* the innovation is made but *before* the manager decides on his effort level, there does not necessarily be an inefficiency.

Given this structure of game, we can determine the value of the firm under managerial delegation, $V^M(n)$. As usual we solve the game with backward induction. We first state a useful result.

Lemma 1. *Let α be the bargaining share of the manager. The ex-post payoff to the manager in the bargaining game is given by*

$$P = \alpha\left(1-\delta^{\sigma(1-\mu)}\right)\frac{\gamma-1}{\gamma}e^{\sigma}Y(t)\,n = \alpha\left(1-\delta^{\sigma(1-\mu)}\right)\pi\left([e_C]_0^{\mu}, [e_{NC}]_{\mu}^{1}, n; Y(t)\right), \tag{27}$$

where

$$e = exp\left(\int_0^{\mu} ln\left(e_C(i)\right)di + \int_{\mu}^{1} ln\left(e_{NC}(i)\right)di\right).$$

Proof. See Section 6.3 in the Appendix. □

[13] A more detailed microfoundation of the game is as follows: in stage 4, when the manager decides on his effort, we can think of him learning the details of the firm (firm-specific knowledge). Let $[\hat{e}(i)]_{i=\mu}^{1}$ be the level of learning. The cost of learning are given by (26). Given a level of learning $\hat{e}(i)$, the cost of *providing* effort is then given by

$$\Gamma(e,\hat{e}) = \begin{cases} 0 & \text{if } e \leq \hat{e} \\ \infty & \text{if } e > \hat{e} \end{cases}.$$

Hence, in stage 4, the manager essentially invests in the capability of providing effort and once these costs are sunk, she can provide these at zero marginal costs. She can however threaten to not show up to work.

Hence, ex-post the manager receives a constant share of profits. This share however is partly under the control of the firm through the exertion of monitoring effort δ. In particular through its choice of δ, the firm can govern the degree of managerial *authority*

$$\vartheta \equiv \alpha \left(1 - \delta^{\sigma(1-\mu)}\right), \tag{28}$$

and we will have firm directly choose ϑ. We will see that the distribution of endogenous authority across firm ϑ will be a crucial object in the analysis. By choosing ϑ, the firm experiences the well-known trade-off of assigning property rights: monitoring the manager intensely will reduce managerial authority ϑ and hence the degree of hold-up the firm faces ex-post. However, it will also reduce the managers' allocation of effort in non-contractual tasks. In particular, given his payoff P and the cost function C^M (see (26)) the optimally condition for the managers' effort in non-contractual tasks is given by

$$\vartheta \frac{\gamma - 1}{\gamma} \sigma e^{\sigma} \frac{1}{e^{NC}(i)} Y(t) n = v_L n Q(t). \tag{29}$$

As all tasks are symmetric, we have $e^{NC}(j) = e^{NC}$ for $j \in [\mu, 1]$. Similarly, the optimal contract will specify symmetric effort levels for the contractual activity, i.e. $e_C(i) = e_C$ for $i \in [0, \mu]$. Hence, total managerial effort is given by

$$e = exp\left(\int_0^{\mu} ln\left(e_C(i)\right) di + \int_{\mu}^{1} ln\left(e_{NC}(i)\right) di\right) = e_C^{\mu} e_{NC}^{1-\mu},$$

which is a simple Cobb-Douglas "production function" where the Cobb-Douglas shares are determined by the contractual environment. From (29) we hence get the optimal amount of effort in non-contractual tasks as

$$e_{NC} = \left[\vartheta \frac{\sigma}{v_L} \frac{\gamma - 1}{\gamma} \frac{Y(t)}{Q(t)} e_C^{\mu\sigma}\right]^{\frac{1}{1-(1-\mu)\sigma}}. \tag{30}$$

(30) clearly shows the two instruments, the firm has to govern managerial incentives in non-contractual tasks. First of all, it can increase the marginal return to effort by giving the manager a large degree of authority ϑ. While this increase effort provision, it also reduces the share of profits the firm eventually keeps. Secondly, the firm can exploit the fact that contractual and non-contractual effort are complements, i.e. e_{NC} is increasing in e_C. Intuitively: by forcing the managers to work particularly hard in those tasks where contracts can be written, the firm increases managerial incentives in other tasks. While such behavior will emerge in countries with bad institutions, it will of course be inefficient as the marginal return between contractual and non-contractual tasks will not be equalized.

With Lemma 1 at hand we can now formally define the value of the firm. As the firm makes a take-it-or-leave it offer to the manager taking his outside opportunity as given, the value of hiring a manager is implicitly defined by the optimal contracting problem

$$V^M(n,t) \equiv \max_{\tau, \vartheta, [e^C(i)]_0^{\mu}} (1-\vartheta) \pi \left([e_C]_0^{\mu}, [e_{NC}]_{\mu}^1, n; Y(t)\right) - \tau \tag{31}$$

subject to

$$z \times V^{NM}(1) \leq \vartheta \pi \left([e_C]_0^{\mu}, [e_{NC}]_{\mu}^1, n; Y(t)\right) - v_L n Q(t) \left(\left[\int_0^{\mu} e_C(i)\, di\right] + \left[\int_{\mu}^1 e_{NC}(i)\, di\right]\right) + \tau \tag{32}$$

$$\tau \geq -\chi Y(t) \tag{33}$$

$$\vartheta \leq \alpha \tag{34}$$

$$e_{NC}(i) = \arg\max_{e_{NC}(i)} \left\{\vartheta \pi \left(e_C, e_{NC}, n\right) - v_L n Q(t) \left[\int_{\mu}^1 e_{NC}(i)\, di\right]\right\}, \tag{35}$$

where

$$\pi\left([e_C]_0^\mu, [e_{NC}]_\mu^1, n; Y(t)\right) = \frac{\gamma-1}{\gamma} exp\left(\int_0^\mu ln\left(e_C(i)\right) di + \int_\mu^1 ln\left(e_{NC}(i)\right) di\right)^\sigma Y(t)\, n.$$

Hence, the firm has three control variables to govern the relationship with its manager: it assigns effort in contractual tasks, it choses the degree of authority and it distributed surplus via the ex-ante payment τ. When doing so however, it is subject to three constraints. First of all, it has to satisfy the participation constraint of the manager (32). Secondly, we assume tat managers face liquidity constraints (33), i.e. managers do not have deep enough pockets, which would allow them to simply buy the firm. Finally, (35) is the managers' incentive constraint for non-contractual effort, the solution of which is simply (30).[14]

Before characterizing the optimal contract, note that the participation constraint (32) will always be binding. The reason is the presence of some contractual tasks. As the marginal product of contractual tasks is positive, the firm will always simply have the manager work more in such activities until the manager's participation constraint binds.

To characterize the optimal contract under *imperfect* contractual enforcement, consider first Proposition 1.

Proposition 1. *Consider the setup above and suppose $\mu = 1$, i.e. contracts can be perfectly enforced. Then:*

1. Managerial effort is given by
$$e^{PC} = \left(\frac{\sigma}{\nu_L}\frac{\gamma-1}{\gamma}\frac{Y(t)}{Q(t)}\right)^{\frac{1}{1-\sigma}}$$

2. Flow profits $\pi^M(n,t;\mu)$ are given by
$$\pi^M(n,t;1) = \frac{\gamma-1}{\gamma}\left(e^{PC}\right)^\sigma Y(t)\, n$$

3. The value of the firm $V^M(n,t;\mu)$ is given by
$$\begin{aligned} V^M(n,t;1) &= (1-\sigma)\frac{\gamma-1}{\gamma}\left(e^{PC}\right)^\sigma Y(t)\, n - zV^{NM}(1) \\ &= (1-\sigma)\frac{\gamma-1}{\gamma}\left(e^{PC}\right)^\sigma Y(t)\, n - z(1-\sigma)\pi^{NM}Y(t), \end{aligned} \tag{36}$$

   where π^{NM} is given in (24).

Proof. See Appendix. □

Proposition 1 is very useful because it shows that our model (essentially) boils down to the canonical model by Klette and Kortum (2004) in case contracts are perfect.[15] To see this, note that 36 implies that

$$\Delta(n,t;1) \equiv \frac{V(n+1,t;1) - V(n,t;1)}{Q(t)} = (1-\sigma)\frac{\gamma-1}{\gamma}\left(\frac{\sigma}{\nu_L}\frac{\gamma-1}{\gamma}\right)^{\frac{\sigma}{1-\sigma}}\left(\frac{Y(t)}{Q(t)}\right)^{\frac{1}{1-\sigma}}. \tag{37}$$

[14] The restriction for managerial authority ϑ (see (28)) simply reflects that the firm can only reduce managerial authority by monitoring. For our main results we will focus on the case of $\alpha = 1$, which is analytically attractive. For that case (28) will never be binding. The general case of $\alpha < 1$ is contained in the Appendix.

[15] The slight difference to Klette and Kortum (2004) (hence the "essentially") is firms' choice of organizational form. As firms' choice of either hiring a manager or running the firm as self-employed entrepreneurs will be size dependent, innovation incentives will be non-constant around that cutoff. We will come back to this in Section 3.5 below.

As $\frac{Y(t)}{Q(t)}$ is constant in a stationary equilibrium, (37) and (16) imply that innovation incentives are also constant. As far as firm dynamics are concerned however, our model is then simply the Klette and Kortum (2004) model.

The interesting case in our model is of course the one, where contracts are *not* perfect. If that is the case, owners have to elicit managerial effort through the distribution of authority ϑ and their assignment of contractual effort. Firms' optimal policies are contained in the following Proposition.

Proposition 2. *Consider the setup above and let $\alpha = 1$. Let $\vartheta(n,\mu)$ be the endogenous allocation of authority. Then:*

1. Managerial effort in contractual and non-contractual tasks is given by

$$\frac{e_C}{e^{PC}} = \left(\vartheta(n,\mu)\left(\frac{1-(1-\mu)\sigma}{\vartheta(n,\mu)-(1-\mu)\sigma}\right)^{1-(1-\mu)\sigma}\right)^{\frac{1}{1-\sigma}} \geq 1 \tag{38}$$

$$\frac{e_{NC}}{e^{PC}} = \left(\vartheta(n,\mu)\left(\frac{1-(1-\mu)\sigma}{\vartheta(n,\mu)-(1-\mu)\sigma}\right)^{\mu\sigma}\right)^{\frac{1}{1-\sigma}} \leq 1. \tag{39}$$

2. Flow profits are given by

$$\pi^M(n,t;\mu) = \left(\left(\frac{1-(1-\mu)\sigma}{\vartheta(n,\mu)-(1-\mu)\sigma}\right)^{\mu}\vartheta(n,\mu)\right)^{\frac{\sigma}{1-\sigma}}\pi^M(n,t;1) \tag{40}$$

3. The value of the firm is given by

$$V^M(n,t;\mu) = \left[1-\sigma\vartheta(n,\mu)\left(1+\frac{\mu(1-\vartheta(n,\mu))}{\vartheta(n,\mu)-(1-\mu)\sigma}\right)\right]\pi^M(n,t;\mu) - z(1-\sigma)\pi^{NM}Y(t). \tag{41}$$

In particular, effort levels (e_C, e_{NC}), profits π^M and values V^M are equal to their perfect contract benchmark whenever $\vartheta(n,\mu) = 1$.

Proof. See Appendix. □

Proposition 2 contains two sets of important results. First of all: incomplete contracts introduce distortions into managerial effort provision. As seen from (38) and (39), effort is inefficiently provided in *both* types of tasks. While effort in contractual activities is too high, there is too little effort in the non-contractual part of managerial activities. (40) and (41) then show the consequences of this intra-firm distortion: Both flow profits π^M and the firms' value V^M are below their perfect contracts benchmark. While the latter is not surprising (with full contracting, the firm could of course have replicated the imperfect contract allocation), the second part of Proposition 2 has a less obvious implication: contractual frictions *only* matter in far as they reduce managerial authority below unity. More precisely: As long as the firm provides full authority to the manager, $\vartheta = 1$, the allocations replicate the perfect contract allocations irrespective of μ. The reason is the usual property rights intuition: Full authority makes the manager the residual claimant to the firm. As the owners' input into joint production is realized before bargaining takes place, the intuition from Grossman and Hart (1986) suggests to assign property rights to the manager. Contractual frictions therefore matter *only* in as far as firms are unwilling to delegate authority to their managerial personnel. Wether this is the case, is the content of the following proposition.

