Version: NBER–EFJK growth group, February 2012

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

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

Abstract

What determines the choice of technology within a country? 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 the decision to underwrite technology adoption. Can such a theory help to explain the differences in TFP and establishment-size distributions between India, Mexico and the U.S.? Some 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

---

** Preliminary and incomplete*: This is a report on some research in progress, which 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 employ 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 project 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 due to private information problems. And, in some countries, it may be difficult even 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 tell the truth. 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 may restrict the type of investment projects that can be profitably financed. The design of incentive schemes may be severely circumscribed, though. It may not be 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 type of technologies that can be funded.

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 productivity, 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.

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 that financial development has 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 circumscribe the set of technologies that it can adopt. In the benchmark model, new firms enter the economy every period with blueprints for production opportunities. A new firm will go to 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, however; extending the Greenwood, Sanchez and Wang (2010) analysis to dynamic contracts 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. 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 for ventures differ across firms. 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 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. Greenwood, Sanchez and Wang (2010) also allow for blueprints to differ across firms, but again this is done in a static setting. Placing things into a dynamic framework brings to the fore some new and important considerations.

For certain blueprints it may not be feasible for any intermediary to offer a lending

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

contract, given the state of the financial system. In particular, a backloading strategy of the sort analyzed in Clementi and Hopenhayn (2006) will not work. This may happen when the blueprint requires a lot of up-front investments before much pertinent information about the state of the project is revealed to the intermediary. Monitoring may be needed to make a lending contract viable, but it may be too expensive to undertake.

When monitoring can't be used, intermediaries will have to rely on incentive schemes to fund technologies. This leads to the second stage of the theoretical analysis. Think about backloading strategies: These redirect payoffs for the firm away from the start of the project toward the end, where they are contingent upon performance. This requires that the intermediary has some ability to control the publicly acknowledged cash flows of the firm. In some countries this can't be done effectively. Thus, for certain technologies it may not be possible to write a lending contract. The benchmark model is extended, in the second stage of the analysis, to study the situation in which it is costly to control the publicly acknowledged profit stream of the firm. It turns out that the ability to monitor helps with the costly cash-flow control problem as well.

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.[2]

[2] There is other quantitative work examining the link between economic development and financial development. For example, Buera et al. (2011) focus on the importance of borrowing constraints. Limited investor protection is emphasized by Castro et al. (2009). Greenwood, Sanchez and Wang (2010b) take 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

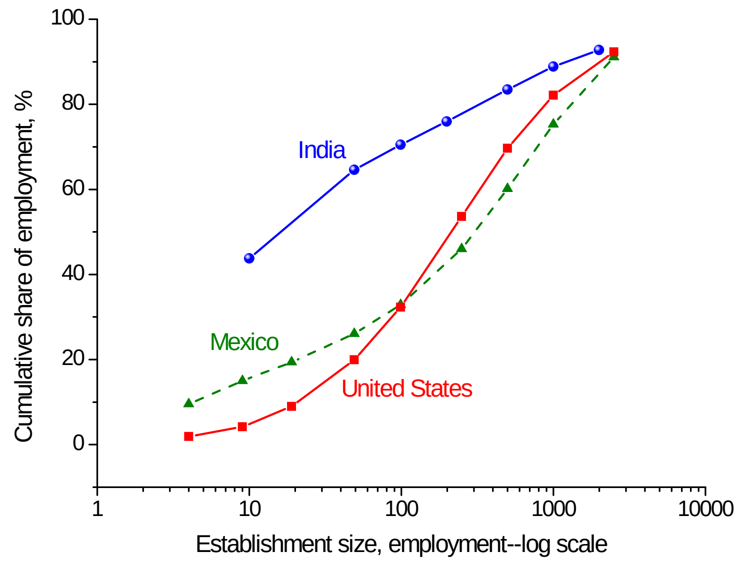

Figure 1: Establishment Size Distribution of Employment; India, Mexico and the U.S. Data sources for all figures are presented in the Appendix.

Hsieh and Klenow (2010) document some interesting differences in firms across these three countries. The average firm size is much smaller in Mexico than the U.S. and is much smaller in India than Mexico. In addition, the level of labor productivity follows a similar pattern. Figure 1 plots the overall cumulative distribution for establishment size in these countries. (See the Appendix for all descriptions of the data used in the paper.) Look at the left-hand side of the plot: The share of employment in establishments with fewer than 10 employees is bigger in Mexico than the U.S. and is larger in India relative to Mexico. Second, the largest firms in the U.S. are much older than those in Mexico or India. Figure 2 displays this fact. It plots the complementary cumulative distribution of employment by age–i.e., it graphs one minus the cumulative distribution of employment by age. Focus on the right-hand side of the graph: The share of employment that establishments older than 30 years contribute is bigger in the U.S. than in India or Mexico. These data suggest that these countries are using very different technologies.

stressed by Townsend and Ueda (2010). All of this research is very different in nature from what is pursued here.

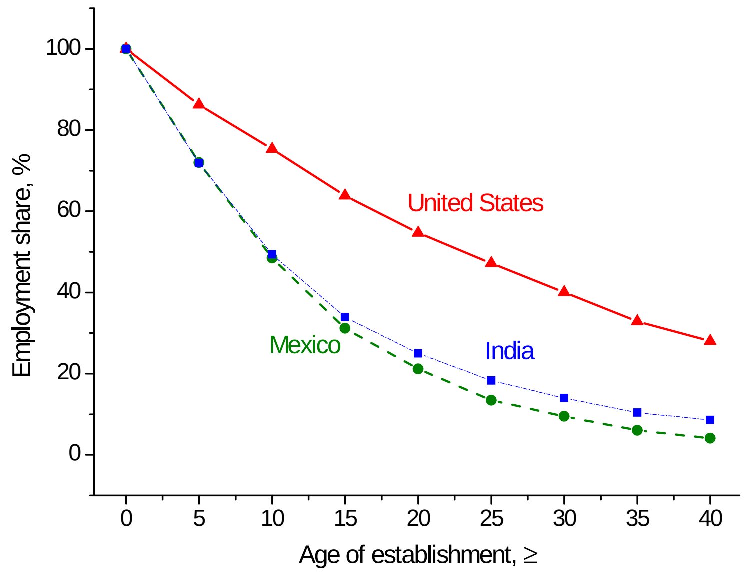

Figure 2: Distribution of Employment by Establishment Age, India, Mexico and the U.S.

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 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 need monitoring to implement. The second is a less profitable, intermediate-level technology that calls for smaller up-front investments relative to the timing of information. 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 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? Yes is the answer.

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 firms 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 it may be too costly to implement the type of backloading strategies required to finance the technologies used in Mexico and the U.S. 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 firms.

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 firm starts off with a blueprint for a 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 must decide which types of projects to finance. They borrow funds from consumers/workers in the economy at a fixed rate of return. Intermediation is competitive. Last, 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 into the background.

3 Ventures

Each period, new firms enter into 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 $\tau \equiv \{\theta_0, \theta_1, ..., \theta_S, \phi\}$. The firm enters into 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 happens, 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 3 illustrates potential productivity paths for a firm over its lifetime.

In the t-th 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 $t-1$.

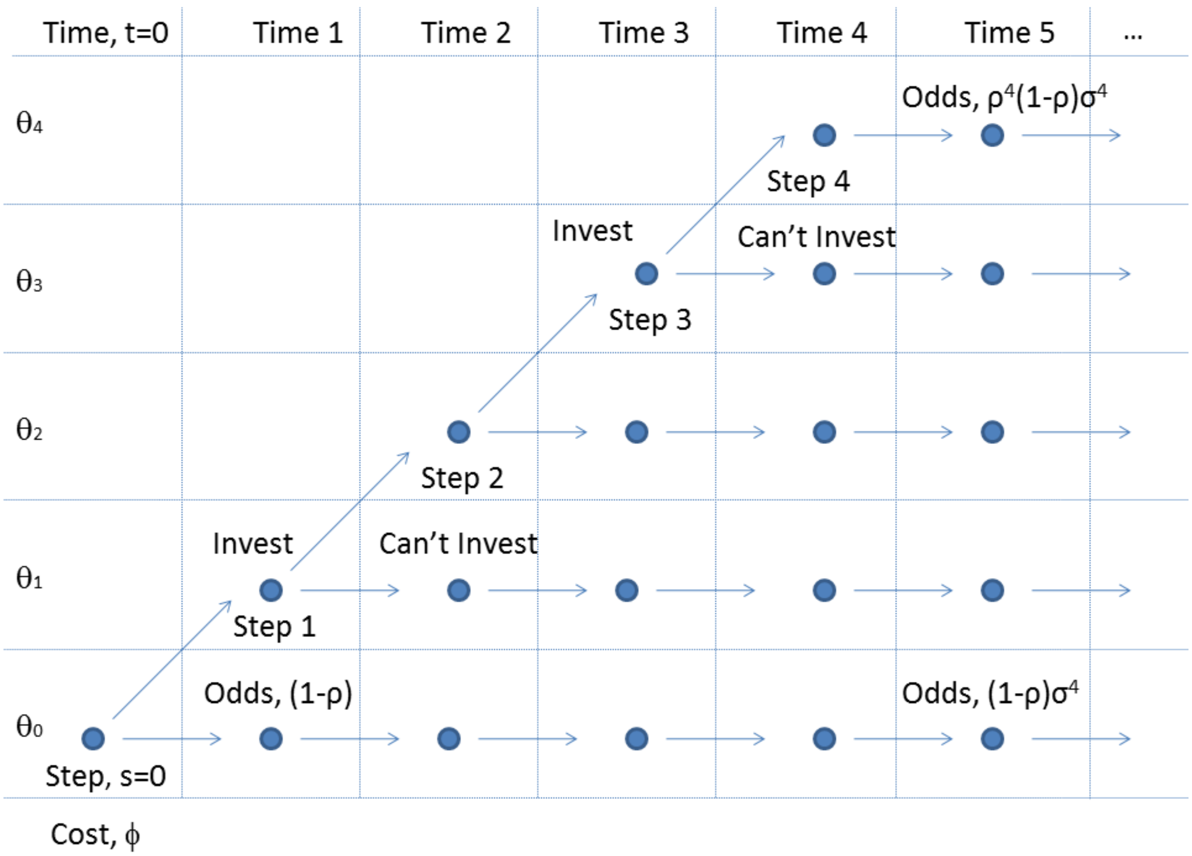

Figure 3: Possible productivity paths for a venture over its lifespan

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

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

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 a 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 supplies working capital to ventures. Intermediaries enter into financial contracts with new firms. 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 upon 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.[3] 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:

$$C(k, p; w, z) = w(\frac{k}{z})^{\gamma}(\frac{1}{1-p} - 1)p, \text{ with } \gamma > 1. \tag{4}$$

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 $C(k, 0; w, z) = 0$ and $C_1(0, p; w, z) = C_2(k, 0; w, z) = 0$; as $p \to 1$, both $C(k, 1; w, z) \to \infty$ and $C_2(k, 1; w, z) \to \infty$. Second, the marginal and total costs of monitoring are increasing in the wage rate, w. That is, $C_3(k, p; w, z) > 0$ and $C_{23}(k, p; w, z) > 0$. This is a desirable property if labor must be used for monitoring. Third, the cost is increasing and convex in the size of the project as measured by amalgamated input k; i.e., $C_1(k, p; w, z) > 0$ and $C_{11}(k, p; w, z) > 0$. A bigger 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 w and z will be suppressed when not needed to keep the notation simple.

