Version: September 2012

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 there could be many factors, the efficiency of the country's financial system may 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 a firm's decision to adopt a technology. Can such a theory help to explain the differences in total factor productivity and establishment-size distributions across India, Mexico, and the U.S.? Applied analysis suggests that answer is yes.

Keywords: Costly cash-flow control; costly state verification; dynamic contract theory; economic development; establishment-size distributions; financial intermediation; India, Mexico, and the U.S.; monitoring; productivity; technology adoption; underwriting; ventures

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

*This research was previously circulated under the title "Underwriting Ventures: Trust but Verify." *Comments are welcome.* Please contact Hal Cole at colehl⊗sas.upenn.edu

1 Introduction

Why do countries use different production technologies? How does the adoption of a production technology affect a nation's income and total factor productivity (TFP)? Surely, all nations should adopt best-practice technologies, which produce the highest levels of income. Yet, this doesn't happen. To paraphrase Lucas (1990): Why doesn't technology flow from rich to poor countries?

The hypothesis entertained here is that the efficiency of financial markets plays an important role in technology adoption. Investing in new technologies is risky business. Advanced technologies require a great deal of funding and the payoff from any investment is uncertain. Compounding the problem is the fact that investors in a venture have access to a more limited set of information than the developers of the venture do. Therefore, there is scope for funds to be misappropriated as a result of private information problems. And, in some countries, it may even be difficult to control the use of publicly acknowledged funds.

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 an expensive activity, however, and in some places this cannot be done effectively. In this circumstance 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, though. Sometimes it is not possible for an intermediary to exert the desirable level of control over a firm's publicly acknowledged revenue streams. As a result, a contract cannot be written with the necessary reward structure required to ensure the likelihood of a successful investment. That is, there may be issues associated with the costly control of cash flows, and this will further limit the profitability of some technologies.

Long ago, Schumpeter theorized that financial development is important for economic development. Indeed, King and Levine (1993) find strong evidence that financial development is important for capital accumulation, economic growth, and productivity gains.

Subsequent research by King and Levine (1994) suggests that differences in productivities, and not factor supplies, are likely to explain differences in incomes across countries. This sentiment is echoed by Prescott (1998), who calls for a theory of TFP. Such differences in productivities can emerge from differences in technology adoption, which may in turn be affected by disparities in financial development; this hypothesis is pursued here.

Of course, other factors influence the choice of technologies that a country uses. A country's resource endowment is important. For example, Caselli and Coleman (2006) develop a model in which countries with large endowments of skilled labor tend to have lower skill premiums and, consequently, are more likely to pick skill-intensive production technologies. From their analysis, it is clear that differences in resource endowments alone cannot explain cross-country differences in productivity and income; indeed, this is in accord with the message in King and Levine (1994). Government policies that discourage or promote technologies are significant factors in technology adoption, too. Considerations such as these are neglected in the current analysis, which focuses in a single-minded fashion on the impact of financial development on technology adoption.

1.1 The Theoretical Analysis

A dynamic costly state verification model of venture capital is developed, with several unique features, to address the question of interest. The theory is put forth in two stages. In the first stage the benchmark model for the analysis is presented. This stage emphasizes the importance of monitoring for implementing advanced technologies. Countries differ in their ability to monitor effectively. The inability of a country to monitor efficiently will limit the profitability of certain types of investments. In the benchmark model, new firms enter the economy every period. They are free to pick a blueprint for a venture from a set of production opportunities. They can only start a single venture, though. A new firm will ask an intermediary to underwrite its venture. A firm's blueprint is represented by a non-decreasing stochastic process that describes movement up a productivity ladder. A firm's position on the ladder is private information. Intermediaries can audit the returns of a firm.

A distinguishing feature of the developed 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. This flexible auditing technology is borrowed from Greenwood, Sanchez, and Wang (2010), who extend the well-known costly state verification framework developed in important work by Townsend (1979) and Williamson (1986). They study static contracts. Extending the Greenwood, Sanchez, and Wang (2010) analysis to dynamic contracts brings to the fore some new and important considerations and involves resolving some tricky issues.[1]

The assumed structure of a productivity ladder implies that there is persistence in the firm's private information. This is a difficult problem, as readers familiar with Fernandez and Phelan (2000) will know. This is made manageable here by assuming that a stall at a rung on the ladder is an absorbing state.[2] The dynamic contract offered by an intermediary to a firm is a function of the latter's blueprint and the state of the country's financial system. The contract specifies a state-contingent plan outlining the advancement of funds from the intermediary to the firm, the intermediary's auditing strategy, and the payments from the firm back to the intermediary.

The developed costly state verification model is embedded into a general equilibrium framework. Intermediation is competitive. Another novel feature of the analysis is that blueprints differ across prospective ventures. Some blueprints have productivity profiles that offer exciting profit opportunities. Others are more mundane. This is operationalized by assuming that there are differences in the positions of the rungs on the productivity ladders, as well as in the odds of stepping between rungs. Blueprints also differ in the capital investment that they require. Some may require a substantial amount of investment before much information about the likely outcome is known. A new firm is free to implement the blueprint that it wishes, but it can put only one into effect. Greenwood, Sanchez, and

[1] Wang (2005) examines costly state verification with dynamic contracts in a Townsend-style (1979) setting.

[2] A similar insight is exploited in Golosov and Tsyvinski (2006).

Wang (2010) allow for blueprints to differ across firms, but again this is done in a static setting.

The nature of the blueprint, input prices, and state of the 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 can arise because inputs prices are two high and/or because the level of monitoring needed to make the project viable is simply too expensive given the state of the financial system.

When monitoring is expensive, intermediaries must rely on incentive schemes that backload payments to the firm to fund technologies, as has been made well known by Clementi and Hopenhayn (2006). These strategies redirect payoffs for the firm away from the start of the project toward the end, where they are contingent on performance. This requires that the intermediary has some ability to control the publicly acknowledged cash flows of the firm. This leads to the second part of the analysis. Here limitations on the control of cash flows are considered. In some countries this cannot be done effectively because the operators of a firm can always take part of the observed cash flow unless they can be enticed not to do so. As a result, for certain technologies it may not be possible to write a lending contract even though the technology would be very profitable given input prices. This arises both because limitations on cash-flow control reduce the amount available to compensate the lenders and because payments to the firm cannot be efficiently channeled to reduce the incentive to misreport earnings.

Thus, as discussed, 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. It seems reasonable to postulate that financial sector productivity may differ across countries, just as the efficiency of the non-financial sector does. It also seems likely that financial sector productivity grows over time within a country, too.

1.2 The Applied Analysis

To evaluate the ability of the theory to account for the data, the analysis focuses on three countries at very different levels of development and wealth: India, Mexico, and the U.S.[3] Hsieh and Klenow (2010) document some interesting differences in establishments across these three countries. The average establishment size is much smaller in Mexico than in the U.S. and is much smaller in India than in Mexico (Table 1). In addition, the level of labor productivity follows a similar pattern. Also, the share of employment contributed by younger (older) establishments is much larger (smaller) in India and Mexico than in the U.S. These facts suggest that these countries are using very different technologies.

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

---

[3] There is other quantitative work examining the link between economic development and financial development. For example, Buera, Kaboski, and Shin (2011) focus on the importance of borrowing constraints. Limited investor protection is emphasized by Castro, Clementi, and MacDonald (2009). Greenwood, Sanchez, and Wang (forth.) apply the Greenwood Sanchez, and Wang (2010) static contract model, discussed above, to the international data. The role that financial intermediaries play in producing ex ante information about investment projects is stressed by Townsend and Ueda (2010). All of this research is very different in nature from what is pursued here.

Table 1: Stylized Facts: India, Mexico, and the U.S.

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

The applied analysis proceeds in two phases, which parallel the theoretical development. First, a comparison is made between the choice of technology in Mexico and in the U.S. This comparison emphasizes the importance that monitoring plays for funding advanced technologies. To execute the analysis, a stylized version of the model is used in which there are only two production technologies available. The first is an advanced technology that promises to be highly profitable. Its blueprint requires substantial investment before much information about the state of productivity is known. Therefore, this project will require monitoring to implement. The second is a less profitable, intermediate-level technology that calls for smaller upfront investment. It can be implemented with a backloading strategy alone. To put some discipline on the analysis, factor prices are chosen to match the Mexican and U.S. economies. Capital is more expensive in Mexico, but labor is much cheaper. Labor is also less efficient in Mexico. Thus, ex ante, it is not clear whether the total cost of inputs is more or less expensive in Mexico than in the U.S. On net, it turns out that the costs of production are lower in Mexico than in the U.S. Thus, on first appearance, the advanced technology should be more profitable in Mexico than the U.S. The structure is parameterized so that it matches the above stylized facts about the Mexican and U.S. establishment size distributions. The question is this: Can an equilibrium be constructed where the U.S. will use the first technology and Mexico the second? The answer is yes.