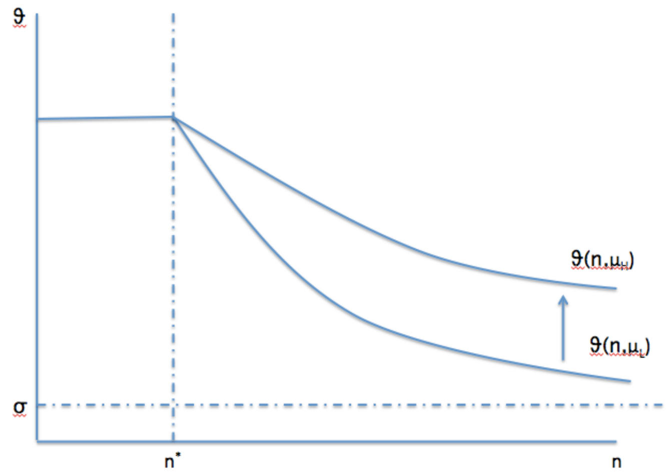


Figure 8: The optimal level of authority $\vartheta(n,\mu)$

Proposition 3. *Consider the setup above and define a critical level of firm size n^* as*

$$n^* = \left(\frac{\nu_L}{\nu_H}\right)^{\frac{\sigma}{1-\sigma}} \left(z + \frac{\chi}{(1-\sigma)\,\pi^{NM}}\right), \tag{42}$$

where π^{NM} is a constant given in (24). The optimal degree of authority $\vartheta(n,\mu)$ satisfies

$$\vartheta(n,\mu) = \begin{cases} 1 & \text{if } n \leq n^* \\ \vartheta^C(n,\mu) & \text{if } n > n^* \end{cases},$$

where $\vartheta^C(n,\mu)$ is implicitly defined by

$$\frac{n^*}{n} = \frac{\vartheta - \sigma}{1-\sigma}\left[\left(\frac{1-(1-\mu)\sigma}{\vartheta^C(n,\mu)-(1-\mu)\sigma}\right)^{1-\sigma(1-\mu)} \vartheta^C(n,\mu)\right]^{\frac{1}{1-\sigma}}. \tag{43}$$

Note that $\vartheta^C(n^*,\mu) = 1$ so that $\vartheta(n,\mu)$ is continuous. Furthermore,

- $\frac{\partial \vartheta^C(n,\mu)}{\partial n} < 0$, i.e. bigger firms grant less authority
- $\frac{\partial \vartheta^C(n,\mu)}{\partial \mu} > 0$, i.e. better contracts induce firms to grant more authority
- $\lim_{n\to\infty} \vartheta(n,\mu) = \sigma$, i.e. the limit of managerial authority does not depend on the legal system.

Proof. See Appendix. □

Proposition 2 showed that firms only achieve "x-efficiency" when they grant full authority to their managers. Proposition 3 then shows that full efficiency is not only not achieved by all firms in the economy but that large firms have higher incentives to distort managerial incentives and that this is particularly the case in environments, where contractual frictions are large. A summary of Proposition 3 is contained in Figure 8 below. The first salient feature of the optimal allocation of authority is that full authority is granted in small firms (with $n < n^*$). The intuition is as follows: Allocational efficiency (within the firm) requires that $\vartheta = 1$. While this maximizes production efficiency, it also implies that the entirety of firms' profits will be allocated to the manager once the bargaining stage is reached. The reason why firms might still be willing to accept such contract, is of course the possibility of ex-ante transfers τ. By requiring the manager to *pay* the firm in the beginning of the period, the firm essentially solves the contractual friction by simply selling the firm and letting the manager, who now has the efficient incentives, maximize profits. The result is that managerial effort levels, and hence profits and firm values, are equal to their perfect contract counterparts.

Such transfers however are limited by managerial liquidity constraints (see (33)). As large firms will need to set *higher* ex-ante transfers (because managers will earn higher rents through bargaining ex-post), managerial liquidity constraints will become binding eventually. The critical size of the firm when such liquidity constraints are binding is therefore precisely n^*. As expected, looser credit limits (χ large) will increase this cutoff. Furthermore, better outside options (z large) will also make liquidity constraints less of a problem. The reason is that better outside options ensure that managers require relatively payments to even participate in the contract. The required payment to make managers indifferent will therefore be lower. Finally, liquidity constraints will become tighter, the more profits are to be earned (π^{NM} large and ν_L low) as it requires firms to extract even higher payments ex-ante.

Once liquidity constraints are binding, firms are unable to limit managerial hold-up though ex-ante transfers. Hence, they have to rely on the other two instruments. Either thy can require managers to work inefficiently long hours in contractual activities. Or they can limit managerial authority, reduce their exposure to ex-post hold-up but at the same time non-contractual incentives. Proposition 3 and the optimal contractual effort e_C (see (38)) show that optimal behavior implies that both margins are distorted. Hence, larger firms are particularly hurt by imperfect contracts as agency problems become more severe. From a dynamic perspective, this implies that contractual problems will reduce innovation incentives because firms cannot seamlessly expand.

A second implication of Proposition 3 is that better contracts will *increase* managerial authority. In terms of Figure 8, the authority schedule shifts up. This is not obvious. As only non-contractual activities benefit from more authority one could have expected *more* authority to be granted in environments, where such authority is more valuable. However, the opposite the case. The reason for this result is precisely the distortion of contractual effort, whereby an excessive usage of those tasks is used to incentive non-contractual effort. If μ is large, this "overproduction" is costly and the firm has to remunerate the manager of putting the effort in low-marginal-product activities. Hence, the firm shifts from incentivizing the managers with contractual effort to incentive schemes relying on authority.

The implications of such optimal delegation behavior on firms' payoff are contained in Figure 9. We depict the value function as a function of firm size for different levels of the contractual environment. As shown in Proposition 2, the case of perfect contracts corresponds to a linear value function and constant innovation incentives. As the contractual environment deteriorates, managerial authority declines and with it production efficiency. Hence, the value function becomes *concave* as larger firms will be especially affected by contractual frictions. This resonates will with the empirical findings of Bloom et al. (2010) and Bloom and Reenen (2010). Note also that all value functions lie on top of each other for sufficiently small firms ($n < n^*$): the informal allocation of managerial authority is a good substitute for formal contracts as long as firms are small.

3.5 Innovation Incentives

To determine firms' final innovation incentives we of course have to take account of their binary organizational choice to either hire a manager or provide managerial services themselves. As there are no dynamic ramifications in switching organization design, this is a simple static trade-off. Hence,

$$V(n,t;\mu) = \max\left\{V^{NM}(n,t), V^{M}(n,t)\right\},$$

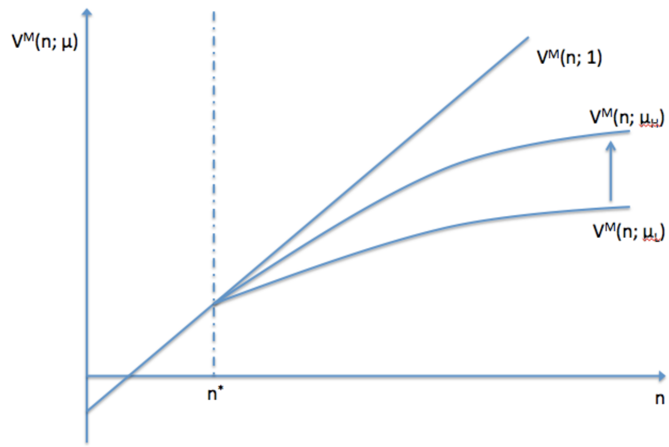


Figure 9: The value function $V(n, \mu)$

where V^{NM} and V^M are given in (23) and (41) respectively. Using these expressions we get that

$$\begin{aligned} V^M(n,t;\mu) &= h(\vartheta(n,\mu),\mu)\,\pi^M(n,t;1) - z(1-\sigma)\,\pi^{NM}Y(t) \\ &= \frac{h(\vartheta(n,\mu),\mu)}{1-\sigma}\left(\frac{v_H}{v_L}\right)^{\frac{\sigma}{1-\sigma}} V^{NM}(n) - zV^{NM}(1), \end{aligned} \tag{44}$$

where

$$h(\vartheta,\mu) = \left(1-\sigma\vartheta(n,\mu)\left(1+\frac{\mu(1-\vartheta(n,\mu))}{\vartheta(n,\mu)-(1-\mu)\sigma}\right)\right)\left(\left(\frac{1-(1-\mu)\sigma}{\vartheta(n,\mu)-(1-\mu)\sigma}\right)^{\mu}\vartheta(n,\mu)\right)^{\frac{\sigma}{1-\sigma}}$$

is decreasing in n as long as $\mu < 1$. (44) show concisely the three forces that determine the pattern of organizational form. The last term $z \times V^{NM}(1)$ is the *opportunity cost* term. As managers have to be remunerated for their time but owners can perform managerial duties "while owning the firm", employing a manager has a fixed costs. The firms term $h(\vartheta(n),\mu)\left(\frac{v_H}{v_L}\right)^{\frac{\sigma}{1-\sigma}} n^{\frac{\varepsilon-1}{1-\sigma}}$ combines two counteracting forces. The first component, $h(\vartheta(n,\mu),\mu)$, is akin to an *imperfect contract tax* as firms will endogenously distort managerial incentives and firm efficiency. As shown in Propositions (2) and (3), large firms will do so particularly strongly so that this tax is size-dependent and increasing in firm size. The second term, $\left(\frac{v_H}{v_L}\right)^{\frac{\sigma}{1-\sigma}}$, is the *specialization benefit*. As owners do not have time to specialize in their managerial tasks, they can provide them less efficiently. It is the tradeoff between these two forces and the fixed opportunity costs, which determine firms' organizational choice. In particular, imperfect contracts can stand in the way of an efficient division of labor by making the interaction between owners and managers so inefficient that the technological gains from specialization are dominated.

As $h(.)$ is increasing and concave in n, (44) shows that there are three patterns of organizational choice that can prevail in equilibrium:

1. No firm hires a manger
2. There is cut-off n^M such that all firm with $n > n^M$ hire managers
3. There are two cut-offs n_L^M and n_H^M such that all firms with $n_L^M < n < n_H^M$ hire managers.

We consider this last case as somewhat pathological as it implies that optimal authority depreciates so quickly in firm size that the *biggest* firms are again better off not hiring any managers. Hence, we are going to impose Assumption 1, which ensures that this is not the case.

Assumption 1. *Suppose that* $\left(\sigma, \mu, \frac{\nu_H}{\nu_L}\right)$ *are such that*

$$\left(\left(\frac{1-(1-\mu)\,\sigma}{\mu}\right)^{\mu}\sigma^{1-\mu}\right)\left(\frac{\upsilon_H}{\upsilon_L}\right) > 1.$$

As $\vartheta(n,\mu) \to \sigma$, Assumption 1 implies that $\frac{h(\vartheta(n,\mu),\mu)}{1-\sigma}\left(\frac{\upsilon_H}{\upsilon_L}\right)^{\frac{\sigma}{1-\sigma}} \to 1$ from above. Hence, the marginal product has a higher return for firms with managerial delegation.

The final pattern of organizational choice is given in Figure 10 below.

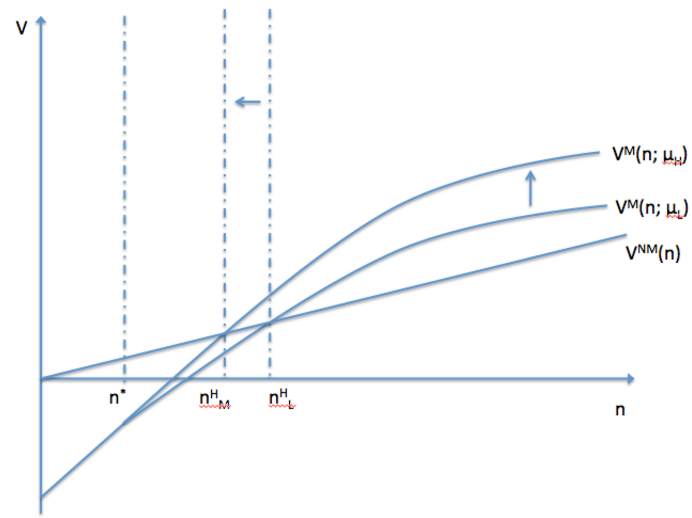


Figure 10: Managerial Demand

We depict two managerial value functions (under different contractual environments) and the value of the self-employed firm, which is linear and independent of μ. As better contracts increase the value of delegation, the value function with managerial decentralization shifts to the lefts as μ increases. Hence, the marginal firm with managerial delegation will be smaller and - holding the firm size production fixed - the share of firms that opt to delegate authority will be larger.

