Version: December 2014 (revised)

Why Doesn't Technology Flow from Rich to Poor Countries?*

Harold L. Cole, Jeremy Greenwood, and Juan M. Sanchez

Abstract

What determines the technology that a country adopts? While many factors affect technological adoption, the efficiency of the country's financial system may also play a significant role. To address this question, a dynamic contract model is embedded into a general equilibrium setting with competitive intermediation. The ability of an intermediary to monitor and control the cash flows of a firm plays an important role in the technology adoption decision. Can such a theory help to explain the differences in total factor productivity and establishment-size distributions across India, Mexico, and the United States? A quantitative illustration suggests the answer is yes.

Keywords: Costly cash-flow control; costly state verification; dynamic contract theory; economic development; establishment-size distributions; finance and development; financial intermediation; India, Mexico, and the United States; monitoring; productivity; self-finance; technology adoption; ventures

Affliations: University of Pennsylvania and Federal Reserve Bank of St. Louis

*Address correspondence to Hal Cole at colehl⊗sas.upenn.edu.

1 Introduction

Why do countries use different production technologies? Surely, all nations should adopt best-practice technologies, which produce the highest levels of income. Yet, this does not happen. To paraphrase Lucas (1990): Why doesn't technology flow from rich to poor countries? The premise here is that the efficiency of financial markets plays a vital role in technology adoption. If a country's financial markets affect its technology adoption, then it is a small step to argue that they will affect the nation's total factor productivity (TFP) and income.

Investing in new technologies is risky. Advanced technologies require substantial funding and the payoff from any investment is uncertain. Compounding the problem is the fact that investors in a venture have more limited information than the developers of the venture. Therefore, there is scope to misappropriate funds due to differences in information. In some countries, it may even be difficult to control the use of publicly acknowledged funds due to poor rule of law for contractual disputes.

Financial institutions play an important role in constructing mechanisms that ensure investments are used wisely. They do this by both monitoring firms and implementing reward structures that encourage firms to truthfully reveal their profits so that investors can be fairly compensated. Monitoring firms is expensive. Efficient monitoring requires sound accounting systems, the timely release of information, the presence of financial analysts, and the like. Monitoring cannot be done effectively in some places. In such circumstances, intermediaries must rely primarily on incentive schemes to ensure honesty. This restricts the profitability of certain types of investment projects. The design of incentive schemes may be severely circumscribed, however. It may not be possible for an intermediary to exert the desirable level of control over a firm's publicly acknowledged revenue streams. In some places contractual disputes may be complicated, costly, and require lengthy settlement times. These factors circumscribe what can be effectively delimited in a contract. As a result, it may be infeasible to write a contract that can impose the reward structure required to ensure the likelihood of a successful investment.

1.1 Finance and Development: A Brief Literature Review

Earlier work has drawn a connection between finance and the adoption of technologies. For example, Greenwood and Jovanovic (1990) allow for two technologies: a primitive one with a low, certain rate of return and an advanced one, with a higher expected but uncertain rate of return. By pooling risks intermediaries reduce the vagaries associated with the advanced technology. There are fixed costs associated with intermediation, so only the wealthy choose to use this channel. Banerjee and Duflo (2005) present a stylized model where more advanced technologies require larger investments in terms of fixed costs. Given the presence of borrowing constraints, countries such as India lack the wherewithal to finance advanced technologies. They suggest this as a potential explanation for the productivity gap between India and the United States.

Within the context of a two-sector model where technologies may differ, Buera, Kaboski, and Shin (2011) quantitatively examine the link between financial development and economic development. They emphasize the importance of borrowing and enforcement constraints. Greenwood, Sanchez, and Wang (2010, 2013) allow for an infinite number of technologies. Better intermediation prunes the ones with low returns from the economy. In all of these papers, technologies differ in a simple way. The prototypical setting is similar to Greenwood and Jovanovic (1990): Better technologies have higher expected levels of productivity, are riskier, and usually involve a higher fixed cost in terms of adoption.

The decision to finance a venture is likely to depend on the nature of the technology in a more deep-rooted manner. Selling drinks on the street is much different than launching rockets into space. The former requires a small investment that yields returns relatively quickly and with little risk. The latter requires years of funding before any returns are realized and there is tremendous risk associated with financing such ventures. To capture this notion, technologies are given a richer representation than is conventionally assumed. Each technology has its own nonstationary stochastic process for total factor productivity (TFP) over its life cycle. The investments involved at each event in the life cycle may differ too. A long-term project is likely to require a long-term relationship between a firm and an

intermediary.[1] Thus, a dynamic contracting perspective is taken.

Why is this important? First, the structure of the technology adopted and the effects of financial structure are likely to be inextricably linked. Consider a model where entrepreneurs are constrained by some initial level of wealth and can borrow only a fixed limited amount per period on short-term markets. Intuitively, one would expect a firm to be much more capable of self-financing a project over time if the profile for TFP is flat, implying a flat profile of capital, as opposed to one where productivity perpetually grows in a convex manner requiring ever-increasing levels of investments. Midrigan and Xu (2014) argue that with stationary AR1 productivity shocks the capital required by a firm can be accumulated reasonably quickly by self-financing [see also Moll (2014) for an analysis of how the degree of persistence in technology shocks and the ability to self-finance interact]. Midrigan and Xu (2014) conclude that the impact of intermediation on technology adoption is as important as its impact on capital accumulation for explaining cross-country TFP differences.

Suppose alternatively that the entrepreneur has limited funding of his own but can enter into a long-term financial contract with an intermediary. Long-term contracts offer, relative to short-term ones, the potential to obtain efficient levels of capital quickly. This can be done by using backloading strategies, where the rewards to owners of firms are delayed until the desired outcomes are obtained. In fact, when productivity shocks are independently and identically distributed over time, contracts can be designed such that the deviations from first-best allocations are relatively small, as noted in Marcet and Marimon (1992). Here it is shown that this is no longer the case when the return structure offered from a technology is generalized. Now, it may be impossible to write contracts that allow for certain technologies to be funded. This is a potential explanation for why technologies do not flow from rich to poor countries. If an investment cannot be funded with a long-term contract, then it cannot be funded with a sequence of short-term contracts because a long-term contract can always be written to mimic a succession of short-term ones. Short-term contracts do not

[1] As a factual matter, a lot of lending is long term in nature. For example, the average maturity of U.S. corporate debt is around 13 years.

allow lenders to commit to extended punishment strategies, such as withholding future funds based on a bad report, or auditing cash flows over some probationary period of time and seizing them if malfeasance is detected. For example, in Buera, Kaboski, and Shin (2011) an entrepreneur who defaults gains full access to the credit market in the subsequent period. Long-run punishment strategies are important for achieving efficient contracts.

Second, as an empirical matter, Luttmer (2011) has emphasized that it is difficult for the standard model of technology (where TFP shocks follow a stationary AR1 process) to mimic the phenomenal growth of large firms such as Amazon, General Motors, Microsoft, Walmart, etc. Hsieh and Olken (2014) present evidence suggesting labor productivity increases with employment in China, India, and Indonesia. This is inconsistent with the standard model of production, which would predict constant labor productivity.[2] As discussed later, the fact that employment over the life cycle of plants in developed countries differs so much from lesser-developed ones reflects, in part, that they are operating fundamentally different technologies.

1.2 The Theoretical Analysis

A dynamic costly state verification model of venture capital is developed. The model has multiple unique features. First, production technologies are represented in a more general way than in the usual finance and development literature. Entrepreneurs start new firms every period. There is a menu of potential technologies that can be operated. Entrepreneurs select a blueprint from this menu of technologies, but they can start only a single venture. A firm's blueprint is represented by a non-decreasing stochastic process that describes movement up a productivity ladder. Some blueprints have productivity profiles that offer exciting profit opportunities; others are more mundane. This is operationalized by assuming there are differences in the positions of the rungs on the productivity ladders, as well as in the

[2] With a Cobb-Douglas production function the average product of labor is proportional to the marginal product of labor. If labor markets are competitive, then the marginal product of labor, and hence its average product, will be equalized across firms.

odds of stepping between rungs. Blueprints also differ in the required capital investment. Some may require substantial investment before much information about the likely outcome is known.

A start-up firm will ask an intermediary to underwrite its venture. The financial contract between the new firm and intermediary is long term in duration, unlike most of literature which assumes short term contracts. A contract specifies a state-contingent plan over the life cycle of the project, outlining the advancement of funds from the intermediary to the firm and the payments from the firm back to the intermediary. A firm's position on a productivity ladder is private information. Since the flow of funds depends on reports by the firm to the intermediary, there is an incentive for the firm to misrepresent its position to the intermediary. Intermediaries can audit the returns of a firm, as in the prototypical costly state verification paradigms of Townsend (1979) and Williamson (1986).

A distinguishing feature of the contracting framework is that the intermediary can pick the odds of a successful audit. The cost of auditing is increasing and convex in these odds. This cost is also decreasing in the productivity of a country's financial sector. Another unique feature of the analysis is the notion of poor cash-flow control. Specifically, it is assumed that some fraction of a firm's cash flow can never be secured by the intermediary via contractual means due to a poor rule of law in a country. The analysis allows for a new firm to self-finance some of the start-up costs of the venture at the time of writing a contract.

Several propositions are proved. It is established that in general the intermediary pays the firm its rewards only if it reaches the top of the productivity ladder. Additionally, the intermediary will audit all reports of a failure to move up the ladder. Auditing reduces the incentive to lie. When there is poor cash-flow control, the intermediary will also have to provide rewards even when the firm fails to move up the ladder. This reduces its ability to backload. The nature of the blueprint, a country's input prices, and the state of its financial system will determine the profitability of a project. For certain blueprints it may not be feasible for any intermediary to offer a lending contract that will make the project profitable. This situation can arise because *given the structure of the technology ladder*: (i) Input prices

are too high, (ii) the level of monitoring needed to make the project viable is simply too expensive given the efficiency of the financial system, or (iii) poor cash-flow control makes it impossible to implement enough backloading. It is shown that an entrepreneur starting a new venture should commit all of his available funds to the project. When the firm self-finances some of the start-up costs, there is less incentive to cheat on the contract in an attempt to avoid paying some of the fixed costs. Not surprisingly, if the new firm's funds are large enough, then the project will be financed in the first-best manner. Interestingly, if there is a distribution of internal funds across new firms in a country, then there may be a corresponding distribution over the technologies adopted by these firms. Thus, the state of a nation's financial system will have an impact on the type of ventures that will be financed. Financial sector efficiency will affect a nation's income and TFP. Therefore, a link between finance and development is established.

1.3 The Quantitative Illustration

While the analysis is primarily theoretical in nature, the *potential* of the financial mechanism developed here to explain cross-country differences in income is illustrated. The quantitative illustration is not intended as a formal empirical assessment of the theory outlined here or as a means to discriminate between this and other financial mechanisms.[3] The applied analysis focuses on three countries at very different levels of development and wealth: India, Mexico, and the United States. Hsieh and Klenow (2014) document some interesting facts about differences in establishments across these three countries. The average establishment size is much smaller in Mexico than in the United States, and is much smaller in India than in Mexico, as Table 1 shows. (All data sources are discussed in Appendix 14.) In addition, the

[3]For example, Buera, Kaboski, and Shin (2011) and Midrigan and Xi (2014) focus on the importance of borrowing constraints. Limited investor protection is emphasized by Antunes, Cavalcanti, and Villamil (2008) and Castro, Clementi, and MacDonald (2009). Greenwood, Sanchez, and Wang (2013) apply the static contract model of Greenwood Sanchez, and Wang (2010) to the international data. The role of financial intermediaries in producing ex ante information about investment projects is stressed by Townsend and Ueda's (2010) work on Thailand.

STYLIZED FACTS: INDIA, MEXICO, AND THE UNITED STATES

Statistics	*U.S.*	*Mexico*	*India*
Output per worker	1.00	0.33	0.12
TFP	1.00	0.46	0.24
Average establishment size	1.00	0.55	0.11
Employment share, age $\leq$ 10 yr.	0.25	0.52	0.51
$\ln(\text{TFP}_{age>35}) - \ln(\text{TFP}_{age<5})$	2.23	0.51	0.30

Table 1: See Appendix 14 for all sources.

level of labor productivity follows a similar pattern. The share of employment contributed by younger (older) establishments is also much larger (smaller) in India and Mexico than in the United States. These facts suggest that these countries are using very different technologies.

One interpretation of these stylized facts is that the United States uses technologies with a higher level of TFP than does Mexico. This leads to larger establishments in the United States than in Mexico. Mexico, in turn, uses technologies that are more productive than those chosen in India, implying that Mexican establishments are larger than Indian ones. Additionally, TFP in a U.S. establishment increases faster with age than in a Mexican one. This results in an employment profile that rises more steeply with age in the United States than in Mexico. The same story applies when comparing Mexico with India. Some facts supporting this story are presented in Table 1.

To undertake the quantitative illustration, a stylized version of the model is used where there are only three production technologies available: advanced, intermediate, and entry level. Each project has a different blueprint. The advanced technology promises high returns. When the project is successful, the time path of productivity has a convex shape. This implies that growth in output and profits materialize toward the end of the project's life cycle. The project requires large up-front investment. The entry-level technology has a lower expected return. Output and profits follow a concave time path. The project's returns are therefore more immediate. It requires less start-up investment. The intermediate technology lies between these two. To put some discipline on the analysis, factor prices are chosen to match the Indian, Mexican, and U.S. economies. Labor is much less expensive in India

than in Mexico, which in turn is less expensive than in the United States. Thus, on first appearance, the advanced technology should be more profitable in India than in Mexico and more profitable in Mexico than in the United States. The question is this: Can an equilibrium be constructed where the United States will use the first technology, Mexico the second, and India the third? And can such a structure match the above stylized facts about the Indian, Mexican, and U.S. economies, including the observations on establishment-size distributions?

The answer is yes. The advanced and intermediate technologies cannot be implemented when monitoring is not efficient and/or when there is a significant cash-flow problem. Are these two factors important? Some cross-country regression evidence is presented suggesting that they are. Additionally, the quantitative illustration is consistent with an assumption that a firm's start-up wealth is zero. Even so, the framework predicts that the ratio of private debt to GDP will rise with GDP. Why is this important? The observed concordance of this ratio with GDP is often interpreted as indicating firms in poor countries rely more on internal start-up funds than those in rich nations. The current analysis suggests that this arises, in part, because of differences across countries in the pattern of firms' cash flows related to variations in the patterns of technology adoption. These differences in technology adoption arise, to some extent, from variations in financial structures. Finally, financial development plays a very important role in economic development. It explains a significant portion of the differences in cross-country income, primarily through the technology adoption channel and not through capital deepening. Still, it does not explain the majority of the differences in incomes among India, Mexico and the United States.

2 The Environment

At the heart of the analysis is the interplay between firms and financial intermediaries. This interaction is studied in steady-state general equilibrium. Firms produce output in the economy. They do so using capital and labor. New firms are started by entrepreneurs.

The entrepreneur selects a blueprint for his firm from a portfolio of plans. He can operate only one type of project. Implementing this blueprint requires working capital. While an entrepreneur may have some personal funds, in general this working capital is obtained from financial intermediaries. Projects differ by the payoff structures they promise. For example, some projects may offer low returns but are ones that will materialize quickly with reasonable certainty and without much investment. Others may promise high returns. These projects may be risky in the sense that the odds are high that the returns are unlikely to materialize, plus the ventures may require extended periods of finance. Intermediaries borrow funds from consumers/workers in the economy at a fixed rate of return. Intermediation is competitive. The structure of a financial contract offered by an intermediary will depend on the type of venture being funded, the fraction of the start-up costs of the project the entrepreneur can self-finance, input prices, and the state of the financial system. Of course, an entrepreneur will choose the most profitable blueprint to implement. For certain blueprints it is sometimes not possible for an intermediary to offer a financial contract that will generate positive profits. Finally, in addition to supplying intermediaries with working capital, consumer/workers provide firms with labor. Consumer/workers own the intermediaries. In equilibrium, intermediaries will earn zero profits. Since consumer/workers play an ancillary role in the analysis, they are relegated to the background.[4]

3 Ventures

The theory of entrepreneurship here is simple. Each period there is a fixed amount of risk-neutral entrepreneurs that can potentially start new firms. Entrepreneurs differ by type, $\mathfrak{t} \in \mathcal{T}$, and in the amount of funds they have, $f \in \mathcal{F} \equiv [0, \overline{f}]$. Let the (non-normalized) distribution for potential type-$\mathfrak{t}$ entrepreneurs over funds be represented by $\Phi_{\mathfrak{t}}(f) : \mathcal{F} \rightarrow [0, 1]$. Projects also differ by type, $\tau \in \mathcal{T}$. A type-$\mathfrak{t}$ entrepreneur can start up and run a project of

[4] It also does not matter whether the analysis is considered as modeling (i) a closed economy in a steady state where the real interest rate rate earned by savers is equal to the rate of time preference or (ii) a small, open economy where savers can borrow or lend at some fixed real interest rate.

type $\tau \leq \mathfrak{t}$. Think about higher levels of τ as corresponding to more advanced technologies. Thus, an entrepreneur that can run technology $\nu \in \mathcal{T}$ can also operate any simpler one $\tau < \nu$. The entrepreneur faces a disutility cost, ε_τ, measured in terms of consumption, connected with operating technology τ. Envision ε_τ as representing the disutility of acquiring the skills necessary for operating a technology or as the disutility associated with running it.[5] An entrepreneur can only operate one firm at a time.

A new firm started by an entrepreneur can potentially produce for T periods, indexed by $t = 1, 2, \cdots, T$. There is a setup period denoted by $t = 0$. Here the firm must incur a fixed cost connected with entry that is represented by ϕ. Associated with each new firm is a productivity ladder $\{\theta_0, \theta_1, ..., \theta_S\}$, where $S \leq T$. As mentioned earlier, the firm's blueprint or type is denoted by τ. This indexes the vector $\{\theta_0, \theta_1, ..., \theta_S, \phi\}$. An entrepreneur selects the type of the blueprint for his firm, τ, from a portfolio of available plans, $\mathcal{T}$. Again, only one plan can be implemented. The firm enters a period at some step on the productivity ladder from the previous period, denoted by θ_{s-1}. With probability ρ it moves up the ladder to the next step, θ_s. At time $s-1$ the firm can invest in new capital for period s. This investment is made before it is known whether θ_{s-1} will move up in period s to θ_s. With probability $1-\rho$ the project stalls at the previous step θ_{s-1}, implying that the move up the ladder was unsuccessful. If a stall occurs, then the project remains at the previous level, θ_{s-1}, forever after.[6] Capital then becomes locked in place and cannot be changed. At the end

[5] The determination of who becomes an entrepreneur is of secondary importance for the analysis undertaken here. Interested readers are referred to the work of Buera, Kaboski, and Shin (2011) and Guner, Ventura, and Xu (2008) to see how such a consideration could be appended onto the current analysis. Abstracting from this factor allows the current work to focus on the novel aspects of the analysis.

[6] This assumption avoids the persistence private information analyzed in Fernandez and Phelan (2000). By extending the analysis undertaken in Appendix 13.1, it can also be deduced that in the current setting productivity can actually follow any deterministic process subsequent to a failure to move up the ladder. A similar insight is exploited in Golosov and Tsyvinski's (2006) work on disability insurance. In their analysis, with no real productive activity, a person is either able or disabled according to a two-state Markov chain. Disability is an absorbing state. The similarity with the current analysis ends there. Their work lies in the realm of new public finance rather than dynamic contracting (there is no contract): In their analysis there

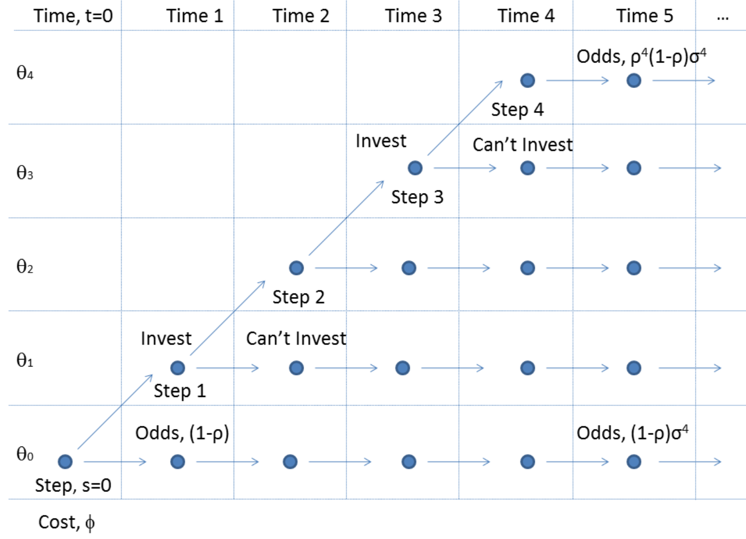

Figure 1: Possible productivity paths for a venture over its lifetime.

of each period, the firm faces a survival probability of σ. Assume that an entrepreneur dies with his firm. Figure 1 illustrates potential productivity paths for a firm over its lifetime. The ladder is somewhat reminiscent of Aghion and Howitt (1992).

In the tth period of its life, the firm will produce output, o_t, according to the diminishing-returns-to-scale production function

$$o_t = \theta_s [\widetilde{k}_t^{\omega} (\chi l_t)^{1-\omega}]^{\alpha}, \text{ with } 0 < \alpha, \omega < 1,$$

where $\widetilde{k}_t$ and l_t are, respectively, the inputs of physical capital and labor that it employs. Here χ is a fixed factor reflecting the productivity of labor in a country; this factor will prove useful for calibrating the model. Denote the rental rate for physical capital by r and the wage for labor by w. The firm finances the input bundle, $(\widetilde{k}_t, l_t)$, that it will hire in period t using working capital provided by the intermediary in period $t-1$.