Attention is directed toward India in the second phase of the applied analysis. India has a much lower income and productivity level than Mexico. It also has much lower labor costs, which imply a much lower cost of production. This latter fact suggests, on face value, that the potential profits from implementing advanced technologies in India are extremely large. Additionally, one would expect that Indian establishments should be large when inputs are inexpensive. Yet, they are very small. So, why doesn't India adopt either the U.S. technology or, more importantly, the Mexican technology, which does not require extensive monitoring? Here, the analysis focuses on the question of costly cash-flow control. To examine this, a third entry-level technology is added to the menu of blueprints. It turns out that the type of backloading/monitoring strategies required to finance the technologies used in Mexico and the U.S. are not profitable for intermediary to employ in India. Thus, India must use the unproductive third technology. The analysis is undertaken at the observed levels of Indian factor prices and the parameterized structure matches, in a rough sense, the above stylized facts about Indian establishments.

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. A new firm starts with a blueprint for a project. It chooses its blueprint from a portfolio of plans. It can operate only one type of project. Implementing this blueprint requires working capital. This funding is obtained from financial intermediaries. Projects differ by the payoff structures that they promise. For example, some projects may offer low returns, but 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 there are high odds 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 that is being funded, input prices, and the state of the financial system. A firm will choose the most profitable blueprint to implement, of course. Sometimes for certain blueprints it is not possible for an intermediary to offer a financial contract that will generate positive profits. Finally, in addition to supplying intermediaries with working capital, consumers/workers provide firms with labor. Since consumers/workers play an ancillary role in the analysis, they are relegated to the background.

3 Ventures

Each period, new firms enter the economy. A new firm 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 denoted by ϕ. Associated with each new firm is a productivity ladder $\{\theta_0, \theta_1, ..., \theta_S\}$, where $S \leq T$. Denote the firm's blueprint or type by $\tau \equiv \{\theta_0, \theta_1, ..., \theta_S, \phi\}$. A new firm selects the type of its blueprint, τ, from a portfolio of available plans, $\mathcal{T}$; it can implement only one plan. 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 is done 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. Capital then becomes locked in place and cannot be changed. At the end of each period, the firm faces a survival probability of σ. Figure 1 illustrates potential productivity paths for a firm over its lifetime.

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 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 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 $t-1$.

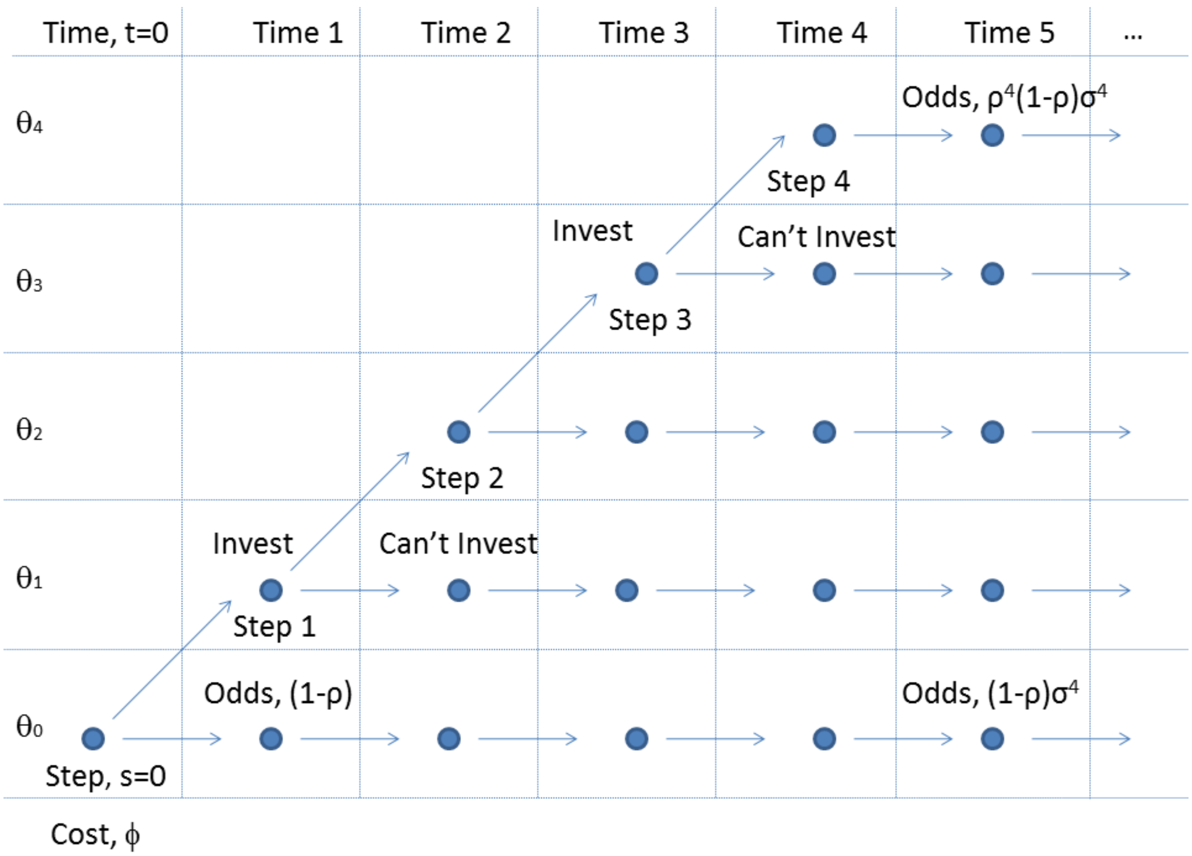

Figure 1: Possible productivity paths for a venture over its lifetime.

Focus on the *amalgamated* input, $k_t \equiv \widetilde{k}_t^{\omega} l_t^{1-\omega}$. The minimum cost of purchasing k units 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 will be referred to as the working capital for the firm, even though strictly speaking it should be multiplied by q. The rental rate, r, comprises 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 a new firms in order to supply working capital for the latter's 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 costlessly observe 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.[4] 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 properties that are worth noting. 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 $C(k, 1; q, z) \to \infty$ and $C_2(k, 1; q, z) \to \infty$. 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 that 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 measured here by z. (The dependence of C on q and z is suppressed when not needed in order to simplify the notation.)

[4] 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 $\theta_r > \theta_s$.

5 The Contract Problem

The contract problem between a firm and an intermediary will now be formulated. To prepare for this, 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 a firm and intermediary will stipulate the following for each step and 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, $p(s,t)$. 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 firm and intermediary can be expressed as

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

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

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

The objective function in (P2) gives 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 born firm. Equation (6) is the limited liability constraint for the firm. The intermediary cannot take more than the firm produces at the step and date combination (s,t).

The incentive constraint for a firm is specified by (7). This constraint is imposed on the firm only at each date and state 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 12.1. Here 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.

The left-hand side of the constraint gives 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 gives 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 it 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 in (u, u) are embodied in a multiplicative manner in the $\Pr(s, t)$ terms and these will cancel out of both sides of (7). Thus, the unconditional 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 or not there will be a stall 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 at each node. 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 stall in productivity 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 in Figure 2. All working capitals within a horizontal box take the same value. Think of a plant as having a putty-clay structure: In the event of a stall, all inputs become locked in.