Note that Figure 10 implicitly assumes that $n^* < n_M^H$. In general the relationship between n_M^H and n^* depends of course on parameters. However, we think of $n^* < n_M^H$ as the natural case. The reason is as follows: Proposition 3 implies that $\vartheta = 1$ for all $n < n^*$. According to Propositions 2 and 1 we then get

$$\begin{aligned} V^M(n,t;\mu) &= (1-\sigma)\,\pi^M(n,t;1) - z\,(1-\sigma)\,\pi^{NM}Y(t) = \left[\left(\frac{\nu_H}{\nu_L}\right)^{\frac{\sigma}{1=\sigma}} n - z\right](1-\sigma)\,\pi^{NM}Y(t) \\ &= \left[\left(\frac{\nu_H}{\nu_L}\right)^{\frac{\sigma}{1=\sigma}}(n-n^*) + \frac{\chi}{(1-\sigma)\,\pi^{NM}}\right](1-\sigma)\,\pi^{NM}Y(t)\,. \end{aligned}$$

A natural benchmark is the case of $\chi = 0$, i.e. managers will never *pay* firms ex-ante to provide managerial services. This case seems natural to us because either managers are sufficiently cash-constrained or frictions in the rule of law might make managers unwilling to pay their employer upfront - with imperfect enforcement, there is always a risk the the firm simply hires a new manager. Such contracts at least do not seem to occur frequently in the data.[16] This however directly implies that $V^{M}(n,t;\mu)<0$ for all $n<n^{*}$. The reason is intuitive: By construction of n^{*}, firm with $n<n^{*}$ will be run efficiently in that they will chose $\vartheta=1$. Then however, all the surplus will go to the manager ex-post. If ex-ante payments are ruled out, firms will not be able to appropriate any of these returns and have negative payoffs. Hence, such firm will clearly be better off running the firm on their own. Because we want to not rule out this benchmark case of $\chi=0$, we will restrict our analysis to the case of $n^{*}<n_{M}^{H}$.

In this paper we are of course mostly interested in the dynamic ramifications of contractual frictions. The crucial object for innovation incentives is of course the slope of the value function (see (16)). The implications of changes in the contractual environment on firms' incentives to grow are contained in Figure 11 below.

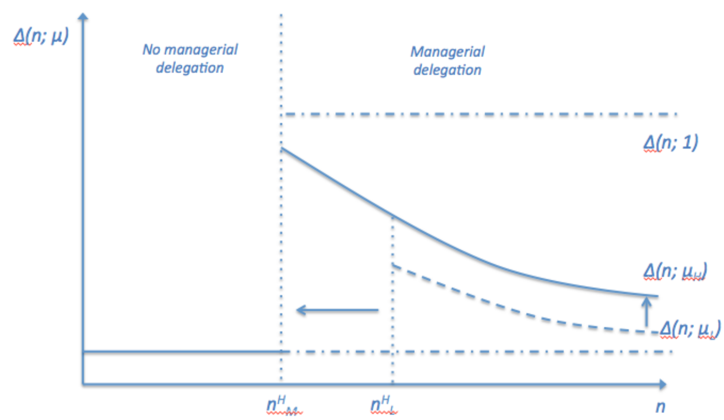


Figure 11: Contracts and the Incentives to Expand

Consider first the solid line, which represents the slope of the value function in an environment plagued by contractual frictions ($\Delta(n;\mu_{L})$). The marginal firm hiring a manager has n_{L}^{M} products. Until that point, the firm is managed by its owner and innovation incentives are both constant and low. This follows from (16) and (23), which imply that

$$\Delta(n,t)=(1-\sigma)\frac{\gamma-1}{\gamma}\left(\frac{\sigma}{\nu_{L}}\frac{\gamma-1}{\gamma}\right)^{\frac{\sigma}{1-\sigma}}\left(\frac{Y(t)}{Q(t)}\right)^{\frac{1}{1-\sigma}}\propto\left(\frac{1}{\nu_{L}}\right)^{\frac{\sigma}{1-\sigma}}. \tag{45}$$

If agents *owning* firms are relatively bad in *managing* firms, ν_{L} will be large and both profit margins and innovation incentives be low. As n_{M}^{L} the firm reaches the critical size to bring an outside manager into the firm. Through the lens of this model, such managerial upscaling is akin to the adoption of a technology: the firm incurs some fixed costs but it benefits from higher returns on the margin. Hence, innovation incentives will initially be high around the cutoff. However, as the firm grows, contractual frictions become more and more severe and through limited managerial authority the firm is less and less able to reap the benefits of managerial specialization. This concavity of the value function (see Figure 10) implies that marginal returns are declining in size. However, Assumption 1 ensures that such marginal returns are bounded from below by (45).

[16]However, one interpretation of such ex-ante payments are long-term contracts with a backloaded profile. Hence, managers might get compensated less than their marginal product in the early stages of their career and will then see steep wage growth during their life-cycle.

Now suppose that contracts improve. This will shift both the marginal return function and reduce the cutoff of "entering" firms from n_M^L to n_M^H. Such a change is depicted in Figure 11 as the dotted line $(\Delta(n;\mu_H))$. Intuitively: because better contracts enable firms to employ their managers more efficiently, innovation incentives increase holding firm size constant. It is this mechanism, that will explain the differences in the life-cycle between India and the US: Indian firms will have little incentives to grow as imperfect contracts will depress innovation incentives of large firms. Note the importance of the size-dependence: small firms will *not* be affected by contractual frictions in either country as they do not delegate decision power to outside managers anyway. Hence, the drag of contractual frictions will be especially severe for large firms. Dynamically, firms will of course anticipate that they will run into such decreasing returns. Hence, in equilibrium only few firms will grow large in India. Finally, consider the perfect contracts benchmark, which is also depicted in Figure 11. As argued above (see (37)), with perfect contracts, the size-dependence of innovation incentives disappears as the value function will be linear. The adoption-cutoff will be reduced to n_M^{PC} and innovation incentives going forward will be constant as in the standard model of Klette and Kortum (2004).

Conceptually, we think of the two schedules in Figure 11 as representing India and the US: While firms in the US have high innovation and growth incentives and hence expand, firms in India anticipate that they will not be able to run large firm efficiently as the legal system makes it hard to use managers efficiently, i.e. without any distortionary monitoring. Hence, firms in India have less incentives to grow. It is this lack of selection however, which keeps the inefficient low types in the market. In the next section we will now study the quantitative bite of this mechanism in a calibrated version of the model.

4 Quantitative Exercise

We now take the basic model to the data. Our quantitative aim is twofold. On the positive side, we want to show that the model is able to match the basic facts we set out to explain. On the normative side, our goal is to study the welfare implications of some common policies adopted in developing countries, such as small business subsidies.

4.1 Data

We are using two major sources of datasets. The first source is the Annual Survey of Industries (ASI) and the second is the National Sample Survey (NSS). For a more detailed description and some descriptive statics, we refer to Hsieh and Olken (2014). We also complement these datasets using the summary statistics on the US economy from Hsieh and Klenow (2011).

The ASI is an annual survey of manufacturing enterprises. It covers all factories registered under Sections 2m(i) and 2m(ii) of the Factories Act, 1948 i.e. those factories employing 10 or more workers using power; and those employing 20 or more workers without using power. The survey also covers bidi and cigar manufacturing establishments registered under the Bidi & Cigar Workers (Conditions of Employment) Act, 1966 with coverage as above. All electricity undertakings engaged in generation, transmission and distribution of electricity registered with the Central Electricity Authority (CEA) are also covered under ASI irrespective of their employment size. Defense establishments, oil storage and distribution depots, restaurants, hotels, cafes and computer services and the technical training institutes, etc. are excluded from the survey. The primary unit of enumeration in the survey is a factory in the case of manufacturing industries, a workshop in the case of repair services, an undertaking or a licensee in the case of electricity, gas & water supply undertakings and an establishment in the case of bidi & cigar industries. The NSS is a socio-economic survey covering different aspects. Every 5 years, a random sample of production units are sampled (Schedule 2.2). Importantly, this survey covers the population of producers, including small and informal producers not covered in the ASI. The production units covered in this survey sector has roughly about one-third share in the total contribution by the manufacturing sector in the GDP.

4.2 Parameters and Moments

In the model, we have the following parameters to calibrate: $\mu, \sigma, \zeta, \chi, \kappa, \gamma, \theta, \alpha$, and z. The parameters are calibrated by targeting the moments reported in Table 4.2. The calibrated parameters are given in Table 4.2.

Parameter	Description	Value
ζ	innovation elasticity	0.5
σ	curvature of efficiency	0.03
χ	pledgability	0.2
z	entry rate	0.027
κ	outside option	0.1
γ	innovation step size	1.24
θ	innovativeness	2.12
α	share of high type	0.02
μ	share of contractable tasks	0.35

Table 2: Parameter Calibration for Indian Firms

MMoment	Data	Model
$\frac{\text{Avg Employment for } n>5}{\text{Avg Employment for } n\leq 5}$	7.96	7.42
$\frac{\text{Value Added - Labor Costs}}{\text{Gross Output}}$	0.17	0.22
Employment Share of firms with $n \geq 5$	0.32	31
Growth rate	1.1%	1.3%
Life Cycle Profile	Figure	Figure

Table 3: Parameter Calibration for Indian Firms

As seen in Table 4.2, the model does a good job of matching the targets. The resulting Indian life-cycle profile is shown in Figure 12 and 13. The prediction of the model is that while the overall life-cycle profile of firms is completely flat as seen in the data (Figure 12), there is a subgroup of firms that have a positively-sloped profile (Figure 13). However, since these firms are constrained through the imperfect contractual environment, these firms do not grow sufficiently to push out the low-type firms, which, in the model, do not see any growth.

4.3 Comparison to the US

In this section, our goal is to compare Indian firms to US firms and diagnose the underlying differences. Our goal is to test the quantitative power of the contractual frictions in explaining the observed cross-country differences. For this, we recalibrate two parameters of the model (μ and z) to the US economy, keeping the rest of the parameters fixed at the Indian levels. These parameters for U.S. are chosen to match life cycle properties and employment share of 20+ firms (see Figure 14). The resulting parameters are reported in Table 4.3.

Figure 12: Life Cycle of Indian Firms

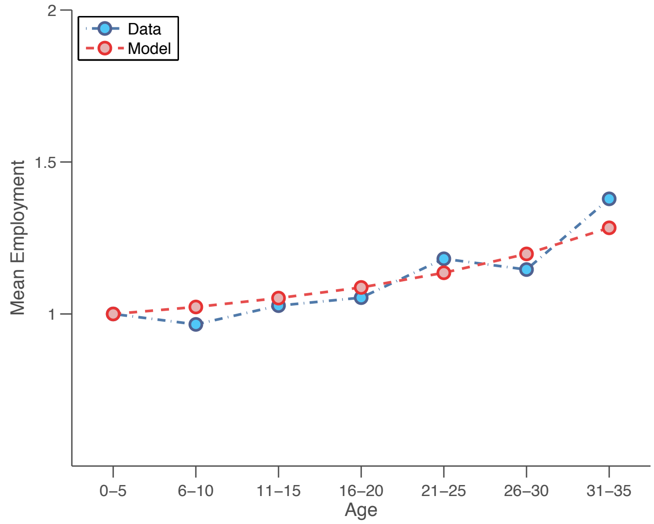


Figure 13: Life Cycle of Indian Firms

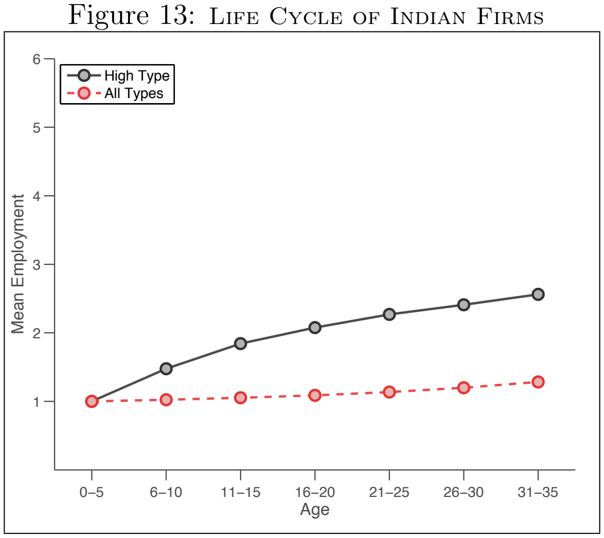
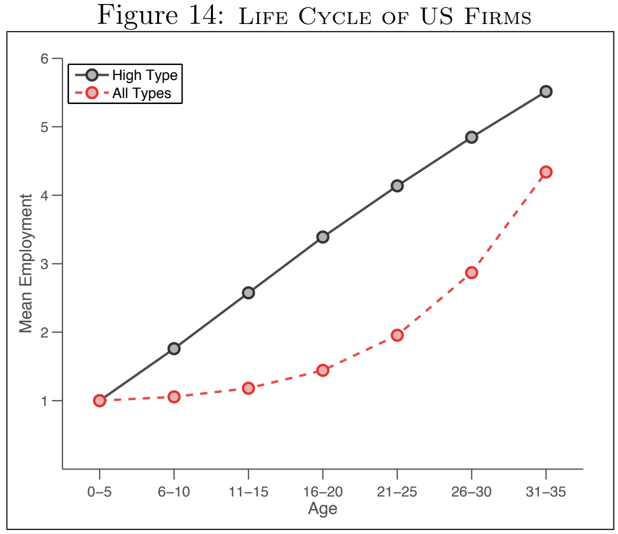


Figure 14: Life Cycle of US Firms



Notes: Figure 13 plots the life-cycle of Indian firms. The red dotted line shows the overall economy whereas the solid line depicts only the high-type firms. Figure 14 plots the same figure for the US firms.