Focus on the *amalgamated* input, $k_t \equiv \widetilde{k}_t^{\omega} l_t^{1-\omega}$. The minimum cost of purchasing k units

is no productivity ladder, no investment in capital, no monitoring, no cash-flow control problem, no start-up funding, no technology adoption decision, and so forth.

of the amalgamated input will be

$$[(\frac{w}{r}\frac{\omega}{1-\omega})^{-\omega}\chi^{\omega-1}w+(\frac{w}{r}\frac{\omega}{1-\omega})^{1-\omega}\chi^{\omega-1}r]k=\min_{\widetilde{k}_t,l_t}\{r\widetilde{k}+wl:\widetilde{k}^{\omega}(\chi l)^{1-\omega}=k\}. \tag{P1}$$

Thus, the cost of purchasing one unit of the amalgam, q, is given by

$$q=(\frac{w}{r}\frac{\omega}{1-\omega})^{-\omega}\chi^{\omega-1}w+(\frac{w}{r}\frac{\omega}{1-\omega})^{1-\omega}\chi^{\omega-1}r. \tag{1}$$

The cost of the intermediary providing k units of the amalgamated input is then qk. This represents the working capital, qk, provided by the intermediary to the firm. In what follows, k is referred to as the working capital for the firm, even though strictly speaking it should be multiplied by q. The rental rate, r, consists of the interest and depreciation linked with the physical capital. It is exogenous in the analysis: In a steady state, the interest rate will be pinned down by savers' rate of time preference, modulo country-specific distortions such as import duties on physical capital. The wage rate, w, will also have an interest component built into it. The wage will be endogenously determined. Hence, the cost of purchasing one unit of the amalgam, q, will be dictated by the equilibrium wage rate, w, via (1).

Finally, it is also easy to deduce that the quantities of physical capital and labor required to make k units of the amalgam are given by

$$\widetilde{k}=(\frac{w}{r}\frac{\omega}{1-\omega})^{1-\omega}\chi^{\omega-1}k \tag{2}$$

and

$$l=(\frac{w}{r}\frac{\omega}{1-\omega})^{-\omega}\chi^{\omega-1}k. \tag{3}$$

4 Intermediaries

Intermediation is a competitive industry. An intermediary borrows from consumers/workers and enters into financial contracts with new firms to supply working capital for the latters' ventures. The entrepreneur starting a new firm may have some personal funds of his own, f. He can choose to use some or all of its funds to finance part of the venture. At the time of the

contract, the intermediary knows the firm's productivity ladder, $\{\theta_0, \theta_1, ..., \theta_S\}$, and its fixed cost, ϕ. The contract specifies, among other things, the funds that the intermediary will invest in the firm over the course of its lifetime and the payments that the firm will make to the intermediary. These investments and payments are contingent on reports that the firm makes to the intermediary about its position on the productivity ladder. The intermediary cannot observe without cost the firm's position on the productivity ladder. Specifically, in any period t of the firm's life, it cannot see o_t or θ_s.

Now, suppose that in period t the firm reports that its productivity level is θ_r, which may differ from the true level θ_s.[7] The intermediary can choose whether it wants to monitor the firm's report. The success of an audit in detecting an untruthful report is a random event. The intermediary can choose the odds, p, of a successful audit. Write the cost function for monitoring as follows:

$$C(k, p; q, z) = q(\frac{k}{z})^2(\frac{1}{1-p} - 1)p. \tag{4}$$

This cost function has four noteworthy properties. First, it is increasing and convex in the odds, p, of a successful audit. When $p = 0$, both $C(k, 0; q, z) = 0$ and $C_1(0, p; q, z) = C_2(k, 0; q, z) = 0$; as $p \to 1$, both $C(k, p; q, z) \to \infty$ and $C_2(k, p; q, z) \to \infty$. Second, the marginal and total costs of monitoring are increasing in the price of the amalgam, q; that is, $C_3(k, p; q, z) > 0$ and $C_{23}(k, p; q, z) > 0$. This is a desirable property if the amalgamated input must be used for monitoring. Third, the cost is increasing and convex in the size of the project as measured by the amalgamated input k; that is, $C_1(k, p; q, z) > 0$ and $C_{11}(k, p; q, z) > 0$. A larger scale implies there are more transactions to monitor. Detecting fraud will be harder. Fourth, the cost of monitoring is decreasing in the productivity of the financial sector, which is represented here by z. (The dependence of C on q and z is suppressed when not needed to simplify the notation.)

[7] It is assumed that the firm shows to the intermediary a level of output that would correspond to the report θ_r. If $\theta_r < \theta_s$, then the intermediary must hide some of its output. Note that it is not feasible to make a report where $\theta_r > \theta_s$. Footnote 8 continues this discussion.

5 The Contract Problem

The contract problem between an entrepreneur and an intermediary is now formulated. In preparation, note that the probability distribution for the firm surviving until date t with a productivity level s is given by

$$\Pr(s,t) = \begin{cases} \rho^s \sigma^{s-1}, & \text{if } s = t, \\ \rho^s (1-\rho)\sigma^{t-1}, & \text{if } s < t, \\ 0, & \text{if } s > t. \end{cases} \tag{5}$$

The discount factor for both firms and intermediaries is denoted by β.

A financial contract between an entrepreneur and intermediary will stipulate the following for each step/date pair, (s,t): (i) the quantities of working capital to be supplied by the intermediary to the firm, $k(s,t)$; (ii) a schedule of payments by the firm to the intermediary, $x(s,t)$; and (iii) audit detection probabilities, $p(s,t)$. The contract also specifies the amount of funding, $\widetilde{f}$, that the entrepreneur will invest in the project. Take the entrepreneur as turning over these funds to the intermediary at the start of the project. Because a large number of competitive intermediaries are seeking to lend to each firm, the optimal contract will maximize the expected payoff of the firm, subject to an expected non-negative profit constraint for the intermediary. The problem is formulated as the truth-telling equilibrium of a direct mechanism because the revelation principle applies. When a firm is found to have misrepresented its productivity, the intermediary imposes the harshest possible punishment: It shuts the firm down. Since the firm has limited liability, it cannot be asked to pay out more than its output in any period. The contract problem between the entrepreneur and intermediary can be expressed as

$$v = \max_{\left\{k(s,t),x(s,t),p(s,t),\widetilde{f}\right\}} \sum_{t=1}^{T} \sum_{s=0}^{\min\{t,S\}} \beta^t \left[\theta_s k(s,t)^\alpha - x(s,t)\right] \Pr(s,t) + f - \widetilde{f}, \tag{P2}$$

subject to

$$\theta_s k(s,t)^\alpha - x(s,t) \geq 0, \text{ for } s = \{0, \cdots, \min\{t,S\}\} \text{ and all } t, \tag{6}$$

$$\sum_{t=u}^{T} \sum_{s=u}^{\min(t,S\}} \beta^t \left[\theta_s k(s,t)^\alpha - x(s,t)\right] \Pr(s,t)$$

$$\geq \sum_{t=u}^{T} \sum_{s=u}^{\min\{t,S\}} \beta^t \left[\theta_s k(u-1,t)^\alpha - x(u-1,t)\right] \prod_{n=u}^{t} \left[1 - p(u-1,n)\right] \Pr(s,t),$$

$$\text{for all } u \in \{1, ..., S\}, \tag{7}$$

$$k(t,t) = k(t-1,t), \text{ for all } t \leq S, \tag{8}$$

$$\begin{aligned} &k(s-1,t) = k(s-1,s), && \text{for } 1 \leq s < S \text{ and } t \geq s+1, \\ &k(S,t) = k(S,S), && \text{for } t > S, \end{aligned} \tag{9}$$

and

$$\sum_{t=1}^{T} \sum_{s=0}^{\min\{t,S\}} \beta^t \left[x(s,t) - C\left(p(s,t), k(s,t)\right) - qk(s,t)\right] \Pr(s,t) - \phi + \widetilde{f} \geq 0, \tag{10}$$

$$f - \widetilde{f} \geq 0. \tag{11}$$

The objective function in (P2) represents the expected present value of the profits for the firm. This is simply the expected present value of the gross returns on working capital investments, minus the payments that the firm must make to the intermediary. The maximized value of this is denoted by v, which represents the value of a newly formed firm. The value of running the firm to an entrepreneur, v, will be a function of: the amount of funds the entrepreneur possesses, f; the price of inputs, q; the state of the financial system, ψ and z; and the type of technology that is being operated, τ (note that τ has been suppressed in the above contracting problem to ease notation). Equation (6) is the limited liability constraint for the firm. The intermediary cannot take more than the firm produces at the step/date combination (s,t).

The incentive constraint for a firm is specified by (7). This constraint is imposed on the firm only at each state/date combination where there is a new productivity draw. Since no information is revealed at dates and states where there is not a new productivity draw, the firm can be treated as not making a report and hence as not having an incentive constraint

at such nodes. The validity of this is established in Appendix 13.1. There a more general problem is formulated where reports are allowed at all dates and times. These reports are general in nature and can be inconsistent over time or infeasible; for example, the firm can make a report that implies that it lied in the past. This general problem has a single time-1 incentive constraint that requires the expected present value to the firm from adopting a truth-telling strategy to be at least as good as the expected present value to the firm from any other reporting strategy. It is shown that any contract that is feasible for this more general formulation is also feasible for the restricted problem presented above and vice versa. This establishes the validity of imposing S stepwise incentive constraints along the diagonal of Figure 1.[8]

The left-hand side of the constraint shows the value to the firm when it truthfully reports that it currently has the step/date pair (u, u), for all $u \in \{1, ..., S\}$. The right-hand side denotes the value from lying and reporting that the pair is $(u-1, u)$ or that a stall has occurred. Suppose that the firm lies at time u and reports that its productivity is $u-1$. Then, in period $t \geq u$ the firm will keep the cash flow $\theta_s k(u-1, t)^{\alpha} - x(u-1, t)$, provided that it is not caught cheating. The odds of the intermediary not detecting this fraud are given by $\prod_{n=u}^{t}[1 - p(u-1, n)]$, since the intermediary will engage in auditing from time u to t. One would expect that in (7) the probabilities for arriving at an (s, t) pair should be conditioned on starting from the step/date combination (u, u). This is true; however, note that the initial odds of landing at (u, u) are embodied in a multiplicative manner in the $\Pr(s, t)$ terms and these will cancel out on both sides of (7). Thus, the unconditional

[8]Do incentive constaints need to be imposed to prevent the firm from "lying upward?" The answer is no. To see this, suppose that the firm stalls at time u and therefore finds itself at the node $(u-1, u)$. Would it have an incentive to lie upward and claim that it did not stall? By reporting that it was at node (u, u) the firm might hope to receive a higher level of working capital, $k(u+1, u+1) > k(u, u)$ from the intermediary. In the current setting, this is impossible because the intermediary could simply demand that the firm shows output in the amount $\theta_u k(u, u)$. This is greater than what the firm can actually produce at its current node, $\theta_{u-1} k(u, u)$. Alternatively, the intermediary could ask for a payment, $x(u, u)$, larger than $\theta_{u-1} k(u, u)$. This discussion is continued in footnotes 9 and 11.

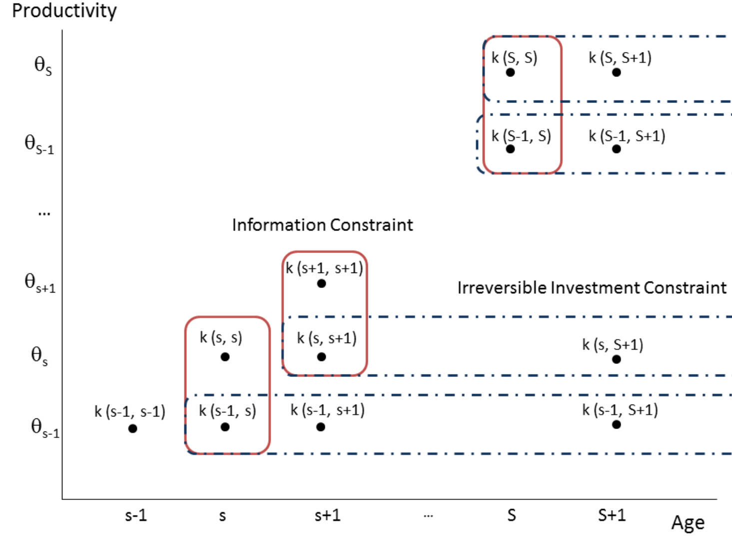

Figure 2: The information and irreversibility constraints.

probabilities, or the $\Pr(s,t)$'s, can be used in (7).

Note that in each period $t-1$, when there is not a stall, the contract will specify a level of working capital for the next period, t. This is done before it is known whether there will be a stall in the next period. Therefore, the value of the working capital in the state where productivity grows, $k(t+1,t)$, will equal the value in the state where it does not, $k(t,t)$. This explains equation (8).

The information constraint is portrayed in Figure 2 by the vertical boxes defined by the solid lines. The two working capitals within each vertical box must have the same value. Equation (9) is an irreversibility constraint on working capital. Specifically, if a productivity stall occurs in period s, working capital becomes locked in at its current level, $k(s-1,s)$. The irreversibility constraint is illustrated by the horizontal boxes drawn with the dashed-dotted lines in Figure 2. All working capitals within a horizontal box take the same value. Envision a plant as having a putty-clay structure: In the event of a stall, all inputs become locked in.

The penultimate constraint (10) stipulates that the intermediary expects to earn positive

profits from its loan contract. For an (s,t) combination the intermediary will earn $x(s,t) - C(p(s,t), k(s,t)) - qk(s,t)$ in profits after netting out both the cost of monitoring and raising the funds for the working capital investment. The intermediary must also finance the up-front fixed cost for the project. This is represented by the term ϕ in (10). Finally, equation (11) is the self-financing constraint. It simply states that the new firm cannot invest more in the venture than it has.

The contract between the entrepreneur and the intermediary specifies a plan for investment, monitoring, and payments such that the firm always truthfully reports productivity. This plan generally leads to a suboptimal level of investment due to the need to provide incentives so that the firm will always report the true state of productivity. Intuitively, one might think that this incentive problem will be reduced if the entrepreneur uses some of his own money to start up the firm. In fact, the entrepreneur should invest everything in his project. This yields an expected gross return on investment at least as great as the $1/\beta$ that the entrepreneur can earn from depositing his funds in a savings account with an intermediary.

Lemma 1 *(Go all in) It is weakly efficient to set* $\widetilde{f} = f$*.*

Proof. See Appendix 13.3. ■

Suppose that the firm reports at time $t = u$ that the technology has stalled at step $u-1$. If the incentive constraint is binding at step u, then the intermediary should monitor the firm over the remainder of its life. As the right-hand side of (7) shows, this monitoring activity reduces the firm's incentive to lie. In fact, a feature of the contract is that the firm will *never* lie, precisely because the incentive constraint (7) always holds.

Lemma 2 *(Trust but verify) Upon a report by the firm at time* u *of a stall at node* $(u-1, u)$*, for* $u = 1, 2, \cdots, S$*, the intermediary will monitor the project for the remaining time,* $t = u, u+1, \cdots, T$*, contingent upon survival, if and only if the incentive constraint (7) binds at node* (u, u)*.*

Proof. See Appendix 13.5. ■

How should the intermediary schedule the flow of payments owed by the firm, $x(s,t)$? To encourage the firm to always tell the truth the intermediary should backload the rewards that the firm can earn. In particular, it is (weakly) optimal to let the firm realize all of its awards only upon arrival at the terminal node (S,T). The intermediary should take away all the cash flow from the firm before this terminal node by setting $x(s,t) = \theta_s k(s,t)^\alpha$ for $(s,t) \neq (S,T)$. It should then give the firm at node (S,T) all of the expected accrued profits from the project. This amounts to a negative payment from the firm to the intermediary at this time so that $x(S,T) \leq 0$. The profits from the enterprise will amount in expected present-value terms to $\sum_{t=1}^{T} \sum_{s=0}^{\min\{t,S\}} \beta^t \left[\theta_s k(s,t)^\alpha - C\left(p(s,t), k(s,t)\right) - qk(s,t)\right] \Pr\left(s,t\right) - \phi + f \geq 0$. There may be other payment schedules that are equally efficient but *none* can dominate this one.[9]

Lemma 3 *(Backloading) An optimal payment schedule from the firm to the intermediary,* $\{x(s,t)\}$*, is given by*

1. $x(s,t) = \theta_s k(s,t)^\alpha$*, for* $0 \leq s \leq S$*,* $s \leq t$*,*$1 \leq t \leq T$*, and* $(s,t) \neq (S,T)$*;*

2.

$$x(S,T) = \theta_S k(S,T)^\alpha - \{\sum_{t=1}^{T} \sum_{s=0}^{\min\{t,S\}} \beta^t \left[\theta_s k(s,t)^\alpha - C\left(p(s,t), k(s,t)\right) - qk(s,t)\right] \Pr\left(s,t\right) - \phi + f\}/[\beta^T \Pr\left(S,T\right)] \leq 0.$$

Proof. See Appendix 13.6. ■

[9]Return to the issue of lying upward, as discussed in footnote 8. Now, imagine an alternative setting where the firm gets a private signal about whether it is going to move up the ladder at time u before the intermediary provides any working capital. Would it like to lie upwards in an attempt to receive more working capital? The intermediary can structure a payment schedule that renders this option unattractive. When the firm fails to produce the $\theta_u k(u,u)$ units of output in period u consistent with the positive signal, then the intermediary should simply take all of actual output, or $\theta_{u-1} k(u,u)$, forever after.

5.1 Discussion

The solution to the above contract problem shares some features common to dynamic contracts, but it also has some properties that are quite different. The current setting allows for a nonstationary, non-decreasing process for TFP, or for the θ's. In fact, the θ's could be allowed to drop after a stall, as long as the descent is deterministic (or any shocks are public knowledge). The steps on the ladder need not be equally spaced. The odds of moving up the ladder and the probabilities of survival could also be expressed as functions of s and t. The theory is developed in terms of the left-hand side of (5), $\Pr(s,t)$, which is a general function of s and t. Additionally, it would make no difference if investors had to incur a fixed cost every time they attempt to move up the ladder.

The contract problem (P2) is presented in its primitive sequence space form as opposed to the more typical recursive representation. This is more transparent, given the structure adopted here for the economic environment. First, the binding pattern of the incentive constraints may be quite complicated. In particular, it could bind at node (s,s), not bind at node $(s+1,s+1)$, and bind again at some node $(s+j,s+j)$ for $j>1$. This depends on the assumed structure for the productivity ladder.

Second, the fact that productivity can step up only to the next rung on the ladder or remain on the current step complicates matters. It inserts history dependence into the problem and implies that private information about the true value of the shock may persist into the future. In the recursive representation, the intermediary would pick, each period, the continuation payoffs for the firm subject to a promise-keeping constraint, say, as in Clementi and Hopenhayn (2006). When the firm lies, it will have a different belief about how the future will evolve vis-à-vis the intermediary, given the history dependence in the shock structure. The intermediary must also choose a continuation payoff to govern this situation, as in the recursive formulation of Fernandez and Phelan (2000). This payoff places an upper bound on the value of lying in the future. As a result, it forms the basis of a threat-keeping constraint that the problem must also incorporate, a feature not required in Clementi and Hopenhayn (2006). The sequence space form turns out to be more intuitive for the problem at hand.

6 The Contract with Costly Cash-Flow Control

The theory developed up to this point stresses the role of monitoring in designing an efficient contract. The ability to monitor reduces the incentive of the firm to misrepresent its current situation and misappropriate funds, which makes it easier for the intermediary to recover its investment and to finance technology adoption. When monitoring is very costly, an intermediary must rely primarily on a backloading strategy to create the incentives for truthful behavior. As will be seen in the quantitative illustration, which is the subject of Section 9, it may not be possible to finance certain technologies absent the ability to monitor effectively. This is most likely to happen when a project has a large up front investment and promises payoff streams tilted toward the end of the venture's lifetime. This is the case in the Mexico/United States example studied in Section 9. Here, Mexico has an inefficient monitoring technology relative to that of the United States. Thus, it is not able to adopt the advanced technology used by the United States, which has a large fixed cost and a convex productivity profile. This occurs despite the fact that production cost is lower in Mexico. Instead, Mexico uses a less-productive technology, with a lower fixed cost and a concave productivity schedule, which can be financed using a backloading strategy that requires little monitoring.

The cost of production in some countries is much lower than in Mexico. These lower production costs should imply bigger profits that, in turn, will make it easier for the intermediary to recover its investment. The intermediary could promise the firm these extra profits at node (S, T), which will increase the incentive effects of backloading. Maybe such countries could implement the U.S. technology at their lower cost of production. If not, then what prevents them from using the Mexican technology? After all, it requires little in the way of monitoring services.

An extension to the baseline theory that provides one possible answer is now developed. The premise is that it is very costly for intermediaries in some countries to force firms to pay out all of their publicly acknowledged output. Perhaps a fraction of output inherently benefits the operators of firms in the form of perks, kickbacks, nepotism, and so on. The

intermediary can offer enticements to the operators of firms so they will not do this, of course, but this limits the types of technology that can be implemented. The extended model is applied in Section 9 to India, where labor costs are extremely low.

6.1 Adding Costly Cash-Flow Control

Assume that a firm can openly take the fraction ψ of output due to weak institutional structures. The intermediary cannot recover this output unless it catches the operators of the firm lying about the firm's state during an audit.[10] The intermediary must design the contract in a manner such that the retention of output will be dissuaded. How does this affect the contract presented in (P2)?

Before characterizing the optimal contract for the extended setting, two observations are made:

1. The intermediary wants to design a contract that dissuades the firm from trying to retain the fraction ψ of output at a node. To accomplish this, the payoff at any node from deciding not to retain part of output must be at least as great at the payoff from retaining a portion of output.

2. A retention request is an out-of-equilibrium move. Therefore, it is always weakly efficient for the intermediary to threaten to respond to a retention by lowering the firm's payoff to the minimum amount possible.