[3] 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, $x(s,t)$; (iii) audit detection probabilities, $p(s,t)$. Because there is a large number of competitive intermediaries 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 principal 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

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

$$\sum_{t=u}^{T}\sum_{s=u}^{\min(t,S)}\beta^{t}\left[\theta_{s}k(s,t)^{\alpha}-x(s,t)\right]\Pr\left(s,t\right)$$
$$\geq\sum_{t=u}^{T}\sum_{s=u}^{\min\{t,S\}}\beta^{t}\left[\theta_{s}k(u-1,t)^{\alpha}-x(u-1,t)\right]\prod_{n=u}^{t}\left[1-p\left(u-1,n\right)\right]\Pr\left(s,t\right),$$
$$\text{for all } u\in\{1,...,S\}, \tag{7}$$

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

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

and

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

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. The validity of this is established in Appendix 11.1. 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 which implies that it lied in the past. This general problem has a single time-1 incentive constraint that requires the expected present-value to firm from adopting a truth-telling strategy to be at least as good as the expected present-value to 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 3.

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 isn't 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 out 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, $t-1$, where there isn't 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 doesn't, $k(t, t)$. This explains equation (8). The information constraint is portrayed in Figure 4 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 at 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 4. All working capitals within a horizontal box take the same value. Think about 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

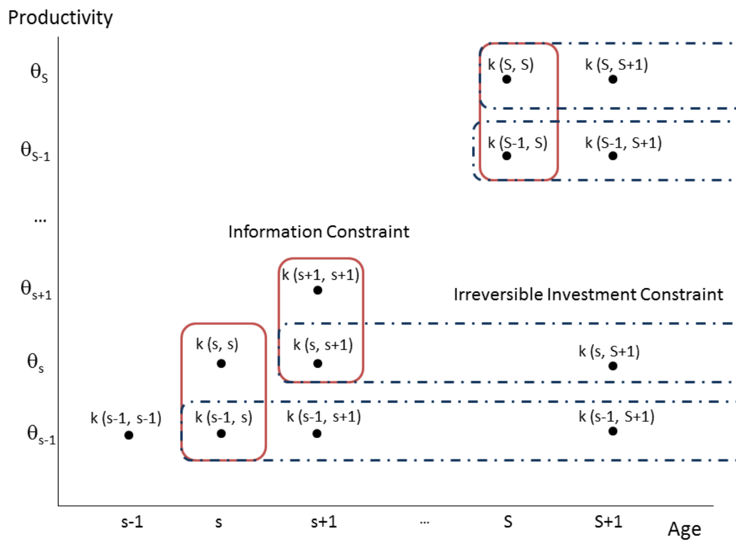

Figure 4: The information and irreversibility constraints

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 up-front fixed costs 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 level $u - 1$. 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 of a stall at step $u \in \{1, 2, \cdots, S\}$ 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 the stalled step.*

Proof. See Appendix 11.3. ■

Furthermore, in this situation the intermediary should take all of the firm's output for as long as the project operates. This reduces the incentive to lie, as can be seen from (7).

Lemma 2 *(Seize everything) Suppose the incentive constraint (7) binds at some step* $u \in \{1, 2, \cdots, S\}$*. Upon a report by the firm of a stall at this step the intermediary will take everything, so long as the firm continues to operate. This is done by setting* $x(u-1,t) = \theta_{u-1}k(u-1,t)^{\alpha}$*, for* $t = u, u+1, \cdots, T$*.*

Proof. See Appendix 11.4. ■

Something stronger than the above lemma can be said. Focus on Figure 5. Consider the last date t where both productivity could have grown and the incentive constraint (7) binds. The contract specifies that everything should be taken at the nodes below the (t,t) node in the figure. This includes all the (i,i) nodes, for $i < t$, where the firm *truthfully* reports that the good state has occurred. This is done as part of a backloading strategy. Here the intermediary provides the firm with all of its rewards somewhere in the trapezoid that has its bottom left-hand corner lying at the (t,t) node. This encourages the firm to tell the truth.

Lemma 3 *(Backloading) Let* μ_u *be the Lagrange multiplier associated with the incentive constraint (7) at the node* (u,u) *and* $\lambda_{i,l}$ *be the multiplier linked with the limited liability constraint (6) at node* (i,l)*. Assume that a new firm expects to make profits, so that* $v > 0$ *in (P2). Suppose* $\mu_u > 0$ *for some* u *and let* $t \equiv \max_u\{u : \mu_u > 0\}$*. Then,* $\lambda_{i,l} > 0$*, for all* $i < t$ *and all* $l \geq i$*.*

Proof. See Appendix 11.5. ■

Remark 1 *It is possible for* $\lambda_{u,u} > 0$ *even when* $\mu_u = 0$ *(for* $u < t$*).*

Last, 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 at some point $\mu_{s+1} = 0$ for all $s \geq m$. In other words, the incentive constraint no longer binds after period $m+1$. Will investment be efficient from then on? Yes, is the answer.

Lemma 4 *(Efficient investment) Let* $\mu_{s+1} = 0$ *for all* $s \geq t \equiv \max_u\{u : \mu_u > 0\}$*. Investment will be efficiently undertaken upon arriving at the date/state combination* (t,t)*.*

Proof. See Appendix 11.6. ■

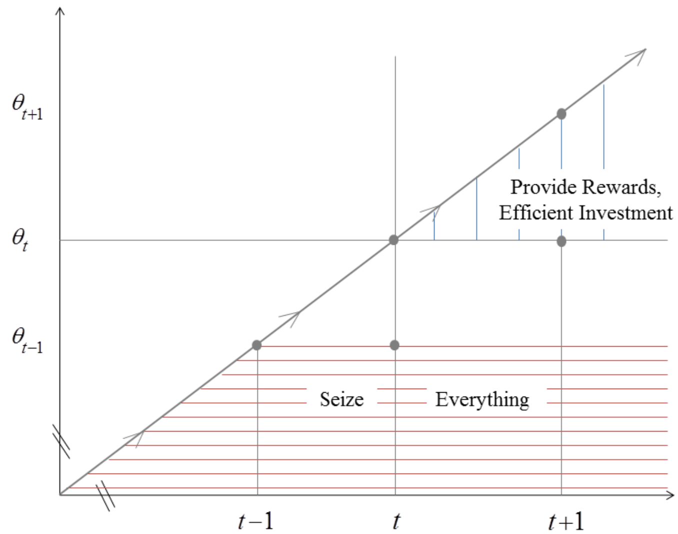

Figure 5: Backloading. The firm will begin to receive payments only after the node (t, t) is reached. Investment is undertaken efficiently from that node on

5.1 Discussion

The solution to the above contract problem shares some features that are common to dynamic contracts, but has some properties that are quite different as well. 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. Lemma 3 states that the intermediary should take everything up until the *last* point in time when the incentive constraint binds. This is true even though the incentive constraint may be slack at times before this date. Compare this with Clementi and Hopenhayn (2006, Lemma 2) where the incentive constraint always binds in those states when the lender takes everything. In Clementi and Hopenhayn (2006) once promised expected utilities hit an upper bound, $\widetilde{V}$, capital accumulation is optimal thereafter. The state-date combination (t, t) plays this role here. The firm earns nothing until the node is reached. This happens with probability $\Pr(t, t)$. Therefore, the firm must expect to earn $v/\Pr(t, t)$ from this node on, where again

v is defined in (P2).

The focus of the analysis is on competitive intermediation. Therefore, the intermediary will earn zero profits, as stated by (10). Using the definition for v in (10) yields

$$v = \sum_{l=1}^{T} \sum_{i=0}^{\min\{l,S\}} \beta^l [\theta_i k(i,l)^\alpha - qk(i,l) - C\left(p(i,l), k(i,l)\right)] \Pr\left(i,l\right) - \phi.$$

By Lemma 1, monitoring will occur only for those state/date combinations (i,l), for $i < t$ and all $l \geq i$, where the incentive constraint binds. Therefore, this expression can be rewritten as

$$\begin{aligned} v \quad = \quad & \sum_{l=1}^{T} \sum_{i=0}^{\min\{l,t-1\}} \beta^l [\theta_i k(i,l)^\alpha - qk(i,l) - C\left(p(i,l), k(i,l)\right)] \Pr\left(i,l\right) \\ & + \sum_{l=t}^{T} \sum_{i=t}^{\min\{l,S\}} \beta^l [\theta_i k(i,l)^\alpha - qk(i,l)] \Pr\left(i,l\right) - \phi. \end{aligned}$$

The contract specifies that the firm earns all of the expected rents from the project, after the cost of raising the capital and monitoring are netted out. It is easy to see that v is an endogenous variable, unlike the upper bound in Clementi and Hopenhayn (2006). It depends on the time path for working capital and monitoring. It also is a function of the equilibrium level of wages, w, which is implicitly embedded in the q terms present in (P2).

After the state $t \equiv \max_u\{\mu_u > 0\}$ has been attained, any feasible time path of payments $\{x(i,l)\}$, for $i \geq t$ and $l \geq i$, from the firm to the intermediary that generates an expected present value of

$$\sum_{l=t}^{T} \sum_{i=t}^{\min\{l,S\}} \beta^l x(i,l) \Pr\left(i,l\right) = \sum_{l=t}^{T} \sum_{i=t}^{\min\{l,S\}} \beta^l \theta_i k(i,l)^\alpha - v$$

is permissible under the contract. This can be deduced by noting that only the expected present value of these payments will enter into the objective function in (P2) and the constraints (7) and (10). Specifically, suppose one is given a solution $\{x(i,l)\}$, for $i \geq t$ and $l \geq i$, in problem (P2). Any other sequence of payments with the same expected present value will also satisfy the constraints (6) to (10) and give the same value for the objective function in (P2). This is not true for the $x(i,l)$'s when $i < t$ and $l \geq i$, because the $x(i,l)$'s terms will enter only the first i incentive constraints (7)–or in those where $u \leq i$.

Last, 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. The fact that productivity can only step up to the next rung or remain on the current step complicates things. 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 A Two-Period Example

A simple two-period example illustrating the contract setup is now presented. It shows how the shape of the productivity profile and the size of the fixed cost connected with a blueprint influence the form of the contract. It also illustrates the importance that monitoring may play in making a contract feasible. A blueprint, b, is described by the quadruple $b \equiv \{\theta_0 = 0, \theta_1 > 0, \theta_2 \geq \theta_1, \phi \geq 0\}$. A venture's survival will be guaranteed, implying $\sigma = 1$. 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. Finally, the cost of monitoring is assumed to be *prohibitive*; in particular, set $z = 0$. 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 costs of the venture.