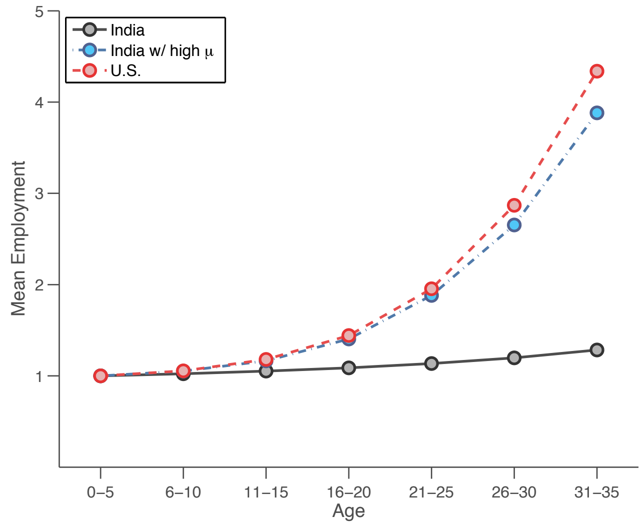
Parameter	**U.S.**	**India**
μ	0.70	0.35
z	0.0015	0.027

Table 4: Parameter Calibration for US Firms

Our calibration depicts also a quite sizable difference in the contractual environment between the two countries. To see the economic importance of μ and z, Figure 15 plots various scenarios: The solid line shows the baseline Indian economy. The dotted-blue line just introduces μ^{us} into the Indian economy while preserving z.

Figure 15 shows that changing the value of μ alone has a strong effect on firm dynamics. In particular,

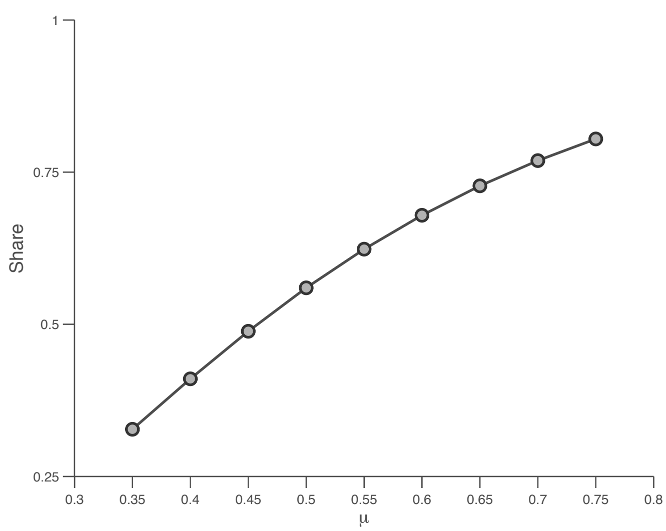
Figure 15: Improving the Contractual Environment (μ) vs Firm's Life Cycle



Notes: This figure plots the calibrated life-cycle dynamics of Indian and US firms. After calibrating the Indian economy, the US economy is recalibrated by changing only the entry rate z and contractual parameter μ. The intermediate blue line shows a hypothetical Indian economy with the US contractual environment $\mu^{ind} = \mu^{us}$.

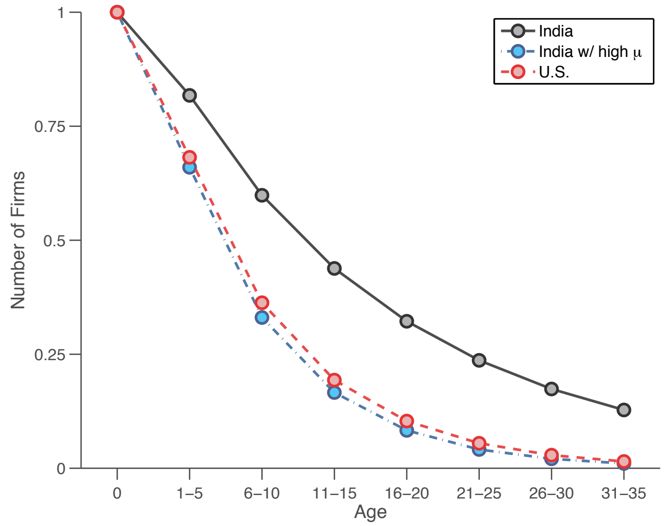
it shows that allowing for higher fraction contractible tasks (setting $\mu^{ind} = \mu^{us}$) can explain more than 80% of the observed employment increase of older firms. This is achieved owing to the fact that firm owners can offer a better contract to their manager, and thereby induce more effort by the manager. This in turn increases the incentives to expand the firm and increases its size. According to Figure 15, the remaining gap is explained by the difference in entry rate.

Figure 16: Employment Share of 5+ Firms vs μ



To see how the selection is important in the model, Figure 4.3 plots the employment share of firms that are 5+ years old for different values of μ. The figure implies that expansion of the firms due to better

Figure 17: Number of Low-type Firms by Age



contractual environment leads to the reallocation of workers from the low-types to high-types, and as high-types get older, they increase their relative employment share in the economy. This margin explains the higher share of employment held by older firms in the US compared to the Indian firms.

Figure 4.3 also shows how selection differs between India and U.S. through the lens of the model, by plotting the share of low-type firms within a cohort as it ages. The stark decline in the US is a sign of the strong selection as opposed to the weak selection margin in the Indian economy. Any policy that makes the low-type firms live longer would slow down this selection process and thus have a detrimental impact on the aggregate productivity. The following section will analyze the role of such a policy directly.

4.4 Small Business Subsidies

(To Be Written)

5 Conclusion

This paper studies the reasons behind the stark differences in firm dynamics across countries as documented in Hsieh and Klenow (2011). We focus on manufacturing firms in India and analyze the stagnant firm behavior. We show that the poor life-cycle behavior in the India could be explained by the lack of firm selection, wherein firms with little growth potential survive, because innovative firms do not expand sufficiently quickly to replace them. Our theory stresses the role of imperfect managerial contracts as the main cause for the insufficient expansion by the high-type firms. We show that if managerial effort provision is non-contractible, firms will endogenously limit managerial authority to reduce the extent of hold-up. As large producers will have a higher incentive to put such inefficient monitoring policies in place, the returns to innovation decline rapidly. Improvements in the degree of contract enforcement will therefore raise the returns to growth and increase the degree of creative destruction. This argument is in line with the empirical findings of Bloom and Reenen (2007, 2010).

Our analysis so far suggests the following conclusions: First, the steepness of the life-cycle growth trajectory of US firms (conditional on survival) reflects larger incumbent innovation incentives in the US, driven by an efficient managerial technology that allows firms to scale up easily. The fact that US firms can write better managerial contracts allows them to incentivize their managerial personnel and expand into new products without facing increasing marginal managerial costs. Second, US aggregate productivity growth is mostly driven by incumbents' innovation incentives that induce rapid growth of successful firms and early exit of unsuccessful ones. Indian firms, by contrast, simply earn too little infra-marginal rents to generate sufficient innovation incentives for steep life-cycle growth. Through the lens of our model, this is due to an imperfect contractual environment. Third, the policies that aim to support small businesses might indeed be detrimental to aggregate productivity, as these policies slow down the creative destruction and selection process that is needed in India.

References

Acemoglu, D., U. Akcigit, N. Bloom, and W. Kerr: 2012, 'Innovation, Reallocation and Growth'. *Working Paper*.

Acemoglu, D., P. Antras, and E. Helpman: 2007, 'Contracts and Technology Adoption'. *American Economic Review* **97**(3), 916–943.

Aghion, P. and P. Howitt: 1992, 'A Model of Growth through Creative Destruction'. *Econometrica* **60**(2), 323–351.

Akcigit, U. and W. Kerr: 2010, 'Growth Through Heterogeneous Innovations'. Technical report, NBER Working Paper 16443.

Bloom, N., B. Eifert, D. McKenzie, A. Mahajan, and J. Roberts: 2010, 'Does management matter: Evidence from India'.

Bloom, N. and J. V. Reenen: 2007, 'Measuring and Explaining Management Practices Across Firms and Countries'. *Quarterly Journal of Economics*.

Bloom, N. and J. V. Reenen: 2010, 'Why Do Management Practices Differ across Firms and Countries?'. *Journal of Economic Perspectives* **24**(1), 203–224.

Bloom, N., R. Sadun, and J. V. Reenen: 2009, 'The Organization of Firms Across Countries'. Technical report, NBER Working Paper 15129.

Buera, F., J. Kaboski, and Y. Shin: 2011, 'Finance and Development: A Tale of Two Sectors'. *American Economic Review* **101**(5), 1964–2002.

Caselli, F. and N. Gennaioli: 2012, 'Dynastic Management'. *Economic Inquiry*.

Cole, H., J. Greenwood, and J. Sanchez: 2012, 'Why Doesn't Technology Flow from Rich to Poor Countries?'. Technical report, Working Paper.

Collard-Wexler, A., J. Asker, and J. De Loecker: 2011, 'Productivity Volatility and the Misallocation of Resources in Developing Economies'. *NBER Working Paper* **17175**.

Gollin, D.: 2008, 'Nobody's Business but My Own: Self Employment and Small Enterprise in Economic Development'. *Journal of Monetary Economics* **55**(2), 219–233.

Grossman, G. and E. Helpman: 1991, 'Quality Ladders in the Theory of Growth'. *Review of Economic Studies* **58**(1), 43–61.

Grossman, S. J. and O. Hart: 1986, 'The costs and benefits of ownership: A theory of vertical and lateral integration'. *Journal of Political Economy* **94**(4), 691–719.

Guner, N., G. Ventura, and Y. Xu: 2011, 'Macroeconomic Implications of Size Dependent Policies'. *Review of Economic Dynamics* **11**(4), 721–744.

Hopenhayn, H.: 2012, 'On the Measure of Distortions'. *Working Paper.*

Hsieh, C.-T. and P. Klenow: 2009, 'Misallocation and Manufacturing TFP in China and India'. *Quarterly Journal of Economics* **124**(2), 771–807.

Hsieh, C.-T. and P. Klenow: 2011, 'The Life-Cycle of Plants in India and Mexico'. Working Paper.

Hsieh, C.-T. and B. A. Olken: 2014, 'The Missing "Missing Middle"'. *Journal of Economic Perspectives.*

Hurst, E. and B. Pugsley: 2012, 'What Do Small Businesses Do'. *Brookings Papers on Economic Activity.*

Jones, C.: 2013, 'Misallocation, Economic Growth and Input-Output Economics'. In: D. Acemoglu, M. Arrelano, and E. Dekel (eds.): *Advances in Econometrics: 10th World Congress*, Vol. 2. Cambridge University Press.

Klette, T. J. and S. Kortum: 2004, 'Innovating Firms and Aggregate Innovation'. *Journal of Political Economy* **112**(5), 986–1018.

Lentz, R. and D. T. Mortensen: 2008, 'An Empirical Model of Growth through Product Innovation'. *Econometrica* **76**(6), 1317–1373.

Lucas, R. E.: 1978, 'On the Size Distribution of Business Firms'. *The Bell Journal of Economics* **9**(2), 508–523.

Midrigan, V. and D. Y. Xu: 2010, 'Finance and Misallocation: Evidence From Plant-Level Data'. NBER Working Paper 15647.

Moll, B.: 2010, 'Productivity Losses from Financial Frictions: Can Self-Financing Undo Capital Misallocation?'. Working Paper.