These two observations lead to a *no-retention constraint* at each node (s, t) on the design

[10] The assumption that the intermediary can take everything when the firm is caught lying in an audit is not necessary. It serves to separate the incentive constraints from the no-retention constraints, (12) and (13), presented below. As a result, the incentive constraints (7) are not directly affected by this extension. This feature is a virtue for both analytical clarity and simplicity.

of the contract:

$$
\begin{aligned}
& \sum_{j=t}^{T} \beta^{j} \left[\theta_s k(s,j)^{\alpha} - x(s,j) \right] \Pr(s,j) \\
\geq \;\; & \psi \sum_{j=t}^{T} \beta^{j} \theta_s k(s,j)^{\alpha} \Pr(s,j)\,, \text{ for } 1 \leq s \leq S, s < t,\ 2 \leq t \leq T \text{ (off-diagonal node)}
\end{aligned} \tag{12}
$$

and

$$
\begin{aligned}
& \sum_{t=u}^{T} \sum_{s=u}^{\min\{t,S\}} \beta^{t} \left[\theta_s k(s,t)^{\alpha} - x(s,t) \right] \Pr(s,t) \\
\geq \;\; & \psi \sum_{t=u}^{T} \sum_{s=u}^{\min\{t,S\}} \beta^{t} \theta_s k(u-1,t)^{\alpha} \Pr(s,t)\,, \text{ for } 1 \leq u \leq S \text{ (diagonal node).}
\end{aligned} \tag{13}
$$

The first constraint (12) applies to the case of a stall at state s. Here, productivity is stuck at θ_s forever. The second constraint (13) governs the situation where the firm can still move up the productivity ladder. If the firm exercises its retention option, then the intermediary will keep the capital stock at $k(u-1,t)$; that is, it will no longer evolve with the state of the firm's productivity. Equation (9) then implies that the capital stock is locked in.

To formulate the contract problem with costly cash-flow control, simply append the no-retention constraints (12) and (13) to problem (P2). Lemma 2 still holds. Thus, the intermediary will again monitor the firm for the rest of its life whenever it claims that technological progress has stalled (if and only if the incentive constraint at the stalled step is binding). The payment schedule $\{x(s,t)\}$ now takes a different form. In the baseline version of the model, it is always optimal to make all payments to the firm at the terminal node (S,T) to relax the incentive constraints. The retention option precludes this, however. To discourage the firm from exercising its retention option, it pays for the intermediary to make additional payments, $N(s,T)$, to the firm at the terminal date T for all steps $s \leq S$ on the ladder, provided the firm does not exercise its retention option at any time before T. This payment should equal the expected present value of what the firm would receive if it

exercised the retention option. Thus,

$$N(s,T) = \psi \frac{\sum_{t=s+1}^{T} \beta^t \theta_s k(s,t)^\alpha \Pr(s,t)}{\beta^T \Pr(s,T)}, \text{ for } 0 \leq s \leq S. \tag{14}$$

Hence, Lemma 3 now appears as Lemma 4. Observe how the necessity to provide retention payments reduces the size of the reward, $-x(S,T)$, that the intermediary can give to the firm if and when it reaches the end of the ladder or node (S,T). Thus, retention payments reduce the intermediary's ability to redirect the firm's rewards (or cash flow) to the top of the ladder.[11]

Lemma 4 *(Backloading with retention payments) An optimal payment schedule from the firm to the intermediary, $\{x(s,t)\}$, is given by*

1. $x(s,t) = \theta_s k(s,t)^\alpha$, *for* $0 \leq s \leq S, 1 \leq t < T$, *and* $s \leq t$;

2. $x(s,T) = \theta_s k(s,T)^\alpha - N(s,T)$, *for* $0 \leq s < S$;

3. $$\begin{aligned} x(S,T) &= \theta_S k(S,T)^\alpha - \{\sum_{t=1}^{T} \sum_{s=0}^{\min\{t,S\}} \beta^t \left[\theta_s k(s,t)^\alpha - C\left(p(s,t), k(s,t)\right) - qk(s,t)\right] \Pr(s,t) \\ &\quad - \sum_{s=0}^{S} \beta^T N(s,T) \Pr(s,T) - \phi + f\} / [\beta^T \Pr(S,T)], \end{aligned}$$

   where $N(s,T)$ is specified by (14).

Proof. See Appendix 13.6. ■

Backloading the retention payments helps to satisfy the incentive constraint. To understand this, suppose that the firm lies and declares a stall at node (u,u). The intermediary

[11] Return to the setting discussed in footnote 9, where the firm gets a private signal about whether it will move up the ladder before the intermediary supplies any working capital. Would it like to lie upward? If the firm can retain some output upon a stall, it may like to do this. The intermediary would then need to provide an incentive not to lie upward. It could do this by providing a payment for reporting a stall. This is another source of leakage in the contract, similar to the retention payments.

will audit the firm from then on. Recall the intermediary can recover all output if detects a lie at some node (u, t), where $t \geq u$. Some firms will indeed stall and find themselves at node $(u-1, u)$. Under the old contract, a stalled firm would receive nothing because $x(u-1, t) = \theta_{u-1} k(u-1, t)^{\alpha}$ for all $t > u-1$. This firm can exercise its retention option and take $\psi \theta_{u-1} k(u-1, t)^{\alpha}$ for $t > u-1$. Now a firm that is at node (u, u), but declares that it is at $(u-1, u)$, would also like to claim this part of output. It can potentially do this provided it is not caught. To mitigate this problem, the intermediary gives the firm the accrued value of these retentions, $N(u-1, T)$, at the end of the contract, or time T, assuming that the latter survives. This reduces the incentive for a firm to lie and declare a stall at node (u, u). A deceitful firm will receive the payment $N(u-1, T)$ only if it successfully evades detection along the entire path from u to T. This happens with odds $\prod_{n=u}^{T}[1 - p(u-1, n)]$.

Note how the intermediary's ability to monitor interacts with the firm's potential to retain output. The expected value of the retention payment from lying at (u, u) is $N(u-1, T) \prod_{n=u}^{T}[1 - p(u-1, n)]$, for all $u \in \{1, ..., S\}$. When monitoring is very effective, it is difficult for a masquerading firm to capture this payment, which reduces the incentive to lie. When monitoring is ineffective, it is easy to do this. The incentive to lie is then higher.

Finally, when is investment efficient or when will it match the level that would occur in a world where the intermediary can observe the firm's shock without cost?[12] Suppose that after some state/date combination (t^*, t^*) along the diagonal of the ladder that neither the incentive nor no-retention constraints, (7) and (13), ever bind again. Will investment be efficient from then on? Yes.

Lemma 5 *(Efficient investment) Suppose that neither the incentive nor the diagonal-node*

[12] An implication of the irreversible investment assumption is now discussed. Suppose that capital can be freely adjusted. Then, after a stall has been declared the intermediary could withdraw all working capital from the venture. The ability of the firm to retain funds would not matter. Alternatively, suppose the intermediary cannot reduce the level of working capital but that the firm can use some of the funds it retains to increase inputs. To dissuade this possibility the intermediary could (i) monitor the firm more heavily and take everything (including the additional inputs) if it detects malfeasance and/or (ii) offer a larger retention payment to entice the firm not to retain some of its output.

no-retention constraints ever bind after node (t^*, t^*) for $t^* < S$. Investment will be efficiently undertaken on arriving at the state/date combination (t^*, t^*). (That is, the capital stock will be at its efficient level from period $t^* + 1$ on.)

Proof. See Appendix 13.7. ■

7 Self-Financed Start-Up Funding

The self-financing of projects is discussed in this section. To highlight some points related to self-financing, per se, the environment outlined previously is simplified slightly. In particular, assume that there is one type of entrepreneur that can operate any type of project. To map this into the developed structure, let $\mathfrak{t} = \overline{\tau} = \max_{\tau \in T}$, $\Phi(f) = \Phi_{\overline{\tau}}(f)$ (the highest type), and $\Phi_\tau(f) = 0$ for $\tau \neq \overline{\tau}$ (no entrepreneurs of other types).

The higher the level of funds that an entrepreneur possesses, then the greater is the fraction of the project that he can self-finance. This circumvents the informational problem. At some point, the first-best allocations can be achieved.

Lemma 6 *(Efficient self-finance) There exists a level of self-financed start-up funding, $\widehat{f}$, such that the first-best allocations obtain.*

Proof. See Appendix 13.8. ■

At a given level of factor prices some technologies, $\tau \in \mathcal{T}$, will be able to produce a higher level of potential output than others. In particular, suppose that the following condition holds.

Condition 7 *(Technology ranking) Assume that, at some particular input price, q, technology υ yields a higher first-best level of expected discounted profits than technology τ whenever $\upsilon > \tau$. (Note that the ranking of technologies may change as the input price changes.)*

Now, an entrepreneur is free to choose any technology he likes. He will select the one, τ^*, that maximizes his surplus. That is,

$$\tau^* = \arg\max_{\tau \in \mathcal{T}} \upsilon(f; \tau) - \varepsilon_\tau. \tag{15}$$

As an entrepreneur's level of funds increases, allowing him to self-finance his project better, the incentive problem disappears. The entrepreneur should start favoring more advanced technologies (higher τ's) over less advanced ones.

Proposition 8 *(Technology switching) Suppose that at some level of wealth, $f_{\tau,\upsilon}$, technology $\tau < \upsilon$ is chosen by an entrepreneur because it maximizes the value of his firm. There exists a set of wealth levels, $\mathcal{F}_{\tau,\upsilon} \subset [f_{\tau,\upsilon}, \widehat{f}_{\upsilon}]$ such that the more advanced technology υ is preferred to technology τ whenever $f \in \mathcal{F}_{\tau,\upsilon}$ and technology τ is preferred to the more advanced technology υ whenever $f \in [f_{\tau,\upsilon}, \widehat{f}_{\upsilon}] - \mathcal{F}_{\tau,\upsilon}$.*

Proof. Appendix 13.9. ■

Given that there is a distribution of funds across new entrepreneurs, as represented by $\Phi(f)$, some entrepreneurs may prefer to use one type of technology while others pick different ones. Thus, in general, multiple technologies may be used in an economy.

Corollary 1 *(Coexisting technologies) It is possible for multiple technologies to coexist in an economy.*

Proof. Appendix 13.10. ■

Some simple two-period examples illustrating the contracting setup are presented in Appendix 15.

8 Equilibrium

There is one unit of labor available in the economy. This must be split across all operating firms. Recall that a firm's type is given by $\tau \in \mathcal{T}$, which indexes the vector $\{\theta_0, \theta_1, ..., \theta_S, \phi\}$ connected with a particular productivity ladder and fixed cost. Again, the technologies are ordered so that higher τ's correspond with more advanced technologies. An entrepreneur of type $\mathfrak{t} \in \mathcal{T}$ can potentially start a new firm of type $\tau \leq \mathfrak{t}$. He incurs the disutility cost, ε_{τ}, (measured in terms of consumption) to operate a type-τ firm. Entrepreneurs may differ by

the level of funds, f, that they bring to the project. The (non-normalized) distribution for potential type-$\mathfrak{t}$ entrepreneurs over funds is represented by $\Phi_{\mathfrak{t}}(f) : [0, \overline{f}] \rightarrow [0, 1]$.

Clearly an entrepreneur will operate the technology that offers the largest surplus. The choice for a type-$\mathfrak{t}$ entrepreneur with f in funds is represented by $\tau^{*}(\mathfrak{t}, f)$, where

$$\tau^{*}(\mathfrak{t}, f) = \arg\max_{\tau \leq \mathfrak{t}}[v(f; \tau) - \varepsilon_{\tau}]. \tag{16}$$

It may be the case that this entrepreneur does not want to operate any type of project, because $v(f, \tau^{*}) < \varepsilon_{\tau^{*}}$. Let the indicator function $I_{\tau}(\mathfrak{t}, f)$ denote whether a type-$\mathfrak{t}$ entrepreneur with f in funds will operate (or match with) a type-τ venture. It is defined by

$$I_{\tau}(\mathfrak{t}, f) = \begin{cases} 1, & \text{if } \tau = \arg\max_{\tau \leq \mathfrak{t}}[v(f; \tau) - \varepsilon_{\tau}] \text{ and } v(f; \tau) - \varepsilon_{\tau} \geq 0, \\ 0, & \text{otherwise.} \end{cases} \tag{17}$$

Note that the entrepreneur's type, $\mathfrak{t}$, will not influence his decision about how much working capital and labor to hire. Represent the working capital and labor used at an (s, t) node in a type-τ firm, operated by an entrepreneur with f in funds, by $k(s, t; \tau, f)$ and $l(s, t; \tau, f)$, respectively.

The labor market clearing condition for the economy then reads

$$\sum_{\tau \in \mathcal{T}} \sum_{\mathfrak{t}=1}^{\tau} \int I_{\tau}(\mathfrak{t}, f) \sum_{t=1}^{T} \sum_{s=1}^{\min\{t,S\}} [l(s, t; \tau, f) + l_{m}(s, t; \tau, f)] \Pr(s, t)\, d\Phi_{\mathfrak{t}}(f) = 1, \tag{18}$$

where $l_{m}(s, t; \tau, f)$ is the amount of labor that an intermediary will spend at node (s, t) monitoring a type-τ venture operated by an entrepreneur with funds f. Every period some firms die; this death process is subsumed in the probabilities $\Pr(s, t)$. The quantity of the amalgamated input used in monitoring, $k_{m}(s, t; \tau, f)$, is given by

$$k_{m}(s, t; \tau, f) = [\frac{k(s, t; \tau, f)}{z}]^{2}[\frac{1}{1 - p(s, t; \tau, f)} - 1]p(s, t; \tau, f) \text{ [cf. (4)]}, \tag{19}$$

which implies a usage of labor in the following amount:

$$l_{m}(s, t; \tau, f) = (\frac{w}{r}\frac{\omega}{1 - \omega})^{-\omega}\chi^{\omega - 1}k_{m}(s, t; \tau, f) \text{ [cf. (3)]}. \tag{20}$$

A definition of the competitive equilibrium under study is now presented to crystallize the discussion so far.

Definition 1 *For a given steady-state cost of capital, r, a stationary competitive equilibrium is described by (a) a set of working capital allocations, $k(s,t;\tau,f)$, labor allocations, $l(s,t;\tau,f)$ and $l_m(s,t;\tau,f)$, and monitoring strategy, $p(s,t;\tau,f)$; (b) a set of optimal matches between entrepreneurs and technologies represented by $I_\tau(\mathfrak{t},f)$; and (c) an amalgamated input price, q, and wage rate, w, all such that*

1. *The working capital financing program, $k(s,t;\tau,f)$, and the monitoring strategy, $p(s,t;\tau,f)$, specified in the financial contract maximize the value of a type-τ venture for an entrepreneur with f in funds, as set out by (P2), given the amalgamated input price, q. [Here (P2) should be amended to include the no-retention constraints (12) and (13).]*

2. *The set of optimal matches between entrepreneurs and technologies, as represented by $I_\tau(\mathfrak{t},f)$,is specified by (17).*

3. *A type-τ venture operated by an entrepreneur with f in funds hires labor, $l(s,t;\tau,f)$, to minimize its costs in accordance with (P1), given wages, w, and the size of the loan, $k(s,t;\tau,f)$, offered by the intermediary. [This implies that $l(s,t;\tau,f) = \{(w/r)[\omega/(1-\omega)]\}^{-\omega}k(s,t;\tau,f)$.]*

4. *The amount of labor, $l_m(s,t;\tau,f)$, used to monitor a venture is given by (20) in conjunction with (19).*

5. *The price of the amalgamated input, q, is dictated by w in accordance with (1).*

6. *The wage rate, w, is determined so that the labor market clears, as written in (18).*

9 A Quantitative Illustration: The Choice of Technology in India, Mexico, and the United States

Why might one country choose a different set of production technologies than another country? A quantitative illustration is presented to show that the financial mechanism proposed

here offers some promise for explaining cross-country differences in technology adoption and, hence, income. There are many reasons, of course, why countries may adopt different technology: differences in the supplies of labor or natural resources that create a comparative advantage for certain types of firms; government regulations, subsidies, or taxes that favor certain forms of enterprise over others; and the presence of labor unions and other factors that may dissuade certain types of business. While these are valid reasons, the focus here is on differences in the efficiency of the financial system. This is done without apology, because abstraction is a necessary ingredient for theory. A formal empirical assessment of the mechanism, and a comparison with other explanations (including financial ones), is beyond the scope of this work.

In the quantitative illustration, an entrepreneur is free to adopt one of three technologies: advanced, intermediate, and entry level. Additionally, it is assumed that entrepreneurs have no start-up funds of their own ($f = 0$).[13] The advanced technology has a (convex) productivity ladder that grows faster than the intermediate one (which has a concave ladder), which in turn grows faster than the entry-level technology (which also has a concave ladder). The fixed cost for the advanced technology is bigger than that of the intermediate one, which is larger than that of the entry-level technology. The advanced technology, with its convex payoff structure and high fixed cost, is difficult to implement without monitoring at high factor prices. It is also difficult to adopt at low factor prices when there is a costly cash-flow control problem. The entry-level technology with its very low fixed cost is easy to implement in the absence of monitoring and when there is a costly cash-flow control problem. A country's choice of technology depends on its factor prices and the state of its

[13] Assume that there are many entrepreneurs capable of running each technology; that is, set $\Phi_t(0) = \mathfrak{f}_\tau$, where $\mathfrak{f}_\tau$ is some large number. This implies that there is free-entry into running a firm, as in Hopenhayn and Rogerson (1993). Assuming that only one technology, τ, will be run in a country, the labor-market clearing condition (18) now appears as

$$\sum_{t=1}^{T} \sum_{s=1}^{\min\{t,S\}} [l(s,t;\tau,0) + l_m(s,t;\tau,0)] \Pr(s,t)\, \mathfrak{e}_\tau \mathfrak{f}_\tau = 1,$$

where $\mathfrak{e}_\tau \leq \mathfrak{f}_\tau$ is the equilibrium fraction of potential type-τ entrepreneurs that run a firm.

financial system. An equilibrium is constructed in which the United States will adopt the advanced technology, Mexico selects the intermediate one, and India chooses the entry-level technology.

Since the focus here is on the long run, let the length of a period be 5 years and set the number of periods to 10, so that $T = 10$. Given this period length, the discount factor is set so $\beta = 0.98^5$, slightly below the 3 percent return documented by Siegal (1992). This is a conservative choice since it gives backloaded long-term contracts a better chance. The weight on capital in the production function, ω, is chosen so that $\omega = 0.33$. A value of 0.85 is assigned to the scale parameter, α. According to Guner, Ventura, and Xu (2008), this lies in the range of recent studies.

9.1 Estimating the Input Prices

A key input into the analysis is the price for the amalgamated input, q. Start with Mexico and the United States. The price of this input in Mexico relative to the United States is what is important. Normalize this price to be 1 for the United States, so that $q^{US} = 1$. (A superscript attached to a variable, either MX or US, denotes the relevant country of interest; viz, Mexico or the United States.) This can be done by picking an appropriate value for U.S. labor productivity, χ^{US}, given values for the rental rate on capital, r^{US}, and the wage rate, w^{US}. How to do this is discussed below. Is the price for this input more or less expensive in Mexico? On the one hand, wages are much lower in Mexico. On the other hand, capital is more expensive and labor is less productive. Hence, the answer is unclear ex ante. Estimating the price of the input in Mexico, q^{MX}, requires using formulas (1), (2), and (3) in conjunction with an estimate of the rental price of capital in Mexico, r^{MX}, the wage rate, w^{MX}, and the productivity of labor, χ^{MX}.

How is q^{US} set to 1? First, the rental rate on capital, r^{US}, is pinned down. To do this, suppose that the relative price of capital in terms of consumption in the United States is 1. Thus, $p_k^{US}/p_c^{US} = 1$, where p_k^{US} and p_c^{US} are the U.S. prices for capital and consumption goods. Assume that interest plus depreciation in each country sums to 10 percent of the

cost of capital. Hence, set $r^{US} = (1.10^5 - 1) \times (p_k^{US}/p_c^{US}) = 1.10^5 - 1$, which measures the cost of capital in terms of consumption. Second, a value for the wage rate, w^{US}, is selected. This is obtained by dividing the annual payroll by the number of employees in all establishments in the manufacturing sector using the 2008 Annual Survey of Manufactures. Thus, $w^{US} = 47,501$. Last, given the above data for r^{US} and w^{US}, the value for χ^{US} that sets q^{US} equal to 1 can be backed out using equation (1). This implies $\chi^{US} = 96,427$.

Turn now to Mexico. What is the value of q^{MX}? Determining this value requires knowing r^{MX}, w^{MX}, and χ^{MX}. First, a value for the rental price of capital, r^{MX}, is determined. The relative price of capital is estimated (from the Penn World Table) to be about 21 percent higher in Mexico than in the United States. Therefore, $(p_k^{MX}/p_c^{MX})/(p_k^{US}/p_c^{US}) = 1.21$, where p_k^{MX} and p_c^{MX} are the Mexican prices for capital and consumption goods. Therefore, $r^{MX} = (1.10^5 - 1) \times (p_k^{MX}/p_c^{MX}) = (1.10^5 - 1) \times (p_k^{US}/p_c^{US}) \times [(p_k^{MX}/p_c^{MX})/(p_k^{US}/p_c^{US})] = r^{US} \times [(p_k^{MX}/p_c^{MX})/(p_k^{US}/p_c^{US})] = (1.10^5 - 1) \times 1.21$. This gives the rental price of capital in terms of consumption for Mexico.

Second, a real wage rate is needed for Mexico, or a value for w^{MX} is sought. Again, this is pinned down using data on annual payroll and the total number of workers in manufacturing establishments; in this case, the data come from Mexico's National Institute of Statistics and Geography (INEGI). The result is $w^{MX} = 21,419$ once Mexican pesos are converted to U.S. dollars on a purchasing power parity basis.

Third, what is the productivity of labor in Mexico? A unit of labor in Mexico is taken to be 55 percent as productive as in the United States, following Schoellman (2012). So set $\chi^{MX} = 0.55 \times \chi^{US} = 53,035$. Finally, by using the obtained values for r^{MX}, w^{MX}, and χ^{MX} in equation (1), it then follows that $q^{MX} = 0.9371$. The upshot is that the amalgamated input is 6 percent less expensive in Mexico relative to the United States.

Move now to India. The rental price of capital in India, r^{IN}, is about 23 percent higher in India than in the United States (from the Penn World Table). Therefore, $r^{IN} = (1.10^5 - 1) \times 1.23$. The real wage rate for India, w^{IN}, will be chosen to approximate the output per worker in the manufacturing sector relative to the United States. As a result, $w^{IN} = 7,000$, which

is about 15 percent of the U.S. wage rate. Finally, what is the productivity of labor in India? A unit of labor in India is taken to be 35 percent as productive as in the United States. Here 1.6 years of education are added to the number in Barro and Lee (2013) to adjust their aggregate number upward to reflect the higher level of education in the manufacturing sector. The procedure developed in Schoellman (2011) is then used to obtain a measure of labor productivity. This leads to $\chi^{IN} = 33,750$. Finally, by plugging the obtained values for r^{IN}, w^{IN}, and χ^{IN} into equation (1), it follows that $q^{IN} = 0.6$.