Finally, (10) stipulates that the intermediary expects to earn positive profits from its loan

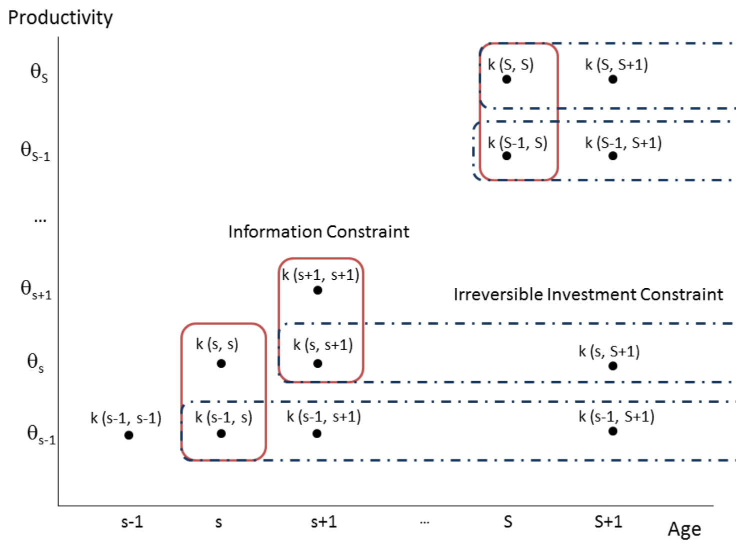

Figure 2: The information and irreversibility constraints.

contract. For an (s,t) combination the intermediary will earn $x(s,t) - C\left(p(s,t), k(s,t)\right) - qk(s,t)$ 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 upfront fixed cost for the project. This is represented by the term ϕ in (10).

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 can be seen from the right-hand side of (7), 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. So, given this, the intermediary can be sure that the firm is always telling the truth.

Lemma 1 *(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 12.3. ■

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 it should backload the rewards that the firm can earn. In particular, it is optimal to let the firm realize all of its awards only upon arrival at the terminal node (S, T). The intermediary should take all cash flow away 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 $\sum_{t=1}^{T} \sum_{s=0}^{\min\{t,S\}} \beta^t \left[\theta_s k(s,t)^\alpha \ - C\left(p(s,t), k(s,t)\right) \ - qk(s,t)\right] \Pr(s,t) - \phi \geq 0$. There may be other payment schedules that are equally efficient, but *none* can dominate this one.

Lemma 2 *(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. $$x(S,T) = \theta_S k(S,T)^\alpha - \{\sum_{t=1}^{T} \sum_{s=0}^{\min\{t,S\}} \beta^t \left[\theta_s k(s,t)^\alpha - C\left(p(s,t), k(s,t)\right) - qk(s,t)\right] \Pr(s,t) - \phi\} / [\beta^T \Pr(S,T)] \leq 0.$$

Proof. See Appendix 12.4. ■

5.1 Discussion

The solution to the above contract problem shares some features that are common to dynamic contracts, but 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, so long as the descent is deterministic. The steps on the ladder do not have to 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. Observe that the theory is presented 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 try 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 will depend on the assumed structure for the productivity ladder.

Second, the fact that productivity can only step up to the next rung 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 Fernandez and Phelan (2000). This payoff places an upper bound on the value of lying in the future. 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 so far 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. This 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 analysis, 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 upfront investment and promises payoff streams that are tilted toward the end of the venture's lifetime. This is the case in the Mexico/U.S. example studied in Section 9. Here, Mexico has an inefficient monitoring technology relative to that of the U.S. So, it is not able to adopt the advanced technology used by the U.S., 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 lower fixed cost and a concave productivity schedule, which can be financed using a backloading strategy that requires little monitoring.

Now, there are countries in the world where the cost of production 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 is preventing them from using the Mexican one? After all, it requires little in the way of monitoring services.

An extension to the baseline theory is developed now that provides one possible answer. The idea is that in some countries it is very costly for intermediaries to force firms to pay out all of their publicly acknowledged output. Perhaps a fraction of output inherently goes to the benefit of 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 Extending the Theory

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, during an audit, the operators of the firm lying about the firm's state.[5] The intermediary must design the

[5] 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, (11) and (12), that will be presented below. As a result, the incentive constraints (7) are not directly effected by this

contract in a manner such that the retention of output will be dissuaded. How will this affect the contract presented in (P2)?

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

1. The intermediary desires 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 of the contract:

$$\begin{aligned} & \sum_{j=t}^{T} \beta^t \left[\theta_s k(s,j)^\alpha - x(s,j)\right] \frac{\Pr(s,j)}{\Pr(s,t)} \\ \geq \quad & \psi \sum_{j=t}^{T} \beta^t \theta_s k(s,j)^\alpha \frac{\Pr(s,j)}{\Pr(s,t)}, \text{ for } 1 \leq s \leq S, s < t, \, 2 \leq t \leq T \text{ (off-diagonal node),} \end{aligned} \tag{11}$$

and

$$\begin{aligned} & \sum_{t=u}^{T} \sum_{s=u}^{\min\{t,S\}} \beta^t \left[\theta_s k(s,t)^\alpha - x(s,t)\right] \frac{\Pr(s,t)}{\Pr(u,u)} \\ \geq \quad & \psi \sum_{t=u}^{T} \sum_{s=u}^{\min\{t,S\}} \beta^t \theta_s k(u-1,t)^\alpha \frac{\Pr(s,t)}{\Pr(u,u)}, \text{ for } 1 \leq u \leq S \text{ (diagonal node).} \end{aligned} \tag{12}$$

The first constraint (11) applies to the case where a stall has occurred at state s. Here, productivity is stuck at θ_s forever. The second constraint (12) governs the situation where

extension. This feature is a virtue for both analytical clarity and simplicity.

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 just append the no-retention constraints (11) and (12) to problem (P2). Lemma 1 still holds. So, once again the intermediary will 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 encourage the firm not to exercise 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 that 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,

$$N(s,T) = \psi \frac{\sum_{t=s+1}^{T} \beta^t \theta_s k(s,t)^\alpha \Pr(s,t)}{\beta^T \Pr(s,T)}, \text{ for } 0 \leq s \leq S. \tag{13}$$

Hence, Lemma 2 now appears as Lemma 3.

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* $0 \leq s \leq S, 1 \leq t < T$, *and* $s \leq t$;
2. $x(s,T) = \theta_s k(s,T)^\alpha - N(s,T)$, *for* $0 \leq s < S$;
3.

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

where $N(s,T)$ is specified by (13).

Proof. See Appendix 12.4. ■

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 will audit firm from then on. Recall that if the intermediary detects a lie at some node (u, t), where $t \geq u$, it can recover all output. 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 so long as 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 $\prod_{n=u}^{T}[1 - p(u-1, n)]$.

Notice 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 will be difficult for a masquerading firm to capture this payment. This reduces the incentive to lie. When monitoring is ineffective, it will be easy to do this. The incentive to lie will then be higher.

Finally, when is investment efficient or when will it match the level that would be observed in a world where the intermediary can costlessly observe the firm's shock? Suppose that after some date/state combination (t^*, t^*) along the diagonal of the ladder that neither the incentive or no-retention constraints, (7) and (12), ever bind again. Will investment be efficient from then on? Yes.

Lemma 4 *(Efficient investment) Suppose that neither the incentive nor the diagonal-node no-retention constraints ever bind after node* (t^*, t^*) *for* $t^* < S$. *Investment will be efficiently undertaken on arriving at the date/state combination* (t^*, t^*).

Proof. See Appendix 12.5. ■

7 Some Two-Period Examples

Some simple two-period examples illustrating the contract setup are now presented.[6] 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 illustrate the importance that monitoring and retention play in the design of a contract. Finally, a connection is drawn between the productivity ladder and the aggregate age distribution of employment.

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.

7.1 The Importance of Monitoring

The first example focuses on the importance of monitoring. 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 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?

[6] The examples section can be skipped for the reader who wants to move on the applied analysis.

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 $\pi/(\beta^2\rho^2)$ 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 $\pi/(\beta^2\rho^2)\geq\theta_2-\theta_1$, or

$$\theta_2\leq\theta_1+\pi/(\beta^2\rho^2). \tag{14}$$

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. 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 $\pi/\beta^2\rho^2$, dominates the payoff from lying and claiming $\theta=\theta_0=0$. Thus, it must transpire that $\pi\geq\beta\rho\theta_1+\beta^2\rho\theta_1$, implying that

$$\theta_1\leq\pi/[(1+\beta)(\rho\beta)]. \tag{15}$$

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 $\phi\geq0$. This implies the restriction $\beta\rho\theta_1+\beta^2\rho(1-\rho)\theta_1+\beta^2\rho^2\theta_2-\pi\geq0$, which can be rewritten as

$$\theta_2\geq\pi/(\beta^2\rho^2)-\{\beta\rho[1+\beta(1-\rho)]/(\beta^2\rho^2)\}\theta_1. \tag{16}$$

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, $\beta\rho\theta_1+$

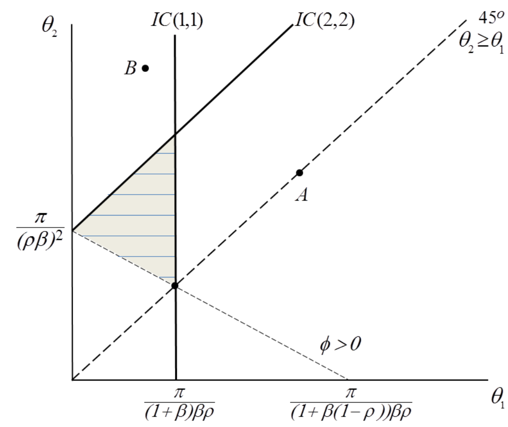

Figure 3: Set of implementable first-best allocations.