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

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

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

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

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

Because monitoring is prohibitively expensive, backloading is the only way to satisfy the incentive constraints at nodes $(2,2)$ and $(1,1)$. Backloading implies that the firm receives a return of $\pi/(\beta^2\rho^2)$ 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

$$\theta_2 \leq \theta_1 + \pi/(\beta^2\rho^2). \tag{11}$$

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

$$\theta_1 \leq \pi/[(1+\beta)(\rho\beta)]. \tag{12}$$

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

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

$$\theta_2 \geq \pi/(\beta^2\rho^2) - \{\beta\rho[1+\beta(1-\rho)]/(\beta^2\rho^2)\}\theta_1. \tag{13}$$

To understand the impact of variations in the fixed cost, set $\phi = 0$. It is a simple matter to show that both incentive constraints must hold. In this situation all of the returns from the project will be given to the firm. The payoff from lying arises solely from the possibility of evading the fixed cost. As ϕ increases the "first-best" gross profits of the blueprint, $\beta\rho\theta_1 + \beta^2\rho(1-\rho)\theta_1 + \beta^2\rho^2\theta_2$, increase 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 6 plots the two incentive constraints (11) and (12), the 45-degree line and the fixed-cost constraint (13). 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. When θ_1 is high, then either θ_2 must relatively small or ϕ relatively large, in order 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 (12) 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–i.e., 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

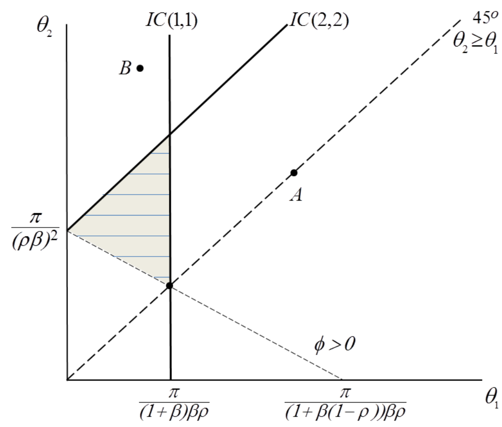

Figure 6: Set of implementable first-best allocations

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 $(2, 2)$. I.e., there are no profits, only losses, that the intermediary can redirect to node $(2, 2)$ from the other nodes on the tree. By lying the firm avoids these losses. To implement such a point monitoring must be used. 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 Equilibrium

There is one unit of labor available in the economy. This must be split across all operating firms. Recall that each new firm is connected with a productivity ladder $\{\theta_0, \theta_1, ..., \theta_S\}$ and 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. Suppose that type is distributed across firms according to the cumulative distribution function $F : \mathcal{T} \to [0, 1]$. 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. Not all type-τ new firms may receive funding. In particular, for some type-τ firms there may not exist a solution to the contract defined by (P2) that allows them to earn non-negative profits (so that $v(\tau) \geq 0$). Let $\mathcal{A}(w)$ denote the active set of ventures. This set depends on the equilibrium wage rate, w, which influences the price of the amalgamated input, q through (1). The active set is defined by

$$\mathcal{A}(w) \equiv \{\tau : v(\tau) \geq 0\}. \tag{14}$$

The labor-market clearing condition for the economy then reads

$$\sum_{t=1}^{T} \sum_{s=1}^{\min\{t,S\}} \int_{\mathcal{A}(w)} [l(s, t; \tau) + l_m(s, t; \tau)] \Pr(s, t)\, dF(\tau) = 1, \tag{15}$$

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) and is given by

$$l_m(s, t; \tau) = [\frac{k(s, t; \tau)}{z}]^{\gamma}[\frac{1}{1 - p(s, t; \tau)} - 1]p(s, t; \tau) \text{ [cf (4)]}. \tag{16}$$

A definition of the competitive equilibrium under study will now be 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 ventures, $\mathcal{A}(w) \sqsubseteq \mathcal{T}$; (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 maximizes the value of a type-τ venture, as set out by (P2), given the amalgamated input price q.*
2. *A venture is funded only if it is contained in the active set, $\mathcal{A}(w)$, as specified by (14), where $v(\tau)$ is determined by (P2).*
3. *A type-τ venture hires labor, $l(s,t;\tau)$, so as 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 (16).*
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 (15).*

8 The Choice of Venture in Mexico and the U.S.: A Quantitative Exploration

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; the presence of labor unions and other factors that may dissuade certain types of business. While these are valid reasons, the focus here will be on differences in the efficiency of the financial system. This is done without apology, because abstraction is a necessary ingredient for theory.

The idea is this: The precise form of a financial contract between a firm and intermediary will depend on both the type of technology that is being bankrolled and the efficiency of the financial system. Specifically, the details of the contract will be a function of the productivity ladder for the project. For some ladders, backloading the reward structure will be enough to guarantee that a viable contract between a firm and intermediary can be written. For others, monitoring will have to employed. The desirability or feasibility of monitoring will depend on the state of a country's financial system.

In what follows, a numerical example will be constructed where Mexico *chooses* to adopt a different technology from the U.S. In particular, the American (or advanced) technology will offer a productivity profile that grows much faster than the Mexican (which represents an intermediate-level technology) contour. Financing the American technology requires a level of monitoring that only an efficient financial system can undertake. The Mexican technology does not require this. The example is constructed so that the framework matches the size distribution of establishments observed in Mexico and the U.S. It also replicates the pattern of employment by age. These two facts discipline the assumed productivity profiles. In the equilibrium constructed, it is not desirable to finance the Mexican technology in America, given the state of the U.S. financial system and American input prices. Likewise, it is not worthwhile to underwrite the American technology in Mexico, using the latter's financial system and input prices.

Since the focus here is on the long run, let the length of a period be 5 years and set the number of periods be 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 thing to do, 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 Xi (2008) this lies in the range of recent studies.

8.1 Estimating the input price, q

A key input into the analysis is the price for the amalgamated input, q. 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 American 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 American prices for capital and consumption goods. Assume that interest plus depreciation in each country sum 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}, will be selected. This will be obtained dividing the annual payroll by the number of employees in all establishment in the manufacturing sector using the 2008 Annual Survey of Manufactures. Thus, $w^{US} = 47,501$. Last, given the above datums 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}? To determine this requires knowing r^{MX}, w^{MX}, and χ^{MX}. First, a value for the rental price of capital, r^{MX}, will be determined. The relative price of capital is estimated, from Penn World Tables, to be about 21 percent higher in Mexico than 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 will be pinned down using data on annual payroll and the total number of workers in manufacturing establishment; in this case the data come from INEGI. The result is $w^{MX} = 21,419$ once Mexican pesos are converted to US dollars using

PPP. 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 (2011). So set $\chi^{MX} = 0.55 \times \chi^{US} = 53,035$. Finally, by plugging in the obtained values for r^{MX}, w^{MX}, and χ^{MX} into 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.

8.2 Parameterizing the Technology Ladder

There will be 9 unique rungs on the technology ladder, the last three being the same. The U.S. uses an advanced technology whose productivity ladder (along the diagonal) is described by

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

Mexico uses a traditional technology represented 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 generic process describing the odds of stalling is given by

$$\rho_s = \overline{\rho}_0 + \overline{\rho}_1 s, \text{ for } s = 1, \cdots, 9.$$

The parameter values for this process are different in Mexico and from those in the U.S. The probability of surviving (until age t) is the same for both 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. 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. Last, an upper bound on working capital is imposed. This is denoted by $\overline{k}$.

8.3 Calibrating the Technology Ladder

First, the survival probabilities are taken 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, several stylized facts about establishment-size distributions in Mexico and the U.S. The first stylized fact is the Lorenz curve for employment by establishment in each country, which is represented by a collection of points. The next fact is the complementary distribution of employment by establishment age in each nation, again characterized by set of points. The last fact is the mean size of an establishment in Mexico and the U.S. So, let D^j proxy for the j-th data target for the model and $M^j(p)$ represent model's prediction for this data target as a function of the parameter vector $p \equiv \{\overline{\theta}_0, \overline{\theta}_1, \overline{\theta}_2, , \overline{\theta}_3, \phi, \overline{\rho}_0, \overline{\rho}_1, \overline{k}\}$. The parameter vector p is chosen in the following fashion:

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

Figure 7 shows the salient features of the technologies used in Mexico and U.S. 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 convex one, as can be seen from Figure 7. Note that the ascent is much steeper for a U.S. firm than a Mexican one. The chances of stall are higher with the Mexican technology. For Mexico they rise with a move up the ladder, while for the U.S. they decline. The survival processes are the same in each country. The parameter values used in numerical example are presented in Table I.

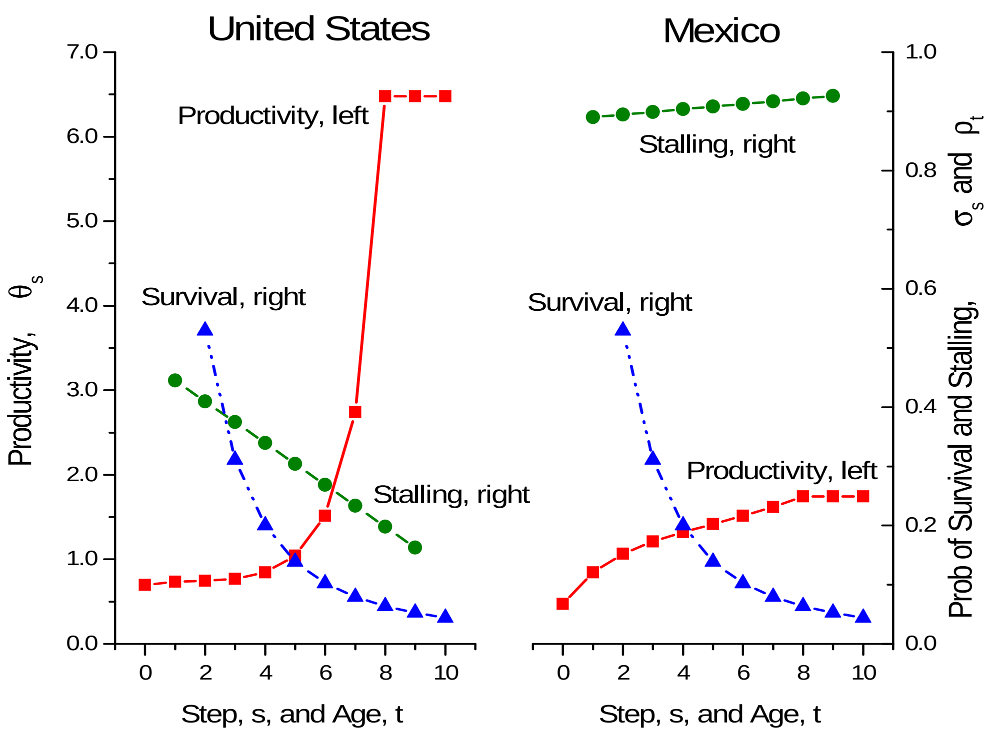

Figure 7: Productivity, survival and stalling in Mexico and the U.S. (model). The diagram displays the assumed productivity ladder (left axis) for Mexico and the U.S.. It also illustrates the probability profiles for survival and stalling (right axis).