9.2 Parameterizing the Technology Ladder

There are nine unique rungs (eleven) on the technology ladder; the last three are the same. The generic productivity ladder is described by

$$\theta_s = \ln[\overline{\theta}_0 + \overline{\theta}_1(s+1) + \overline{\theta}_2(s+1)^2 + \overline{\theta}_3(s+1)^3], \text{ for } s = 0, \cdots, 9.$$

The parameter values for this ladder are different for India, Mexico, and the United States. The odds of stalling are fixed over the age of a firm and are given by ρ. This differs by technology.

The probability of surviving (until age t) is also allowed to differ across technologies. The survival probabilities follow the process

$$\sigma_t = \sigma_{t-1}[1 - (\overline{\sigma}_0 + \overline{\sigma}_1 t + \overline{\sigma}_2 t^2)]^5, \text{ for } t = 2, \cdots, 10, \text{ with } \sigma_1 = 1.$$

This structure characterizing the odds of survival and stalling can easily be added onto the theory developed, as discussed in Section 5.1. Finally, an upper bound on working capital is imposed. This is denoted by $\overline{k}$ and is common across technologies.[14]

[14]This upper bound prevents the scale of a venture becoming unrealistically large as input prices drop to low levels. That is, the upper bound forces decreasing returns to bite more sharply at some point than the adopted Cobb-Douglas representation of the production function allows. This could be due to span of control or other problems.

9.3 The Choice of Technology in India, Mexico, and the United States

A quantitative illustration is now provided where India, Mexico and the United States all *choose* to adopt different production technologies. The U.S. (or advanced) technology offers a productivity profile that grows much faster with age than the Mexican contour (which represents an intermediate-level technology). The start-up cost for the U.S. technology is higher than the Mexican one. Even though Mexican factor prices are slightly lower than in the United States, it is not profitable to operate the advanced technology in Mexico. This is because financing the advanced technology (at Mexican factor prices) requires a level of efficiency in monitoring that is too high for the Mexican financial system. Without efficient monitoring it is not possible for financiers to recover the cost of investment. The intermediate-level technology does not require such a high level of monitoring efficiency. The productivity profile for the Indian technology is lower and flatter than the Mexican one. The cost of production in India ($q^{IN} = 0.6$) is much less expensive than in Mexico ($q^{MX} = 0.94$) and the United States ($q^{US} = 1$). Therefore, at first glance, one would expect that India could easily adopt the technology used in Mexico. But adopting the Mexican technology is not feasible for India because of a costly cash-flow control problem. This factor prevent financiers from recovering their up front investment. Thus, India is forced to adopt the entry-level technology with a relatively flat productivity schedule but low fixed cost. The quantitative illustration is constructed so that the framework matches the size distribution of establishments by age that is observed for India, Mexico, and the United States. It also replicates the average size of firms in these three countries—in fact, for the United States the entire size distribution is fit. These four sets of facts discipline the assumed productivity profiles.

9.3.1 Calibrating the Technology Ladders

First, the survival probabilities are obtained from the Indian, Mexican, and U.S. data. In particular, a polynomial of the specified form is fit to the data from each country. It turns out that these survival probabilities are remarkably similar for Mexico and the United States. So, assume that they are the same.

Second, this leaves the parameters for describing productivity and the odds of a stall along the diagonal. These parameters are selected so that the model fits, as well as possible, several stylized facts about the Indian, Mexican, and U.S. economies. These facts are output per worker, average plant size, the average growth in TFP over a plant's life, the (complementary) cumulative distribution of employment by establishment age, and the private-debt-to-GDP ratio. The (complementary) distribution of employment by establishment age is characterized by a set of points. For the United States alone, the establishment size distribution in Lorenz-curve form is also added to the collection of stylized facts. So, let D^j proxy for the jth data target for the model and $M^j(p)$ represent the model's prediction for this data target as a function of the parameter vector $p \equiv \{\overline{\theta}_0, \overline{\theta}_1, \overline{\theta}_2, \overline{\theta}_3, \phi, \rho, \overline{k}\}$. The parameter vector p is chosen for each country in the following fashion:

$$\min_p \sum_j [D^j - M^j(p)]^2.$$

The parameter values used in the simulation are reported in Table 2.

Figure 3 shows the salient features of the technologies used in India, Mexico, and the United States. The productivity of a firm increases with a move up the ladder. The U.S. ladder has a convex/concave profile, while the Indian one is concave. The Mexican ladder lies between the other two. Note that the ascent is much steeper for a U.S. firm than a Mexican one. The productivity profile for the Indian is lower and flatter than the Mexican one. The survival rate is higher for younger establishments in India than for plants in either Mexico or the United States (recall that the survival rates for the latter two countries are the same). The shape of the technology ladder, or the θ's, is identified from the age distribution of employment in each country. In Appendix 15.3, the two-period example presented earlier

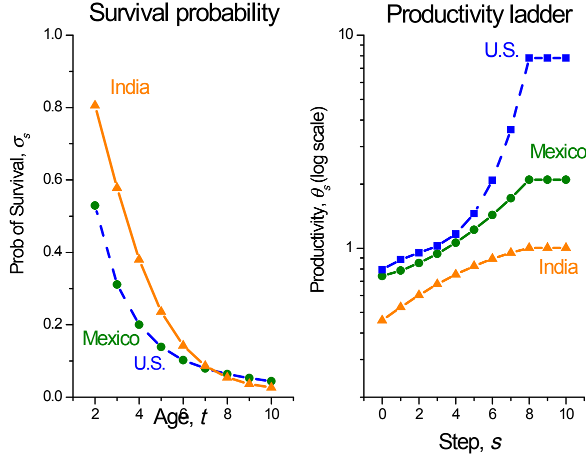

Figure 3: Productivity and survival in India, Mexico, and the United States (model). The diagram displays the assumed productivity ladders (right panel) for India, Mexico, and the United States. It also illustrates the probability profiles for survival (left panel).

is used to illustrate how this is done.

9.3.2 Establishment Size Distributions

The model matches the U.S. establishment-size distribution very well, as seen in Figure 4, which plots this distribution in Lorenz-curve form. However, the model overpredicts the share of small establishments in employment. Mexican plants are about half the size of U.S. plants. The model mimics this feature of the data well, as shown in Table 3. Plants in India are even smaller, about 10 percent of the size of American plants; the model predicts 7 percent.

Figure 5 plots the model's fit for the complementary cumulative distributions of employment by age for the three countries; that is, it graphs one minus the cumulative distribution of employment by age.[15] Establishments older than 30 years account for a smaller fraction

[15] The data on establishments in India are problematic for at least two reasons. First, India has a large informal sector. Therefore, using statistics containing information about only the formal sector might be misleading. Second, the large differences between sectors in India—mainly agriculture versus manufacturing—imply that statistics computed at the aggregate level may not be close to those computed for manufacturing

Parameter Values

Parameter	*Value*		
	United States	Mexico	India
Discount factor, β	0.98^5	0.98^5	0.98^5
Production function, scale, α, capital's share, ω	0.8, 0.33	0.8, 0.33	0.8, 0.33
Capital, upper bound, $\overline{k}$	12	12	12
Fixed cost, ϕ	0.255	0.015	0
Labor efficiency, χ	$96,427$	$53,035$	$33,750$
Pr Stall, $1-\rho$	0.31	0.6	0.6
Steps along ladder, the σ's (productivity)	See Fig. 3	See Fig. 3	See Fig. 3
Pr survival at time t, the σ's	See Fig. 3	See Fig. 3	See Fig. 3
Input price, q	1.0	0.94	0.6
Monitoring efficiency, z	1.75	0.50	0.50
Retention, ψ	0.06	0.06	0.48

Table 2: The parameter values used in the simulations.

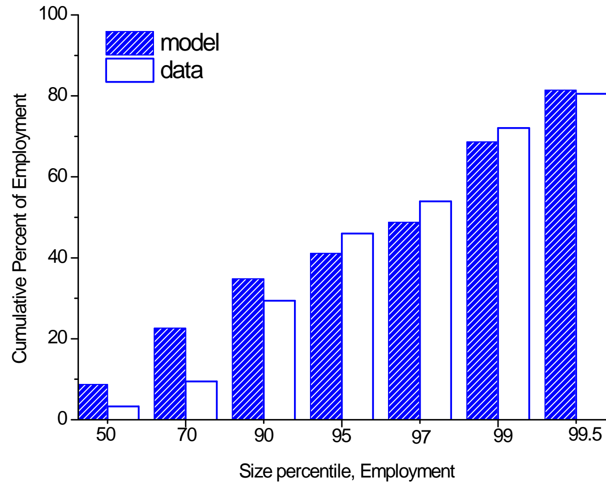

Figure 4: U.S. establishment-size distribution in Lorenz-curve form: data and model. Data sources for all figures are presented in Appendix 14.

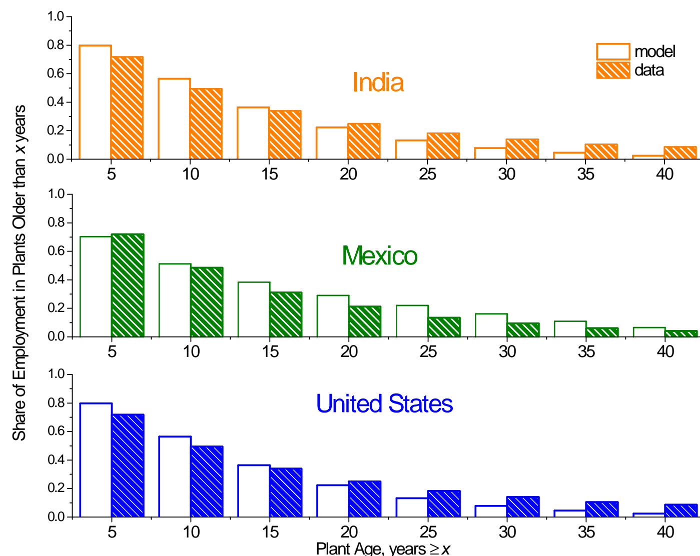

Figure 5: The (complementary) cumulative distributions of employment by age (≥) for plants in India, Mexico, and the U.S., data and model.

of employment in India or Mexico relative to the United States, as the right sides of the graphs show. The calibrated framework mimics the share of employment by age for Indian firms (the top panel) very well. The size of old Indian plants in the model is slightly too small, though. Next, consider Mexico (the middle panel). The fit is good, but the model has a little difficulty matching the size of young plants in Mexico; for example, the model overpredicts (underpredicts) the employment share for establishments older (younger) than 10 years. Now switch to the United States (the bottom panel). The model matches the share of employment by age for the United States very well. Still, it does not quite capture the fact that some old firms in the United States are very large. Finally, note from Table 3 that the model's predictions about the relationship between employment and establishment age are captured using TFP profiles for plants that grow at roughly the correct rates for India, Mexico, and the United States. That is, employment grows with age faster in an American plant than in an Indian one because TFP grows faster in a plant in the United States compared with one in India.

alone.

9.3.3 Productivity

Can the above framework generate sizable differences in productivity between India, Mexico, and the United States, due to differences in technology adoption, which are in turn induced by differences in financial markets? Before proceeding, some definitions are needed. Aggregate output in a country is given by

$$\mathbf{o}(\tau) = \sum_{t=1}^{T} \sum_{s=1}^{\min\{t,S\}} o(s,t;\tau) \Pr(s,t;\tau),$$

where $o(s,t;\tau)$ represents a firm's production at the (s,t) node when it uses the τ technology. Note that the odds of arriving at node (s,t) are now also a function of τ. In a similar vein, define the aggregate labor amounts of labor and capital that are hired by

$$\mathbf{l}(\tau) = \sum_{t=1}^{T} \sum_{s=1}^{\min\{t,S\}} l(s,t;\tau) \Pr(s,t;\tau),$$

$$\mathbf{k}(\tau) = \sum_{t=1}^{T} \sum_{s=1}^{\min\{t,S\}} k(s,t;\tau) \Pr(s,t;\tau),$$

where $k(s,t;\tau)$ and $l(s,t;\tau)$, respectively, denote the quantities of capital and labor that a firm will hire at node (s,t) when it uses the τ technology.

Labor productivity in a country reads $\mathbf{o}(\tau)/\mathbf{l}(\tau)$. As can be seen, the model performs will in replicating the fact that productivity in Mexico is only one-third of productivity in the United States. Indian productivity is only one-tenth of the American level. The model duplicates this as well. Likewise, a measure of TFP can be constructed. In particular, TFP is defined as $\mathbf{o}(\tau)/[\mathbf{k}(\tau)^{\kappa}\mathbf{l}(\tau)^{1-\kappa}]$, where κ is capital's share of income and is set to $1/3$. The framework mimics excellently the facts that Indian and Mexican TFPs are 46 and 24 percent, respectively, of the U.S. level.

Stylized Facts for India, Mexico, and the United States

Statistics	*U.S.*		*Mexico*		*India*	
	Data	Model	Data	Model	Data	Model
Output per worker	1.00	1.00	0.33	0.31	0.12	0.15
TFP	1.00	1.00	0.46	0.40	0.24	0.25
Average firm size	1.00	1.00	0.55	0.67	0.11	0.14
Debt-to-output ratio	1.65	1.83	0.24	0.08	0.24	0.08
Employment share, age $\leq$ 10 yr	0.25	0.21	0.52	0.49	0.51	0.44
$\ln(\text{TFP}_{age>35}) - \ln(\text{TFP}_{age<5})$	2.23	2.10	0.51	0.33	0.30	0.16

Table 3: Stylized Facts, Data Versus Model.

10 Why Doesn't Technology Flow from Rich to Poor Countries?

What determines the technology a nation will use? Can differences in cash-flow control and monitoring justify the adoption of less productive technologies, even when input prices are substantially less expensive (implying that the advanced technology would be very profitable in the absence of any contracting frictions)? As it turns out, there is a wide range of values for ψ and z that are consistent with the United States adopting the advanced technology, Mexico the intermediate one, and India the entry-level technology. Some diagrams are developed next to show this.

For technology τ to operate in a country with a financial system characterized by (ψ, z) requires that[16]

$$v(q; \psi, z, \tau) - \varepsilon_\tau = 0, \tag{21}$$

and

$$v(q; \psi, z, t) - \varepsilon_t \leq 0, \text{ for } t \neq \tau \text{ and } \tau, t \in \{IN, MX, US\}. \tag{22}$$

Recall that ε_τ is the cost to an entrepreneur of running technology τ. Equation (21) is the zero-profit condition for technology τ, while equation (22) ensures that it is not profitable for an entrepreneur to deviate and operate one of the other two technologies. Focus on the

[16] This condition is the analog of (17) for the simulated economy.

choice between the advanced and intermediate technology. If an equilibrium occurs where the advanced technology is operated, then condition (21) implicitly describes an equilibrium price function defined by $v(q; \psi, z, US) - \varepsilon_{US} = 0$. Write this relationship as $q^{US} = Q(z; \psi, \tau = US)$. Similarly, when the intermediate technology is adopted, the condition $v(q; \psi, z, MX) - \varepsilon_{MX} = 0$ will specify a price locus $q^{MX} = Q(z; \psi, \tau = MX)$. Figure 6 plots the two price schedules, which result from the simulation, as a function of monitoring efficiency, z, when $\psi = 0.06$ (the value in Mexico and the United States). For either technology, as monitoring becomes more efficient investment will increase, which will drive up wages and hence q. The function $Q(z; \psi, \tau = US)$ moves up faster with z than the function $Q(z; \psi, \tau = MX)$ because the advanced technology responds more to shifts in the efficiency of monitoring, z, than does the intermediate one.

Now focus on the points to the right of the vertical line in Figure 6. In this region, the advanced technology is adopted so equilibrium prices will lie on the dashed line. It is not profitable for an entrepreneur to deviate and operate the intermediate technology. To see why, note that along the lower solid line an entrepreneur would earn zero profits from operating the intermediate technology. Thus, at higher prices he would incur a loss. Suppose, counterfactually, that the intermediate technology is adopted in equilibrium; then input prices would be on the solid line. Here, an entrepreneur should deviate and run the advanced technology. This occurs because along the higher dashed line the entrepreneur earns zero profits from the advanced technology. So, clearly, he would earn positive profits at the lower prices on the solid line. Therefore, an equilibrium where the intermediate technology is operated is not deviation proof. Observe the wide range of z's that are consistent with adopting either technology.

Figure 7 shows the adoption zones in (ψ, z) space for each technology. That is, it illustrates the combinations of ψ and z that are consistent with the adoption of each technology.[17]

[17]Note that the units of z are not pinned down in the analysis. For example, one could just as easily rewrite the monitoring cost function as $C(k, p; q, \widetilde{z}) = qk^2 \exp(-2\widetilde{z})^2(\frac{1}{1-p} - 1)p$. The solution to the contracting problem would be exactly the same, but $\widetilde{z}$ would have the same units as $\ln z$. So, think of z as an ordinal measure. Pinning down the units would require information about the relationship between the inputs and

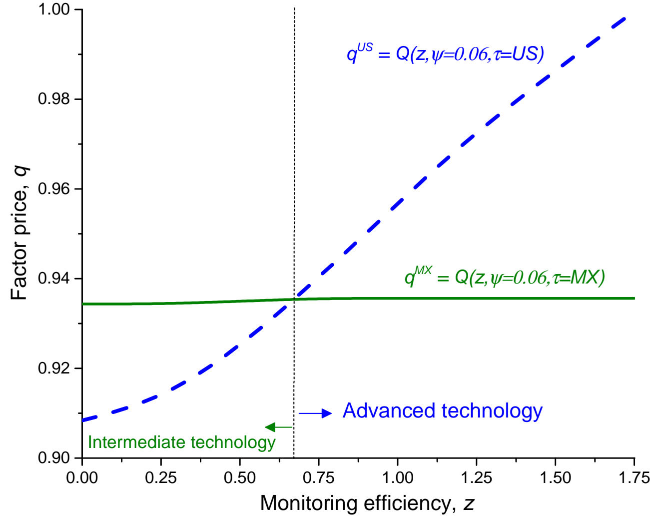

Figure 6: The choice between the advanced and intermediate technologies as a function of monitoring efficiency.

The diagram takes into account that as ψ and z change, so does q. In other words, it is done in general equilibrium. Focus on the boundary between the advanced and intermediate technologies, shown in the left panel of the figure. This "zooming in" spotlights Mexico and the United States. There is a trade-off between ψ and z. Higher levels for ψ, which imply poorer cash-flow control, can be compensated for by higher values of z or by greater efficiency in monitoring, at least up to a point. The points labeled "Mexico" and "USA" indicate the values for ψ and z that are used for these two countries in the simulation. At these two points, the cost of the amalgamated input, q, is the same as in the data for Mexico and the United States; that is, 0.94 and 1.0.

The right panel of Figure 7 portrays the boundary between the entry-level and intermediate technologies. Again, there is trade-off between ψ and z but now it is not as steep because monitoring is less important for these technologies. The point labeled "India" indicates the values for ψ and z for this country in the simulation. At this point, the price for the amalgamated input is 0.6, the value observed in India. The retention problem in India

outputs of the monitoring activity.

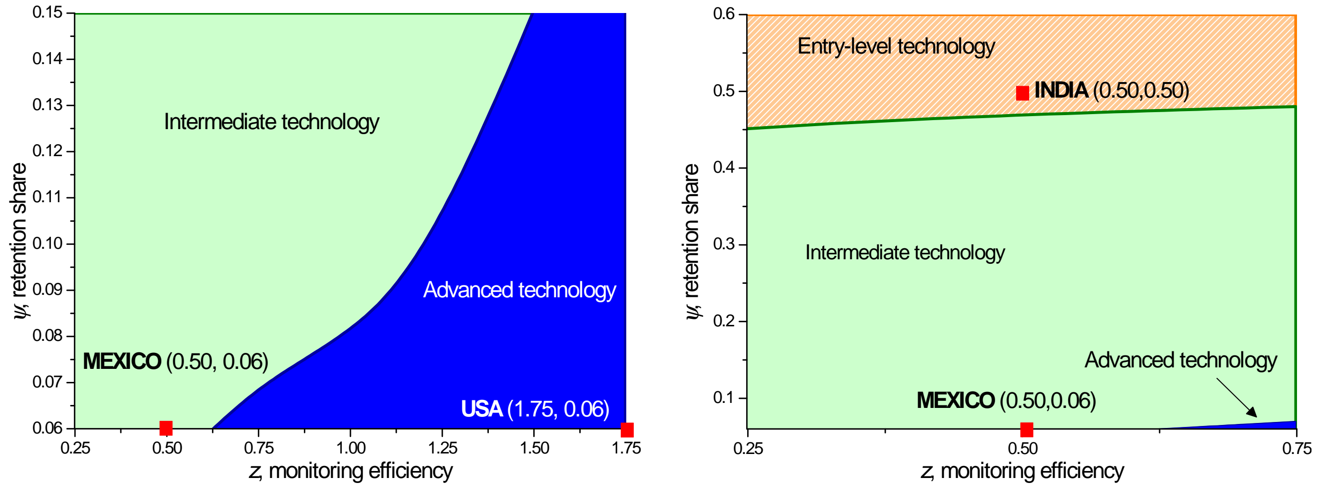

Figure 7: Adoption zones in general equilibrium for the advanced, intermediate, and entry-level technologies.

is so severe that it is far removed from being able to adopt the advanced technology. Last, note that each technology has large adoption zones. Thus, there are many combinations of ψ and z that are consistent with the adoption of the particular technology. In this sense, the analysis is quite robust.

Finally, by how much would the fixed cost, ϕ, for the advanced technology need to lowered so that India and Mexico would adopt it, holding fixed their price for the amalgamated input? India would adopt the advanced technology at 77 percent of the U.S. value and Mexico at 88 percent. Note that because the price of amalgamated input is lower in India and Mexico, firms in these countries would produce more output than in the United States. Thus, the fixed-cost-to-output ratio in India is 43.4 percent of the U.S. one, while for Mexico it is 84.9. If q were allowed to change, then the fixed cost in each country would need to be lowered further to entice adoption. This experiment shows that the fixed costs would need to be lowered considerably to switch the pattern of adoption.