Penrose, E. T.: 1959, *Theory of the Growth of Firms.* New York: J. Wiley & Sons.

Peters, M.: 2013, 'Heterogeneous Mark-Ups, Growth and Endogenous Misallocation'. Working Paper.

Powell, M.: 2012, 'Productivity and Credibility in Industry Equilibrium'. Technical report, Working Paper.

Restuccia, D. and R. Rogerson: 2008, 'Policy Distortions and Aggregate Productivity with Heterogenous Establishments'. *Review of Economic Dynamics* **11**(4), 707–720.

Romer, P. M.: 1990, 'Endogenous Technological Change'. *Journal of Political Economy* **98**(5), 71–102.

6 Appendix

6.1 Replicating Hsieh and Olken (2014)

Figure 18 below replicates the analysis of Hsieh and Olken (2014) for the firm size distribution in India. Despite the different sample (we use data from 1995, they use data in 2010), the general shape of the distribution looks similar. It is apparent that essentially all firms in India have at most 10 employees and that the the distribution is unimodal and smoothly declines. We also do not find any evidence for a missing middle.

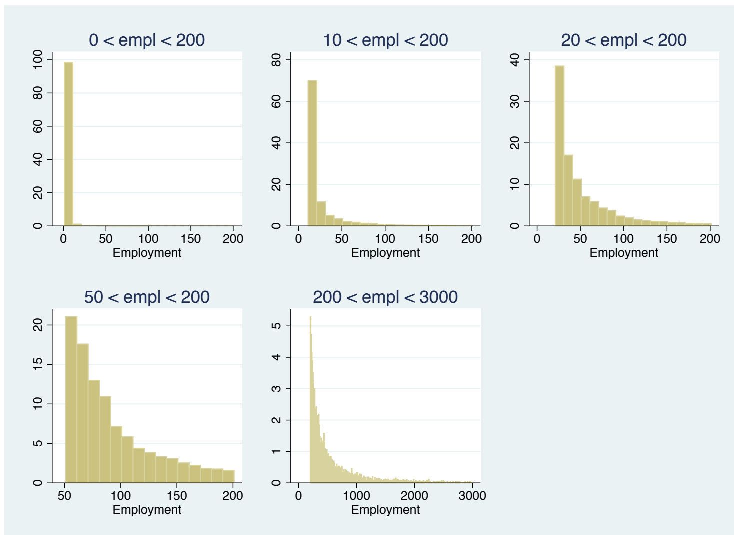


Figure 18: The firms size distribution in India

6.2 Additional Evidence for Section 2.1

Figure 19 plots the share of firms in ASI by age. It is seen that the share of ASI firms is very low — on average about 1%. It is slightly increasing by age, which we interpret as NSS firms having a higher probability of exit. Interestingly, there are few old firms in the ASI. While this might be due to measurement issues, it also hints at the importance of cohort effects in that the only remaining firms that entered the economy in 1955 are not part of the formal manufacturing sector.

In Figure 20 we provide some robustness to Figure 3 by replicating the basic pattern for different cutoff for the upper tail of the firm observations in the data. As in Figure 3 it is seen that irrespective of the precise cutoff, the tail actually fans out, i.e. there does exist a subset of firms that do grow.

Figure 21 addresses the concern of selection. , i.e. what is the share of firms after sampling weights are taken into account. If sampling weights were unity (or uncorrelated with employment) the share of firms was simply equal to the share of observations. Figure 21 shows that the firm size is negatively correlated with the sampling weights (so that the share of top firms is very small). More importantly, there is not much systematic correlation across age. Hence, the selection across age is unlikely to play a large role in the strong upward sloping pattern shown in Figures 20 and 3.

In Figure (7) we showed that the profitability of managerial resources is particularly high in large firms. However, Figure (7) did not include any controls. Table 6.2 below contains the results of regressions of the form

$$ln\left(APM_i\right) = \alpha + \beta ln\left(l_i\right) + x_i'\gamma + u_i,$$

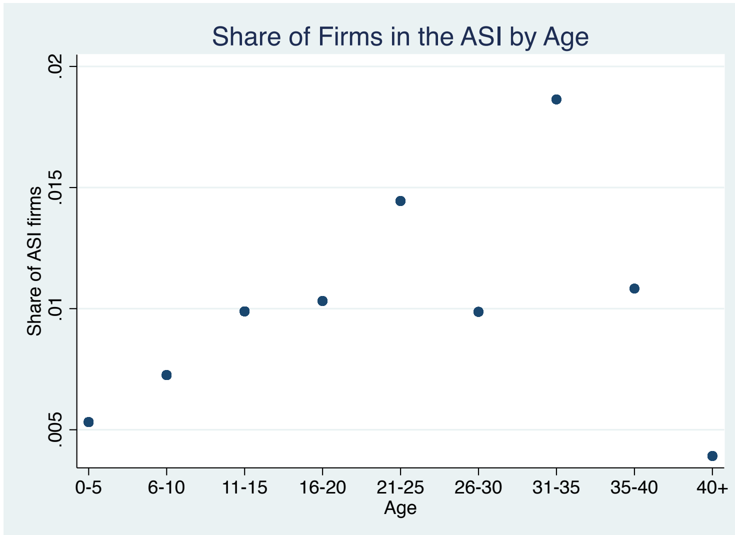


Figure 19: The share of ASI firms by age group

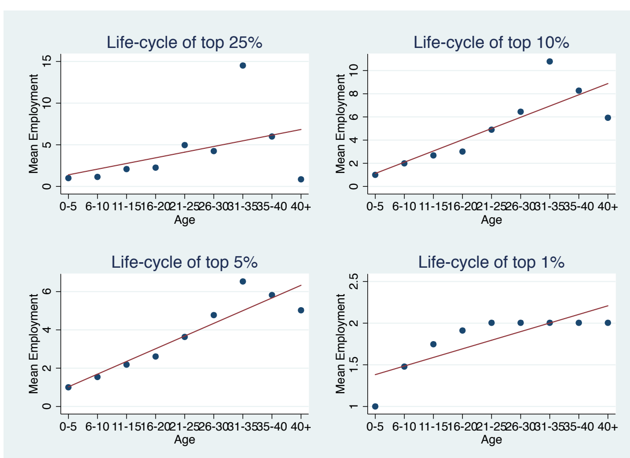


Figure 20: The life-cycle of the top firms in India

Figure 21: The evolution of the top firms in India

where APM is the average product of managers, measured either as value added per manager ($\frac{va}{m}$) or value added per managerial wage bill ($\frac{va}{w_m m}$), and l_i denotes total employment and x_i contains various fixed effects as controls. As seen in Table 6.2, the coefficient β is positive and highly significant, and hence consistent with Figure (7). Columns one to three contain the simple regression with different fixed effects. In column four we include the firm-specific population weights as weights for the regression.

6.3 The Bargaining Share

After the ex-ante payments are received and the manager decided how many firm-specific tasks to learn, he decides how much efficiency units actually to provide in the share of tasks, which are non-contractable. Let P be the bargained wage for the manager. If the parties do not agree, the manager does not get his payment P. The firm however can still force the manager to supply $\delta\hat{e}$ units of effort in the set of non-contractable activities. Hence, the payoff of the firm if no agreement is reached is given by

$$\begin{aligned}\pi^{NA}\left([e_C(i)]_0^{\mu}, [\hat{e}_{NC}(i)]_{\mu}^{1}, n, \delta\right) &= \frac{\gamma-1}{\gamma}\left(exp\left(\int_0^{\mu} ln\left(e_C(i)\right)di + \int_{\mu}^{1} ln\left(\delta e_{NC}(i)\right)di\right)\right)^{\sigma} Y(t)\, n \\ &= \frac{\gamma-1}{\gamma}\left(exp\left(\int_0^{1} ln\left(e(i)\right)di + ln\left(\delta^{1-\mu}\right)\right)\right)^{\sigma} Y(t)\, n \\ &= \frac{\gamma-1}{\gamma}\delta^{(1-\mu)\sigma}\left(exp\left(\int_0^{1} ln\left(e(i)\right)di\right)\right)^{\sigma} Y(t)\, n \\ &= \delta^{(1-\mu)\sigma}\pi\left([e_C(i)]_0^{\mu}, [\hat{e}_{NC}(i)]_{\mu}^{1}, n\right),\end{aligned}$$

where π denotes the profit of the firm if an agreement is reached. The manager in contrast does not get anything, because his skills are firm-specific and it is too late for him to get matched with a different employer. If the parties agree, the manager gets P and the firms get $\pi - P$. Hence, Nash bargaining implies that the payment P maximizes a geometric weighted average of the respective surplus. Letting α be the bargaining weight of the manager, P is given by

$$P^* = \arg\max_P \left(\pi\left([e(i)], n\right) - P - \pi^{NA}\left([e(i)], n, \delta\right)\right)^{1-\alpha} P^{\alpha}.$$

	Dep Variable: Average product per manager $ln(\frac{va_i}{m_i})$			
	(1)	(2)	(3)	(4)
log employment	0.327***	0.339***	0.363***	0.406***
	(0.009)	(0.009)	(0.009)	(0.013)
Industry FE	x	x	x	x
State FE		x	x	x
Age FE			x	x
Population weights				x
Observations	16,418	16,418	16,418	16,418
R-squared	0.24	0.27	0.28	0.26

	Dep Variable: Average product per managerial spending $ln(\frac{va_i}{w_m m_i})$			
	(1)	(2)	(3)	(4)
log employment	0.045***	0.058***	0.085***	0.131***
	(0.008)	(0.008)	(0.008)	(0.011)
Industry FE	x	x	x	x
State FE		x	x	x
Age FE			x	x
Population weights				x
Observations	16,284	16,284	16,284	16,284
R-squared	0.19	0.21	0.22	0.24

Table 5: The average product of managerial inputs and firm size

This implies that

$$\begin{aligned} P &= \alpha\left[\pi\left(\left[e\left(i\right)\right], n\right) - \pi^{NA}\left(\left[e\left(i\right)\right], n, \delta\right)\right] \\ &= \alpha\left(1-\delta^{\sigma}\right)\pi\left(\left[e\left(i\right)\right], n\right) \qquad (46) \\ &= \alpha\left(1-\delta^{\sigma}\right)\frac{\gamma-1}{\gamma}\hat{e}^{\sigma}Y\left(t\right)n, \qquad (47) \end{aligned}$$

where $e = exp\left(\int_0^{\mu} ln\left(e_C\left(i\right)\right)di + \int_{\mu}^{1} ln\left(e_{NC}\left(i\right)\right)di\right)$. (47) is the expression in (27).

6.4 Proof of Propositions 1 and 2

Using Lemma 1 and writing (30) as $e_{NC} = e_{NC}\left(\vartheta, e_C\right)$, we can rewrite (31) as

$$V^F\left(n\right) \equiv \max_{\tau,\vartheta,e^C}\left(1-\vartheta\right)\pi\left(e_C, e_{NC}, n\right) - \tau \qquad (48)$$

subject to

$$\begin{aligned} \vartheta\pi\left(e_C, e_{NC}, n\right) - C_{NC} - C_C + \tau &= z\left(1-\sigma\right)\pi^{NM}Y\left(t\right) \quad [\phi] \\ \tau &\geq -\chi Y\left(t\right) \quad [\eta] \qquad (49) \\ \vartheta &\leq 1 \quad [\rho], \qquad (50) \end{aligned}$$

where $C_C\left(e_C\right) = \mu v n Q\left(t\right)e^C$, and ϕ, η and ρ are the respective Lagrange multipliers, where we already imposed $\alpha = 1$. The three first order conditions are

$$
\begin{aligned}
0 &= (1-\vartheta)\left[\frac{\partial \pi}{\partial e_C}+\frac{\partial \pi}{\partial e_{NC}}\frac{\partial e_{NC}}{\partial e_C}\right]+\phi\left[\vartheta\left[\frac{\partial \pi}{\partial e_C}+\frac{\partial \pi}{\partial e_{NC}}\frac{\partial e_{NC}}{\partial e_C}\right]-\frac{\partial C_{NC}(e_C)}{\partial e_{NC}}\frac{\partial e_{NC}}{\partial e_C}-\frac{\partial C_C(e_C)}{\partial e_C}\right]\\
0 &= -\pi(e_C,e_{NC},n)+(1-\vartheta)\frac{\partial \pi}{\partial e_{NC}}\frac{\partial e_{NC}}{\partial \vartheta}+\phi\left[\pi(e_C,e_{NC},n)+\vartheta\frac{\partial \pi}{\partial e_{NC}}\frac{\partial e_{NC}}{\partial \vartheta}-\frac{\partial C_{NC}(e_C)}{\partial e_{NC}}\frac{\partial e_{NC}}{\partial \vartheta}\right]-\rho\\
0 &= -1+\phi+\eta,
\end{aligned}
$$

with the complementary slackness conditions

$$
\begin{aligned}
(\tau+LC)\,\eta &= 0 \qquad (51)\\
(\alpha-\vartheta)\,\rho &= 0 \qquad (52)
\end{aligned}
$$

and $\phi>0$. Note that managerial optimality implies that

$$
0=\vartheta\frac{\partial \pi}{\partial e_{NC}}-\frac{\partial C_{NC}(e_{NC})}{\partial e_{NC}},
$$

so that the optimality conditions read

$$
\begin{aligned}
0 &= (1-\vartheta)\left[\frac{\partial \pi}{\partial e_C}+\frac{\partial \pi}{\partial e_{NC}}\frac{\partial e_{NC}}{\partial e_C}\right]+\phi\left[\vartheta\frac{\partial \pi}{\partial e_C}-\frac{\partial C_C(e_C)}{\partial e_C}\right] \qquad (53)\\
0 &= -\pi(e_C,e_{NC},n)+(1-\vartheta)\frac{\partial \pi}{\partial e_{NC}}\frac{\partial e_{NC}}{\partial \vartheta}+\phi\pi(e_C,e_{NC},n)-\rho \qquad (54)\\
0 &= -1+\phi+\eta. \qquad (55)
\end{aligned}
$$

Hence, (53)-(55), (51), (52) and (32) are 6 equations, which we can solve for the 6 unknowns $(\tau, e_C, \vartheta, \eta, \phi, \rho)$.