TABLE I: PARAMETER VALUES

Parameter	*Value*	
	U.S.	Mexico
Discount Factor	$\beta = 0.98^5$	$\beta = 0.98^5$
Prod Function–scale, capital's share	$\alpha = 0.85$, $\omega = 0.33$	$\alpha = 0.85$, $\omega = 0.33$
Labor, efficiency	$\chi = 96,427$	$\chi = 53,035$
Ladder, state parameters	$\overline{\theta}_0 = -0.506, \overline{\theta}_1 = 0.205$	$\overline{\theta}_0 = 0.646, \overline{\theta}_1 = 1.079$
	$\overline{\theta}_2 = -0.07, \overline{\theta}_3 = 0.0085$	$\overline{\theta}_2 = -0.135, \overline{\theta}_3 = 0.009$
Pr Stall, parameters	$\overline{\rho}_0 = 0.445, \overline{\rho}_1 = -0.035$	$\overline{\rho}_0 = 0.89, \overline{\rho}_1 = 0.0045$
Pr Survival, time t	$\overline{\sigma}_0 = 0.14$, $\overline{\sigma}_1 = -0.022$	$\overline{\sigma}_0 = 0.14$, $\overline{\sigma}_1 = -0.022$
	$\overline{\sigma}_2 = 0.00115$	$\overline{\sigma}_2 = 0.00115$
Capital, upper bound	$\overline{k} = 3$	$\overline{k} = 25$
Fixed Cost	$\phi = 0.07$	$\phi = 0.005$
Monitoring, function	$\gamma = 2.0$	$\gamma = 2.0$
Monitoring, efficiency	$z = 10$	$z = 1$
Input Price	$q = 1.0$	$q = 0.937$

8.4 Establishment-Size Distributions in Mexico and the U.S.

The expected profitability for a Mexican firm using the U.S. (advanced) technology is plotted in Figure 8 as a function of the productivity in financial sector, z. The plot is based on the prevailing level of Mexican factor prices summarized by q. Expected profits, v, have a concave shape in z. Observe that profitability begins to descend rapidly at some point as z falls. Eventually, profits hit zero (somewhere around $z = 1.2$). Thus, there is a lower bound on financial sector productivity that will sustain the adoption of the U.S. technology in Mexico. The intermediate-level technology is used when z lies below this number. Of course, a similar picture could be plotted for the advanced technology in U.S. The American level of z would support the use of this technology at American factor prices.

Figures 9 and 10 plot the model's fit for the Mexican and U.S. establishment-size distrib-

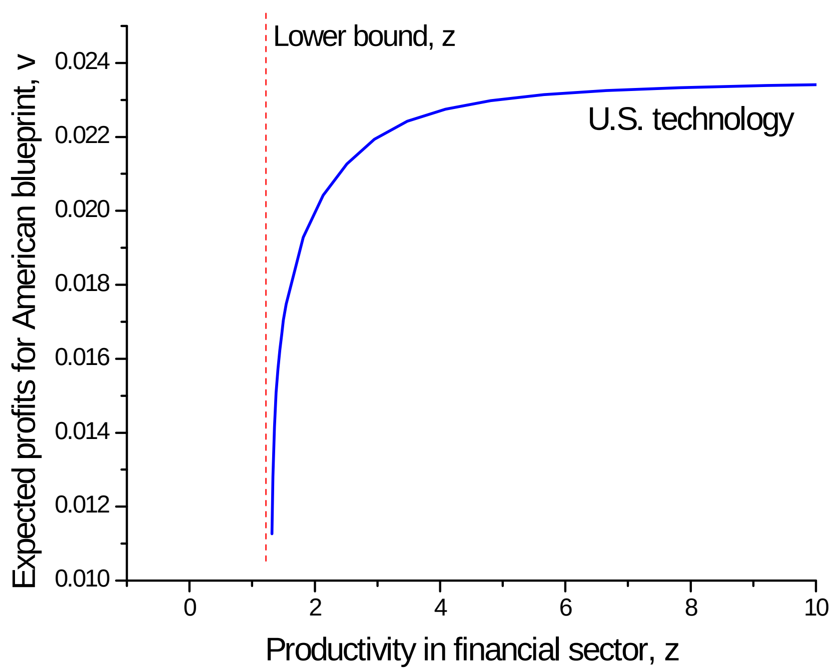

Figure 8: The expected profitability of the U.S. (or advanced) technology for Mexican firms, v, as a function of financial sector productivity, z

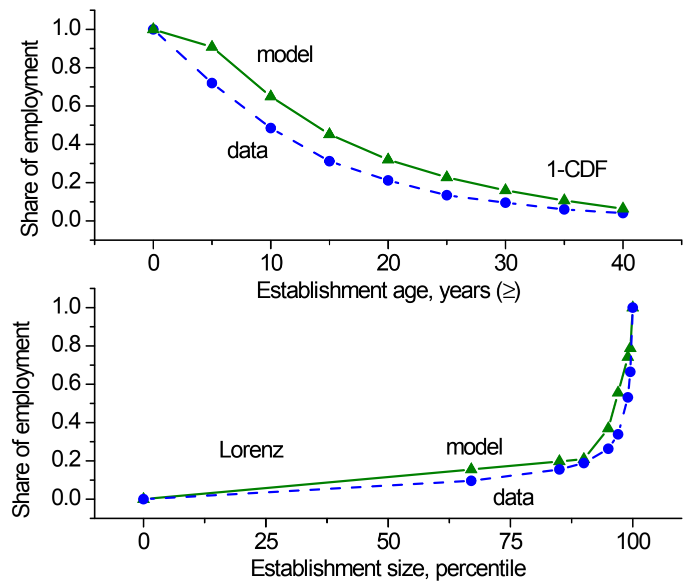

Figure 9: Mexican establishment-size distribution, data and model. The lower panel plots the Lorenz curve for Mexican establishments. The upper panel charts the complementary cumulative distribution of employment by age.

utions. Take Mexico first or Figure 9. Focus on the lower panel. The model does an excellent job matching the Lorenz curve for establishment size in Mexico. The model tends to overpredict the share of small firms in employment. The framework mimics less well the share of employment by age. This is shown in the upper panel, which charts the complementary cumulative distribution by age. The fit isn't bad 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 5 years. Now switch to the U.S. Examine the lower panel in Figure 10. Again, the model overpredicts the share of small establishments in employment, but more so in the U.S. than in Mexico. The model does a superb job matching the share of employment by age for the U.S. Still, it can't quite capture the fact that some old firms in the U.S. are very large.

The framework has the potential to match the stylized facts mentioned in the introduction. The average size of establishments in the model is smaller in Mexico than the U.S. The

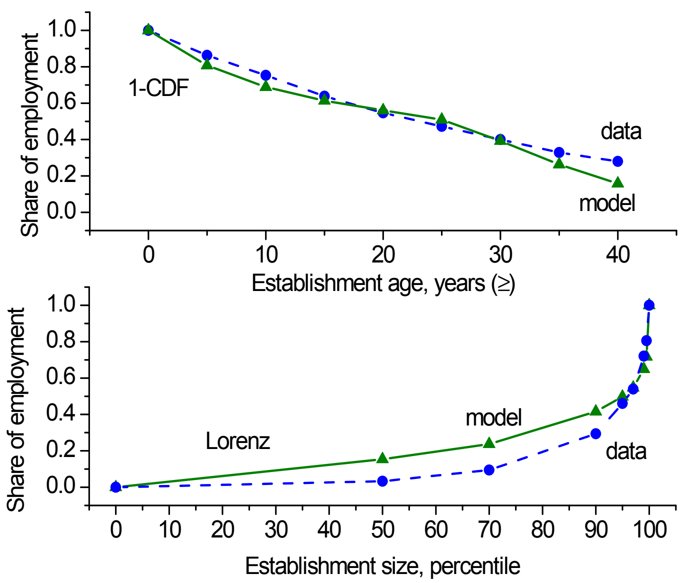

Figure 10: U.S. establishment-size distribution, data and model. The lower panel plots the Lorenz curve for U.S. establishments. The upper panel charts the complementary cumulative distribution of employment by age.

model does an excellent job matching the mean size of employment, as can be seen from Table II. Establishments older than 30 years account for a smaller fraction of employment in Mexico relative to the U.S. The model matches this fact, too, but as can been seen there is still room for improvement, as old establishments' share of employment in the model is higher than in the Mexican data.

TABLE II: STYLIZED ESTABLISHMENT FACTS

	Average Size		*Size, ≥ 30 yrs (fraction)*	
	Data	Model	Data	Model
Mexico	19.9	18.4	0.10	0.15
U.S.	41.9	41.9	0.40	0.39

8.5 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 technology τ. 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 $\mathbf{o}(\tau)/\mathbf{l}(\tau)$. The model predicts that it will be much higher in the U.S. than in Mexico, as can be seen in Table III. In particular, the model calculates that it should be 3.2, which actually is close to what is found in the data. 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. Here, $\kappa = 0.33$. The model predicts that measured TFP in the U.S. should be 2.3 times Mexican TFP. In the data, it is 2. One could also ask by what factor U.S. productivity would drop if the U.S. were forced to adopt the Mexican production technology, due to inefficient financial markets. This is also displayed in the table. U.S. productivity would fall by 45 percent (in terms of ln differences). Likewise, by how much would Mexican productivity rise if it could adopt the U.S. technology? Specifically, let Mexico have the same level of efficiency in its financial sector as the U.S. does. This would lead to a 31 percent increase in its level of productivity.

Table III: Measures of Productivity

	U.S./Mexico		*Counterfactuals*	
	data	model	*U.S.*	*Mexico*
Labor Productivity–$\mathbf{o}(\tau)/\mathbf{l}(\tau)$	2.96	3.24	0.64	1.36
TFP–$\mathbf{o}(\tau)/[\mathbf{k}(\tau)^{\kappa}\mathbf{l}(\tau)^{1-\kappa}]$	1.99	2.33	0.64	1.36

9 The Contract with Costly Cash-Flow Control

In the above example, Mexico was not able to adopt the technology employed by the U.S., because this technology required monitoring of the firm by the intermediary. So, instead they used a less productive technology that could be financed using a backloading strategy alone. Now, there are countries in the world where the cost of production is much lower than in Mexico. Maybe they could they 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 doesn't require monitoring services. Again, why doesn't technology flow from rich to poor countries? An extension to the baseline theory is developed now that provides one possible answer. The revised model is then applied to India, where labor costs are extremely low.

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, etc., unless the intermediary pays very large enforcement costs to prevent this. 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.

9.1 Extending the Theory

Assume that a firm can openly take the fraction ψ of output, due to weak institutional structures. An intermediary can do two things to thwart this. First, it can stop this retention of funds, but at a prohibitively high fixed cost, ϵ. Assume that this fixed cost exceeds the expected present value of profits for the project in the full information world. Second, it can design the contract in a manner so that this is dissuaded. How will this effect 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 contract:

$$\sum_{j=t}^{T} \beta^t \left[\theta_s k(s,j)^\alpha - x(s,t)\right] \frac{\Pr(s,j)}{\Pr(s,t)}$$
$$\geq \psi \sum_{j=t}^{T} \beta^t \theta_s k(s,j)^\alpha \frac{\Pr(s,j)}{\Pr(s,t)}, \text{ for } s<t \text{ (off-diagonal node on ladder)}, \tag{17}$$

and

$$\sum_{t=u}^{T} \sum_{s=u}^{\min\{t,S\}} \beta^t \left[\theta_s k(s,t)^\alpha - x(s,t)\right] \frac{\Pr(s,t)}{\Pr(u,u)}$$
$$\geq \psi \sum_{t=u}^{T} \sum_{s=u}^{\min\{t,S\}} \beta^t \theta_s k(u-1,t)^\alpha \frac{\Pr(s,t)}{\Pr(u,u)}, \text{ for } s=t=u \text{ (diagonal node)}. \tag{18}$$

The first constraint (17) applies to the case where a stall has occurred at state s. Here, productivity is stuck at θ_s forever. The second constraint (18) governs to the situation where the firm can still move up the productivity ladder. If the firm exercises its retention option then the intermediary will keep the capital stock at $k(u-1,t)$; i.e., it will no longer evolve with the state of the firm's productivity. Equation (9) then implies that the capital stock is locked in.