$\beta^2\rho(1-\rho)\theta_1 + \beta^2\rho^2\theta_2$, 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.

Figure 3 plots the two incentive constraints (14) and (15), the 45-degree line, and the fixed-cost constraint (16). The shaded triangle 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.

Consider a point, such as A, where $\theta_1 = \theta_2$ and $\phi < \beta\pi$. In this case, the incentive constraint (15) collapses to $\theta_1 \leq \theta_1 - \phi/[(1+\beta)(\rho\beta)]$. Then, the first-best allocation cannot be supported if $\phi > 0$. 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. 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 $\rho\beta k(1,1) + \rho\beta^2\theta_1 - \phi$. To maintain indifference between these two strategies, set $\rho(\beta + \beta^2)k(1,1) = \rho\beta k(1,1) + \rho\beta^2\theta_1 - \phi$. This implies $k(1,1) = \theta_1 - \phi/(\rho\beta^2) < \theta_1$. The condition that $\phi < \beta\pi$ guarantees that $k(1,1) > 0$. This assumption ensures that the fixed cost, ϕ, is not too large, so that the promise of future profits from telling the truth exceeds the gains from lying and avoiding the fixed cost. When $\phi > \beta\pi$ 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 binds, so that $\theta_2 \geq \theta_1 + \pi/(\beta\rho)^2$. This implies that $\theta_1 < \phi/[\beta\rho(1+\beta)]$. 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, ϕ. 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 be obtained at point A, too. 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.

7.2 Costly Cash-Flow Control

The second example focuses on how costly cash-flow control influences the design of the contract. The features of the previous example are retained but now $\psi \geq 0$.

7.2.1 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 4 illustrates the no-retention constraint. It slopes downward.

Move back in time to period 1, specifically to node $(1,1)$. If the firm transits to node $(2,2)$ it will earn in profits 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 4.

7.2.2 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 4 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.

7.2.3 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 left in the left panel of Figure 4 shows those (θ_1, θ_2) combinations that satisfy the second-period no-retention contraint, $RC(2,2)$, 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

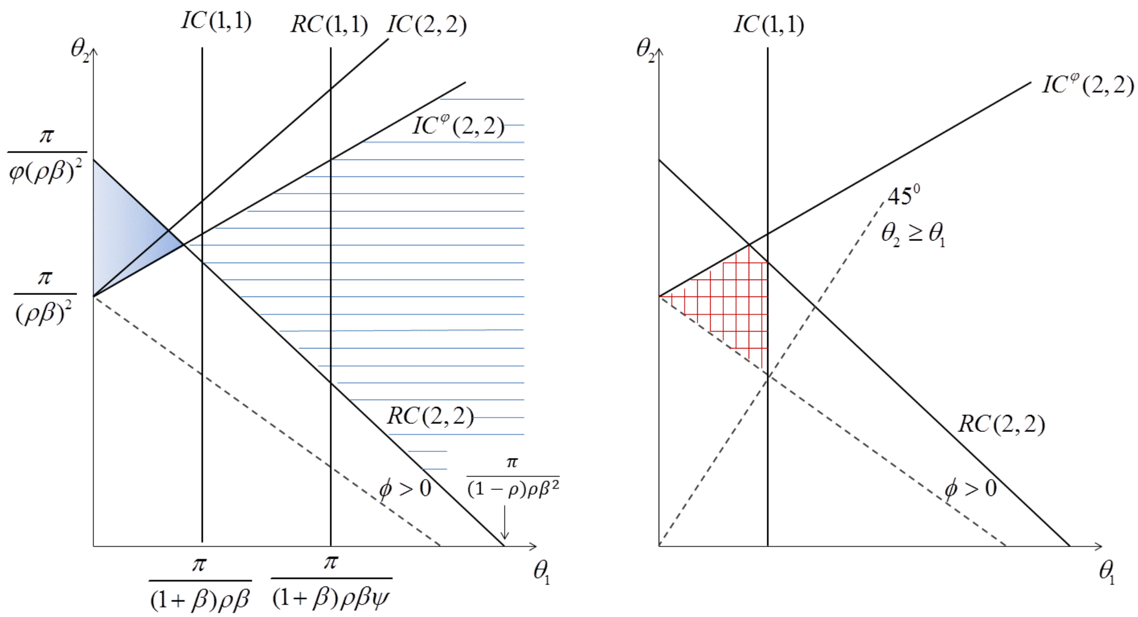

Figure 4: Set of implementable first-best allocations with costly cash-flow control. The left-hand side panel portrays the no-retention constraints. The set of first-best allocations that can be implemented is shown in the right-hand side panel.

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 makes the incentive problem worse 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.

Turn now to the panel on the right in Figure 4, which 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. It lies within the old triangle.

7.3 Identifying the Productivity Ladder

The two-period example is now used to illustrate the connection between the productivity ladder and the aggregate age distribution of employment. 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 firms and old 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 $\theta_1 \lesseqqgtr \sigma[(1-\rho)\theta_1 + \rho\theta_2]$. Suppose that $\theta_1 \geq \theta_2$. Then, clearly $\sigma\rho\theta_1 \geq \sigma^2\rho(1-\rho)\theta_1 + \sigma^2\rho^2\theta_2$ because $\sigma \leq 1$. In this situation, a young firm hires more than an old firm (since $\theta_1 \geq \theta_2$). There are fewer old firms around (as $\sigma \leq 1$). So, clearly employment by old firms must be less than employment by young ones.

Therefore, to have old firms accounting for more employment than young ones, when survival is not 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 U.S. 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.

8 Equilibrium

There is one unit of labor available in the economy. This must be split across all operating firms. Recall that each new firm selects some productivity ladder $\{\theta_0, \theta_1, ..., \theta_S\}$, which is associated with a fixed cost, ϕ. Call this the firm's type. Denote the firm's type by $\tau \equiv \{\theta_0, \theta_1, ..., \theta_S, \phi\} \in \mathcal{T}$, which indexes a particular productivity ladder. Likewise, represent the working capital and labor used by a type-τ firm at an (s,t) pair by $k(s,t;\tau)$ and $l(s,t;\tau)$, respectively.

Now a new firm is free to pick a project τ from the set of potential blueprints, $\mathcal{T}$. A firm may operate only one type of venture. Clearly, the firm will choose the type τ that *maximizes* the value of the firm, $v(\tau)$. The type that maximizes the value of the firm will depend on the price of the amalgamated input, q, which is a function of the equilibrium wage, w, through (1). Thus, the set of active venture types, $\mathcal{A}(w)$, is defined by[7]

$$\mathcal{A}(w) \equiv \{x : x = \arg\max_{\tau \in \mathcal{T}} v(\tau) \text{ and } v(x) \geq 0\}. \tag{17}$$

Suppose that # new firms enter the economy each period. Every period some firms die. This death process is subsumed in the probabilities $\Pr(s,t)$. The labor market clearing condition for the economy then reads

$$\# \sum_{t=1}^{T} \sum_{s=1}^{\min\{t,S\}} \int_{\mathcal{A}(w)} [l(s,t;\tau) + l_m(s,t;\tau)] \Pr(s,t) = 1, \tag{18}$$

where $l_m(s,t;\tau)$ is the amount of labor that an intermediary will spend monitoring a type-τ venture at node (s,t). The quantity of the amalgamated input used in monitoring, $k_m(s,t;\tau)$, is given by

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

which implies a usagage of labor in the amount

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

[7] Theoretically speaking, it is possible for more than one technology to be chosen in a country, provided that the technologies yield the same level of profits. In the simulations this does not happen.