It is useful to characterize everything in terms of the multiplier on the liquidity constraint η and the level of authority ϑ, which are at the heart of the contracting problem. Using (55) to substitute for ϕ and (30) to get

$$
\begin{aligned}
\frac{\partial \pi}{\partial e_{NC}}\frac{\partial e_{NC}}{\partial e_C} &= \frac{\partial ln(\pi)}{\partial ln(e_{NC})}\frac{\pi}{e_{NC}}\frac{\partial ln(e_{NC})}{\partial ln(e_C)}\frac{e_{NC}}{e_C}=(1-\mu)\sigma\frac{\pi}{e_{NC}}\frac{\mu\sigma}{1-(1-\mu)\sigma}\frac{e_{NC}}{e_C}\\
&= \frac{(1-\mu)\sigma}{1-(1-\mu)\sigma}\frac{\partial \pi}{\partial e_C},
\end{aligned}
$$

we can rewrite (53), the optimality condition for e_C, as

$$
\begin{aligned}
0 &= (1-\vartheta)\left[\frac{\partial \pi}{\partial e_C}+\frac{(1-\mu)\sigma}{1-(1-\mu)\sigma}\frac{\partial \pi}{\partial e_C}\right]+(1-\eta)\left[\vartheta\frac{\partial \pi}{\partial e_C}-\frac{\partial C_C(e_C)}{\partial e_C}\right]\\
&= \left[\frac{(1-\vartheta)}{1-(1-\mu)\sigma}+(1-\eta)\vartheta\right]\frac{\partial \pi}{\partial e_C}-(1-\eta)\frac{\partial C_C(e_C)}{\partial e_C},
\end{aligned}
$$

so that

$$
\frac{\partial C_C(e_C)}{\partial e_C} = \left[\vartheta+\frac{(1-\vartheta)}{(1-(1-\mu)\sigma)(1-\eta)}\right]\frac{\partial \pi}{\partial e_C} \qquad (56)
$$

Note that (56) uniquely determines $e_C = e_C(\vartheta,\eta)$. As the cost function and the profit function have constant elasticity, we get that

$$
\begin{aligned}
C_C &= \frac{\partial C_C(e_C)}{\partial e_C}e_C=\left[\vartheta+\frac{(1-\vartheta)}{(1-(1-\mu)\sigma)(1-\eta)}\right]\frac{\partial \pi}{\partial e_C}e_C\\
&= \left[\vartheta+\frac{(1-\vartheta)}{(1-(1-\mu)\sigma)(1-\eta)}\right]\sigma\mu\pi.
\end{aligned}
$$

Hence, the participation constraint implies that transfers are given by

$$\begin{aligned}\tau &= z(1-\sigma)\pi^{NM}Y(t) - (\vartheta\pi - C_{NC} - C_C) \\ &= z(1-\sigma)\pi^{NM}Y(t) - \left((1-(1-\mu)\sigma)\vartheta\pi - \left[\vartheta + \frac{(1-\vartheta)}{(1-(1-\mu)\sigma)(1-\eta)}\right]\sigma\mu\pi\right) \\ &= z(1-\sigma)\pi^{NM}Y(t) - \left((1-\sigma)\vartheta - \left(\frac{(1-\vartheta)\sigma\mu}{(1-(1-\mu)\sigma)(1-\eta)}\right)\right)\pi, \end{aligned} \quad (57)$$

so that the value of the firm is given by

$$\begin{aligned}V^M(n) &= (1-\vartheta)\pi - \tau \\ &= \left[(1-\sigma)\vartheta + \left[1 - \frac{\sigma\mu}{(1-(1-\mu)\sigma)(1-\eta)}\right](1-\vartheta)\right]\pi - z\pi^{NM}Y(t). \end{aligned} \quad (58)$$

Note that $\pi = \pi(e_C, e_{NC})$ is fully determined given ϑ and η. Hence, (58) determines the value of the firm given (ϑ, η). It is also clear from above that we can solve for ρ from (52).

Suppose first that the liquidity constraint is indeed slack. From (54) the optimal level of authority solves

$$\rho = (1-\vartheta)\frac{\partial\pi}{\partial e_{NC}}\frac{\partial e_{NC}}{\partial\vartheta} > 0.$$

Hence, (52) implies that $\vartheta = 1$. (56) then implies that $\frac{\partial C_C(e_C)}{\partial e_C} = \frac{\partial\pi}{\partial e_C}$, i.e. the marginal returns of contractual investments are equal to the marginal costs. Hence, using (30), we get

$$\begin{aligned}\frac{\partial C_C(e_C)}{\partial e_C} = \mu v_L Q(t)n &= \sigma\mu\frac{\pi}{e_C} = \sigma\mu\frac{\frac{\gamma-1}{\gamma}\left(e_C^{\mu}e_{NC}^{1-\mu}\right)^{\sigma}Y(t)n}{e_C} \\ &= \sigma\mu\frac{\gamma-1}{\gamma}\frac{e_C^{\mu\sigma}\left[\frac{\sigma}{v_L}\frac{\gamma-1}{\gamma}\frac{Y(t)}{Q(t)}e_C^{\mu\sigma}\right]^{\frac{(1-\mu)\sigma}{1-(1-\mu)\sigma}}Y(t)n}{e_C} \\ &= \nu_L\mu Q(t)n\frac{\left[\frac{\sigma}{v_L}\frac{\gamma-1}{\gamma}\frac{Y(t)}{Q(t)}e_C^{\mu\sigma}\right]^{\frac{1}{1-(1-\mu)\sigma}}}{e_C}.\end{aligned}$$

Rearranging terms, yields

$$e_C = e_{NC} = e^{FB} = \left(\frac{\sigma}{v_L}\frac{Y(t)}{Q(t)}\frac{\gamma-1}{\gamma}\right)^{\frac{1}{1-\sigma}},$$

so that

$$\begin{aligned}\pi &= \pi^{FB} = \frac{\gamma-1}{\gamma}\left(e^{FB}\right)^{\sigma}Y(t)n \\ &= \frac{\gamma-1}{\gamma}\left(e_{NM}(1)\left(\frac{v_H}{v_L}\right)^{\frac{1}{1-\sigma}}\right)^{\sigma}Y(t)n \\ &= \left(\frac{v_H}{v_L}\right)^{\frac{\sigma}{1-\sigma}}\frac{\gamma-1}{\gamma}\left(e_{NM}(1)\right)^{\sigma}Y(t)n \\ &= \left(\frac{v_H}{v_L}\right)^{\frac{\sigma}{1-\sigma}}\pi^{NM}Y(t)n.\end{aligned}$$

The value of the firms follows directly from (58). This solution is the correct one, as long as the liquidity constraint is indeed slack. This proves Proposition 1.

From (57) we get that

$$\begin{aligned}\tau &= zV^{NM}(1)-(1-\sigma)\pi(n)\\ &= z(1-\sigma)\pi^{NM}Y(t)-(1-\sigma)\left(\frac{\nu_H}{\nu_L}\right)^{\frac{\sigma}{1-\sigma}}\pi^{NM}Y(t)n\end{aligned}$$

Hence, this solution is feasible as long as $\tau > -\chi Y(t)$, i.e.

$$\begin{aligned}z(1-\sigma)\pi^{NM}Y(t)-(1-\sigma)\left(\frac{\nu_H}{\nu_L}\right)^{\frac{\sigma}{1-\sigma}}\pi^{NM}Y(t)n &\geq -\chi Y(t)\\ z-\left(\frac{\nu_H}{\nu_L}\right)^{\frac{\sigma}{1-\sigma}}n &\geq -\frac{\chi}{(1-\sigma)\pi^{NM}},\end{aligned}$$

which yields

$$n\leq\frac{1}{\left(\frac{\nu_H}{\nu_L}\right)^{\frac{\sigma}{1-\sigma}}}\left(z+\frac{\chi}{(1-\sigma)\pi^{NM}}\right)=n^*.$$

For all $n>n^*$ the liquidity constraint is binding. From (57) we now have

$$\tau = z(1-\sigma)\pi^{NM}Y(t)-\left((1-\sigma)\vartheta-\left[\frac{(1-\vartheta)\sigma\mu}{(1-(1-\mu)\sigma)(1-\eta)}\right]\right)\pi=-\chi Y(t).$$

Rearranging terms yields

$$\begin{aligned}z(1-\sigma)\pi^{NM}Y(t)+\chi Y(t) &= \left((1-\sigma)\vartheta-\left[\frac{(1-\vartheta)\sigma\mu}{(1-(1-\mu)\sigma)(1-\eta)}\right]\right)\pi\\ \left(z+\frac{\chi}{(1-\sigma)\pi^{NM}}\right) &= \frac{\left((1-\sigma)\vartheta-\left[\frac{(1-\vartheta)\sigma\mu}{(1-(1-\mu)\sigma)(1-\eta)}\right]\right)}{1-\sigma}\frac{\pi}{Y(t)\pi^{NM}},\end{aligned}$$

so that

$$n^* = \frac{1}{\left(\frac{\nu_H}{\nu_L}\right)^{\frac{\sigma}{1-\sigma}}}\left(z+\frac{\chi}{(1-\sigma)\pi^{NM}}\right)=\frac{\left((1-\sigma)\vartheta-\left[\frac{(1-\vartheta)\sigma\mu}{(1-(1-\mu)\sigma)(1-\eta)}\right]\right)}{1-\sigma}\frac{\pi}{\left(\frac{\nu_H}{\nu_L}\right)^{\frac{\sigma}{1-\sigma}}Y(t)\pi^{NM}} \quad (59)$$

$$= \frac{\left((1-\sigma)\vartheta-\left[\frac{(1-\vartheta)\sigma\mu}{(1-(1-\mu)\sigma)(1-\eta)}\right]\right)}{1-\sigma}\frac{\frac{\gamma-1}{\gamma}\left(e_C^{\mu}e_{NC}^{1-\mu}\right)^{\sigma}Y(t)n}{\left(\frac{\nu_H}{\nu_L}\right)^{\frac{\sigma}{1-\sigma}}Y(t)\frac{\gamma-1}{\gamma}e_{NM}^{\sigma}} \quad (60)$$

$$\frac{n^*}{n} = \frac{\left((1-\sigma)\vartheta-\left[\frac{(1-\vartheta)\sigma\mu}{(1-(1-\mu)\sigma)(1-\eta)}\right]\right)}{1-\sigma}\left(\frac{1}{\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1}{1-\sigma}}}\left(\frac{e_C}{e_{NM}}\right)^{\mu}\left(\frac{e_{NC}}{e_{NM}}\right)^{1-\mu}\right)^{\sigma}. \quad (61)$$

We also note that $\vartheta<1$ by construction. Hence, $\rho=0$ so that the optimal level of authority is given by (54)

$$0=-\pi(e_C,e_{NC},n)+(1-\vartheta)\frac{\partial\pi}{\partial e_{NC}}\frac{\partial e_{NC}}{\partial\vartheta}+(1-\eta)\pi(e_C,e_{NC},n).$$