In the baseline version of the model, it is always weakly efficient to make all payments to the firm at node (S,T) in order relax the incentive constraint. The retention option precludes this, however. In order to encourage the firm not to exercise its retention option, it pays for the intermediary to make additional payments, $c(s,T)$, to the firm at the terminal date T for all steps $s<S$ on the ladder, provided that the latter does not exercise its retention option at any time before T. This payment should equal the expected presented value of what the firm would receive if it exercised the retention option. Thus,

$$c(s,T) = \psi \frac{\sum_{t=s}^{T} \beta^t \theta_s k(s,t)^\alpha \Pr(s,j)}{\beta^T \Pr(s,T)}, \text{ for } s<T. \tag{19}$$

Lemma 5 *(Backloaded retention payments) It is weakly efficient to set:*

1. $x(s,t) = \theta_s k(s,t)^\alpha$, *for all* $t<T$,
2. $x(s,T) = \theta_s k(s,T)^\alpha - c(s,T)$, *for all* $s<S$, *where* $c(s,T)$ *is defined in (19).*

Proof. See Appendix 11.7. ■

Backloading the retention payments helps to satisfy the incentive constraint. To understand this, suppose that the firm lies and declares at stall at node (u, u). What happens if the intermediary detects this lie at some node (u, t), where $t \geq u$? Specifically, can the firm now retain the fraction ψ of output? No, is the answer. The intermediary can recover this lost output at the cost ϵ. It is important to note that such an "off-the-equilibrium-path" event will never actually transpire, because the contract is constructed so that the firm will never lie.[4] 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, he intermediary gives the firm the accrued value of these retentions, $c(u-1, T)$, 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 only receive the payment $c(u-1, T)$ if it successfully evades detection along the entire path from u to T. This happen with odds $\prod_{n=u}^{T}[1 - p(u-1, n)]$.

In the new setting the incentive compatibility constraint will read

$$\sum_{t=u}^{T} \sum_{s=u}^{\min(t,S\}} \beta^t \left[\theta_s k(s,t)^{\alpha} - x(s,t)\right] \Pr(s,t)$$
$$\geq \sum_{t=u}^{T} \sum_{s=u}^{\min\{t,S\}} \beta^t \left[\theta_s k(u-1,t)^{\alpha} - x(u-1,t)\right] \prod_{n=u}^{t} [1 - p(u-1,n)] \Pr(s,t), \tag{20}$$

for all $u \in \{1, ..., S\}$. Notice how the intermediary's ability to monitor interacts with the firm's potential to retain output. When monitoring is very effective it will be difficult for

[4] An alternative interpretation is that the intermediary is lending to many firms. A few of these firms, perhaps because they do not believe the intermediary is really committed to the optimal contract will misreport their productivity level. In this case, the intermediary would only want to use the option of paying ε and taking all of the output for this small subset of the set of borrowers in order to establish credibility with the rest of the firms.

a masquerading firm to retain output. 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. Because the lefthand sides of (18) and (20) are the same, they can be combined them into a single unified constraint:

$$\sum_{t=u}^{T}\sum_{s=u}^{\min(t,S)} \beta^t \left[\theta_s k(s,t)^\alpha - x(s,t)\right] \Pr(s,t)$$
$$\geq \max\{\sum_{t=u}^{T}\sum_{s=u}^{\min\{t,S\}} \beta^t \lambda\theta_s k(u-1,t)^\alpha \Pr(s,t)\,,$$
$$\sum_{t=u}^{T}\sum_{s=u}^{\min\{t,S\}} \beta^t \left[\theta_s k(u-1,t)^\alpha - x(u-1,t)\right] \prod_{n=u}^{t}\left[1-p(u-1,n)\right] \Pr(s,t)\} \tag{21}$$

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

The lemma below takes stock of discussion.

Lemma 6 *(The Contract Problem with Costly Control) An efficient contract is given by maximizing the payoff to the firm as given by the objective function in (P2) subject to: (i) the restrictions on payments outlined in Lemma 5; (ii) the incentive-cum-retention constraint (21); (iii) the capital accumulation constraints, (8) and (9); (iv) the zero-profit condition for the intermediary, (10).*

9.2 The Two-Period Example, Continued

To better understand how the contract will change, return to the two-period example presented earlier. The firm now has the ability to retain the fraction ψ of output at any node on the ladder. Suppose that the firm finds itself at node $(2,2)$. If it tells the truth then it will receive $[\pi - \beta^2\rho(1-\rho)\psi\theta_1]/(\beta^2\rho^2)$. Its payoff has been reduced by the amount $\beta^2\rho(1-\rho)\psi\theta_1/(\beta^2\rho^2)$. This is because the intermediary must give a firm that stalls in period 2 the amount $\beta^2\psi\theta_1$, in present value terms. Such a stall happens with probability $\rho(1-\rho)$. (Note that when a firm stalls in period 1 it cannot retain any output since $\theta_0 = 0$.) This payoff must exceed what the firm will get if it either (i), falsely declares a stall or (ii), decides to keep the fraction ψ of period-2 output.

When the firm lies it can now pocket $\theta_2 - \theta_1 + \psi\theta_1$. The firm can retain $\psi\theta_1$ units of output when it stalls at state one, because this income is too expensive for the intermediary to recover. 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 Figure 11 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 $\psi > \rho$. If the firm decides to keep the fraction ψ of second-period output it will get $\psi\theta_2$. Thus, retention constraint 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 Figure 11 illustrates the retention constraint. It slopes downwards and is located above the $\phi > 0$ constraint, since $\pi/(\psi\beta^2\rho^2) \geq \pi/(\beta^2\rho^2)$ and $\pi/\{[1+\beta(1-\rho)]\beta\rho\} < \pi/[(1-\rho)]\beta\rho$.

Move back in time to node $(1,1)$. If the firm transits to node $(2,2)$ it will earn in profits in the amount $[\pi - \beta^2\rho(1-\rho)\psi\theta_1]/(\beta^2\rho^2)$. This occurs with probability ρ. If it moves to $(1,2)$ then it will receive $\psi\theta_1$. Therefore, its expected discounted profits from telling the truth are $\beta\rho[\pi - \beta^2\rho(1-\rho)\psi\theta_1]/(\beta^2\rho^2) + (1-\rho)\beta\psi\theta_1 = \beta\rho\pi$. The profits from lying will be $(1+\beta)(\theta_1 + \psi\theta_0) = (1+\beta)\theta_1$, because $\theta_0 = 0$. Therefore, the period-1 incentive constraint is the same as before:

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

The period-1 retention constraint dictates that $\beta\rho\pi \leq (1+\beta)\psi\theta_1$, or that $\theta_1 \leq \pi/[(1+\beta)(\rho\beta\psi)]$. Observe that the period-1 retention constraint will be automatically satisfied when the period-1 incentive constraint holds. Hence, the old $IC(1,1)$ curve will still apply for period 1.

Figure 11 illustrates the upshot of the above analysis. Once again the shaded area illustrates the values of θ_1 and θ_2 where the first-best allocation can be supported using a

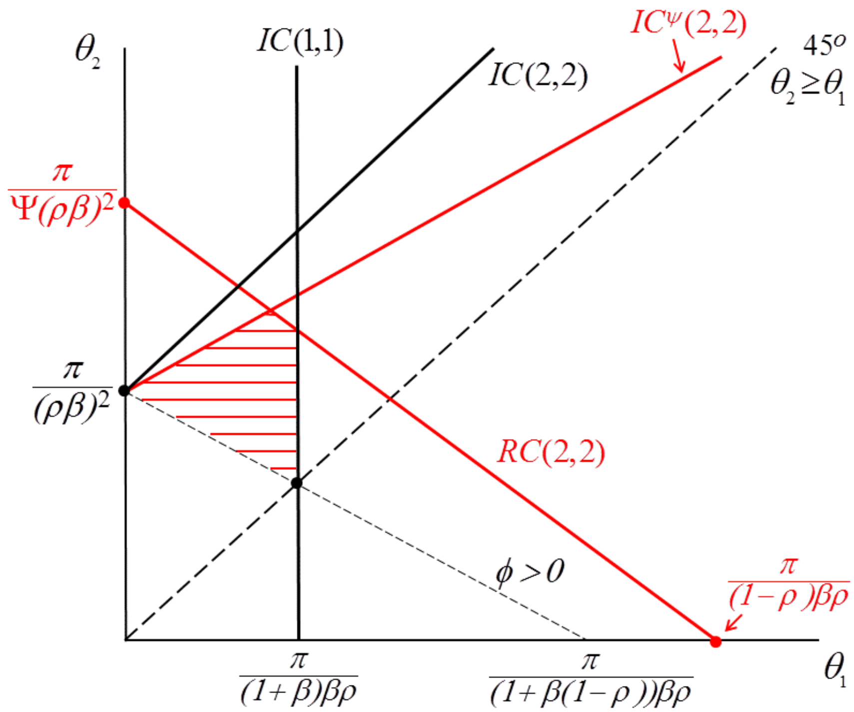

Figure 11: Set of implementable first-best allocations with leaky contracts

backloading strategy. This area has shrunk due to the leakage problem. It lies within the old triangle.

9.3 The Choice of Venture in India

As might be expected, the cost of producing in India is much less expensive than in Mexico and the U.S. The input price for India, q^{IN}, is obtained following a similar approach to the one used for Mexico and the U.S. The data for India is problematic for at least two reasons. First, India has a large informal sector. Therefore, using statistics containing information only about 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 rental price of capital, r^{IN}, will be determined using information on the relative price of capital taken from the Penn World Tables. This is about 23 percent higher in India than the U.S. 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 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 $q^{IN} = 0.6$.

The technology ladder for India is fit to the data in the manner used for Mexico and the U.S. There is not enough data to construct a useful Lorenz curve for employment by establishment, though. So, now the model is calibrated to match just two stylized facts: the average size of establishments and the complementary distribution of employment by establishment age. The upshot of the calibration procedure is displayed in Figure 12, which shows both the Indian and Mexican technology 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 establishment in India. Recall that the survival rates are obtained directly from data. The resulting parameters values are displayed in Table IV.