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, $k(s,t;\tau)$, labor allocations, $l(s,t;\tau)$ and $l_m(s,t;\tau)$, and monitoring strategy, $p(s,t;\tau)$, for all $s = 1,...,S$, $t = s = 1,...,T$; (b) a set of active venture types, $\mathcal{A}(w) \sqsubseteq \mathcal{T}$; (c) an amalgamated input price, q, and wage rate, w, all such that:*

1. *The working capital financing program, $k(s,t;\tau)$, and the monitoring strategy, $p(s,t;\tau)$, specified in the financial contract maximize the value of a type-τ venture, as set out by (P2), given the amalgamated input price, q. [Here (P2) should be amended to include the no-retention constraints (11) and (12).]*

2. *A type τ-venture is chosen only if it is contained in the active set, $\mathcal{A}(w)$, as specified by (17), where $v(\tau)$ is determined by (P2).*

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

4. *The amount of labor, $l_m(s,t;\tau)$, 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 Applied Analysis: The Choice of Technology in India, Mexico, and the U.S.

Why might one country choose a different set of production technologies than another nation? There are many reasons, of course: 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.[8] While these

[8] Think of differences in factor supplies as being captured in the current analysis by differences in the cost of the amalgamated input, q.

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.

In the applied analysis a nation is free to adopt one of three technologies: advanced, intermediate, and entry level. 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 the one for the intermediate one, which is larger than the one for the entry-level technology. The advanced technology, with its convex payoff structure and high fixed cost, is hard 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. Which technology a country chooses depends on its factor prices and the state of its financial system. An equilibrium is constructed in which the U.S. 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 0.15 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 Price, q

A key input into the analysis is the price for the amalgamated input, q. Start with Mexico and the U.S. The price of this input in Mexico relative to the U.S. is what is important. Normalize this price to be 1 for the U.S., 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

U.S.) 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 U.S. 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 U.S. 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. Next, 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 U.S., 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 U.S.

Move on now to India. The rental price of capital in India, r^{IN}, is about 23 percent higher in India than in the U.S. (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 U.S. 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 U.S. Here 1.6 years of education are added to the number in Barro and Lee 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 (out of eleven) on the technology ladder, the last three being 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 U.S. 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 admitted into the theory developed, as is discussed in Section 5.1. Last, an upper bound on working capital is imposed. This is denoted by $\overline{k}$ and is common across technologies.[9]

9.3 The Choice of Technology in Mexico and the U.S.

A numerical example is now presented in which Mexico *chooses* to adopt a different technology from the U.S. In particular, the U.S. (or advanced) technology offers a productivity profile that grows much faster than the Mexican contour (which represents an intermediate-level technology). Financing the U.S. technology requires a level of monitoring that only an efficient financial system can undertake. The Mexican technology does not require this. The Mexico/U.S. comparison using the baseline model that abstracts away from the costly cash-flow control problem. So, $\psi = 0$ in both countries. The example is constructed so that the framework matches the size distribution of establishments by age that is observed for Mexico and the U.S. It also replicates the average size of firms in these two countries—in fact, for the U.S. the entire size distribution is fit. These two facts discipline the assumed productivity profiles. In the equilibrium constructed, it is not desirable to finance the Mexican technology in the U.S., given the state of the U.S. financial system and U.S. input prices. Likewise, it is not worthwhile to underwrite the U.S. technology in Mexico given the latter's financial system and input prices.

9.3.1 Calibrating the Mexican and U.S. Technology Ladders

First, the survival probabilities are from the 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 each country. 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 will be selected so that the model fits, as well as possible,

[9] 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.

several stylized facts about the U.S. and Mexican 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 labor productivity in the banking sector (deposits divided by employment in banking). The (complementary) distribution of employment by establishment age is characterized by a set of points. For the U.S. alone, the establishment size distribution in Lorenz-curve form is also added to the collection of styled 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 5 shows the salient features of the technologies used in Mexico and the U.S. (India is discussed later.) The productivity of a firm rises with a move up the ladder. The U.S. ladder has a convex/concave profile, while the Mexican has a concave one as Figure 5 shows. Note that the ascent is much steeper for a U.S. firm than a Mexican one. The survival processes are the same in each country.

Table 2: Parameter Values

Parameter	*Value*		
	U.S.	Mexico	India
Discount factor	$\beta = 0.98^5$	$\beta = 0.98^5$	$\beta = 0.98^5$
Prod. function–scale, capital's share	$\alpha = 0.8,\ \omega = 0.33$	$\alpha = 0.8,\ \omega = 0.33$	$\alpha = 0.8,\ \omega = 0.33$
Capital, upper bound	$\overline{k} = 12$	$\overline{k} = 12$	$\overline{k} = 12$
Fixed cost	$\phi = 0.29$	$\phi = 0.031$	$\phi = 0.001$
Labor, efficiency	$\chi = 96,427$	$\chi = 53,035$	$\chi = 33,750$
Pr Stall, parameters	$1 - \rho = 0.31$	$1 - \rho = 0.6$	$1 - \rho = 0.6$
Ladder, state parameters	θ's–see Fig. 5	See Fig. 5	See Fig. 5
Pr Survival, time t	σ's–see Fig. 5	"	"
Input price	$q = 1.0$	$q = 0.94$	$q = 0.6$
Monitoring efficiency	$z = 25$	$z = 0.25$	$z = 0.25$
Retention	$\psi = 0$	$\psi = 0$	$\psi = 0.425$

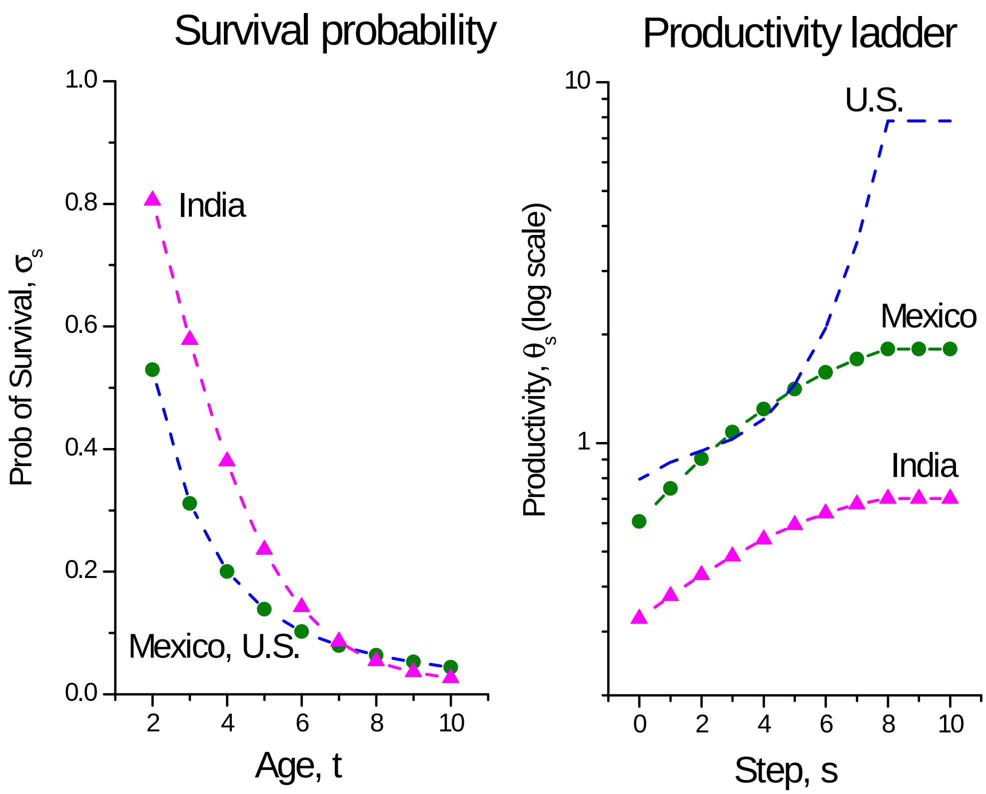

Figure 5: Productivity and survival in India, Mexico, and the U.S. (model). The diagram displays the assumed productivity ladders (right panel) for India, Mexico, and the U.S. It also illustrates the probability profiles for survival (left panel).