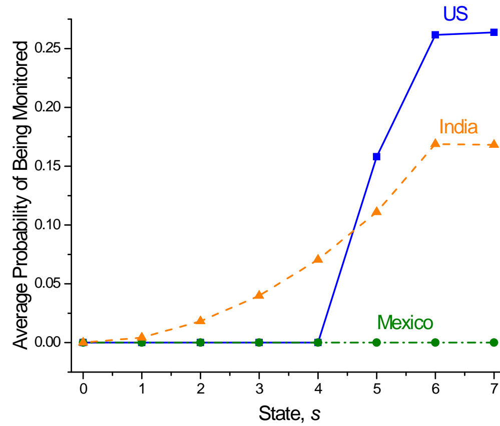

Figure 8: Average monitoring probability by state, $[\sum_{t\geq s} \Pr(s,t)\, p(s,t)] / \sum_{t\geq s} \Pr(s,t)$.

10.1 The Role of Monitoring

Figure 8 illustrates the average odds of being monitored by state, s. As can be seen, for the U.S. technology the odds of being monitored after declaring a stall rise steeply toward the top of the ladder. The U.S. technology ladder is very convex, so the payoffs from lying are greatest at the last steps. At step $s = 7$, the average monitoring probability (across all $t \geq 7$) exceeds 25 percent. Monitoring is also done for the Indian ladder, but for a different reason. Monitoring helps with the cash-flow control problem. With monitoring someone who lies is more likely to be caught. They can then no longer retain part of the firm's cash flow. Very little monitoring is done for the Mexican ladder because monitoring is inefficient and expensive and the cash-flow control problem is small.

11 The Role of Financial Frictions

11.1 Empirical evidence on the availability of financial information and the cost of enforcing contracts

Is the ability of a nation's financial system to produce information and enforce contracts important for its level of output and TFP? Some evidence on this question is presented now. The data used in this section are discussed in Appendix 14. Bushman et al. (2004) construct an index measuring financial transparency in firms across countries. The index is based on six series for each country. The first series measures disclosures about R&D, capital investments, accounting methods, and whether disclosures are broken down across geographic locations, product lines, and subsidiaries. The second measure reflects information about corporate governance, such as the identity and remuneration of key personnel and the ownership structure of the firm. The quality of the information provided by the accounting principles adopted is captured in the third measure. The frequency and timeliness of financial reporting is given in the fourth series. The amount of private information acquisition by private analysts is captured by the number of analysts in a country following large firms. This constitutes the fifth series. The last series proxies for the quality of financial reporting by the media. Bushman et al. (2004) aggregate up these six series using factor analysis into a single index of financial transparency, dubbed "info" here. This variable can be thought of as reflecting z in current analysis. Figure 9 presents scatterplots showing how GDP and TFP are related to this index representing the production of financial information. Both GDP and TFP are positively associated with the series measuring the production of financial information. The relationship is quite tight.

Next, an index is constructed that measures the cost of enforcing contracts in various countries. The underlying data are obtained from The World Bank's *Doing Business* database. In particular, three series are used. The first measures the cost of settling a business dispute. The second series records the number of procedures that must be filed to resolve a dispute. The number of days required to settle a dispute constitutes the last index. These

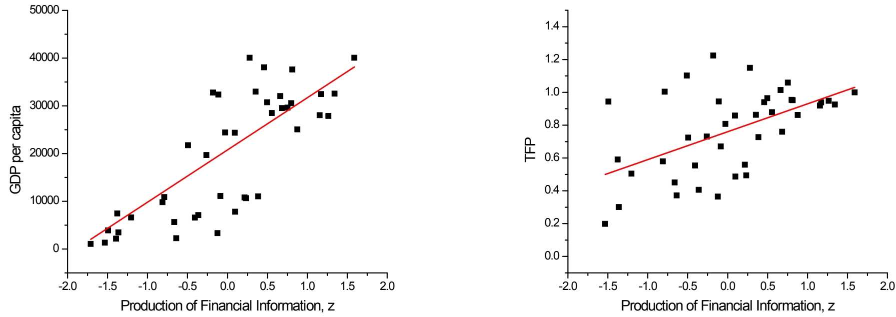

Figure 9: The relationship between the production of financial information, on the one hand, and GDP per capita (left panel) and TFP (right panel) on the other hand.

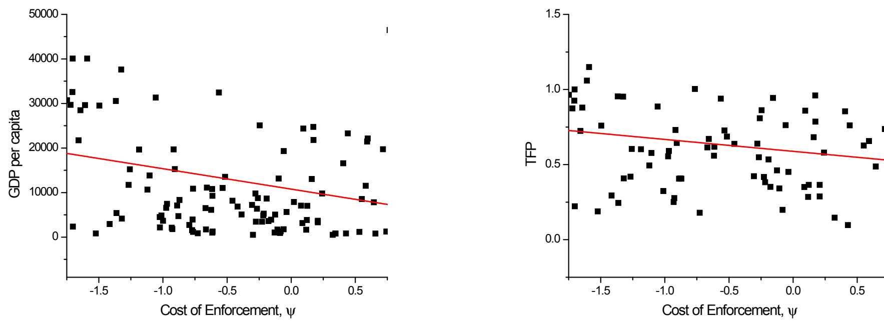

Figure 10: The relationship between the cost of enforcing contracts, on the one hand, and GDP per capital (left panel) and TFP (right panel) on the other hand.

three series are aggregated up using factor analysis into a single index reflecting the cost of contract enforcement, call "enfor." This variable is taken as a measure of ψ. Figure 10 presents scatterplots showing the relationship of GDP and TFP to this index. Both GDP and TFP are negatively related to the cost of contract enforcement. The relationship between the cost of contract enforcement, on the one hand, and GDP or TFP, on the other, is cloudier than the relationship between the production of financial information and either of the latter two variables. Still, the relationships plotted in Figure 10 are statistically significant (at the 1 percent level).

Table 4 presents the results of some regression analysis. This analysis is only intended

Regression Results

Variable	ln(GDP)		ln(TFP)	
info, z	0.69^{***} (0.11)	0.60^{***} (0.13)	0.20^{***} (0.06)	0.20^{***} (0.08)
enfor, ψ	-0.37^{***} (0.07)	-0.33^{***} (0.08)	-0.16^{***} (0.04)	-0.16^{***} (0.05)
findev		$0.00^{\sim}$ (.00)		$-0.00^{\sim}$ (0.00)
constant	9.27^{***} (0.10)	9.07^{***} (0.21)	-0.47^{***} (0.06)	-0.46^{***} (0.12)
R^2	0.75	0.76	0.51	0.51
observations	42	42	40	40

Table 4: *** denotes significance at the 1 percent level and ~ indicates a lack of significance at the 10 percent level. The numbers in parentheses are standard errors.

for illustrative purposes.[18] In particular, both GDP and TFP are positively related with "info" and negatively associated with "enfor." They are also economically and statistically significant. If Kenya increased its financial transparency to the U.S. level, then its GDP and TFP would rise by 215 and 62 percent respectively. Similarly, by reducing the cost of enforcing contracts to the U.S. level, Bangladesh could increase its GDP and TFP by 159 and 69 percent, respectively. Interestingly, when a traditional measure of the efficiency of financial intermediation is added to these regressions, the private-credit-to-GDP ratio (labeled "findev"), it is insignificant in both an economic and a statistical sense. Two measures of collateral requirements and a measure of access to financial markets were also used as the third variable. They, too, are insignificant.

11.2 The Role of External Finance

> "The entrepreneur does not save in order to obtain the means which he needs, nor does he accumulate any goods before he begins to produce."
>
> "The entrepreneur is never the risk bearer. ... The one who gives credit comes to grief if the undertaking fails." Joseph A. Schumpeter (1961, pp. 136 and 137).

[18] A more careful analysis would proceed along the line of the papers surveyed in Levine (2005) and would constitute a paper in its own right.

Schumpeter (1961) believed that finance provided nutrients necessary for entrepreneurial activity to flourish. The current analysis stresses the role of information and the costs of enforcing contracts as key factors in determining the amount of financial capital that can be raised. For simplicity, it has downplayed the importance of start-up funds that the entrepreneurs themselves may have. (Recall that in the quantitative analysis $f = 0$.) From the perspective of the theory developed here, start-up funds from outside parties, such as venture capitalists, should be treated as originating from the intermediary. This does not dismiss as unimportant the wherewithal an entrepreneur brings to a venture, but simply states that the quantitative analysis ignores it. Does this omission matter?

The importance of external finance in the literature is usually gauged by cross-country comparisons of measures of private credit to GDP. The private-credit-to-GDP ratio rises with GDP. This fact can be interpreted as indicating that an entrepreneur's own start-up funds are more important in developing countries vis à vis developed ones, since private debt is smaller in the former countries. Another interpretation is that developed countries use more advanced technologies than developing countries and that these technologies require more funding than less advanced ones. Is the model consistent with observed data on private-debt-to-GDP ratios in India, Mexico and the United States?

How should the amount of debt extended by the intermediary to the firm be measured in the current setting? The notation of debt is state-contingent here. Let $d(s,t)$ represent the amount of debt that the firm owes the intermediary at node (s,t). Suppose that $\{\widehat{x}(s,t)\}$ is some schedule of payments that the firm has committed to paying the intermediary. Many payment schedules can potentially support the *optimal* contract. For a particular schedule, define the debt at node (s,t), which is connected with the payment schedule $\{\widehat{x}(s,t)\}$, by

$$d(s,t+1) = [d(s,t) - \widehat{x}(s,t)]\frac{\Pr(s,t)}{\beta \Pr(s,t+1)}, \text{ for } s < t \text{ and } t < T-1, \tag{23}$$

$$d(s,T) = \widehat{x}(s,T), \text{ for all } s, \tag{24}$$

$$d(s+1,s+1) = [d(s,s) - \widehat{x}(s,s)]\frac{\Pr(s,s)}{\beta\Pr(s+1,s+1)}$$
$$-d(s,s+1)\frac{\Pr(s,s+1)}{\Pr(s+1,s+1)}, \text{ for } 0 < s < S-1, \tag{25}$$

and

$$d(S,S+1) = [d(S,S) - \widehat{x}(S,S)]\frac{\Pr(S,S)}{\beta\Pr(S,S+1)}. \tag{26}$$

The interpretation of (23), which determines debt at off-diagonal points, is relatively straightforward. In state (s,t), the firm owes debt in the amount $d(s,t)$. It discharges the amount $\widehat{x}(s,t)$ so it will then owe $d(s,t) - \widehat{x}(s,t)$. Should it arrive at the next node $(s,t+1)$, it will then be obliged to repay $[d(s,t) - \widehat{x}(s,t)]\{\Pr(s,t)/[\beta\Pr(s,t+1)]\}$ in interest and principal, where the gross state-contingent interest rate is $\Pr(s,t)/[\beta\Pr(s,t+1)]$. Equation (24) is the terminal condition that ensures the debt is paid off at the end node (s,T). The interpretation of (25) is similar, except that as the firm moves up a step along the diagonal, the expected debt at node $(s,s+1)$, which is associated with a stall, is written off the books; the states $(s,s+1)$ and $(s+1,s+1)$ had an odds ratio of $\Pr(s,s+1)/\Pr(s+1,s+1)$. Equation (26) is another terminal condition determining debt at the last step (S,S) along the diagonal. The debt at the step (S,S) must be consistent with the debt at the next off-diagonal step, $(S,S+1)$, along the top of the ladder.

Using the definition for debt at each node, a measure of private debt to GDP can be constructed as

$$\frac{\sum_{t=1}^{T}\sum_{s=0}^{S}\Pr(s,t)\max\{d(s,t),0\}}{\sum_{t=1}^{T}\sum_{s=0}^{S}\Pr(s,t)o(s,t)}.$$

As in the data, only the nonnegative values of $d(s,t)$ are recorded. That is, only those nodes where the firm is indebted to the intermediary are counted. Now, various schedules for $\{\widehat{x}(s,t)\}$ can be constructed that are consistent with the optimal contract. For example, when constructing of $d(s,t)$, let $\widehat{x}(s,t) = x(s,t) - qk(s,t)$. Here the firm can be thought of as using its own cash flow to pay for its inputs instead of surrendering its cash flow to the intermediary and having the intermediary lend the money to buy its inputs. Clearly, this does not change the nature of the contract. This definition for $\widehat{x}(s,t)$ is used below for the

calculation of the private-debt-to-GDP ratios.[19]

What are the model's predictions for the private-debt-to-GDP ratios? These are presented in Table 3. As can be seen, in the data the private-debt-to-GDP ratio is much higher in the United States (1.65) than it is for either India (0.24) or Mexico (0.24). The model captures this fact in a qualitative sense (1.83, 0.08, and 0.08). It does reasonably well in matching the magnitudes for the United States, but it underpredicts the magnitudes for India and Mexico. Again, in the quantitative analysis, the entrepreneur's self-financed start-up funds are set to zero ($f = 0$). This suggests that cross-country differences in the private-debt-to-GDP ratio may reflect differences in technology adoption. That is, more developed countries adopt more advanced technologies that, in turn, require higher levels of financing and hence debt. Selling drinks on the street requires a different level of borrowing than launching rockets into space, so to speak.

11.3 Finance and Development

Imagine endowing India and Mexico with the U.S. financial system. Three questions come to mind: By how much would Mexican and Indian GDP increase? Would this bring them to the U.S. level of development? How much of the gain in output is due to the adoption of new technologies versus capital deepening? To address these questions, let $O(z, \psi; r^{MX}, \chi^{MX})$ represent the level of output that Mexico would produce if it had the financial system proxied for by (z, ψ), *given* the Mexican rental rate on capital, r^{MX}, and the Mexican level of human capital, χ^{MX}. The percentage gain in output from Mexico adopting the U.S. financial system is $100 * [\ln O(z^{US}, \psi^{US}; r^{MX}, \chi^{MX}) - \ln O(z^{MX}, \psi^{MX}; r^{MX}, \chi^{MX})]$. The percentage of the

[19] Just to be clear, when $\widehat{x}(s,t) = x(s,t)$ the firm surrenders all cash flow to the intermediary while the intermediary lends the firm all of the money for its inputs, $qk(s,t)$. (Recall at that in most state/date combinations $x(s,t) = \theta_s k(s,t)^\alpha$.) The two payment schemes $\widehat{x}(s,t) = x(s,t) - qk(s,t)$ and $\widehat{x}(s,t) = x(s,t)$ correspond with different amounts of the firm's cash flow being used to finance the investment, $qk(s,t)$, and hence different levels of external financing. They give the same allocations, however. Clearly, payment schemes, $\widehat{x}(s,t)$, that are convex combinations of the two extremes $x(s,t) - qk(s,t)$ and $x(s,t)$ would work too.

gap in the difference between Mexican and U.S. output that would be closed is measured by $100 * [\ln O(z^{US}, \psi^{US}; r^{MX}, \chi^{MX}) - \ln O(z^{MX}, \psi^{MX}; r^{MX}, \chi^{MX})]/[\ln O(z^{US}, \psi^{US}; r^{US}, \chi^{US}) - \ln O(z^{MX}, \psi^{MX}; r^{MX}, \chi^{MX})]$. Similarly, let $T(z, \psi; r^{MX}, \chi^{MX})$ represent the level of TFP that Mexico would produce if it had the financial system (z, ψ), again given r^{MX} and χ^{MX}. Here TFP is measured in the manner discussed earlier. By standard Solow accounting, the contribution of TFP growth to output growth, when Mexico adopts the U.S. financial system, is just $100*[\ln T(z^{US}, \psi^{US}; r^{MX}, \chi^{MX}) - \ln T(z^{MX}, \psi^{MX}; r^{MX}, \chi^{MX})]/[\ln O(z^{US}, \psi^{US}; r^{MX}, \chi^{MX}) - \ln O(z^{MX}, \psi^{MX}; r^{MX}, \chi^{MX})]$. Do the same thing for India.

Table 5 shows that both Mexico and India could increase their outputs considerably by adopting the U.S. financial system: 46.1 percent and 71.8 percent, respectively. These seemingly large numbers are in accord with those in the quantitative models developed by Buera, Kaboski, and Shin (2011), Greenwood, Wang, and Sanchez (2013), and Townsend and Ueda (2010). Yet, adopting the U.S. financial system would close only 40.0 percent of the gap between Mexican and American incomes and 38.4 percent of the gap for India. This transpires because Mexico and India have lower levels of human capital than the United States and higher prices for physical capital. TFP would jump up by 42.8 percent in Mexico and by 46.4 percent in India. This is a consequence of adopting the U.S. technology.

Interestingly, the capital-to-labor ratio rises by only 9.8 percent in Mexico. In India it moves up by 76.3 percent. The improvement in TFP accounts for 93 percent of Mexican output growth and 65 percent of Indian output growth. This is in line with King and Levine (1994), who report that differences in productivities, and not factor supplies, are likely to explain differences in incomes across countries. The finding here is echoed in Midrigan and Xu (2014), who use a quantitative model and argue that the impact of financial frictions on economic development through the capital-deepening channel, versus a technology adoption one, is likely to be small. Last, the debt-to-output ratio increases by over 300 percent for both India and Mexico. When India and Mexico are endowed with the U.S. financial system they adopt the advanced technology. From (21) it follows that the price of the amalgamated input in both of these countries is the same as in the United States; that is, $q^{IN} = q^{MX} = q^{US} = 1$.

IMPACT OF ADOPTING U.S. FINANCIAL SYSTEM
($z = z^{US}$ AND $\psi = \psi^{US}$)

	Mexico		*India*	
	Increase, %	Gap Closed, %	Growth, %	Gap Closed, %
Output per worker	46.1	40.0	71.8	38.4
TFP	42.8	52.9	46.4	39.9
Capital-to-labor ratio	9.8	10.0	76.3	36.0
Debt-to-output ratio	318.7	100.0	312.3	100.1
	Decomposition of Output Growth, %			
Contribution from TFP	93		65	
Contribution from Capital	7		35	

Table 5: The impact for India and Mexico of adopting the U.S. financial system.

(This does not imply that wages are the same in the three countries because r and χ are different.) Consequently, the same financial contract is offered to entrepreneurs in all three countries, resulting in equal levels of debt relative to output.

12 Conclusions

The role of financial intermediation in underwriting business ventures is investigated here. The analysis stresses the interplay between the structure of technology and the ability of an intermediary to fund it. A dynamic costly state verification model of lending from intermediaries to firms is developed to examine this. The model is embedded into a general equilibrium framework where intermediation is competitive. A firm's level of productivity is private information. The framework has several unique features not found in the literature.

First, the costly state verification model presented has several novel characteristics. As in the conventional costly state verification paradigm, an intermediary is free to audit a firm's returns. The auditing technology imposed here, however, is quite flexible. Specifically, the intermediary can pick the odds of a successful audit. The costs of auditing are increasing and convex in this probability. Additionally, these costs are decreasing in the technological efficiency of the financial system. Additionally, it may not be possible to write a contract that secures, when desired, all of a firm's cash flow. This leakage in cash flow limits the ability

of intermediaries to create incentives for firms that increase the likelihood of a successful venture. Additionally, the analysis allows new firms to supply some of their own funds to help venture start-up.

Second, to stress the nexus between finance and the structure of technology, the latter is given a more general representation than is traditionally found in the finance and development literature. Differences in business opportunities are represented by variations in the stochastic processes governing firms' productivities. A stochastic process is characterized by a non-decreasing movement along a productivity ladder. The positions of the rungs on the ladder and the odds of moving up the ladder differ by the type of venture. A stall on the ladder is an absorbing state.

The form of the technology has implications for finance. Some ventures may have exciting potential for profit, but intermediaries may be required to provide large up-front investments of working capital and have to wait for prolonged periods of time before any potential returns are realized. For such investments, the ability of an intermediary to conduct ex post monitoring and to control cash flows are important for the viability of long-term lending contracts. Monitoring is important for detecting malfeasance. The more efficient the monitoring, the less incentive there will be for a firm to cheat on the financial contract. Likewise, the ability to secure cash flows in the contract is vital for creating incentives using back-loading strategies that improve the odds of a successful venture. The upshot is that the set of desirable technologies within a country is a function of the state of the nation's financial system. Therefore, a country's income and TFP also depend on its financial system.

India, Mexico, and the United States have very different levels of income and TFP. Some evidence is presented suggesting that the three countries use different production technologies. First, Indian production establishments are much smaller than Mexican ones, which in turn are much smaller than U.S. ones. Second, establishment size increases more rapidly with age in the United States than in either India or Mexico. Can differences in technology adoption, due to differences in financial systems, partially explain this? To address this question, the framework is specialized to a situation where there are three technologies: an

advanced technology, an intermediate one, and an entry-level one. An equilibrium is constructed where, given U.S. factor prices and the efficiency of the U.S. financial system, it is optimal to adopt the advanced technology in the United States. Likewise, given Mexican (Indian) factor prices and the Mexican (Indian) financial system, it is best to use the intermediate (entry-level) technology in Mexico (India). This is done while matching the above facts about Indian, Mexican, and U.S. establishment-size distributions, so the analysis has some discipline.

The framework is able to replicate the observed patterns of income and TFP across India, Mexico, and the United States. The analysis predicts that India and Mexico will have lower private-debt-to-GDP ratios than the United States, even though firms in all countries are assumed to have no start-up funds of their own. This transpires from differences in the pattern of investments and cash flows generated by the technologies that firms in the countries adopt. Some cross-country regression analysis is presented that suggests information production and cash flow control are important factors in determining nations' GDPs and TFPs. Finally, the analysis suggests that financial development plays an important role in economic development. Both Mexico and India could increase their GDPs significantly by adopting the U.S. financial system, with the technology adoption channel playing a more important role than the capital accumulation channel. Financial development is important but is not the sole driver of economic development.