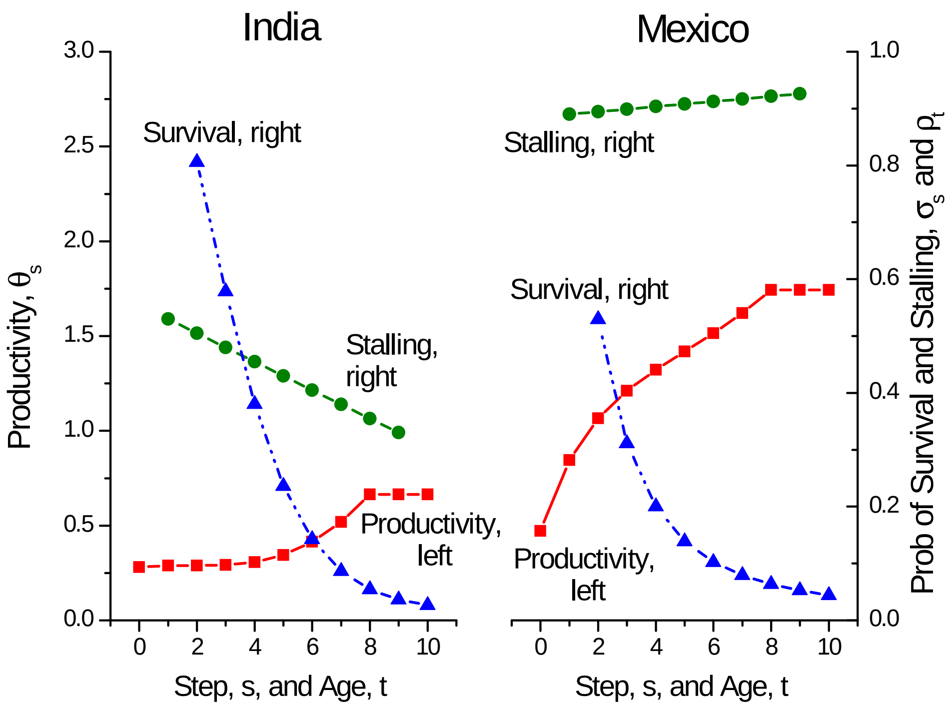

Figure 12: Productivity, survival and stalling in India and Mexico (model). The diagram displays the assumed productivity ladder (left axis) for India and the Mexico.. It also illustrates the probability profiles for survival and stalling (right axis).

Table IV: Parameter Values, India

Parameter	*Value*
Labor, efficiency	$\chi = 33,750$
Ladder, state parameters	$\overline{\theta}_0 = 1.275, \overline{\theta}_1 = 0.069$
	$\overline{\theta}_2 = -0.024, \overline{\theta}_3 = 0.003$
Pr Stall, parameters	$\overline{\rho}_0 = 0.53, \overline{\rho}_1 = -0.025$
Pr Survival, time t	$\overline{\sigma}_0 = 0.015$, $\overline{\sigma}_1 = 0.030$
	$\overline{\sigma}_2 = -0.003$
Capital, upper bound	$\overline{k} = \infty$
Fixed Cost	$\phi = 0.0002$
Monitoring, efficiency	$z = 0.15$
Input Price	$q = 0.6$
Control Parameter	$\psi = 0.4$

Figure 13 shows the combinations of ψ and z required to adopt each of the three technologies, assuming the prevailing level of factor prices in India. That is. it shows the adoption zones for each technology. For any value of ψ, the advanced technology will require a higher level of z than the intermediate one. There is a tradeoff between ψ and z. Higher levels for ψ, which imply poorer control, can be compensated for by bigger values for z, or by greater efficiency in monitoring, at least up to a point. When phi raises to 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 labelled I indicates the value for ψ and z that are used for India in the simulation. (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.)

The model is able to match the stylized facts about firms in India reasonably well. As can been seen from Table V, the average size of firms in the model matches the average size in the data. Figure 14 shows that the calibrated framework mimics the share of employment

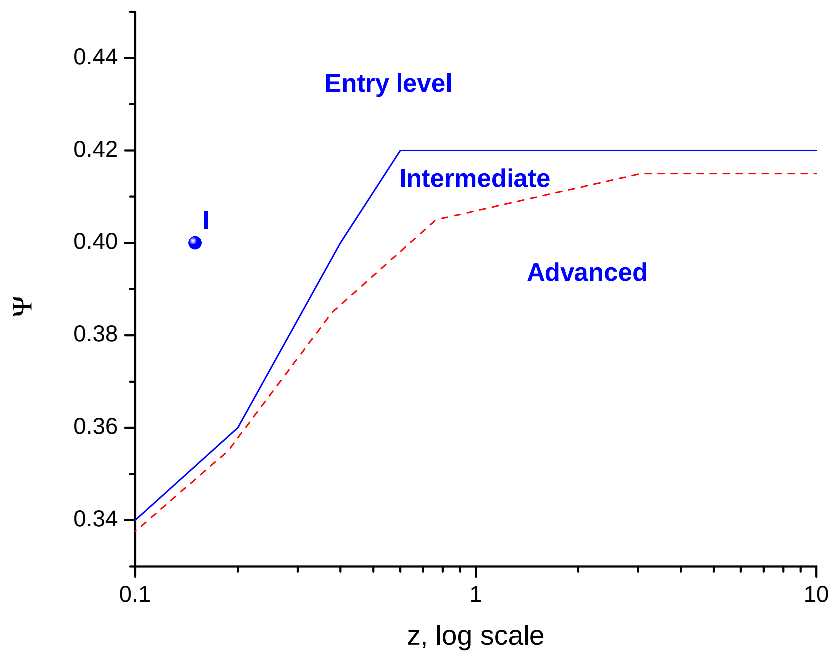

Figure 13: The zones of adoption for India.

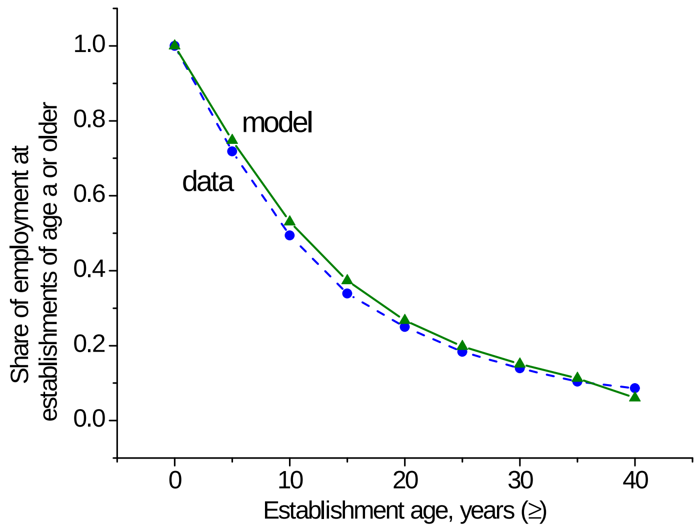

Figure 14: The complementary cumulative distribution of employment by age in India, data and model.

by age for firms perfectly.

Table V: Average Firm Size, India	
data	*model*
7.7	6.3

Can differences in enforcement 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 model predicts that productivity will be much higher in the U.S. than in India, as can be seen in Table VI. In particular, the model calculates that it should be almost 10 times higher, which actually exceeds what is found in the data. Likewise, the model predicts that measured TFP in the U.S. should be 5 times that of India TFP. In the data, it is 4.5.

Table VI: Measures of Productivity

	U.S./India	
	data	model
Labor Productivity–$\mathbf{o}(\tau)/\mathbf{l}(\tau)$	9.21	10.32
TFP–$\mathbf{o}(\tau)/[\mathbf{k}(\tau)^{\kappa}\mathbf{l}(\tau)^{1-\kappa}]$	4.54	5.09

10 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 this. 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 firm. A stall on the ladder is an absorbing state. Some firms may have exciting potential for profit. Perhaps, though, they require large up-front acquisitions of working capital 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 may limit the set of ventures that an intermediary can invest in.

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 payouts to a firm away from the beginning of the project toward the end. 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 will restrict further the set of viable investment opportunities that an intermediary can draw from. The upshot of all of this is that the set of feasible 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: viz, an advanced technology, an intermediate one, and an entry-level one, so to speak. The advanced technology has the potential to deliver high profits. It requires large investments and has considerable scope for financial malfeasance. Therefore, its implementation necessitates 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 employ the intermediate one in Mexico. Monitoring is too inefficient in Mexico and the intermediate technology can be implemented using a backloading strategy alone. 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 the inability to control cash flows, it may be the case that a viable contract cannot be written with the required reward structure that will 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 American ones. Second, the size of an establishment rises more steeply with age in America 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.

11 Appendix

11.1 The General Contract Problem with Reports at all Dates and States

Consider the general contract problem where reports in 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 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, 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 $\mathbf{h}^t$. The period-t level of productivity conditional on a history, $\mathbf{h}^t$, is represented by $\theta(\mathbf{h}^t)$. Last, 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 $\zeta_t^*(\mathbf{h}^t) = h_t(\mathbf{h}^t)$. A reporting strategy is defined by $\zeta \equiv (\zeta_1, \cdots, \zeta_t)$. 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 $\mathbf{h}^t \in \mathcal{H}^t$.

With a slight abuse of 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

$$\theta(\mathbf{h}^t)k(\mathbf{h}^t,t)^\alpha - x(\mathbf{h}^t) \geq 0, \tag{22}$$

$$\sum_{t=1}^{T} \sum_{\mathbf{h}^t \in \mathcal{H}^t} \beta^t \left[\theta(\mathbf{h}^t)k(\mathbf{h}^t,t)^\alpha - x(\mathbf{h}^t)\right] \Pr\left(\mathbf{h}^t\right)$$
$$\geq \max_{\zeta \in \mathcal{S}} \sum_{t=1}^{T} \sum_{\mathbf{h}^t \in \mathcal{H}^t} \beta^t \left[\theta(\mathbf{h}^t)k(\zeta^t(\mathbf{h}^t),t)^\alpha - x(\zeta^t(\mathbf{h}^t))\right] \prod_{n=1}^{t}[1 - p\left(\zeta^n(\mathbf{h}^n)\right)] \Pr\left(\mathbf{h}^t\right), \tag{23}$$

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

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

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 (23) differs from (7). Here a truthful reporting strategy must deliver a payoff in expected present discounted value terms over the entire lifetime of the entire contract that is no smaller than that which could be obtained by an untruthful one. The objective function (P3) and the rest of the constraints (22), (24) to (26) are the direct analogues of those presented in (P2), so they will not be explained.

Turn now to a more restricted problem where the firm is not allowed to make a report that is infeasible; i.e., happens with zero probability. 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

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

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

and (22), (24) and (25).

The lemma presented below holds.

Lemma 7 *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 length of the stretch 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 (P4) 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 righthand side of the incentive constraint for the restricted problem can be no larger than the righthand 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 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 only occur if the optimal solutions for both problems are the same too. ■

It will now be established that problem (P2) delivers the same solution as 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 that which could be obtained by an untruthful one. Problem (P2) has S incentive constraints requiring that reports along the diagonal in Figure 3 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 8 *The contracts specified by problems (P2) and (P4) are the same.*

Proof. First, take the solution to the restricted problem (P4). Suppose this solution violates the incentive constraint (7) in problem (P2) at some node. Evaluate the righthand side of (27), assuming that this is the lone deviation; i.e., insert the truthful reports into the righthand side everywhere else. The righthand side of (27) must then exceed the lefthand one at this deviation. This is a contradiction.