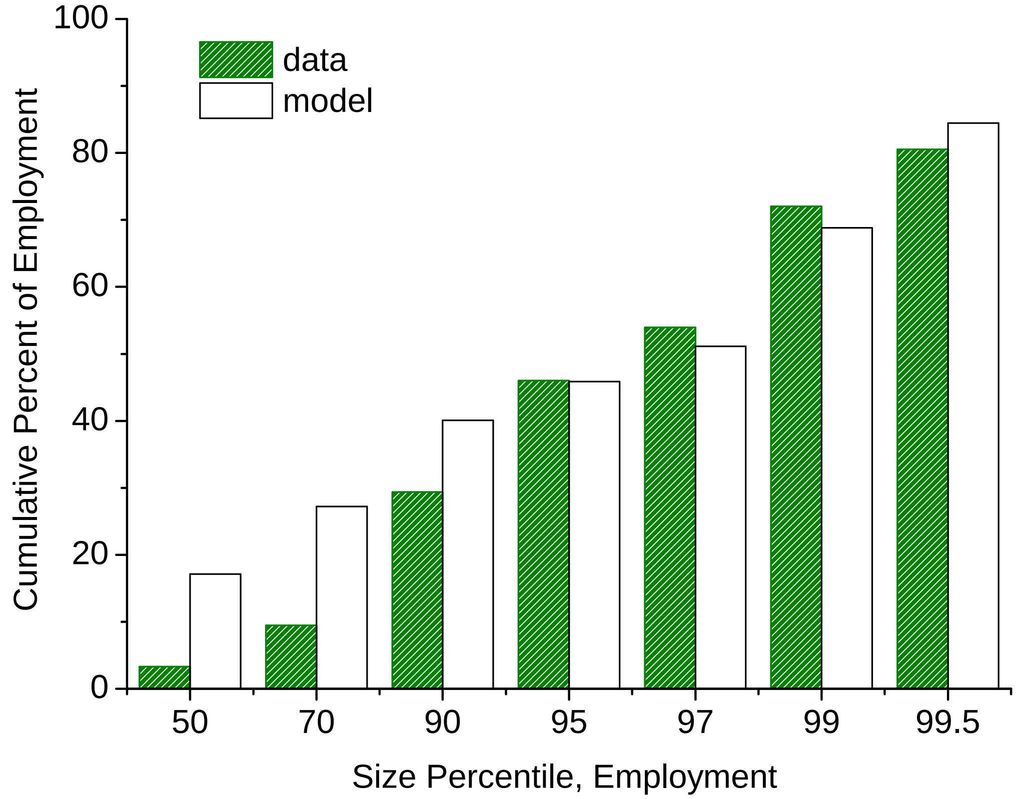

Figure 6: U.S. establishment-size distribution in Lorenz-curve form: data and model. Data sources for all figures are presented in the Appendix.

9.3.2 Establishment Size Distributions in Mexico and the U.S.

The model does a very good job matching the U.S. establishment size distribution. This can be seen from Figure 6, which plots this distribution in Lorenz-curve form. The model overpredicts the share of small establishments in employment, however. Mexican plants are about half the size of U.S. ones. The model mimics this feature of the data well, as shown in Table 3.

Figure 7 plots the model's fit for the Mexican and U.S. complementary cumulative distributions of employment by age—i.e., it graphs one minus the cumulative distribution of employment by age. Establishments older than 30 years account for a smaller fraction of employment in Mexico (or India) relative to the U.S., as can been seen by focusing on the

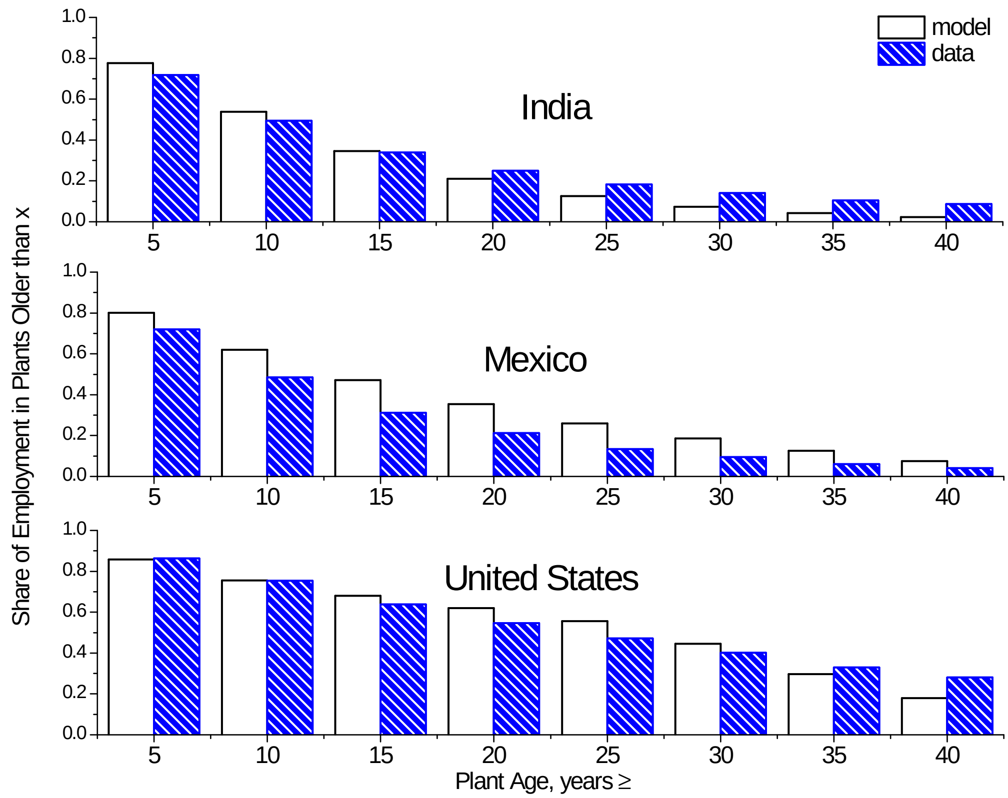

Figure 7: The (complementary) cumulative distributions of employment by age ($\geq$) for plants in India, Mexico, and the U.S., data and model.

right-hand side of the graphs. Consider Mexico first (the middle panel). The fit is not poor but the model has some 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. Plants younger than 10 years account for more than 50 percent of Mexican employment, while in the model they account for just 38 percent. Now switch to the U.S. (the bottom panel in Figure 7). The model does a superb job matching the share of employment by age for the U.S. Still, it does not quite capture the fact that some old firms in the U.S. are very large.

9.3.3 Productivity

Can the above framework generate sizable differences in productivity between Mexico and the U.S., 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 a function of τ too. 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 $l(s,t;\tau)$ 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 does a good job capturing the fact that productivity in Mexico is only a third of productivity in the U.S. 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 does an excellent job mimicking the fact that Mexican TFP is 46 percent of the U.S. level. Last, note 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 Mexico and the U.S.

Table 3: Stylized Facts for India, Mexico, and the U.S.

Statistics	*U.S.*		*Mexico*		*India*	
	Data	Model	Data	Model	Data	Model
Output per worker	1.00	1.00	0.33	0.33	0.12	0.15
TFP	1.00	1.00	0.46	0.46	0.24	0.30
TFP, human capital adjusted	1.00	1.00	0.69	0.68	0.49	0.60
Average firm size	1.00	1.00	0.55	0.50	0.11	0.07
Employment share, age $\leq$ 10 yr	0.25	0.24	0.52	0.38	0.51	0.46
$\ln(\text{TFP}_{age>35}) - \ln(\text{TFP}_{age<5})$	2.23	1.91	0.51	0.45	0.30	0.15

9.4 The Choice of Technology in India

The cost of production in India ($q^{IN} = 0.6$) is much less expensive than in Mexico ($q^{MX} = 0.94$) and the U.S. ($q^{US} = 1$). Therefore, at a minimum, one would expect that India could adopt the technology used in Mexico. Adopting the Mexican technology is not feasible for India due to a costly cash-flow control problem; i.e., for India $\psi > 0$. This prevents the financiers from recovering their upfront investment. Thus, India is forced to adopt an entry-level technology with a flat productivity schedule, but low fixed cost.

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 alone. The technology ladder for India is fit to the data in the manner used for Mexico and the U.S. The upshot of the calibration procedure is displayed in Figure 5, which shows the Indian, Mexican, and U.S. productivity ladders. The main difference between the Indian and Mexican ladders is that the productivity profile for the former is lower and flatter. The survival rate is higher for younger establishments in India than for plants in either Mexico or the U.S. Recall that the survival rates are obtained directly from data. The model is able to match the stylized facts about establishments in India reasonably well. As shown in Table 3, the average size of a plant in India is about 10 percent of one in the U.S.; the model predicts 7 percent. Figure 7 shows that the calibrated framework mimics the share of employment by age for firms very well. Again, this is done with a TFP profile for Indian plants that grows very slowly with plant age, which seems to be the case (see Table 3).