Using

$$\begin{aligned}\frac{\partial \pi}{\partial e_{NC}}\frac{\partial e_{NC}}{\partial \vartheta} &= \frac{\partial ln\left(\pi\right)}{\partial ln\left(e_{NC}\right)}\frac{\pi}{e_{NC}}\frac{\partial ln\left(e_{NC}\right)}{\partial ln\left(\vartheta\right)}\frac{e_{NC}}{\vartheta} \\ &= \frac{\left(1-\mu\right)\sigma}{1-\left(1-\mu\right)\sigma}\frac{\pi}{\vartheta},\end{aligned}$$

we get

$$0=\left[\frac{\left(1-\vartheta\right)}{\vartheta}\frac{\left(1-\mu\right)\sigma}{1-\left(1-\mu\right)\sigma}-\eta\right]\pi\left(e_C,e_{NC},n\right).$$

Hence,

$$\eta=\frac{\left(1-\vartheta\right)}{\vartheta}\frac{\left(1-\mu\right)\sigma}{1-\left(1-\mu\right)\sigma}. \tag{62}$$

Note that (62) defines a schedule $\vartheta=\vartheta\left(\eta\right)$ with $\vartheta'\left(\eta\right)<0$ and

$$\left(1-\mu\right)\sigma\leq\vartheta\left(\eta\right)\leq 1.$$

The optimal level of contractual effort solves (56)

$$\frac{\partial C_C\left(e_C\right)}{\partial e_C}=\left[\vartheta+\frac{\left(1-\vartheta\right)}{\left(1-\left(1-\mu\right)\sigma\right)\left(1-\eta\right)}\right]\frac{\partial \pi}{\partial e_C}.$$

Using (62) we get that

$$\begin{aligned}\vartheta+\frac{\left(1-\vartheta\right)}{\left(1-\left(1-\mu\right)\sigma\right)\left(1-\eta\right)} &= \vartheta\left(1+\frac{\left(1-\vartheta\right)}{\left(\vartheta-\left(1-\mu\right)\sigma\right)}\right) \\ &= \vartheta\left(\frac{1-\left(1-\mu\right)\sigma}{\vartheta-\left(1-\mu\right)\sigma}\right) \\ &\equiv \gamma\left(\vartheta,\mu\right).\end{aligned} \tag{63}$$

Hence,

$$\mu v_L Q\left(t\right)n=\gamma\left(\vartheta,\eta\right)\sigma\mu\frac{\gamma-1}{\gamma}\left(e_C^{\mu}e_{NC}^{1-\mu}\right)^{\sigma}\frac{Y\left(t\right)n}{e_C}.$$

From (30), we get

$$e_{NC}^{(1-\mu)\sigma}=\left[\vartheta\frac{\sigma}{\nu_L}\frac{Y\left(t\right)}{Q\left(t\right)}\frac{\gamma-1}{\gamma}e_C^{\mu\sigma}\right]^{\frac{(1-\mu)\sigma}{1-(1-\mu)\sigma}},$$

so that

$$\begin{aligned}1 &= \gamma\left(\vartheta,\eta\right)\frac{\sigma}{\nu_L}\frac{Y\left(t\right)}{Q\left(t\right)}\frac{\gamma-1}{\gamma}e_C^{\mu\sigma}\left[\vartheta\frac{\sigma}{\nu_L}\frac{Y\left(t\right)}{Q\left(t\right)}\frac{\gamma-1}{\gamma}e_C^{\mu\sigma}\right]^{\frac{(1-\mu)\sigma}{1-(1-\mu)\sigma}}\frac{1}{e_C} \\ &= \gamma\left(\vartheta,\eta\right)\vartheta^{\frac{(1-\mu)\sigma}{1-(1-\mu)\sigma}}\left[\frac{\sigma}{\nu_L}\frac{\gamma-1}{\gamma}\frac{Y\left(t\right)}{Q\left(t\right)}e_C^{\mu\sigma}\right]^{\frac{1}{1-(1-\mu)\sigma}}\frac{1}{e_C}.\end{aligned}$$

Hence,

$$\begin{aligned}e_C^{1-(1-\mu)\sigma} &= \gamma\left(\vartheta,\eta\right)^{1-(1-\mu)\sigma}\vartheta^{(1-\mu)\sigma}\frac{\sigma}{\nu_L}\frac{\gamma-1}{\gamma}\frac{Y\left(t\right)}{Q\left(t\right)}e_C^{\mu\sigma} \\ e_C^{1-\sigma} &= \gamma\left(\vartheta,\eta\right)^{1-(1-\mu)\sigma}\vartheta^{(1-\mu)\sigma}\frac{\sigma}{\nu_L}\frac{\gamma-1}{\gamma}\frac{Y\left(t\right)}{Q\left(t\right)} \\ &= \gamma\left(\vartheta,\eta\right)^{1-(1-\mu)\sigma}\vartheta^{(1-\mu)\sigma}\frac{\nu_H}{\nu_L}\frac{\sigma}{\nu_H}\frac{\gamma-1}{\gamma}\frac{Y\left(t\right)}{Q\left(t\right)} \\ &= \gamma\left(\vartheta,\eta\right)^{1-(1-\mu)\sigma}\vartheta^{(1-\mu)\sigma}\frac{\nu_H}{\nu_L}e_{NM}^{1-\sigma},\end{aligned}$$

so that

$$\left(\frac{e_C}{e_{NM}}\right)^{1-\sigma} = \gamma(\vartheta,\eta)^{1-(1-\mu)\sigma}\,\vartheta^{(1-\mu)\sigma}\frac{\nu_H}{\nu_L}. \tag{64}$$

Similarly,

$$\begin{aligned}
\frac{e_{NC}}{e_{NM}} &= \left[\vartheta\frac{\sigma}{\nu_L}\frac{\gamma-1}{\gamma}\frac{Y(t)}{Q(t)}e_C^{\mu\sigma}\right]^{\frac{1}{1-(1-\mu)\sigma}}\frac{1}{e_{NM}} \\
&= [\vartheta]^{\frac{1}{1-(1-\mu)\sigma}}\left[\frac{\nu_H}{\nu_L}(e_{NM})^{1-\sigma}e_C^{\mu\sigma}\right]^{\frac{1}{1-(1-\mu)\sigma}}\frac{1}{e_{NM}} \\
&= [\vartheta]^{\frac{1}{1-(1-\mu)\sigma}}\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1}{1-(1-\mu)\sigma}}(e_{NM})^{\frac{-\mu\sigma}{1-(1-\mu)\sigma}}[e_C]^{\frac{\mu\sigma}{1-(1-\mu)\sigma}} \\
&= [\vartheta]^{\frac{1}{1-(1-\mu)\sigma}}\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1}{1-(1-\mu)\sigma}}\left(\frac{e_C}{e^{NM}}\right)^{\frac{\mu\sigma}{1-(1-\mu)\sigma}}.
\end{aligned} \tag{65}$$

Hence, we can solve for ϑ from (59) using (62) and the expressions for $\frac{e_{NC}}{e^{FB}}$ and $\frac{e_C}{e^{FB}}$. In particular, total efficiency (relative to the first best) is

$$\begin{aligned}
\frac{e_C^{\mu}e_{NC}^{1-\mu}}{e_{NM}} &= \left(\frac{e_C}{e_{NM}}\right)^{\mu}\left([\vartheta]^{\frac{1}{1-(1-\mu)\sigma}}\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1}{1-(1-\mu)\sigma}}\left(\frac{e_C}{e_{NM}}\right)^{\frac{\mu\sigma}{1-(1-\mu)\sigma}}\right)^{1-\mu} \\
&= [\vartheta]^{\frac{1-\mu}{1-(1-\mu)\sigma}}\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1-\mu}{1-(1-\mu)\sigma}}\left(\frac{e_C}{e_{NM}}\right)^{\mu}\left(\frac{e_C}{e_{NM}}\right)^{\frac{\mu\sigma(1-\mu)}{1-(1-\mu)\sigma}} \\
&= [\vartheta]^{\frac{1-\mu}{1-(1-\mu)\sigma}}\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1-\mu}{1-(1-\mu)\sigma}}\left(\frac{e_C}{e_{NM}}\right)^{\frac{\mu}{1-(1-\mu)\sigma}} \\
&= [\vartheta]^{\frac{1-\mu}{1-(1-\mu)\sigma}}\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1-\mu}{1-(1-\mu)\sigma}}\left(\gamma(\vartheta,\eta)^{\frac{1-(1-\mu)\sigma}{1-\sigma}}\vartheta^{\frac{(1-\mu)\sigma}{1-\sigma}}\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1}{1-\sigma}}\right)^{\frac{\mu}{1-(1-\mu)\sigma}} \\
&= \gamma(\vartheta,\eta)^{\frac{\mu}{1-\sigma}}\,\vartheta^{\frac{1-\mu}{1-(1-\mu)\sigma}\left(1+\frac{\sigma\mu}{1-\sigma}\right)}\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1}{1-(1-\mu)\sigma}\left(1-\mu+\frac{\mu}{1-\sigma}\right)} \\
&= \gamma(\vartheta,\eta)^{\frac{\mu}{1-\sigma}}\,\vartheta^{\frac{1-\mu}{1-\sigma}}\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1}{1-\sigma}}
\end{aligned}$$

Using (63) we get

$$\frac{1}{\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1}{1-\sigma}}}\frac{e_C^{\mu}e_{NC}^{1-\mu}}{e^{FB}} = \left(\left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{\mu}\vartheta\right)^{\frac{1}{1-\sigma}} \tag{66}$$

Hence,

$$\begin{aligned}
\frac{n^*}{n} &= \frac{\left((1-\sigma)\vartheta-\left[\frac{(1-\vartheta)\sigma\mu}{(1-(1-\mu)\sigma)(1-\eta)}\right]\right)}{1-\sigma}\left(\frac{1}{\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1}{1-\sigma}}}\left(\frac{e_C}{e_{NM}}\right)^{\mu}\left(\frac{e_{NC}}{e_{NM}}\right)^{1-\mu}\right)^{\sigma} \\
&= \frac{1}{(1-\sigma)}\left((1-\sigma)\vartheta-\left[\frac{(1-\vartheta)\sigma\mu}{(1-(1-\mu)\sigma)(1-\eta)}\right]\right)\left(\left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{\mu}\vartheta\right)^{\frac{\sigma}{1-\sigma}}.
\end{aligned}$$

Now note that[17]

$$
\begin{aligned}
& (1-\sigma)\vartheta - \left[\frac{(1-\vartheta)\sigma\mu}{(1-(1-\mu)\sigma)(1-\eta)}\right] \\
= & \ (1-\sigma)\vartheta - \frac{(1-\vartheta)\sigma\mu\vartheta}{\vartheta-(1-\mu)\sigma} \\
= & \left(\frac{(1-\sigma)(\vartheta-(1-\mu)\sigma)-(1-\vartheta)\sigma\mu}{\vartheta-(1-\mu)\sigma}\right)\vartheta \\
= & \ \frac{(\vartheta-\sigma)(1-(1-\mu)\sigma)}{\vartheta-(1-\mu)\sigma}\vartheta,
\end{aligned}
$$

so that

$$
\begin{aligned}
\frac{n^*}{n} &= \frac{\vartheta-\sigma}{1-\sigma}\left(\frac{(1-(1-\mu)\sigma)}{\vartheta-(1-\mu)\sigma}\vartheta\right)\left(\left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{\mu}\vartheta\right)^{\frac{\sigma}{1-\sigma}} \\
&= \frac{\vartheta-\sigma}{1-\sigma}\left[\left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{1-\sigma(1-\mu)}\vartheta\right]^{\frac{1}{1-\sigma}}. \qquad (67)
\end{aligned}
$$

This equation determines

$$
\vartheta = \vartheta\left(\frac{n}{n^*},\mu\right).
$$

The function $\vartheta = \vartheta\left(\frac{n}{n^*},\mu\right)$ is decreasing in $\frac{n}{n^*}$.