Second, consider the solution to problem (P2). Assume that this solution violates the incentive constraint (27) for problem (P4). This implies that at some nodes (s, s) along the diagonal in Figure 3 it pays to tell lies. Choose the first such state/time pair (s, s), denoted by (s^*, s^*). The path to this point is unique. Truthful reports were told before this point, too. From this point on, the firm cannot report going further up the ladder. Hence, it cannot tell any further lies. Evaluate the righthand side of (27) at this lone deviation from the truth. The righthand can only exceed the lefthand side if the present-value of the path after making a report of $s^* - 1$ exceeds this present-value after making a report of s^*. Again, the truth was optimal before this point. But, this implies that (7) must have been violated at node (s^*, s^*). ■

11.2 First-order conditions for Problem (P2)

Represent the multipliers associated with the constraints (6) to (10) as follows: (i) $\lambda_{s,t}\beta^t \Pr(s,t)$ is the multiplier attached to the limited liability constraint (6); (ii) μ_u denotes the multiplier attached to the incentive constraint (7), with $\mu_u \equiv 0$ for $u > S$ (a region where the constraint does not apply); (iii) $\iota_{s,t}\xi\beta^t \Pr(s,t)$ is defined to be the multiplier linked with the information constraint (8), where $\iota_{t,t} = -\iota_{t-1,t}$ and $\iota_{s,t} \equiv 0$ for $s \neq t-1, t$; (iv) $\kappa_{s-1,t}\xi\beta^t \Pr(s-1,t)$ is the multiplier connected with the irreversibility constraint (9), where $\kappa_{s-1,t} \equiv 0$ for $t < s+1$ (a zone where this constraint is not applicable) and $I(s,t)$ signifies an indicator function that returns a value of 1 when $t = s+1$ and a value of 0 otherwise; (v) ξ represents the multiplier attached to the zero-profit condition (10). The first-order conditions for $k(s,t)$, $x(s,t)$ and $p(u-1,l)$ are now listed.

$$\begin{aligned}
&k(s,t): \\
&\beta^t \Pr(s,t)\,\theta_s \alpha k(s,t)^{\alpha-1} + \lambda_{s,t}\beta^t \Pr(s,t)\,\theta_s \alpha k(s,t)^{\alpha-1} + \sum_{u=1}^{s} \mu_u \beta^t \Pr(s,t)\,\theta_s \alpha k(s,t)^{\alpha-1} \\
&\qquad - \mu_{s+1}\beta^t \alpha k(s,t)^{\alpha-1} \prod_{n=s+1}^{t} [1 - p(s,n)] \sum_{j=s+1}^{\min\{t,S\}} \Pr(j,t) + \iota_{s,t}\xi\beta^t \Pr(s,t) \\
&\qquad + \kappa_{s,t}\xi\beta^t \Pr(s,t) - I(s,t) \sum_{j=s+2}^{T} \kappa_{j,t}\xi\beta^t \Pr(s,j) \\
&\qquad - \xi\beta^t \Pr(s,t)\,[C_2(p(s,t),k(s,t)) + q] = 0, \text{ for } t = 1,\cdots,T \text{ and } s \leq \min\{t,S\}.
\end{aligned} \tag{29}$$

$x(s,t)$:

$$-\beta^t \Pr(s,t) - \lambda_{s,t}\beta^t \Pr(s,t) - \sum_{u=1}^{s} \mu_u \beta^t \Pr(s,t)$$

$$+\mu_{s+1}\beta^t \prod_{n=s+1}^{t} [1-p(s,n)] \sum_{j=s+1}^{\min\{t,S\}} \Pr(j,t) + \xi\beta^t \Pr(s,t) = 0, \text{ for } t = 1, \cdots, T \text{ and } s \leq \min\{t,S\}. \tag{30}$$

$$\begin{aligned} p(u-1,l) \quad &: \\ &\mu_u \sum_{t=l}^{T} \sum_{s=u}^{\min\{t,S\}} \beta^t \left[\theta_s k(u-1,t)^\alpha - x(u-1,t)\right] \Pr(s,t) \prod_{n=u, n\neq l}^{t} [1-p(u-1,n)] \\ = \quad &\xi\beta^l C_1 \left(p(u-1,l), k(u-1,u)\right) \Pr(u-1,l), \text{ for } u = 1, \ldots, S, \text{ and } l \geq u. \end{aligned} \tag{31}$$

Start with the first-order condition for $k(s,t)$, given by (29). How is it derived? It is easy to see that $k(s,t)$ will not appear in any incentive constraints (7) where $u > s$. With a changes of indices, the right-hand side of (7) can be rewritten as

$$\sum_{t=s+1}^{T} \sum_{j=s+1}^{\min\{t,S\}} \beta^t \left[\theta_j k(s,t)^\alpha - x(s,t)\right] \prod_{n=s+1}^{t} [1-p(s,n)] \Pr(j,t), \text{ for } s = \{0, \cdots, S-1\}. \tag{32}$$

The multiplier associated with this term will be μ_{s+1}. The derivative of this with respect to $k(s,t)$ will be

$$\beta^t \alpha\theta_j k(s,t)^{\alpha-1} \prod_{n=s+1}^{t} [1-p(s,n)] \sum_{j=s+1}^{\min\{t,S\}} \Pr(j,t), \text{ for } t \geq s+1.$$

With a bit of effort, it can now be seen that (29) is the first-order condition for $k(s,t)$.

To derive (30), or the first-order condition for $x(s,t)$, focus on the left-hand side of (7). Again, it is easy to see that $x(s,t)$ will not appear in the left-hand side of any incentive constraints where $u > s$. Turn to the right-hand side. The derivative of (32) with respect to $x(s,t)$ is

$$-\beta^t \prod_{n=s+1}^{t} [1-p(s,n)] \sum_{j=s+1}^{\min\{t,S\}} \Pr(j,t).$$

By making use of these facts it is straightforward to obtain (30).

Equation (31) gives the first-order condition for $p\,(u-1,l)$. To derive this condition, define the term $F(s,u-1,t)$ by

$$F(s,u-1,t) \equiv \beta^t\left[\theta_s k(u-1,t)^\alpha - x(u-1,t)\right]\Pr(s,t)\,,$$

so that the right-hand side of the incentive constraint becomes

$$\sum_{t=u}^{T}\sum_{s=u}^{\min\{t,S\}} F(s,u-1,t)\prod_{n=u}^{t}[1-p\,(u-1,n)].$$

Observe that $p\,(u-1,l)$ will enter the above condition only when $t \geq l$. Differentiating this condition with respect to $p\,(u-1,l)$ gives

$$-\sum_{t=l}^{T}\sum_{s=u}^{\min\{t,S\}} F(s,u-1,t)\prod_{n=u,n\neq l}^{t}[1-p\,(u-1,n)],$$

from which the form of (31) is readily transparent.

11.3 Proof of Trust but Verify

Proof. From (31) it is immediate that if $\mu_u = 0$, then $p(u-1,l) = 0$, because $C_1\left(p(u-1,l),k(u-1,u)\right) > 0$, for $p(u-1,l)$ and $k(u-1,u) > 0$. Furthermore, if $p(u-1,l) = 0$, then $\mu_u = 0$ because $\theta_s k(u-1,t)^\alpha - x(u-1,t) > 0$ and $C_1\left(0,k(u-1,u)\right) = 0$. ■

11.4 Proof of Seize Everything

Proof. By dividing by $\beta^t\Pr(s,t)$, rewrite the first-order condition (30) for $x(s,t)$ as

$$-1-\lambda_{s,t}-\sum_{u=1}^{s}\mu_u+\mu_{s+1}\prod_{n=s+1}^{t}[1-p\,(s,n)]\frac{\sum_{j=s+1}^{\min\{t,S\}}\Pr(j,t)}{\Pr(s,t)}+\xi=0. \tag{33}$$

Take a situation where productivity could have changed in period t. Specifically, either $s = t-1$ or $s = t$. The above condition simplifies for these two cases to

$$-1-\lambda_{t-1,t}-\sum_{u=1}^{t-1}\mu_u+\mu_t[1-p\,(t-1,t)]\frac{\Pr(t,t)}{\Pr(t-1,t)}+\xi=0, \tag{34}$$

and

$$-1 - \lambda_{t,t} - \sum_{u=1}^{t} \mu_u + \xi = 0. \tag{35}$$

Assume that the incentive constraint in period t holds so that $\mu_t > 0$. The equation (35) implies that $-1 - \sum_{u=1}^{t-1} \mu_u + \xi > 0$. For (34) to be satisfied it must then transpire that $\lambda_{t-1,t} > 0$. Therefore, $\mu_t > 0$ implies $\lambda_{t-1,t} > 0$. In other words, when the incentive constraint binds in period t the intermediary will take everything upon a bad report at that time.

Compare the situation where the individual tells the truth at time t to that which happens l periods down the road when he lies in period t. Note that $x(t-1, t+l)$ will be determined by the first-order condition

$$-1 - \lambda_{t-1,t+l} - \sum_{u=1}^{t-1} \mu_u + \mu_t \prod_{n=t}^{t+l} [1 - p(t-1, n)] \sum_{j=t}^{\min\{t+l,S\}} \frac{\Pr(j, t+l)}{\Pr(t-1, t+l)} + \xi = 0.$$

Once again assume that $\mu_t > 0$. As before, this implies that $-1 - \sum_{u=1}^{t-1} \mu_u + \xi > 0$. This implies that $\lambda_{t-1,t+l} > 0$ in the above expression. That is, the intermediary will take everything in period $t + l$ following a declaration of a bad shock in period t, assuming that the incentive constraint is binding. ■

11.5 Proof of Backloading

Proof. To begin with, it will be shown that if either $\lambda_{t,t} > 0$ or $\mu_t > 0$ then $\lambda_{i,i} > 0$ for $i < t$. To do this, assume that either $\lambda_{t,t} > 0$ or $\mu_t > 0$. Suppose to the contrary that $\lambda_{i,i} = 0$ at some state/date node (i, i). Equation (35) would then imply that

$$\sum_{u=1}^{i} \mu_u = \xi - 1.$$

Take some future state/date node, (t, t). Here,

$$-1 - \lambda_{t,t} - \sum_{u=1}^{t} \mu_u + \xi = 0.$$

The above two equations give

$$-\lambda_{t,t}-\sum_{u=i+1}^{t}\mu_u=0 \text{ (for } i<t).$$

This would seem possible only if both terms are zero. Therefore, $\lambda_{i,i}=0$ implies $\lambda_{t,t}=\mu_t=0$, the desired contradiction. Therefore, $\lambda_{i,i}>0$.

Next it will be established that $\lambda_{i,i}>0$ implies $\lambda_{i,l}>0$, for $l>i$. To see this, note that if $\lambda_{i,i}>0$ then equation (35) gives

$$-1-\lambda_{i,i}-\sum_{u=1}^{i}\mu_u+\xi=0.$$

Therefore, $-1-\sum_{u=1}^{i}\mu_u+\xi>0$ because $\lambda_{i,i}>0$. Turn to equation (33), which appears as

$$-1-\lambda_{i,l}-\sum_{u=1}^{i}\mu_u+\mu_i\prod_{n=i+1}^{l}\left[1-p\left(i-1,n\right)\right]\sum_{j=i}^{\min\{l,S\}}\frac{\Pr\left(j,l\right)}{\Pr\left(i-1,l\right)}+\xi=0.$$

Therefore, $\lambda_{i,l}>0$, because $-1-\sum_{u=1}^{i}\mu_u+\xi>0$. Thus, it has been shown that if either $\lambda_{t,t}>0$ or $\mu_t>0$, then $\lambda_{i,i}>0$, for $i<t$, and $\lambda_{i,l}>0$, for $l>i$. Taken together this implies that if either $\lambda_{t,t}>0$ or $\mu_t>0$, then $\lambda_{i,i}>0$ for $i<t$ and $l\geq i$.