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

So, what determines whether a particular technology will be used in a nation? 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)? The right-hand panel of Figure 8 shows the combinations of ψ and z required to adopt each of the three technologies in India, assuming the Indian level of factor prices. That is, it shows the adoption zones in India for each technology. For any value of ψ, the advanced technology will require a higher level of z than for an entry-level and intermediate technology. 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. When ψ rises to a certain level, it is no longer possible to operate the project, regardless of the efficiency level in monitoring or the size of z. The firm can simply retain too much of the cash-flow streams for a viable contract to be written. The point labeled "India" indicates the values for ψ and z that are used for India in the simulation. India would be able to operate the advanced technology at its low factor prices were it not for the retention problem; that is, at $\psi = 0$ India could adopt the advanced technology at its z. Another interesting feature of the diagram is that the entry-level technology simply cannot be run when ψ exceeds 0.46. It is still feasible to operate the advanced technology, however, provided that efficiency in monitoring is high enough.

The left-hand panel of Figure 8 tells a similar story for Mexico. The values for ψ and

z used in the Mexican simulation are shown by the point labeled "Mexico." Since $\psi = 0$ in Mexico, the only factor limiting Mexico from adopting the advanced technology is the efficiency of its financial system. Retention cannot explain the fact that Mexico does not adopt the advanced technology. Observe that the advanced and intermediate technologies can be used in Mexico only when retention is low relative to India. This is because input prices are higher in Mexico than in India. It is interesting to note that the entry-level technology would not be profitable in either Mexico or the U.S. Wages are too high in these nations to operate this unproductive technology. Of course, a similar picture could be plotted showing the combinations of ψ and z that support the use of the advanced technology in the U.S. The U.S. level of z supports the use of this technology at U.S. factor prices (when $\psi = 0$). So, differences in both ψ and z are needed to explain the pattern of technology adoption across India, Mexico, and the U.S.

Are the implied zones for ψ and z required to support the cross-country pattern of adoption reasonable? The analysis presumes that z for the U.S. is higher than that for either India or Mexico and that ψ for India is larger than for either Mexico and the U.S. There does not appear to be direct evidence on the likely values for these variables, but some suggestive facts are shown in Table 4. Most numbers in the table are obtained from the World Bank's *Doing Business* database. Focus first on financial sector efficiency. India and Mexico are approxiately equal in the productivity of their banking sector, as measured by total deposits per employee. Productivity in U.S. banking is much higher. Recall that these facts are targeted. The model matches them well. India and Mexico rank the same among the 189 countries in the database on their ease of obtaining credit. It is much easier to obtain credit in the U.S. These two facts can be interpreted as implying that the U.S. financial sector is more efficient that either the Indian or Mexican ones. Turn to retention. It is much more difficult to recover funds after a bankruptcy, both in terms of the recovery rate and time, in India than in Mexico or the U.S. Additionally, it is much more time consuming to enforce a contract in India. This suggests that it is easier to retain funds in India than in either Mexico or the U.S.

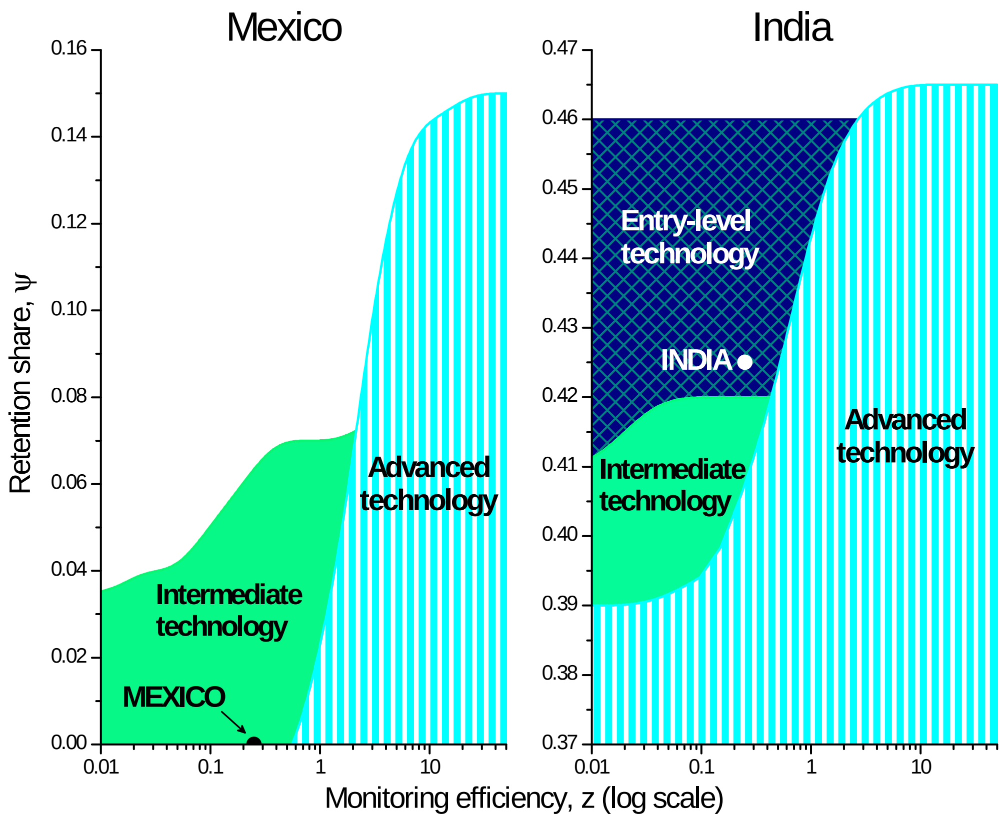

Figure 8: The zones of adoption for Mexico (left panel) and India (right panel).

Table 4: Financial Efficiency and Retention

Variable	*India*	*Mexico*	*U.S.*
Financial Sector Efficiency, z			
z, model	0.25	0.25	25
Getting credit rank	40	40	4
Productivity, deposits	0.37	0.45	1.0
Productivity, deposits–model	0.37	0.49	1.0
Retention, ψ			
ψ, model	0.42	0	0
Recovery rate	20%	67%	82%
Resolving insolvency (years)	7	1.8	1.5
Enforcing contracts (days)	1,420	415	300

11 Conclusions

The role that financial intermediation plays in underwriting business ventures is investigated here. A dynamic costly state verification model of lending from intermediaries to firms is developed to do so. The model is embedded into a general equilibrium framework where intermediation is competitive. A firm's level of productivity is private information. An intermediary is free to audit a firm's returns. 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.

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 position 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. Some ventures may have exciting potential for profit, but a large up-front acquisition of working capital may be required from the intermediary to the firm before much relevant information is revealed to the investors. For these types of investments, the ability of an intermediary to conduct ex post monitoring will be important for the viability of a long-term lending contract. Thus, the inability to monitor investments shapes the set of ventures in which intermediaries and firms will want to invest.

When an intermediary cannot monitor investment projects it must rely on incentive schemes to ensure that certain types of ventures are run profitably. These incentive schemes typically rely on backloading strategies. Such strategies redirect the payoffs to a firm from the beginning of the project toward the end. Ideally, the firm will realize profits only upon the successful consummation of the project. Sometimes it is not possible for the intermediary to control even publicly acknowledged cash flows to the extent needed to implement a successful backloading strategy. This further restricts the profitability of certain types of investment projects. The upshot is that the set of desirable technologies within a country will be a function of the state of the nation's financial system. Therefore, a country's income and TFP will also depend on its financial system.

India, Mexico, and the U.S. have very different levels of income and TFP. Can differences in technology adoption due to differences in financial systems explain this in part? 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. The advanced technology has the potential to deliver high profits. It requires large investments and has considerable scope for financial malfeasance. Therefore, its implementation requires a financial system that can monitor it effectively. 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 U.S. Likewise, given Mexican factor prices and the Mexican financial system, it is best to use the intermediate technology in Mexico. Monitoring is too inefficient in Mexico and the intermediate technology can be implemented by relying primarily on a backloading strategy. Now, suppose that monitoring in India is also prohibitively expensive. Shouldn't India use the intermediate technology? After all, the costs of production in India are very low. The answer here is that it may not be possible to use this technology. Given India's inability to control cash flows, it may be the case that a viable contract cannot be written with the required reward structure to make the intermediate technology profitable in India.

Some evidence is presented suggesting that India, Mexico, and the U.S. 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, the size of an establishment increases more steeply with age in the U.S. than in either India or Mexico. Two numerical examples are developed that mimic these facts about Indian, Mexican, and U.S. establishment size distributions, given the observed differences in factor prices, so the analysis is not without some discipline. The analysis is able to replicate the observed patterns of income and TFP across India, Mexico, and the U.S.