Proof. We need to show that the RHS of (67) is increasing in ϑ for $\sigma < \vartheta < 1$. Let

$$
h(\vartheta) \equiv ln\left(\frac{\vartheta-\sigma}{1-\sigma}\left[\left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{1-\sigma(1-\mu)}\vartheta\right]^{\frac{1}{1-\sigma}}\right).
$$

Then

$$
\begin{aligned}
h'(\vartheta) &= \frac{1}{\vartheta-\sigma} + \frac{1}{1-\sigma}\frac{1}{\vartheta} - \frac{1-\sigma(1-\mu)}{1-\sigma}\frac{1}{\vartheta-(1-\mu)\sigma} \\
&= \frac{1}{1-\sigma}\left(\frac{1-\sigma}{\vartheta-\sigma} + \frac{1}{\vartheta} - \frac{1-\sigma(1-\mu)}{\vartheta-(1-\mu)\sigma}\right) \\
&> \frac{1}{1-\sigma}\left(\frac{1-\sigma+\mu\sigma}{\vartheta-\sigma+\mu\sigma} + \frac{1}{\vartheta} - \frac{1-\sigma(1-\mu)}{\vartheta-(1-\mu)\sigma}\right) \\
&= \frac{1}{1-\sigma}\frac{1}{\vartheta},
\end{aligned}
$$

where the inequality follows from $\frac{\partial\left(\frac{1-\sigma+\Delta}{\vartheta-\sigma+\Delta}\right)}{\partial\Delta} < 0$ so that $\frac{1-\sigma+\mu\sigma}{\vartheta-\sigma+\mu\sigma} < \frac{1-\sigma}{\vartheta-\sigma}$. □

[17] Because,

$$
\begin{aligned}
& (1-\sigma)(\vartheta-(1-\mu)\sigma)-(1-\vartheta)\sigma\mu \\
= & \ (1-\sigma)(1-(1-\mu)\sigma-(1-\vartheta))-(1-\vartheta)\sigma\mu \\
= & \ (1-\sigma)(1-(1-\mu)\sigma)-(1-\sigma)(1-\vartheta)-(1-\vartheta)\sigma\mu \\
= & \ (1-\sigma)(1-(1-\mu)\sigma)-(1-\vartheta)(1-\sigma(1-\mu)) \\
= & \ (\vartheta-\sigma)(1-(1-\mu)\sigma),
\end{aligned}
$$

Now we can solve for the actual effort levels and profits. From (64) and (65) we get

$$\left(\frac{e_C}{e_{NM}}\right)^{1-\sigma} = \frac{\nu_H}{\nu_L}\gamma(\vartheta,\eta)^{1-(1-\mu)\sigma}\vartheta^{(1-\mu)\sigma} = \frac{\nu_H}{\nu_L}\vartheta\left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{1-(1-\mu)\sigma}.$$

Lemma 2. *We have* $e_C \geq \left(\frac{\nu_H}{\nu_L}\right)^{\frac{1}{1-\sigma}} e_{NM}$.

Proof. We need to show that $\vartheta\left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{1-(1-\mu)\sigma} \geq 1$. Define $h(\vartheta) = ln\left(\vartheta\left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{1-(1-\mu)\sigma}\right)$. As $h(1)=0$, it is sufficient to show that $h(.)$ is decreasing in ϑ for all $\vartheta<1$. But

$$h'(\vartheta) = \frac{1}{\vartheta} - \frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma} = \frac{\vartheta-(1-\mu)\sigma-\vartheta+\vartheta(1-\mu)\sigma}{\vartheta(\vartheta-(1-\mu)\sigma)} = -\frac{(1-\vartheta)(1-\mu)\sigma}{\vartheta(\vartheta-(1-\mu)\sigma)} < 0.$$

□

Similarly,

$$\begin{aligned}
\left(\frac{e_{NC}}{e^{FB}}\right)^{1-\sigma} &= [\vartheta]^{\frac{1-\sigma}{1-(1-\mu)\sigma}}\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1-\sigma}{1-(1-\mu)\sigma}}\left(\left(\frac{e_C}{e^{NM}}\right)^{1-\sigma}\right)^{\frac{\mu\sigma}{1-(1-\mu)\sigma}} \\
&= [\vartheta]^{\frac{1-\sigma}{1-(1-\mu)\sigma}}\left(\frac{\nu_H}{\nu_L}\right)^{\frac{1-\sigma}{1-(1-\mu)\sigma}}\left(\frac{\nu_H}{\nu_L}\vartheta\left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{1-(1-\mu)\sigma}\right)^{\frac{\mu\sigma}{1-(1-\mu)\sigma}} \\
&= \frac{\nu_H}{\nu_L}\vartheta\left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{\mu\sigma}.
\end{aligned}$$

Hence,

$$\left(\frac{e_{NC}}{e_C}\right)^{1-\sigma} = \left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{\mu\sigma-1+(1-\mu)\sigma} = \left(\frac{\vartheta-(1-\mu)\sigma}{1-(1-\mu)\sigma}\right)^{1-\sigma}. \tag{68}$$

Lemma 3. *We have* $e_{NC} < e_C$.

Proof. Obvious from (68). □

Lemma 4. *We have* $e_{NC} < e_{FB}$.

Proof. Note that

$$\begin{aligned}
\frac{e_{NC}}{e^{PC}} &= \frac{e_{NC}}{e_C}\frac{e_C}{e^{PC}} \\
&= \frac{\vartheta(n,\mu)-(1-\mu)\sigma}{1-(1-\mu)\sigma}\vartheta(n,\mu)^{\frac{1}{1-\sigma}}\left(\frac{1-(1-\mu)\sigma}{\vartheta(n,\mu)-(1-\mu)\sigma}\right)^{\frac{1-(1-\mu)\sigma}{1-\sigma}} \\
&= \vartheta(n,\mu)^{\frac{1}{1-\sigma}}\left(\frac{1-(1-\mu)\sigma}{\vartheta(n,\mu)-(1-\mu)\sigma}\right)^{\frac{1-(1-\mu)\sigma-1+\sigma}{1-\sigma}} \\
&= \left[\vartheta(n,\mu)\left(\frac{1-(1-\mu)\sigma}{\vartheta(n,\mu)-(1-\mu)\sigma}\right)^{\mu\sigma}\right]^{\frac{1}{1-\sigma}}.
\end{aligned}$$

Note that the term in brackets is increasing as

$$\begin{aligned}\frac{\partial ln\left[\vartheta\left(n,\mu\right)\left(\frac{1}{\vartheta(n,\mu)-(1-\mu)\sigma}\right)^{\mu\sigma}\right]}{\partial\vartheta} &= \frac{1}{\vartheta}-\frac{\mu\sigma}{\vartheta-\left(1-\mu\right)\sigma}\\ &= \frac{\vartheta-\left(1-\mu\right)\sigma-\mu\sigma\vartheta}{\vartheta\left(\vartheta-\left(1-\mu\right)\sigma\right)}\\ &= \frac{\vartheta-\sigma+\sigma\mu\left(1-\vartheta\right)}{\vartheta\left(\vartheta-\left(1-\mu\right)\sigma\right)}.\end{aligned}$$

Hence,

$$\vartheta\left(n,\mu\right)\left(\frac{1-\left(1-\mu\right)\sigma}{\vartheta\left(n,\mu\right)-\left(1-\mu\right)\sigma}\right)^{\mu\sigma}\leq 1\left(\frac{1-\left(1-\mu\right)\sigma}{1-\left(1-\mu\right)\sigma}\right)^{\mu\sigma}=1.$$

□

Total profits are then given by

$$\begin{aligned}\pi\left(n\right) &= \frac{\gamma-1}{\gamma}\left(e_{NC}^{1-\mu}e_{C}^{\mu}\right)^{\sigma}Q\left(t\right)n\\ &= \left(\frac{e_{NC}^{1-\mu}e_{C}^{\mu}}{e_{FB}}\right)^{\sigma}\pi^{FB}\left(n\right)\\ &= \left(\left(\frac{1-\left(1-\mu\right)\sigma}{\vartheta-\left(1-\mu\right)\sigma}\right)^{\mu}\vartheta\right)^{\frac{\sigma}{1-\sigma}}\pi^{FB}\left(n\right). \qquad (69)\end{aligned}$$

where

Lemma 5. *We have* $\pi\left(n\right)<\pi^{FB}$.

Proof. It is easy to that $\frac{\partial ln\left(\left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{\mu}\vartheta\right)}{\partial\vartheta}=\frac{(1-\mu)(\vartheta-\sigma)}{\vartheta(\vartheta-(1-\mu)\sigma)}>0$. Hence, $1=\arg\max_{\vartheta}\left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{\mu}\vartheta$ so that $\left(\frac{1-(1-\mu)\sigma}{\vartheta-(1-\mu)\sigma}\right)^{\mu}\vartheta\leq 1$. □

Finally: need to characterize $V\left(n\right)$. From (58) and (69) we get that

$$\begin{aligned}V^{F}\left(n\right) &= \left[\left(1-\sigma\right)\vartheta+\left[1-\frac{\sigma\mu}{\left(1-\left(1-\mu\right)\sigma\right)\left(1-\eta\right)}\right]\left(1-\vartheta\right)\right]\pi-\kappa Q\left(t\right)\\ &= \left[\left(1-\sigma\right)\vartheta+\left[1-\frac{\sigma\mu}{\left(1-\left(1-\mu\right)\sigma\right)\left(1-\eta\right)}\right]\left(1-\vartheta\right)\right]\left(\left(\frac{1-\left(1-\mu\right)\sigma}{\vartheta-\left(1-\mu\right)\sigma}\right)^{\mu}\vartheta\right)^{\frac{\sigma}{1-\sigma}}\pi^{FB}\left(n\right)-\kappa Q\left(t\right).\end{aligned}$$

From (62) we have that η and ϑ are linked via

$$\eta=\frac{\left(1-\vartheta\right)}{\vartheta}\frac{\left(1-\mu\right)\sigma}{1-\left(1-\mu\right)\sigma}.$$

Substituting this, we get that[18]

$$\begin{aligned}V^{F}\left(n\right) &= \left[\left(1-\sigma\right)\vartheta+\left[1-\frac{\sigma\mu}{1-\frac{1}{\vartheta}\left(1-\mu\right)\sigma}\right]\left(1-\vartheta\right)\right]\left(\left(\frac{1-\left(1-\mu\right)\sigma}{\vartheta-\left(1-\mu\right)\sigma}\right)^{\mu}\vartheta\right)^{\frac{\sigma}{1-\sigma}}\pi^{FB}\left(n\right)-\kappa Q\left(t\right)\\ &= \left[1-\sigma\vartheta\left(1+\frac{\mu\left(1-\vartheta\right)}{\vartheta-\left(1-\mu\right)\sigma}\right)\right]\left(\left(\frac{1-\left(1-\mu\right)\sigma}{\vartheta-\left(1-\mu\right)\sigma}\right)^{\mu}\vartheta\right)^{\frac{\sigma}{1-\sigma}}\pi^{FB}\left(n\right)-\kappa Q\left(t\right).\end{aligned}$$

[18]Note that

$$1-\eta=1-\frac{\left(1-\vartheta\right)}{\vartheta}\frac{\left(1-\mu\right)\sigma}{1-\left(1-\mu\right)\sigma}=\frac{1-\frac{1}{\vartheta}\left(1-\mu\right)\sigma}{1-\left(1-\mu\right)\sigma}.$$

Note first that $\mu = 1$ implies that

$$V^F(n)\big|_{\mu=1} = (1-\sigma)\,\pi^{FB}(n) - \kappa Q(t),$$

which is the first-best outcome.

6.5 Proof of Proposition 3

To show that $\vartheta^C(n,\mu)$ is increasing in μ, define

$$h(\vartheta,\mu) = \left(\frac{1-(1-\mu)\,\sigma}{\vartheta^C(n,\mu)-(1-\mu)\,\sigma}\right)^{1-\sigma(1-\mu)}\vartheta^C(n,\mu).$$

As $\partial_\vartheta h > 0$, $\vartheta^C(n,\mu)$ is increasing in μ if $\partial_\mu h < 0$. Let us define $r = 1-\sigma(1-\mu)$. We then need to show that

$$\frac{\partial}{\partial r}\left[ln\left(\frac{r}{\vartheta^C(n,\mu)-1+r}\right)^r\right] = \frac{\partial r\,[ln(r) - ln(r-(1-\vartheta))]}{\partial r} < 0.$$

But

$$\begin{aligned}
\frac{\partial r\,[ln(r) - ln(r-(1-\vartheta))]}{\partial r} &= [ln(r) - ln(r-(1-\vartheta))] + r\left(\frac{1}{r} - \frac{1}{r-(1-\vartheta)}\right) \\
&= -\left[\left(\frac{r}{r-(1-\vartheta)} - 1\right) - ln\left(\frac{r}{r-(1-\vartheta)}\right)\right] \\
&< 0,
\end{aligned}$$

as $\frac{r}{r-(1-\vartheta)} = x > 1$, and $x - 1 > ln(x)$ for $x > 1$.