Finally, it will be demonstrated that $\lambda_{t,t}>0$ only if $\mu_t>0$. Suppose, to the contrary, that $\lambda_{t,t}>0$, yet $\mu_l=0$ for all $l\geq t$–recall again that $t\equiv\max_u\{u:\mu_u>0\}$. From equation (35),

$$1+\lambda_{l,l}+\sum_{u=1}^{t-1}\mu_u=\xi.$$

From this it is easy to deduce that $\lambda_{l,l}=\lambda_{t,t}$ for all $l>t$. Furthermore, it is easy to show, using (33), that $\lambda_{S,l}=\lambda_{t,t}$ for all $l>S$. Now, if this is true the firm will not make profits in any period. Thus, $v=0$, the desired contradiction. ■

11.6 Proof of Efficient Investment

Proof. By dividing by $\beta^t \Pr(s,t)$, rewrite the efficiency condition for working capital (29) as

$$\theta_s \alpha k(s,t)^{\alpha-1} + \lambda_{s,t}\theta_s \alpha k(s,t)^{\alpha-1} + \sum_{u=1}^{s} \mu_u \theta_s \alpha k(s,t)^{\alpha-1}$$

$$- \mu_{s+1}\alpha k(s,t)^{\alpha-1} \prod_{n=s+1}^{t} [1-p(s,n)] \sum_{j=s+1}^{\min\{t,S\}} \frac{\Pr(j,t)}{\Pr(s,t)} + \iota_{s,t}\xi$$

$$+ \kappa_{s,t}\xi - I(s,t) \sum_{j=s+2}^{T} \kappa_{j,t}\xi \frac{\Pr(s,j)}{\Pr(s,t)}$$

$$- \xi[C_2(p(s,t),k(s,t)) + q] = 0.$$

Consider the choice of $k(s,s+1)$ for $s \geq m$. The above equation will now read

$$\theta_s \alpha k(s,s+1)^{\alpha-1} + \lambda_{s,s+1}\theta_s \alpha k(s,s+1)^{\alpha-1} + \sum_{u=1}^{s} \mu_u \theta_s \alpha k(s,s+1)^{\alpha-1}$$

$$- \mu_{s+1}\alpha k(s,s+1)^{\alpha-1} \prod_{n=s+1}^{s+1} [1-p(s,n)] \sum_{j=s+1}^{\min\{s+1,S\}} \frac{\Pr(j,s+1)}{\Pr(s,s+1)} + \iota_{s,s+1}\xi$$

$$- \sum_{j=s+2}^{T} \kappa_{j,s+1}\xi \frac{\Pr(s,j)}{\Pr(s,t)}$$

$$- \xi[C_2(p(s,s+1),k(s,s+1)) + q] = 0.$$

[Recall that $\kappa_{s,s+1} \equiv 0$ and $I(s,s+1) = 1$] When $\mu_{s+1} = 0$, this constraint will further simplify to

$$\theta_s \alpha k(s,s+1)^{\alpha-1} + \lambda_{s,s+1}\theta_s \alpha k(s,s+1)^{\alpha-1} + \sum_{u=1}^{s} \mu_u \theta_s \alpha k(s,s+1)^{\alpha-1} + \iota_{s,s+1}\xi$$

$$- \sum_{j=s+2}^{T} \kappa_{j,s+1}\xi \frac{\Pr(s,j)}{\Pr(s,t)} - \xi q = 0.$$

[Note that if $\mu_{s+1} = 0$, then $p(s,s+1) = 0$, so that $C_2(p(s,s+1),k(s,s+1)) = 0$.] From equation (34)

$$-1 - \lambda_{s,s+1} - \sum_{u=1}^{s} \mu_u + \mu_{s+1}[1-p(s,s+1)] \frac{\Pr(s+1,s+1)}{\Pr(s,s+1)} + \xi = 0,$$

which simplifies to

$$1 + \lambda_{s,s+1} + \sum_{u=1}^{s} \mu_u = \xi.$$

Using this in the efficiency condition for $k(s, s+1)$ results in

$$\theta_s \alpha k(s, s+1)^{\alpha-1} \xi + \iota_{s,s+1} \xi - \sum_{j=s+2}^{T} \kappa_{j,s+1} \xi \frac{\Pr(s, j)}{\Pr(s, t)} - \xi q = 0,$$

which reduces to

$$\theta_s \alpha k(s, s+1)^{\alpha-1} + \iota_{s,s+1} - \sum_{j=s+2}^{T} \kappa_{j,s+1} \frac{\Pr(s, j)}{\Pr(s, t)} - q = 0.$$

But this is the same efficiency condition that materializes in a world without the private information problem. Note that since the multipliers $\kappa_{j,s+1}$ and $\iota_{s,s+1}$ are forward-looking variables they will take the same values in both worlds. ■

11.7 Proof that Retention Payments are Backloaded

Proof. Suppose that the intermediary desires to make payments to the firm to prevent retentions. As a can be deduced from (17) and (18) all the firm will care about is the expected present value of any payments that an option provides. So, make the payments at the terminal date T. For a node along the diagonal these payments can be assigned to step/date (S, T), because this node has positive probability. For an off-diagonal node (s, t), where $s < t$, the step/date combination (S, T) cannot be reached because a stall has occurred. Thus, the required payments must be made at (s, T). Because the difference between the righthand sides of (17) is increasing in the number of remaining periods, while the lefthand side is zero until date T, it follows that this constraint is the tightest during the first period along a stall path, or $(s, s+1)$ for $s < S$. The payments $c(s, T)$ are the minimum payments necessary to satisfy (17) at this node. This then implies that the retention constraint is slack for all (s, t) where $t > s + 1$. ■

11.8 Data

11.8.1 Figure 1

India: The data for India was obtained using two sources: the Annual Survey of Industries (ASI) for 2007-08, which gathers data on formal sector manufacturing plants, and the National Sample Survey Organization (NSSO) for 2005-06, which collects data on informal sector manufacturing establishments. ASI data can be found online at the Government of India Ministry of Statistics and Programme Implementation, and contains employment and establishment distribution by establishment size. India's employment distribution by establishment size in the informal sector was obtained from Figures 4 and 5 in Hasan and Jandoc (2010), while the establishment distribution by establishment size can be found in Statement 1 of the 2005-06 National Sample Survey Report No.526 from the Ministry of Statistics and Programme Implementation. India's own-account manufacturing enterprises (OAME)–or establishments that operate without any worker employed on a fairly regular basis–are included in the establishment size range of 1 to 10 employees.

INDIA

	Raw Data		Adjusted Data		Cumulative share of	
	EST	EMPL	EST	EMPL	EST	EMPL
All establishments	16,786,874	43,272,366	2,604,298	20,233,009	0	0
Est., 1 to 10 emp.	16,229,052	31,896,645	2,046,476	8,857,288	78.6	43.8
Est., 11 to 49 emp.	454,512	4,210,000	454,512	4,210,000	96.0	64.6
Est., 50 to 99 emp.	80,826	1,200,000	80,826	1,200,000	99.1	70.5
Est., 100 to 199 emp.	10,396	1,100,000	10,396	1,100,000	99.5	76.0
Est., 200 to 499 emp.	7,095	1,525,392	7,095	1,525,392	99.8	83.5
Est., 500 to 999 emp.	2,659	1,087,904	2,659	1,087,904	99.9	88.9
Est., 1,000 to 1,999 emp.	1,257	784,814	1,257	784,814	99.96	92.8
Est., 2,000 to 4,999 emp.	800	908,964	800	908,964	99.99	97.2
Est., 5,000 and more emp.	277	558,647	277	558,647	100.0	100.0
Mean size of est.		2.58		7.77		

Mexico: The data for Mexico for Figure 1 were obtained from Mexico's 2004 Economic Census conducted by the Instituto Nacional de Estadistica y Geografia (INEGI).

MEXICO

	Raw Data		Adjusted Data		Cumulative share of	
	EST	EMPL	EST	EMPL	EST	EMPL
All establishments	328,385	4,198,579	204,293	4,074,487	0	0
Est., 1 to 4 emp.	260,112	512,728	136,020	388,636	66.6	9.5
Est., 5 to 9 emp.	35,548	222,525	35,548	222,525	84.0	15.0
Est., 10 to 19 emp.	13,436	178,781	13,436	178,781	90.6	19.4
Est., 20 to 49 emp.	8,848	272,087	8,848	272,087	94.9	26.1
Est, 50 to 99 emp.	3,945	278,148	3,945	278,148	96.8	32.9
Est., 100 to 249 emp.	3,427	535,197	3,427	535,197	98.5	46.0
Est., 250 to 499 emp.	1,641	574,304	1,641	574,304	99.3	60.1
Est., 500 to 999 emp.	895	617,946	895	617,946	99.7	75.3
Est., 1,000 to 2,499 emp.	439	645,006	439	645,006	100.0	91.1
Est., 2,500 emp. or more	94	361,857	94	361,857	100.0	100.0
Mean size of est.		12.8		19.9		

U.S.: A special request was made to obtain these data. Data for the United States come from the 2002 Economic Census published by the U.S. Census Bureau. They can be obtained using the U.S. Census Bureau's Fact Finder. These are businesses that have no paid employees but are subject to federal income tax in the United States. Data for Mexico and India, however, include this type of establishments. To make Mexico comparable to the U.S., data for Mexico was adjusted to remove nonemployer establishments using the methodology found in Buera et al. (2011). That is, census data were imputed using Mexico's 1998 National Survey of Microenterprises (ENAMIN) to subtract those nonemployer establishments from the group of establishments having 1 to 4 employees. Since ENAMIN is a survey, Buera et al.(2011) scale up the numbers to fit the total population. Thus, 124,092 establishments, accounting for one employee each, were removed from Mexico's raw census data. To make India comparable to the U.S., data for India was adjusted by removing OAME plants from the 1 to 10 employee establishment group in the original census data. The number of OAME plants in India is significant: 14,182,576 establishments that account for 23,039,357 workers. The establishment sizes used in this paper exclude nonemployers and thus incorporate the nonemployer adjustments made to the raw data for Mexico and India.

United States

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

11.8.2 Figure 2

The data for India, Mexico and the U.S. for Figure 2 were obtained from Hsieh and Klenow (2010). The following table shows the statistics used to construct Figure 2.

Hsieh and Klenow (2010) Facts: Mexico and the U.S.

Establishment Age	U.S Empl. Share (2002)	Mexico Empl. Share (2003)	India Empl. Share (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

References

Buera, Franciso J., Kaboski, Joseph P. and Yongseok Shin. 2011. "Finance and Development: A Tale of Two Sectors." *American Economic Review.*

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

Castro, Rui, Clementi, Gian Luca and Glenn M. MacDonald. 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.

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. 2010b. "Quantifying the Impact of Financial Development on Economic Development." Economie d'avant garde, Research Report No. 17, Department of Economics, University of Pennsylvania.

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." *The 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.

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. 2009. "Education Quality and Development Accounting." Unpublished paper, Arizona State University.

Siegel, 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.

Tybout, James R. 2000. "Manufacturing Firms in Developing Countries: How Well Do They Do, and Why?" *Journal of Economic Literature*, 38 (1): 11–44.

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.