13 Appendix: Theory

13.1 The General Contract Problem with Reports at All Dates and States

Consider the general contract problem where reports for all states and dates are allowed. To construct this problem, more powerful notation is needed. To this end, let $\mathcal{H}_t \equiv \{0, 1, \ldots, \min\{t, S\}\}$ represent the set of states that could happen at date t. The set of

all histories for states up to and including date t then reads $\mathcal{H}^t \equiv \mathcal{H}_1 \times \ldots \times \mathcal{H}_t$. Denote an element of $\mathcal{H}^t$, or a history, by $\mathbf{h}^t$. Some of these histories cannot happen. It is not possible for a firm's productivity to advance after a stall, for example. Given this limitation, define the set of feasible or viable histories by $\mathcal{V}^t \equiv \{\mathbf{h}^t \in \mathcal{H}^t : \Pr(\mathbf{h}^t) > 0\}$, where $\Pr(\mathbf{h}^t)$ is the probability of history $\mathbf{h}^t$. To formulate the retention constraints it is useful to define $\mathcal{V}^t(s,j)$ as the set of viable (or feasible) histories that can follow from node (s,j), where $s \leq j \leq t$. The period-t level of productivity conditional on a history, $\mathbf{h}^t$, is represented by $\theta(\mathbf{h}^t)$. Finally, let the state in period j implied by the history h^t read $h_j(\mathbf{h}^t)$ and write the history of states through j as $h^j(\mathbf{h}^t)$.

Let $\zeta_t(\mathbf{h}^t)$ be a report by the firm in period t of its current state to the intermediary, given the true history $\mathbf{h}^t$, where the function $\zeta_t : \mathcal{H}^t \to \mathcal{H}_t$. A truthful report in period t, $\zeta_t^*(\mathbf{h}^t)$, happens when $\zeta_t^*(\mathbf{h}^t) = \zeta_t(\mathbf{h}^t) = h_t(\mathbf{h}^t)$. A reporting strategy is defined by $\zeta^t \equiv (\zeta_1, \cdots, \zeta_t)$. Recall that the firm is unable to report a state higher than it actually has. As a result, the set of all feasible reporting strategies, $\mathcal{S}$, consists of reporting strategies, ζ, such that

1. $\zeta^t(\mathbf{h}^t) \in \mathcal{H}_t$, for all $t \geq 1$ and $\mathbf{h}^t \in \mathcal{H}^t$;

2. $\zeta_t(\mathbf{h}^t) \leq h_t(\mathbf{h}^t)$, for all $t \geq 1$ and $\mathbf{h}^t \in \mathcal{H}^t$.

Taking some liberty with notation, denote the contract elements in terms of the history of reports by $\left\{k(\zeta_t(\mathbf{h}^t), t), x(\zeta_t(\mathbf{h}^t)), p(\zeta_t(\mathbf{h}^t)), \widetilde{f}\right\}_{t=1}^T$. Given this notation, the general contract problem (P3) between the firm and intermediary can be written as

$$\max_{\left\{k(\mathbf{h}^t,t), x(\mathbf{h}^t), p(\mathbf{h}^t), \widetilde{f}\right\}_{t=1}^T} \sum_{t=1}^T \sum_{\mathbf{h}^t \in \mathcal{H}^t} \beta^t \left[\theta(\mathbf{h}^t) k(\mathbf{h}^t, t)^\alpha - x(\mathbf{h}^t)\right] \Pr\left(\mathbf{h}^t\right) + f - \widetilde{f}, \tag{P3}$$

subject to

$$\theta(\mathbf{h}^t) k(\mathbf{h}^t, t)^\alpha - x(\mathbf{h}^t) \geq 0, \tag{27}$$

$$\sum_{t=1}^T \sum_{\mathbf{h}^t \in \mathcal{H}^t} \beta^t \left[\theta(\mathbf{h}^t) k(\mathbf{h}^t, t)^\alpha - x(\mathbf{h}^t)\right] \Pr\left(\mathbf{h}^t\right)$$
$$\geq \max_{\zeta \in \mathcal{S}} \sum_{t=1}^T \sum_{\mathbf{h}^t \in \mathcal{H}^t} \beta^t \left[\theta(\mathbf{h}^t) k(\zeta^t(\mathbf{h}^t), t)^\alpha - x(\zeta^t(\mathbf{h}^t))\right] \prod_{n=1}^t [1 - p(\zeta^n(\mathbf{h}^n))] \Pr\left(\mathbf{h}^t\right),$$

(28)

$$\sum_{t=j}^{T}\sum_{\mathbf{h}^t\in\mathcal{V}^t(s,j)}\beta^t\left[\theta(\mathbf{h}^t)k(\mathbf{h}^t,t)^{\alpha}-x(\mathbf{h}^t)\right]\Pr\left(\mathbf{h}^t\right)$$
$$\geq\psi\sum_{t=j}^{T}\sum_{\mathbf{h}^t\in\mathcal{V}^t(s,j)}\beta^t\theta(\mathbf{h}^t)k(\mathbf{h}^{s-1},s)^{\alpha}\Pr\left(\mathbf{h}^t\right),\ \text{for } s=1,\cdots,S \text{ and } s\leq j\leq T. \tag{29}$$

$$k((\mathbf{h}^{t-1},t),t)=k((\mathbf{h}^{t-1},t-1),t),\ \text{for all } t \text{ where } t-1=h_{t-1}(\mathbf{h}^{t-1}), \tag{30}$$

$$\begin{aligned}k(\mathbf{h}^t,t)&=k((\mathbf{h}^{s-1},s-1),s), &&\text{for all } t>s=h_{s-1}(\mathbf{h}^t)\text{ and } s<S,\\ k(\mathbf{h}^t,t)&=k(\mathbf{h}^S,S), &&\text{for } t>S \text{ and } S=h_S(\mathbf{h}^t),\end{aligned} \tag{31}$$

and

$$\sum_{t=1}^{T}\sum_{\mathbf{h}^t\in\mathcal{H}^t}\beta^t\left[x(\mathbf{h}^t)-C\left(p(\mathbf{h}^t),k(\mathbf{h}^t,t)\right)-qk(\mathbf{h}^t,t)\right]\Pr\left(\mathbf{h}^t\right)-\phi+\widetilde{f}\geq 0, \tag{32}$$

in addition to the self-financing constraint (11). Note how (28) differs from (7). Here a truthful reporting strategy must deliver a payoff in expected present discounted value terms over the entire lifetime of the contract that is no smaller than the one that could be obtained by an untruthful report. The general notation also allows the two no-retention constraints, (12) and (13), to be expressed in the more compact single constraint (29). The objective function (P3) and the rest of the constraints (27), (30) to (32) are the direct analogs of those presented in (P2), so they are not explained.

Turn now to a more restricted problem where the firm is not allowed to make a report that is infeasible; that is, a claim about a zero probability event.[20] The set of restricted reporting strategies, $\mathcal{R}$, consists of all reporting strategies, ζ, such that

1. $\zeta^t(\mathbf{h}^t)\in\mathcal{V}^t$, for all $t\geq 1$ and $\mathbf{h}^t\in\mathcal{V}^t$;

2. $\zeta_t(\mathbf{h}^t)\leq h_t(\mathbf{h}^t)$, for all $t\geq 1$ and $\mathbf{h}^t\in\mathcal{V}^t$.

[20] A similar restriction is made in Kocherlakota (2010).

The restricted contract problem (P4) between the firm and intermediary reads

$$\max_{\{k(\mathbf{h}^t,t),x(\mathbf{h}^t),p(\mathbf{h}^t)\}_{t=1}^T} \sum_{t=1}^{T} \sum_{\mathbf{h}^t \in \mathcal{V}^t} \beta^t \left[\theta(\mathbf{h}^t) k(\mathbf{h}^t,t)^\alpha - x(\mathbf{h}^t)\right] \Pr\left(\mathbf{h}^t\right) + f - \widetilde{f}, \tag{P4}$$

subject to

$$\sum_{t=1}^{T} \sum_{\mathbf{h}^t \in \mathcal{V}^t} \beta^t \left[\theta(\mathbf{h}^t) k(\mathbf{h}^t,t)^\alpha - x(\mathbf{h}^t)\right] \Pr\left(\mathbf{h}^t\right)$$
$$\geq \max_{\zeta \in \mathcal{R}} \sum_{t=1}^{T} \sum_{\mathbf{h}^t \in \mathcal{V}^t} \beta^t \left[\theta(\mathbf{h}^t) k(\zeta^t(\mathbf{h}^t),t)^\alpha - x(\zeta^t(\mathbf{h}^t))\right] \prod_{n=1}^{t} [1 - p\left(\zeta^n(\mathbf{h}^n)\right)] \Pr\left(\mathbf{h}^t\right), \tag{33}$$

$$\sum_{t=1}^{T} \sum_{\mathbf{h}^t \in \mathcal{V}^t} \beta^t \left[x(\mathbf{h}^t) - C\left(p(\mathbf{h}^t), k(\mathbf{h}^t,t)\right) - qk(\mathbf{h}^t,t)\right] \Pr\left(\mathbf{h}^t\right) - \phi + f \geq 0, \tag{34}$$

and (27), (29), (30), and (31) in addition to (11).

The lemma presented below holds.

Lemma 9 *The contracts specified by problems (P3) and (P4) are the same.*

Proof. It will be demonstrated that any contract that is feasible for problem (P3) is also feasible for (P4) and vice versa. Now suppose that $\{k^*(\mathbf{h}^t,t), x^*(\mathbf{h}^t), p^*(\mathbf{h}^t)\}_{t=1}^T$ represents an optimal solution to the general problem (P3). A feasible solution for the restricted problem (P4) will be constructed. To begin with, for reports $\zeta^t(\mathbf{h}^t) \in \mathcal{R}^t$, let

$$\begin{aligned} k^{\sim}(\zeta^t(\mathbf{h}^t),t) &= k^*(\zeta^t(\mathbf{h}^t),t), \\ x^{\sim}(\zeta^t(\mathbf{h}^t)) &= x^*(\zeta^t(\mathbf{h}^t)), \\ p^{\sim}(\zeta^t(\mathbf{h}^t)) &= p^*(\zeta^t(\mathbf{h}^t)), \end{aligned}$$

where a "$\sim$" represents a choice variable in the restricted problem. (Recall that for a truthful report $\zeta^t(\mathbf{h}^t) = \mathbf{h}^t$.)

The general problem also allows for infeasible histories to be reported–that is, for $\zeta^t(\mathbf{h}^t) \in \mathcal{S}^t/\mathcal{R}^t$. For these reports a plausible alternative will be engineered that offers the same payoff to the firm and intermediary and that also satisfies all constraints. To do this, let

$$i = \max_j \zeta^j(\mathbf{h}^t) \in \mathcal{R}^j.$$

Thus, i indexes the duration of feasible reports. Manufacture an alternative plausible history, $\widehat{\zeta}^t(\mathbf{h}^t)$, as follows:

$$\widehat{\zeta}^t(\mathbf{h}^t) = (\zeta^i(\mathbf{h}^t), \underbrace{i, \cdots, i}_{t-i}).$$

Finally, for $\zeta^t(\mathbf{h}^t) \in \mathcal{S}^t/\mathcal{R}^t$ set

$$\begin{aligned} k^{\sim}(\widehat{\zeta}^t(\mathbf{h}^t), t) &= k^*(\zeta^t(\mathbf{h}^t), t), \\ x^{\sim}(\widehat{\zeta}^t(\mathbf{h}^t)) &= x^*(\zeta^t(\mathbf{h}^t)), \\ p^{\sim}(\widehat{\zeta}^t(\mathbf{h}^t)) &= p^*(\zeta^t(\mathbf{h}^t)). \end{aligned}$$

The constructed solution will satisfy all constraints attached to the restricted problem. In particular, a solution to the general problem (P3) will satisfy the incentive compatibility constraint for the restricted problem because $\mathbf{h}^t \in \mathcal{H}^t$ and $\mathcal{R} \subseteq \mathcal{S}$. Therefore, the right-hand side of the incentive constraint for the restricted problem can be no larger than the right-hand side of the incentive constraint for the general problem. Hence, the value of the optimized solution for (P4) must be at least as great as for (P3), since the two problems share the same objective function.

Let $\{k^{\sim}(\mathbf{h}^t, t), x^{\sim}(\mathbf{h}^t), p^{\sim}(\mathbf{h}^t)\}_{t=1}^{T}$ be an optimal solution for the restricted problem (P4). Now, for reports $\zeta^t(\mathbf{h}^t) \in \mathcal{R}^t$, construct a feasible solution to the general problem (P3) as follows:

$$\begin{aligned} k^*(\zeta^t(\mathbf{h}^t), t) &= k^{\sim}(\zeta^t(\mathbf{h}^t), t), \\ x^*(\zeta^t(\mathbf{h}^t)) &= x^{\sim}(\zeta^t(\mathbf{h}^t)), \\ p^*(\zeta^t(\mathbf{h}^t)) &= p^{\sim}(\zeta^t(\mathbf{h}^t)), \end{aligned}$$

where the "*" denotes the quantity in the general problem. The constraints associated with the general problem will be satisfied by this particular solution. Focus on the incentive constraint and take an off-the-equilibrium path report, $\zeta^t(\mathbf{h}^t) \in \mathcal{S}^t/\mathcal{R}^t$. The intermediary can always choose to treat this in the same manner as a report of $(\zeta^i(\mathbf{h}^t), \underbrace{i, \cdots, i}_{t-i})$, with $i = \max_j \zeta^j(\mathbf{h}^t) \in \mathcal{R}^j$, in the restricted problem. Therefore, the value of the optimized solution for (P3) must be at least as great as for (P4). To take stock of the situation, the value of the objective function in problem (P3) must be at least as great as the value returned by problem (P4) and vice versa. Since the objective functions are the same, this can occur only if the optimal solutions for both problems are also the same. ■

Append the no-retention constraints (12) and (13) to problem (P2). It will now be established that the appended version of problem (P2) delivers the same solution as the restricted problem (P4). To do this, the incentive constraint (7) in (P2) must be related to the incentive constraint (33) in (P4). The restricted problem (P4) has just one incentive constraint, which dictates that a truthful reporting strategy must deliver a payoff in expected present discounted value terms over the lifetime of the entire contract that is no smaller than what could be obtained by an untruthful one. Problem (P2) has S incentive constraints requiring that reports along the diagonal in Figure 1 must have payoffs in expected present discounted value terms over the remainder of the contract that weakly dominate those that could be obtained by telling lies.

Lemma 10 *The contracts specified by the appended version of problem (P2) and problem (P4) are the same.*

Proof. The only differences between problems (P2) and (P4) are the incentive constraints, modulo differences in notation used for the states, viz $\mathbf{h}^t$ and (s,t). Knowing $\mathbf{h}^t$ is the same as knowing (s,t), and vice versa, given the structure of the productivity ladder. That is, there is a one-to-one mapping, G, such that $(s,t) = G(\mathbf{h}^t)$ and $\mathbf{h}^t = G^{-1}(s,t)$. It will now be shown that any allocation that satisfies the incentive constraint in one problem must satisfy the incentive constraint in the other. Given this, the two problems must be same.

First, take an allocation $\{k(\mathbf{h}^t, t), x(\mathbf{h}^t), p(\mathbf{h}^t)\}$ that satisfies the incentive constraint (33) for the restricted problem (P4). Consider the *same* allocation $\{k(G(\mathbf{h}^t)), x(G(\mathbf{h}^t)), p(G(\mathbf{h}^t))\}$ for problem (P2). Suppose this allocation violates the incentive constraint (7) in problem (P2) at some node (s^*, s^*). The expected present value of the path following telling the lie at (s^*, s^*) exceeds the expected present value from telling the truth by assumption. The path of truthful reports to this node is unique: There is only one sequence of steps to (s^*, s^*), as is obvious from the structure of the ladder shown in Figure 1. Under a truthful reporting scheme, the paths of potential reports following this node are unique. So too is the path following a lie at (s^*, s^*), because if the firm lies at this node, then it cannot report moving up afterward. Call the path following a lie at (s^*, s^*) the "lie path." Now, in the contract $\{k(\mathbf{h}^t, t), x(\mathbf{h}^t), p(\mathbf{h}^t)\}$ replace the unique potential truthful paths following (s^*, s^*) with the unique lie path. That is, for $t \geq s^*$ replace $\mathbf{h}^t$ with $G^{-1}(s^* - 1, t)$. This must yield a higher expected present value than $\{k(\mathbf{h}^t, t), x(\mathbf{h}^t), p(\mathbf{h}^t)\}$. This is a contraction.

Second, consider some allocation that satisfies the incentive constraint (7) attached to the appended version of problem (P2). Assume that this allocation violates the incentive constraint (33) for problem (P4). This implies that at some nodes (s, s) along the diagonal in Figure 1 it pays to tell lies. Choose the first such state/time pair (s, s), denoted by (s^*, s^*). The path of truthful reports up to this point must be unique. From this point on, the firm cannot report going farther up the ladder. Hence, it cannot tell any further lies. The expected present value of the path following a lie at (s^*, s^*) must exceed the expected present value from telling the truth for (33) to be violated. This implies that (7) must have been violated at node (s^*, s^*), a contradiction. ■

13.2 Proofs for the Contract Problem (P2)

Some lemmas and proofs describing the structure of the optimal contract are now presented. All lemmas and proofs apply to the appended version of problem (P2), where the no-retention constraints (12) and (13) have been added.

13.3 Proof of Go All In

Proof. Let λ be the multiplier associated with the zero-profit constraint (10) and ξ be the multiplier connected with the self-financing constraint (11). The first-order condition linked with $\widetilde{f}$ is

$$-1 + \lambda - \xi = 0.$$

If $\xi > 0$ then the constraint (11) is binding and the result holds automatically. Alternatively, if $\xi = 0$ then $\lambda = 1$. In this situation the firm is indifferent between investing in its own project or placing the funds in a bank. On the one hand, by giving $\widetilde{f}$ to the intermediary the firm lowers its payoff in the objective function by $\widetilde{f}$. On the other hand, this is exactly compensated for by loosening the zero-profit constraint that will result in an decrease in payments from the firm to the intermediary [or the $x(s,t)$'s] in the amount $\widetilde{f}$. ■

Remark 1 *Since $\lambda = 1 + \xi$ and $\xi \geq 0$, it must transpire that $\lambda \geq 1$. This makes intuitive sense. When the entrepreneur hands over wealth to the intermediary, the lowest expected gross return that it can receive is $1/\beta$. This is what a saver earns from depositing funds with the intermediary. This is worth exactly 1 in present value terms.*

13.4 Transformation of Problem (P2) with Self-Financing to Problem (P5) without It

Lemma 1 allows problem (10) with the possibility of self-financing to be converted into an equivalent problem (P5) without self-financing. The latter problem has a larger capitalized value for the fixed costs, $\widehat{\phi}$; specifically, $\widehat{\phi} = \phi - f$.

$$v = \max_{\{k(s,t),x(s,t),p(s,t)\}} \sum_{t=1}^{T} \sum_{s=0}^{\min\{t,S\}} \beta^t \left[\theta_s k(s,t)^{\alpha} - x(s,t)\right] \Pr(s,t), \tag{P5}$$

subject to (6) to (9), the new zero-profit constraint (35) and the no-retention constraints (12) and (13). Note that the self-financing constraint (11) has now been eliminated.

$$\sum_{t=1}^{T} \sum_{s=0}^{\min\{t,S\}} \beta^t \left[x(s,t) - C\left(p(s,t), k(s,t)\right) - qk(s,t)\right] \Pr(s,t) - \underbrace{(\phi - f)}_{\widehat{\phi}} \geq 0. \tag{35}$$

Lemma 11 *(Conversion of problem with self-financing to one without self-financing) The problem with self-financed start-up funds (P2) reduces to problem (P5), where the fixed cost is $\widehat{\phi} = \phi - f$.*

Proof. In line with Lemma 1, set $\widetilde{f} = f$. Use this fact to eliminate $f - \widetilde{f}$ in the objective function and to replace $\widetilde{f}$ with f in the zero-profit condition. ■

Remark 2 *All that matters for the contract is $\phi - f$, given the above lemma. That is, what matters for the contract is the amount of initial funds that the intermediary must put up and this is consistent with many different combinations of ϕ and f. Thus, a project with a fixed cost of ϕ, where the entrepreneur has f in start-up funding will have the same allocations as one where the fixed cost is $\phi - f$, but where the entrepreneur has no start-up funds. Therefore, without cross-country data on ϕ and f separately, it may be difficult to ascertain how much start-up funds matter.*

In what follows the proofs in Sections (13.5), (13.6), and (13.7) refer to the transformed problem (P5).

13.5 Proof of Trust but Verify

Proof. (Sufficiency) It will be shown that the intermediary will monitor the firm at node $(u-1,t)$ (for all $t \geq u$) only if the incentive constraint (7) binds at (u,u). Assume otherwise; that is, suppose to the contrary that the incentive constraint does not bind at (u,u) but that $p(u-1,t) > 0$ for some $t \geq u$. The term $p(u-1,t)$ shows up in only two equations in the appended version of problem (P5): in the zero-profit constraint of the intermediary (35) and on the right-hand side of the incentive constraint (7) at node (u,u). Picture the Lagrangian associated with problem (P5). By setting $p(u-1,t) = 0$, profits to the intermediary can be

increased through the zero-profit constraint (35). This raises the value of the Lagrangian. At the same time, it will have no impact on the maximum problem through the incentive constraint (7) because its multiplier is zero. Therefore, the value of Lagrangian can be raised, a contradiction.

(Necessity) Assume that the incentive constraint (7) binds at (u, u) and that $p(u-1, t) = 0$ for some $t \geq u$. Note that the marginal cost of monitoring is zero at node $(u-1, t)$ since $C_1(0, k(u-1, t)) = 0$. Now increase $p(u-1, t)$ slightly. This relaxes the incentive constraint and thereby increases the value of the Lagrangian. It has no impact on the zero-profit condition (35) as $C_1(0, k(u-1, t)) = 0$. This implies a contradiction because the value of the Lagrangian will increase. ■

13.6 Proof of Backloading: Lemmas 3 and 4

Proof. (Lemma 4, with Lemma 3 a special case) Consider the no-retention constraint (12) at node $(s, s+1)$. Here a stall has just occurred. To satisfy the no-retention constraint at this point the present value of the payments to the firm from there onward must be at least as large as $\psi \sum_{t=s+1}^{T} \beta^t \theta_s k(s,t)^\alpha \Pr(s, j)$. This is what the firm can take by exercising its retention option. This payment, which is necessary, should be made at node (s, T). Thus, at node (s, T) pay the amount $N(s,T) = \psi \sum_{t=s+1}^{T} \beta^t \theta_s k(s,t)^\alpha \Pr(s, j) / [\beta^T \Pr(s,T)]$. Shifting the retention payments along the path $(s, s+1), (s, s+2), \cdots, (s, T-1)$ to the node (s, T), by increasing $x(s, s+1), x(s, s+2), \cdots, x(s, T-1)$ and lowering $x(s, T)$, helps with incentives. It reduces the right-hand side of the incentive constraint (7) at node $(s+1, s+1)$. This occurs because the firm will not receive the retention payment if it is caught lying at some node $(s, s+j)$ for $j > 1$. It has no impact on the right-hand side at other nodes along the diagonal. This shift does affect the left-hand side of (7), for $u < s+1$, and increases it, for $u \geq s+1$. Moreover, if the payments are set according to (2) in the lemma, it follows by construction that the no-retention constraint (12) holds at all nodes (s, t), for $t \geq s+1$. It is not beneficial to pay a retention payment bigger than $N(s,T)$ $= \psi \sum_{t=s+1}^{T} \beta^t \theta_s k(s,t)^\alpha \Pr(s, j) / [\beta^T \Pr(s,T)]$, as will be discussed.