12 Appendix

12.1 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 histories by $\mathcal{F}^t \equiv \{\mathbf{h}^t \in \mathcal{H}^t : \Pr(\mathbf{h}^t) > 0\}$, where $\Pr(\mathbf{h}^t)$ is the probability of history $\mathbf{h}^t$. For formulating the retention contraints it is useful to define $\mathcal{F}^t(s, j)$ as the set of 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:

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

(ii) $\zeta_t(\mathbf{h}^t) \leq h_t(\mathbf{h}^t)$, 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 $\{k(\zeta_t(\mathbf{h}^t), t), x(\zeta_t(\mathbf{h}^t)), p(\zeta_t(\mathbf{h}^t))\}_{t=1}^T$. Given this notation, the general contract problem (P3) between the firm and intermediary can be written as

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

subject to

$$\theta(\mathbf{h}^t) k(\mathbf{h}^t, t)^\alpha - x(\mathbf{h}^t) \geq 0, \tag{21}$$

$$\begin{aligned}&\sum_{t=1}^{T} \sum_{\mathbf{h}^t \in \mathcal{H}^t} \beta^t \left[\theta(\mathbf{h}^t) k(\mathbf{h}^t, t)^\alpha - x(\mathbf{h}^t)\right] \Pr\left(\mathbf{h}^t\right) \\ &\quad \geq \max_{\zeta \in \mathcal{S}} \sum_{t=1}^{T} \sum_{\mathbf{h}^t \in \mathcal{H}^t} \beta^t \left[\theta(\mathbf{h}^t) k(\zeta^t(\mathbf{h}^t), t)^\alpha - x(\zeta^t(\mathbf{h}^t))\right] \prod_{n=1}^{t} [1 - p\left(\zeta^n(\mathbf{h}^n)\right)] \Pr\left(\mathbf{h}^t\right),\end{aligned} \tag{22}$$

$$\begin{aligned}&\sum_{t=j}^{T} \sum_{\mathbf{h}^t \in \mathcal{F}^t(s,j)} \beta^t \left[\theta(\mathbf{h}^t) k(\mathbf{h}^t, t)^\alpha - x(\mathbf{h}^t)\right] \Pr\left(\mathbf{h}^t\right) \\ &\quad \geq \psi \sum_{t=j}^{T} \sum_{\mathbf{h}^t \in \mathcal{F}^t(s,j)} \beta^t \theta(\mathbf{h}^t) k(\mathbf{h}^{s-1}, s)^\alpha \Pr\left(\mathbf{h}^t\right), \text{ for } s = 1, \cdots, S \text{ and } s \leq j \leq T.\end{aligned} \tag{23}$$

$$k((\mathbf{h}^{t-1}, t), t) = k((\mathbf{h}^{t-1}, t-1), t), \text{ for all } t \text{ where } t - 1 = h_{t-1}(\mathbf{h}^{t-1}), \tag{24}$$

$$\begin{aligned} k(\mathbf{h}^t, t) &= k((\mathbf{h}^{s-1}, s-1), s), &&\text{for all } t > s = h_{s-1}(\mathbf{h}^t) \text{ and } s < S, \\ k(\mathbf{h}^t, t) &= k(\mathbf{h}^S, S), &&\text{for } t > S \text{ and } S = h_S(\mathbf{h}^t), \end{aligned} \tag{25}$$

and

$$\sum_{t=1}^{T}\sum_{\mathbf{h}^t\in\mathcal{H}^t}\beta^t\left[x(\mathbf{h}^t)-C\left(p(\mathbf{h}^t),k(\mathbf{h}^t,t)\right)-qk(\mathbf{h}^t,t)\right]\Pr\left(\mathbf{h}^t\right)-\phi\geq 0. \tag{26}$$

Note how (22) 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, (11) and (12), to be expressed in the more compact single constraint (23). The objective function (P3) and the rest of the constraints (21), (24) to (26) are the direct analogues 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, happens with zero probability.[10] The set of restricted reporting strategies, $\mathcal{R}$, consists of all reporting strategies, ζ, such that:

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

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

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

subject to

$$\sum_{t=1}^{T}\sum_{\mathbf{h}^t\in\mathcal{F}^t}\beta^t\left[\theta(\mathbf{h}^t)k(\mathbf{h}^t,t)^{\alpha}-x(\mathbf{h}^t)\right]\Pr\left(\mathbf{h}^t\right)$$
$$\geq\max_{\zeta\in\mathcal{R}}\sum_{t=1}^{T}\sum_{\mathbf{h}^t\in\mathcal{F}^t}\beta^t\left[\theta(\mathbf{h}^t)k(\zeta^t(\mathbf{h}^t),t)^{\alpha}-x(\zeta^t(\mathbf{h}^t))\right]\prod_{n=1}^{t}\left[1-p\left(\zeta^n(\mathbf{h}^n)\right)\right]\Pr\left(\mathbf{h}^t\right), \tag{27}$$

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

$$\sum_{t=1}^{T} \sum_{\mathbf{h}^t \in \mathcal{F}^t} \beta^t \left[x(\mathbf{h}^t) - C\left(p(\mathbf{h}^t), k(\mathbf{h}^t, t)\right) - qk(\mathbf{h}^t, t) \right] \Pr\left(\mathbf{h}^t\right) - \phi \geq 0, \tag{28}$$

and (21), (23), (24), and (25).

The lemma presented below holds.

Lemma 5 *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 of the 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 (11) and (12) 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 (27) 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 6 *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 satifies the incentive constraint in one problem must satisfy the incentive constraint in 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 (27) 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^*). 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 afterwards. 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 satifies the incentive constraint (7) attached to the appended version of problem (P2). Assume that this allocation violates the incentive constraint (27) 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 truthfull 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 (27) to be violated. This implies that (7) must have been violated at node (s^*, s^*), a contradiction. ■

12.2 Proofs for 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 (11) and (12) have been added.

12.3 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 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 (P2): in the zero-profit constraint of the intermediary (10) and on the right-hand side of the incentive constraint (7) at node (u, u). Picture the Lagrangian associated with problem (P2). By setting $p(u-1, t) = 0$, profits to the intermediary can be increased through the zero-profit constraint (10). 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 (10) as $C_1\left(0, k(u-1,t)\right) = 0$. This implies a contradiction because the value of the Lagrangian will increase. ■

12.4 Proof of Backloading—Lemmas 2 and 3

Proof. (Lemma 3—Lemma 2 is a special case) Consider the no-retention constraint (11) at node $(s, s+1)$. Here a stall has just occurred. In order 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 $x(s,s+1), x(s,s+2), \cdots, (s,T-1)$ 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 has does effect 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 (11) 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 (12) 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 (13). 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 (12) 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 1 *(Lemma 2) 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 2 is a special case of Lemma 3.*

12.5 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

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

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 (P2). Now, using the information, irreversibility, and zero-profit constraints, (8), (9), and (10), in conjunction with the solution for the $x(s, t)$'s presented in Lemma 3, 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:

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

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; i.e., 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 1, 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. ■

12.6 Data

12.6.1 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 Organisation (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.

12.6.2 Figure 6

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		

12.6.3 Figure 7

The data for India, Mexico, and the U.S. displayed in Figure 7 are from Hsieh and Klenow (2010). The table below shows the statistics used to construct Figure 7.

Hsieh and Klenow (2010) 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

12.6.4 Table 4

Labor productivity in banking. Labor productivity is calculated for the year 2008. Deposits are divided by employment in the banking sector. Deposits are from the World Bank financial Indicators series "dbagdp" multiplied by GDP in purchasing power parity terms (Penn World Table). Employment data for the U.S. are from the Bureau of Labor Statistics, where all employees for commercial banking are used (Haver code: LAN211A@USECON). Employment data for Mexico are from the Comisión Nacional Bancaria de Valores. Employment data for India are from the Reserve Bank of India: A Profile of Banks. The value for productivity is then normalized by the value for U.S. so that all countries are expressed relative to the productivity in the U.S.

References

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.

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

Greenwood, Jeremy, Wang, Cheng and Sanchez, Juan M. Forthcoming. “Quantifying the Impact of Financial Development on Economic Development.” *Review of Economics Dynamics*. (Special issue on "Misallocation and Productivity," edited by Diego Restuccia and Richard Rogerson.)

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

Hsieh, Chang-Tai and Klenow, Peter. 2010. “The Life Cycle of Plants in India and Mexico.” Presentation, MIT.

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.

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

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.

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 Kenichi Ueda. 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.