Suppose that $x(s,t) < \theta_s k(s,t)^\alpha$ at some node (s,t), for $s \leq t < T$. It will be established that by setting $x(s,t) = \theta_s k(s,t)^\alpha$ the incentive constraint (7) can be (weakly) relaxed. Suppose $t = s$. Then, increase $x(s,s)$ by $\theta_s k(s,s)^\alpha - x(s,s)$ and reduce $x(S,T)$ by $[\theta_s k(s,s)^\alpha - x(s,s)][\beta^{s-T} \Pr(s,s)/\Pr(S,T)]$. In other words, shift the payment to the firm from node (s,s) to node (S,T) while keeping its present value constant. The left-hand sides of the incentive constraints (7), for $u \leq s$, will remain unchanged. For $u > s$, the left-hand sides will increase. The right-hand sides of the incentive constraints will remain constant, however. Thus, this change will help relax any binding incentive constraints. This shift also helps with the no-retention constraints (13) for $u > s$. Next, suppose that $s < t < T$. Presume that a retention payment is made at (s,T) in the amount $N(s,T)$, as specified by (14). It was argued above that a payment of at least this size must be made at node (s,T) to prevent retention at node $(s,s+1)$. It will be argued below that it is not beneficial to pay a higher amount. For the off-diagonal node (s,t), raise $x(s,t)$ by $\theta_s k(s,t)^\alpha - x(s,t)$ and reduce $x(S,T)$ by $[\theta_s k(s,s)^\alpha - x(s,t)][\beta^{t-T} \Pr(s,t)/\Pr(S,T)]$. This change can only increase the left-hand side of the incentive constraints for $u > s$ and has no impact elsewhere. It reduces the right-hand side at node (s,s). The right-hand sides elsewhere are unaffected. This change also helps with the no-retention constraints (13) for $u > s$. Finally, consider the node (s,T), for $s < S$. A similar line of argument can be employed to show that is not optimal to set $x(s,T) < \theta_s k(s,T)^\alpha - N(s,T)$; that is, to pay a retention payment bigger than $N(s,T)$. ■

Corollary 2 *(Lemma 3) If $\psi = 0$, then $x(s,T) = 0$; that is, it is weakly efficient to take all of a firm's output at every node but (S,T). Thus, Lemma 3 is a special case of Lemma 4.*

13.7 Proof of Efficient Investment

Proof. The first step is to define the first-best allocation. The first-best allocation for working capital solves the following problem:

$$\max_{\{k(s,t)\}} \{\sum_{t=1}^{T} \sum_{s=0}^{\min\{t,S\}} \beta^t \left[\theta_s k(s,t)^\alpha - qk(s,t)\right] \Pr(s,t)\} - \phi,$$

subject to the information and irreversibility constraints, (8) and (9). Now, $k(s,t) = k(s,s+1) = k(s+1,s+1)$ for all $t > s$, by the information and irreversibility constraints. This allows the above problem to be recast as

$$\max_{\{k(s,s+1)\}} \{\sum_{t=1}^{T} \beta^t \sum_{s=0}^{\min\{t,S-1\}} [\theta_s k(s,s+1)^\alpha - qk(s,s+1)] \Pr(s,t)$$
$$+ \sum_{s=0}^{S-1} \beta^{s+1} [\theta_{s+1} k(s,s+1)^\alpha - qk(s,s+1)] \Pr(s+1,s+1)\} - \phi.$$

Focus on some $k(s,s+1)$. It will show up in the top line of the objective function whenever $t \geq s+1$. The first-order condition for $k(s,s+1)$ that is connected with this problem reads

$$\sum_{t=s+1}^{T} \beta^t \left[\alpha\theta_s k(s,s+1)^{\alpha-1} - q\right] \Pr(s,t) + \beta^{s+1} \left[\alpha\theta_{s+1} k(s,s+1)^{\alpha-1} - q\right] \Pr(s+1,s+1) = 0.$$

For the second step, focus on the appended version of problem (P5). Now, using the information, irreversibility, and zero-profit constraints, (8), (9), and (35), in conjunction with the solution for the $x(s,t)$'s presented in Lemma 4, the contracting problem can be rewritten as

$$\max_{\{k(s,s+1),p(s,t)\}} \{\sum_{t=1}^{T} \beta^t \sum_{s=0}^{\min\{t,S-1\}} [\theta_s k(s,s+1)^\alpha - C(p(s,t),k(s,s+1)) - qk(s,s+1)] \Pr(s,t)$$
$$+ \sum_{s=0}^{S-1} \beta^{s+1} [\theta_{s+1} k(s,s+1)^\alpha - C(p(s+1,s+1),k(s,s+1)) - qk(s,s+1)] \Pr(s+1,s+1)\} - \phi,$$

subject to the $2S$ incentive and diagonal-node no-retention constraints:

$$\sum_{t=1}^{T} \beta^t \sum_{s=u}^{\min\{t,S-1\}} \left[\theta_s k(s,s+1)^\alpha - C(p(s,t),k(s,s+1)) - qk(s,s+1)\right] \Pr(s,t)$$
$$+\sum_{s=u-1}^{S-1} \beta^{s+1} \left[\theta_{s+1} k(s,s+1)^\alpha - C(p(s+1,s+1),k(s,s+1)) - qk(s,s+1)\right] \Pr(s+1,s+1) - \phi$$
$$-\sum_{s=u}^{u-1} \psi\theta_s k(s,s+1)^\alpha \sum_{t=s+1}^{T} \beta^t \Pr(s,t)$$
$$\geq k(u-1,u)^\alpha \{\sum_{i=u}^{S} (\theta_i - \theta_{u-1}) \{\sum_{j=i}^{T} \beta^j \Pr(i,j) \prod_{n=u}^{j} [1 - p(u-1,n)]$$
$$+ \beta^T \Pr(i,T) \prod_{n=u}^{T} [1 - p(u-1,n)] [\psi \frac{\sum_{t=u}^{S} \beta^t \Pr(u-1,t)}{\beta^T \Pr(u-1,T)}]\},$$

and

$$\sum_{t=1}^{T} \beta^t \sum_{s=u}^{\min\{t,S-1\}} \left[\theta_s k(s,s+1)^\alpha - C(p(s,t),k(s,s+1)) - qk(s,s+1)\right] \Pr(s,t)$$
$$+\sum_{s=u-1}^{S-1} \beta^{s+1} \left[\theta_{s+1} k(s,s+1)^\alpha - C(p(s+1,s+1),k(s,s+1)) - qk(s,s+1)\right] \Pr(s+1,s+1) - \phi$$
$$-\sum_{s=u}^{u-1} \psi\theta_s k(s,s+1)^\alpha \sum_{t=s+1}^{T} \beta^t \Pr(s,t)$$
$$\geq \psi k(u-1,u)^\alpha \sum_{t=u}^{T} \sum_{s=u}^{\min\{t,S\}} \beta^t \theta_s \Pr(s,t),$$

for $u = 1, \cdots, S$. Let ι_u and ν_u represent the multipliers attached to the uth incentive and diagonal-node no-retention constraints, respectively. Now suppose that after some diagonal node (t^*, t^*) that neither the incentive nor diagonal-node no-retention constraints ever bind again; that is, let (t^*, t^*) be the last diagonal node at which one or both of incentive and no-retention constraints bind. Consider one of the incentive or no-retention constraints up to and including node (t^*, t^*). The variable $k(s, s+1)$ will not show up on the right-hand side of any of these constraints. Examine the left-hand side. The variable $k(s, s+1)$ appears in the first line whenever $t \geq s+1$. It does not appear in the third line because $s \leq u-1$.

Therefore, the first-order condition for $k(s,s+1)$ is

$$[1+\sum_{j=1}^{t^*}(\iota_j+\nu_j)]\{\sum_{t=s+1}^{T}\beta^t\left[\alpha\theta_s k(s,s+1)^{\alpha-1}-q\right]\Pr(s,t)$$
$$+\beta^{s+1}\left[\alpha\theta_{s+1}k(s,s+1)^{\alpha-1}-q\right]\Pr(s+1,s+1)\}=0,$$

for $s\geq t^*$. Recall that $p(s,t)=0$ whenever the incentive constraint does not bind by Lemma 2, so that $C_2(0,k(s,s+1))=0$.

Turn to the last step. Divide the above first-order condition by $1+\sum_{j=1}^{t^*}(\iota_j+\nu_j)$. It now coincides with the one for the planner's problem. Thus, investment is efficient. ■

13.8 Efficient Self-Financing

Proof. Let $k^*(s,t)$ denote the allocations that are associated with the efficient investment plan. Set $\widehat{f}$ to

$$\widehat{f}=\sum_{t=1}^{T}\sum_{s=0}^{\min\{t,S\}}\beta^t k^*(s,t)\Pr(s,t)+\phi.$$

It will be shown that the efficient allocation is optimal when $f=\widehat{f}$. To see this, set

$$x(s,t)=p(s,t)=0.$$

This plan for the $p(s,t)$'s and $x(s,t)$'s will satisfy the incentive constraints and limited liability constraints. The zero-profit constraint is also satisfied because the amount that the intermediary initially receives from the entrepreneur, $\widehat{f}$, covers the expected discounted cost of the project. The information and irreversibility constraints are satisfied by construction. The expected discounted return to the firm is

$$\sum_{t=1}^{T}\sum_{s=0}^{\min\{t,S\}}\beta^t\theta_s k^*(s,t)^{\alpha}\Pr(s,t)-\widehat{f},$$

which is of course the level of expected discounted profits occurring in the first-best allocation. Essentially, the firm is turning over to the intermediary sufficient funds to finance to the present value of investments. The firm can then keep the resulting cash flow. ■

Remark 3 *The efficient allocation may be supported at lower levels of self-financed start-up funding than* $\widehat{f}$.

13.9 Proof of Technology Switching

Proof. If at some level of wealth $f_{\tau,\upsilon}$ the more advanced technology υ is not preferred to technology τ, then

$$v(f_{\tau,\upsilon};\upsilon) - v(f_{\tau,\upsilon};\tau) < 0.$$

By the previous lemma, at the level of wealth, $\widehat{f}_\upsilon$, the more advanced technology delivers the first-best level of expected discounted profits. This cannot be replicated by technology τ. Hence,

$$v(\widehat{f}_\upsilon;\upsilon) - v(\widehat{f}_\upsilon;\tau) > 0.$$

Now, $v(f;\tau)$ is continuous in f for τ and υ. By the intermediate value theorem there exists at least one threshold level of wealth, $f^*_{\tau,\upsilon} \in [f_{\tau,\upsilon}, \widehat{f}_\upsilon]$, such that

$$v(f^*_\tau;\upsilon) - v(f^*_\tau;\tau) = 0.$$

Thus, the function $v(f;\upsilon) - v(f;\tau)$ must cross 0 at least once. Now pick

$$\mathcal{F}_{\tau,\upsilon} = \{f : v(f;\upsilon) - v(f;\tau) > 0\}. \tag{36}$$

■

13.10 Proof of Coexisting Technologies

Proof. Suppose that at some level of self-financed start-up funds, f, a firm picks technology, τ^*, in line with (15). Now, consider some more advanced technology $\upsilon > \tau^*$. The firm will prefer υ to τ^* whenever $f \in \mathcal{F}_{\tau^*,\upsilon}$, where $\mathcal{F}_{\tau^*,\upsilon}$ is defined by (36). Technology υ may not maximize $v(f;\tau)$. Denote the optimal technology by υ^*; note $\upsilon^* \neq \tau^*$ since υ^* is preferred to υ, which in turn is preferred to τ^*. Pick any $f' \in \mathcal{F}_{\tau,\upsilon}$. Let υ^* be the optimal technology that is associated with this level of start-up funding. Then, the technologies τ^* and υ^* will coexist whenever there are entrepreneurs with the start-up funds f and f'. ■

14 Appendix: Data

14.1 Table 1

Average establishment size. Data for average establishment size are from different sources for each country. (i) The number for India is based on information obtained from two sources: the Annual Survey of Industries (ASI) for 2007-08, which gathers data on formal sector manufacturing plants, and the National Sample Survey Organization (NSSO) for 2005-06, which collects data on informal sector manufacturing establishments. (ii) The figure for Mexico is calculated using data from Mexico's 2004 Economic Census conducted by INEGI. (iii) The number for the United States is derived from figures published in the 2002 Economic Census published by the U.S. Census Bureau.

14.2 Figure 4

A special request was made to obtain these data. Data for the United States are from the 2002 Economic Census published by the U.S. Census Bureau. They can be obtained using the U.S. Census Bureau's FactFinder. These are businesses that have no paid employees but are subject to federal income tax in the United States.

United States

	Raw Data		Cumulative Share	
	Est	Empl	Est	Empl
All establishments	350,828	14,699,536	0	0
Establishment Size				
1 to 4, employees	141,992	279,481	40.5	1.90
5 to 9	49,284	334,459	54.5	4.18
10 to 19.	50,824	702,428	69.0	8.96
20 to 49	51,660	1,615,349	83.7	19.94
50 to 99	25,883	1,814,999	91.1	32.29
100 to 249	20,346	3,133,384	96.9	53.61
250 to 499	6,853	2,357,917	98.9	69.65
500 to 999	2,720	1,835,386	99.6	82.13
1,000 to 2,499	1,025	1,494,936	99.9	92.30
2,500 or more	241	1,131,197	100.0	100.00
Mean establishment size		41.9		

14.3 Figure 5

The data for India, Mexico, and the United States displayed in Figure 5 are from Hsieh and Klenow (2014). The table below shows the statistics used to construct Figure 5.

Hsieh and Klenow (2014) Facts

	Employment Share		
Establishment Age, yr.	U.S (2002)	Mexico (2003)	India (1994)
<5	0.137	0.280	0.282
5-9	0.110	0.235	0.224
10-14	0.115	0.173	0.155
15-19	0.092	0.100	0.089
20-24	0.074	0.077	0.067
25-29	0.072	0.039	0.043
30-34	0.072	0.035	0.036
35-39	0.049	0.019	0.018
>39	0.280	0.041	0.086

14.4 Section 11

The data used for real GDP and TFP derive from Penn World Table 8. For each country an average value for these series is calculated from 1995 on. The information variable is the

FACTOR1 series presented in Bushman et al. (2004, Appendix B). Three series from The World Bank's *Doing Business* database are aggregated using factor analysis to obtain an index for the cost of enforcing contracts. The series are time (days), cost (% of claims), and procedures (number). For each country an average of these series was taken from 2003 on. Last, the series on financial development are taken from The World Bank's *Global Financial Development* dataset. The series used for "findev" is private credit by deposit money banks and other financial institutions to GDP (%). Here an average from 2005 on is taken. Three other series were also entered as the additional third variable in the regression: viz, firms identifying access to finance as a major constraint (%), loans requiring collateral (%), value of collateral needed for a loan (% of the loan amount). These series had no predictive power in the regressions (albeit they reduced the sample size) and so are omitted from the reporting.

15 Appendix: Some Two-Period Examples

Some simple two-period examples illustrating the contracting setup are presented here. They show how the shape of the productivity profile and the size of the fixed cost connected with a blueprint influence the form of the contract. They also demonstrate the importance that monitoring, retention, and self-financing play in the design of a contract. Last, a connection is drawn between the productivity ladder and survival probabilities, on the one hand, and the aggregate distribution of employment by plant age, on the other.

In all examples, a blueprint, b, is described by the quadruple $b \equiv \{\theta_0 = 0, \theta_1 > 0, \theta_2 \geq \theta_1, \phi \geq 0\}$. Output is produced in accordance with the Leontief production function $o = \min\{\theta, k\}$. The cost of the amalgamated input, q, is set to zero.

15.1 The Importance of Monitoring and Self-Financed Start-Up Funds

The first example focuses on the importance of monitoring and self-finance. To this end, let the cost of monitoring be *prohibitive* and abstract away from the issue of retention; in

particular, set $z = \psi = 0$. A venture's survival is guaranteed, implying $\sigma = 1$. Therefore, a project is financed only when a feasible backloading strategy exists. This strategy must induce the firm to repay the intermediary enough to cover the fixed cost of the venture. The entrepreneur has f in self-financed start-up funds, which he can contribute to financing the project. Therefore, assume $f < \phi$. The entrepreneur turns all of his funds over to the intermediary at the start of the venture.

The first-best production allocation is very easy to compute in the example. Simply set $k(0,1) = k(1,1) = k(0,2) = \theta_1$ and $k(1,2) = k(2,2) = \theta_2$. As a result, the first-best expected profit, π, from implementing the blueprint is

$$\pi \equiv \beta\rho\theta_1 + \beta^2\rho(1-\rho)\theta_1 + \beta^2\rho^2\theta_2 - \phi.$$

Now, focus on the set of blueprints, $\mathcal{B}$, that potentially yield some first-best expected level of profits, π:

$$\mathcal{B}(\pi) \equiv \{\theta_0 = 0, \theta_1 > 0, \theta_2 \geq \theta_1, \phi \geq 0, \beta\rho\theta_1 + \beta^2\rho(1-\rho)\theta_1 + \beta^2\rho^2\theta_2 - \phi = \pi\}.$$

Which blueprints $b \in \mathcal{B}(\pi)$ can actually attain the first-best level of expected profits, π?

Because monitoring is prohibitively expensive, backloading is the only way to satisfy the incentive constraints at nodes $(2,2)$ and $(1,1)$. Backloading implies that the firm receives a return of $(\pi + f)/(\beta^2\rho^2)$ at node $(2,2)$ and nothing elsewhere. (Recall that the intermediary earns zero profits.) If the firm reports θ_1 at node $(2,2)$, or lies, it can pocket $\theta_2 - \theta_1$. Hence, satisfying the incentive constraint at node $(2,2)$ requires that $(\pi + f)/(\beta^2\rho^2) \geq \theta_2 - \theta_1$ or

$$\theta_2 \leq \theta_1 + (\pi + f)/(\beta^2\rho^2). \tag{37}$$

Observe that backloading will work only when the total expected payoff of the project is not too concentrated on the highest productivity state, θ_2. Or, in other words, the productivity profile cannot be too convex.

Next, consider the incentive constraint at node $(1,1)$. By misreporting θ at this node, the firm can guarantee itself $\theta_1 - \theta_0 = \theta_1$ in both periods 1 and 2. By declaring a stall,

however, its capital stock will be locked in at θ_1, in accord with the irreversibility constraint (9). Satisfying the incentive constraint at this node therefore requires that the expected payoff from truthfully reporting $\theta = \theta_1$, in the hope of reaching node $(2, 2)$ and receiving $(\pi + f)/(\beta^2\rho^2)$, dominates the payoff from lying and claiming $\theta = \theta_0 = 0$. Thus, it must transpire that $\pi + f \geq \beta\rho\theta_1 + \beta^2\rho\theta_1$, implying that

$$\theta_1 \leq (\pi + f)/[(1 + \beta)(\rho\beta)]. \tag{38}$$

Hence, when θ_1 is large relative to the project's expected profits, π, it pays for the firm to lie in the first period. The first-best allocation cannot be supported.

There are two additional constraints to consider. First, $\theta_1 \leq \theta_2$, by assumption. Second, recall that $\phi \geq 0$. This implies the restriction $\beta\rho\theta_1 + \beta^2\rho(1 - \rho)\theta_1 + \beta^2\rho^2\theta_2 - \pi \geq 0$, which can be rewritten as

$$\theta_2 \geq \pi/(\beta^2\rho^2) - \{\beta\rho[1 + \beta(1 - \rho)]/(\beta^2\rho^2)\}\theta_1. \tag{39}$$

To understand the impact of variations in the fixed cost, set $\phi = 0$. It is a simple matter to show that both incentive constraints must hold. In this situation, all of the returns from the project will be given to the firm. The payoff from lying arises solely from the possibility of evading the fixed cost. As ϕ increases, the first-best gross profits of the blueprint, $\beta\rho\theta_1 + \beta^2\rho(1 - \rho)\theta_1 + \beta^2\rho^2\theta_2$, rise to keep net profits constant. A larger fraction of the gross profits must be paid back to the intermediary to cover the fixed cost. This makes it harder to satisfy the incentive constraints.

The right panel of Figure 11 plots the two incentive constraints (37) and (38), the 45-degree line, and the fixed-cost constraint (39). The shaded quadrilateral illustrates the values of θ_1 and θ_2 where the first-best allocation can be implemented using a backloading strategy, given the four constraints. Again, a high value of θ_1 will cause the node $(1, 1)$ incentive constraint to bind. Why? When θ_1 is high, then either θ_2 must be relatively small or ϕ relatively large to maintain the fixed level of profits, π. It pays for the firm to lie at node $(1, 1)$ when $k(1, 1) = \theta_1$. Likewise, when θ_2 is large, the incentive constraint at node $(2, 2)$ will bite. The left panel shows how the two incentive constraints shift inward as the

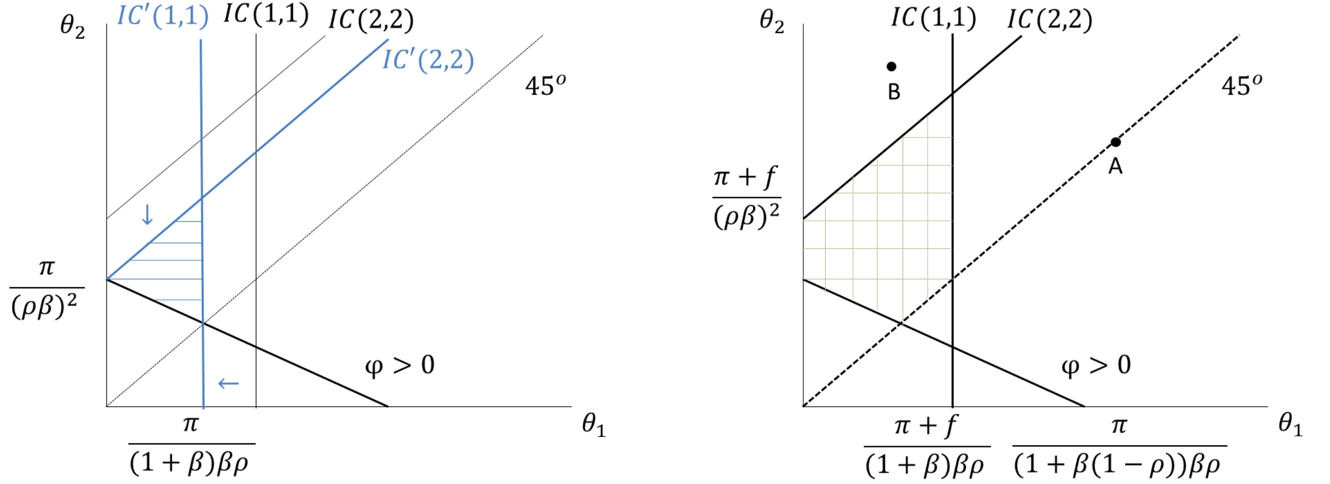

Figure 11: The right panel shows the set of implementable first-best allocations. The left panel illustrates what happens when the level of self-financing drops to zero.

funds that the firm can use for self-financing shrink. The case where $f = 0$ is plotted in the diagram. The set of implementable first-best allocations shrinks from the quadrilateral to the triangle. With self-financing the intermediary funds a smaller fraction of the fixed cost, which reduces the incentive for the firm to lie.

Consider a point, such as A, where $\theta_1 = \theta_2$. In this case, the incentive constraint (38) collapses to $\theta_1 \leq \theta_1 - (\phi - f)/[(1+\beta)(\rho\beta)]$. Then, the first-best allocation cannot be supported if $\phi > f$; that is, when the project cannot be fully self-financed. Hence, the implication of this constraint is that the first-best payoff from the project cannot be supported when the productivity profile is too concave—that is, when θ_2 is close in value to θ_1. Thus, second-best allocations must be entertained. A contract can be written that supports a second-best allocation, provided that $\phi - f < \pi\beta/(1+\beta)$; this condition is explained below. Interestingly, advancing the firm a level of working capital below θ_1 may help to satisfy the first-period incentive compatibility constraint, so that here $k(1,1) = k(1,1) = k(0,2) < \theta_1$. This is because reducing the funding has a larger impact on the payoff to misreporting at node $(1,1)$ than it does to overall profits π, and thereby helps to generate a gradually increasing payoff profile. To see this, suppose that the firm will lie in period 1 when

$\theta = \theta_1$. The expected profits from this lying strategy would be $\rho(\beta+\beta^2)k(1,1)$. Alternatively, the firm could tell the truth. Then, it will receive $\rho\beta k(1,1) + \rho\beta^2\theta_1 - (\phi - f)$. To maintain indifference between these two strategies, set $\rho(\beta+\beta^2)k(1,1) = \rho\beta k(1,1) + \rho\beta^2\theta_1 - (\phi - f)$. This implies $k(1,1) = \theta_1 - (\phi - f)/(\rho\beta^2) < \theta_1$. The condition that $\phi - f < \pi\beta/(1+\beta)$ ensures that the payoff from telling the truth net of the funds used for self-finance is nonnegative, given the level of capital investment, $k(1,1)$. When $\phi - f > \pi\beta/(1+\beta)$, it is not feasible to use such a strategy.

Finally, focus on a point such as B. Now, the incentive constraint at the $(2,2)$ node is violated, so that $\theta_2 \geq \theta_1 + (\pi + f)/(\beta\rho)^2$. This implies that $\theta_1 < (\phi - f)/[\beta\rho(1+\beta)]$. All expected profits derive solely from the return to node $(2,2)$, because the discounted expected value of the returns at nodes $(1,1)$ and $(1,2)$, or $[\beta\rho + \beta^2\rho(1-\rho)]\theta_1$, is insufficient to cover the fixed cost net of the portion funded by the firm, $\phi - f$. Therefore, there are not enough resources available to employ a backloading strategy that will entice the firm to tell the truth at node $(2,2)$. That is, there are no profits–only losses–that the intermediary can redirect to node $(2,2)$ from the other nodes on the tree. The firm avoids these losses by lying. Monitoring must be used to implement such a point. If monitoring is perfectly efficient ($z = \infty$), then the first-best allocations can be supported at point B. When monitoring is efficient, the first-best allocation can also be obtained at point A. Therefore, in economies with poor monitoring the choice set for technologies is limited to those blueprints that can be implemented with backloading strategies. With better monitoring this choice set is expanded to include technologies that cannot be implemented with backloading alone.

15.2 Costly Cash-Flow Control

The second example focuses on how costly cash-flow control influences the design of the contract. To keep things simple, assume that the entrepreneur has no funds available for self-financing start-up costs; that is, set $f = 0$. All of the remaining features of the previous example are retained but now $\psi \geq 0$.

15.2.1 The No-Retention Constraints

The firm now has the ability to retain the fraction ψ of output at any node on the ladder. The nodes $(0,1)$ and $(0,2)$ can be ignored because $\theta_0 = 0$, so there is nothing for the firm to retain here. Focus on the second period first. Suppose that the firm finds itself at node $(1,2)$; that is, it stalls after reaching θ_1. The firm will retain $\psi\theta_1$ units of output here. This event has an expected discounted value of $\beta^2\rho(1-\rho)\psi\theta_1$. Alternatively, consider the case where the firm declares that it has reached node $(2,2)$. Here it will receive the amount $[\pi - \beta^2\rho(1-\rho)\psi\theta_1]/(\beta^2\rho^2)$. Note that the firm's profits have been reduced by the necessity for the intermediary to make a retention payment at node $(2,1)$. The firm can choose to retain the amount $\psi\theta_2$ at node $(2,2)$. Thus, the no-retention constraint at node $(2,2)$ requires that $[\pi - \beta^2\rho(1-\rho)\psi\theta_1]/(\beta^2\rho^2) \geq \psi\theta_2$. This can be rearranged to get

$$\theta_2 \leq -[(1-\rho)/\rho]\theta_1 + \pi/(\psi\beta^2\rho^2).$$

The line $RC(2,2)$ in the left panel of Figure 12 illustrates the no-retention constraint. It slopes downward.

Move back in time to period 1, specifically to node $(1,1)$. If the firm moves to node $(2,2)$, it will earn profits in the amount $[\pi - \beta^2\rho(1-\rho)\psi\theta_1]/(\beta^2\rho^2)$. This occurs with probability ρ. If it moves to node $(1,2)$, then it will receive $\psi\theta_1$. Therefore, its expected discounted profits from telling the truth at node $(1,1)$ are $\beta\rho[\pi - \beta^2\rho(1-\rho)\psi\theta_1]/(\beta^2\rho^2) + (1-\rho)\beta\psi\theta_1$ $= \pi/(\beta\rho)$. If the firm decides to exercise its retention option, it will receive $(1+\beta)\psi\theta_1$. In this circumstance, the intermediary will not increase the working capital to θ_2 (from θ_1). The period-1 no-retention constraint dictates that $\pi/(\beta\rho) \geq (1+\beta)\psi\theta_1$, or that

$$\theta_1 \leq \pi/[(1+\beta)(\rho\beta\psi)].$$

This is shown by the curve $RC(1,1)$ in the left panel of Figure 12.

15.2.2 The Incentive Compatibility Constraints

The incentive compatibility constraints are also affected by the firm's ability to retain cash flow. Consider the incentive constraint at node $(2,2)$ first. As just discussed, when the firm

tells the truth, then it will receive $[\pi - \beta^2\rho(1-\rho)\psi\theta_1]/(\beta^2\rho^2)$. When the firm lies, it can now pocket $\theta_2 - \theta_1 + \psi\theta_1$. Therefore, satisfying the period-2 incentive constraint requires that $[\pi - \beta^2\rho(1-\rho)\psi\theta_1]/(\beta^2\rho^2) \geq \theta_2 - \theta_1 + \psi\theta_1$. This constraint can be rewritten as

$$\theta_2 \leq [(\rho - \psi)/\rho]\theta_1 + \pi/(\beta^2\rho^2).$$

The incentive compatibility constraint is represented in the left panel of Figure 12 by the line $IC^{\psi}(2,2)$. Note that it lies below the old curve $IC(2,2)$, because $(\rho-\psi)/\rho < 1$. In fact, it will slope down when $\psi > \rho$.

Move back in time to node $(1,1)$. The profits from lying will be $(1+\beta)(\theta_1 + \psi\theta_0) = (1+\beta)\theta_1$, because $\theta_0 = 0$. As was mentioned, the expected profits from telling the truth are $\beta\rho\pi$. Therefore, the period-1 incentive constraint is the same as before:

$$\theta_1 \leq \pi/[(1+\beta)(\rho\beta)].$$

Hence, the old $IC(1,1)$ curve will still apply for period 1.

15.2.3 The Upshot

Observe that the period-1 retention constraint will be automatically satisfied when the first-period incentive constraint holds; therefore, it can be dropped from the analysis. Now, the shaded triangle on the far left side in the left panel of Figure 12 shows those (θ_1, θ_2) combinations that satisfy the period-2 no-retention constraint, $RC(2,2)$, but not the incentive compatibility constraint, $IC^{\psi}(2,2)$. The (θ_1, θ_2) combinations that satisfy $IC^{\psi}(2,2)$, but not $RC(2,2)$, are shown by the hatched triangle on the right. Note that the triangle on the left admits higher θ_2/θ_1 ratios than the one on the right. Thus, the no-retention constraint does not penalize convex productivity ladders as much as the incentive constraint does. The fact that the $IC^{\psi}(2,2)$ slopes upward implies that it does not restrict the absolute sizes of θ_1 and θ_2; it is a restriction on how large θ_2 can be relative to θ_1 (for a given expected level of net profits). By contrast, along the $RC(2,2)$ constraint an increase in θ_2 must be met by a decrease in θ_1. If the firm can retain more cash flow in the second period, then the amount

that it can retain in the first period must be decreased, so the payoff from exercising the no-retention option in the second period becomes larger (again, for a given level of expected net profits). Furthermore, the $IC^{\psi}(2,2)$ curve rotates downward as ψ rises. Thus, retention worsens the incentive problem because the payoff from lying increases when it can retain some of the output. Hence, retention further limits the ability to implement convex profiles and makes monitoring even more important.

The right panel in Figure 12 illustrates the upshot of the above analysis. Note that the $RC(2,2)$ constraint is located above the $\phi > 0$ constraint, since $\pi/(\psi\beta^2\rho^2) \geq \pi/(\beta^2\rho^2)$ and $\pi/[(1-\rho)]\beta^2\rho > \pi/\{[1+\beta(1-\rho)]\beta\rho\}$. The hatched area illustrates the values of θ_1 and θ_2 where the first-best allocation can be supported using a backloading strategy. This area has shrunk due to the costly cash-flow control problem and lies within the old triangle.

15.3 Identifying the Productivity Ladder

The two-period example is resurrected here to illustrate the connection between the productivity ladder and survival probabilities, on the one hand, and the aggregate distribution of employment by plant age, on the other. It will be shown that in order for old firms to account for more of aggregate employment than young firms, it must transpire that $\theta_2/\theta_1 > 1$ and that this ratio must rise with the plant death probability $1-\sigma$. The example illustrates how the productivity ladder can be identified in the applied analysis using data on the age distribution of employment. To see this, take the structure of the earlier examples but assume now that $\sigma \leq 1$, $z = \infty$, and $\psi = 0$; thus, a firm's survival is not ensured, monitoring is perfect, and there is no retention. The first-best solution will obtain. Consequently, $k(0,1) = k(1,1) = k(0,2) = \theta_1$ and $k(1,2) = k(2,2) = \theta_2$. Let $k(s,t) = \min\{\widetilde{k}(s,t), l(s,t)\}$ so that the amount of labor used corresponds with the amount of working capital. (Note that $q = 0$ implies that $r = w = 0$.)

Employment by young and old firms in the economy is given by $\sigma\rho\theta_1$ and $\sigma^2\rho(1-\rho)\theta_1 + \sigma^2\rho^2\theta_2$. Note that $\sigma\rho\theta_1 \lesseqqgtr \sigma^2\rho(1-\rho)\theta_1 + \sigma^2\rho^2\theta_2$ as $\theta_1 \lesseqqgtr \sigma[(1-\rho)\theta_1 + \rho\theta_2]$. Therefore, to have old firms accounting for more employment than young ones, when survival is not

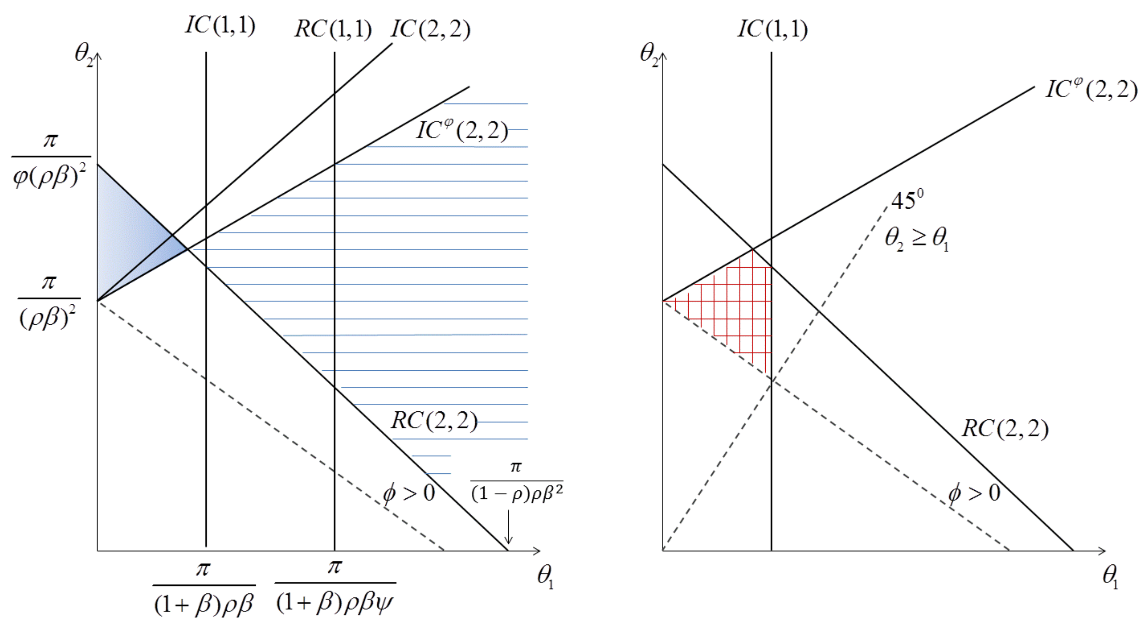

Figure 12: Set of implementable first-best allocations with costly cash-flow control. The left panel portrays the no-retention constraints. The right panel shows the set of first-best allocations that can be implemented.

guaranteed ($\sigma < 1$), it must transpire that $\theta_2 > \theta_1$. In particular, it must happen that $\theta_2 > \theta_1[1 - \sigma(1 - \rho)]/(\sigma\rho)$, where $[1 - \sigma(1 - \rho)]/(\sigma\rho) > 1$ (when $\sigma < 1$). Note that $1 - \sigma(1 - \rho)]/(\sigma\rho)$ is decreasing in σ so that this lower bound for θ_2 will rise as σ falls for a given value of θ_1. In other words, the profile of productivity must become steeper as survival falls.

Now imagine two countries where plants have the same survival rate. Older plants can account for a higher level of employment in one of the countries only if plants there also climb a steeper productivity profile than in the other country. This consideration will be important when comparing plants in the United States with those in Mexico. Alternatively, suppose that in two countries young and old plants have the same aggregate levels of employment. Then, the country with the lower survival rate must also have a steeper productivity profile. This fact will be important when comparing India and Mexico.

References

Aghion, Philippe and Howitt, Peter. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60 (2): 323-351.

Antunes, António, Cavalcanti, Tiago and Villamil, Anne. 2008. "The Effect of Financial Repression and Enforcement on Entrepreneurship and Economic Development." *Journal of Monetary Economics*, 55 (2): 278-297.

Banerjee, Abhijit V., and Duflo, Esther. 2005. "Growth Theory through the Lens of Development Economics." In *Handbook of Economic Growth*, Vol. 1A, edited by Philippe Aghion and Steven N. Durlauf, Amsterdam: Elsevier, 473–552.

Barro, Robert and Lee, Jong-Wha. 2013, "A New Data Set of Educational Attainment in the World, 1950-2010." *Journal of Development Economics*, 104 (C): 184-198.

Buera, Franciso J., Kaboski, Joseph P. and Shin, Yongseok. 2011. "Finance and Development: A Tale of Two Sectors." *American Economic Review.* 101 (5): 1964–2002.

Bushman, Robert M., Piotroski, Joseph D. and Smith, Abbie J. 2004. "What Determines Corporate Transparency." *Journal of Accounting Research*, 42 (2): 207-252.

Caselli, Francesco and Coleman, Wilbur John. 2006. "The World Technology Frontier." *American Economic Review*, 96 (3): 499-522.

Castro, Rui, Clementi, Gian Luca and MacDonald, Glenn M. 2004. "Investor Protection, Optimal Incentives, and Economic Growth." *Quarterly Journal of Economics*, 119 (3): 1131-1175.

Clementi, Gian Luca and Hopenhayn, Hugo. 2006. "A Theory of Financing Constraints and Firm Dynamics." *Quarterly Journal of Economics*, 121 (1): 229-265.

Fernandez, Ana and Phelan, Christopher. 2000. "A Recursive Formulation for Repeated Agency with History Dependence." *Journal of Economic Theory*, 91: 223-247.

Golosov, Mikhail and Tsyvinski, Aleh. 2006. "Designing Optimal Disability Insurance: A Case for Asset Testing." *Journal of Political Economy*, 114 (2): 257-279.

Greenwood, Jeremy and Jovanovic, Boyan. 1990. "Financial Development, Growth, and the Distribution of Income." *Journal of Political Economy*, 98 (5): 1076-1107.

Greenwood, Jeremy, Sanchez, Juan M. and Wang, Cheng. 2010. "Financing Development: The Role of Information Costs." *American Economic Review*, 100 (4): 1875-1891.

Greenwood, Jeremy, Sanchez, Juan M. and Wang, Cheng. 2013. "Quantifying the Impact of Financial Development on Economic Development." *Review of Economics Dynamics*, 16 (1): 194-215.

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

Hopenhayn, Hugo and Rogerson, Richard. 1993. "Job Turnover and Policy Evaluation: A General Equilibrium Analysis." *Journal of Political Economy*, 101 (5): 915-38.

Hsieh, Chang-Tai and Klenow, Peter. 2014. "The Life Cycle of Plants in India and Mexico." Unpublished paper, University of Chicago.

Hsieh, Chang-Tai and Olken, Benjamin A. 2014. "The Missing 'Missing' Middle." *Journal of Economic Perspectives*, 28 (3): 89–108.

King, Robert G. and Levine, Ross. 1993. "Finance and Growth: Schumpeter Might Be Right." *Quarterly Journal of Economics*, 108 (3): 717-737.

King, Robert G. and Levine, Ross. 1994. "Capital Fundamentalism, Economic Development and Economic Growth." *Carnegie-Rochester Conference Series on Public Policy*, 40 (June): 259-292.

Kocherlakota, Narayana R. 2010. *The New Dynamic Public Finance*. Princeton University Press, Princeton, N.J.

Levine, Ross., 2005. "Finance and Growth: Theory and Evidence" In *Handbook of Economic Growth*, Vol. 1A, edited by Philippe Aghion and Steven N. Durlauf, Amsterdam: Elsevier, 865-934.

Luttmer, Erzo G. J. "On the Mechanics of Firm Growth." *Review of Economic Studies*, 78 (3): 1042-1068.

Lucas, Robert E., Jr. 1990. "Why Doesn't Capital Flow from Rich to Poor Countries?" *American Economic Review*, 80 (2): 92-96.

Marcet, Albert and Marimon, Ramon, 1992. "Communication, Commitment, and Growth." *Journal of Economic Theory*, 58 (2): 219-249.

Midrigan, Virgiliu and Xu, Daniel Yi. 2014. "Finance and Misallocation: Evidence from Plant-Level Data." *American Economic Review*, 104 (2): 422-58.

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

Prescott, Edward C. 1998. "Needed: A Theory of Total Factor Productivity." *International Economic Review*, 39 (3): 525-551.

Schoellman, Todd. 2012. "Education Quality and Development Accounting." *Review of Economic Studies*, 79 (1): 388-417.

Schumpeter, Joseph A. 1961. *The Theory of Economic Development: An Inquiry Into Profits, Capital, Credit, Interest, and the Business Cycle*. Harvard University Press, Cambridge, MA.

Siegal, Jeremy J. 1992. "The Real Rate of Interest from 1800-1990: A Study of the U.S. and the U.K." *Journal of Monetary Economics*, 29 (2): 227-252.

Townsend, Robert M. 1979."Optimal Contracts and Competitive Markets with Costly State Verification." *Journal of Economic Theory*, 21 (2): 256-293.

Townsend, Robert M. and Ueda, Kenichi. 2010. "Welfare Gains from Financial Liberalization." *International Economic Review*, 51 (3): 553-597.

Wang, Cheng. 2005. "Dynamic Costly State Verification." *Economic Theory*, 25 (4): 887–916.

Williamson, Stephen D. 1986. "Costly Monitoring, Financial Intermediation, and Equilibrium Credit Rationing."*Journal of Monetary Economics*, 18 (2): 159-79.