Working Paper Series

Financing Development: The Role of Information Costs

WP 08-08R

Jeremy Greenwood
University of Pennsylvania

Juan M. Sanchez
Federal Reserve Bank of Richmond

Cheng Wang
Iowa State University

THE FEDERAL RESERVE BANK OF RICHMOND
RICHMOND ▪ BALTIMORE ▪ CHARLOTTE

Version: August, 2009

Financing Development: The Role of Information Costs*

by

Jeremy Greenwood, Juan M. Sanchez and Cheng Wang

Working Paper No. 08-08R

Abstract

To address how technological progress in financial intermediation affects the economy, a costly-state verification framework is embedded into the standard growth model. The framework has two novel ingredients. First, firms differ in the risk/return combinations that they offer. Second, the efficacy of monitoring depends upon the amount of resources invested in the activity. A financial theory of firm size results. Undeserving firms are over financed, deserving ones under funded. Technological advance in intermediation leads to more capital accumulation and a redirection of funds away from unproductive firms toward productive ones. Quantitative analysis suggests that finance is important for growth.

Keywords: costly-state verification, economic development, financial intermediation, firm size

JEL Nos: E13, O11, O16

Affliations: University of Pennsylvania, Federal Reserve Bank of Richmond, and Iowa State University

*An abridged version of this paper, containing the theory sections, is forthcoming in the *American Economic Review*. A variant of the model with credit rationing is developed in NBER Working Paper No. 13104.

1 Introduction

Financial development "accelerates economic growth and improves economic performance to the extent that it facilitates the migration of funds to the best user, i.e. to the place in the economic system where the funds will earn the highest social return," noted Goldsmith (1969, p. 400) some thirty five years ago. Information production plays a key role in this process of steering of funds to the highest valued users. If the costs of information production drop, then financial intermediation should become more efficient with an associated improvement in economic performance. A theoretical model is developed where reductions in the cost of information processing allow for more efficient capital allocation. Quantitative analysis suggests that cross-country differences in the efficiency of intermediation may account for a substantial, but not the dominant, part of cross-country income differentials.

Some stylized facts of the kind illustrated in Figure 1 were offered by Goldsmith (1969) to suggest that financial intermediation might be important: First, the ratio of business debt to GDP has risen. In 1952 business debt was 30 percent of GDP.[1] Today it is 65 percent. Second, the value of firms relative to GDP has also moved up. In 1951 firms were worth 50 percent of GDP, while today they are valued at 176 percent. This may be evidence of improved intermediation; now it is easier for firms to enter stock and bond markets and raise funds.

Direct measures of the impact of improved efficiency in the financial system on the economy are hard to come by. Improvements in the efficiency of financial intermediation, due to improved information production, are likely to reduce the spread between the internal rate of return on investment in firms and the rate of return on savings received by savers. The spread between these returns reflects the costs of intermediation. This spread will include the costs of ex ante information gathering about investment projects, the ex post information costs of policing investments, and the costs of misappropriation of savers' funds by management, unions, etc., that arise in a world with imperfect information. One may

[1] Data sources are provided in Appendix 14.2.

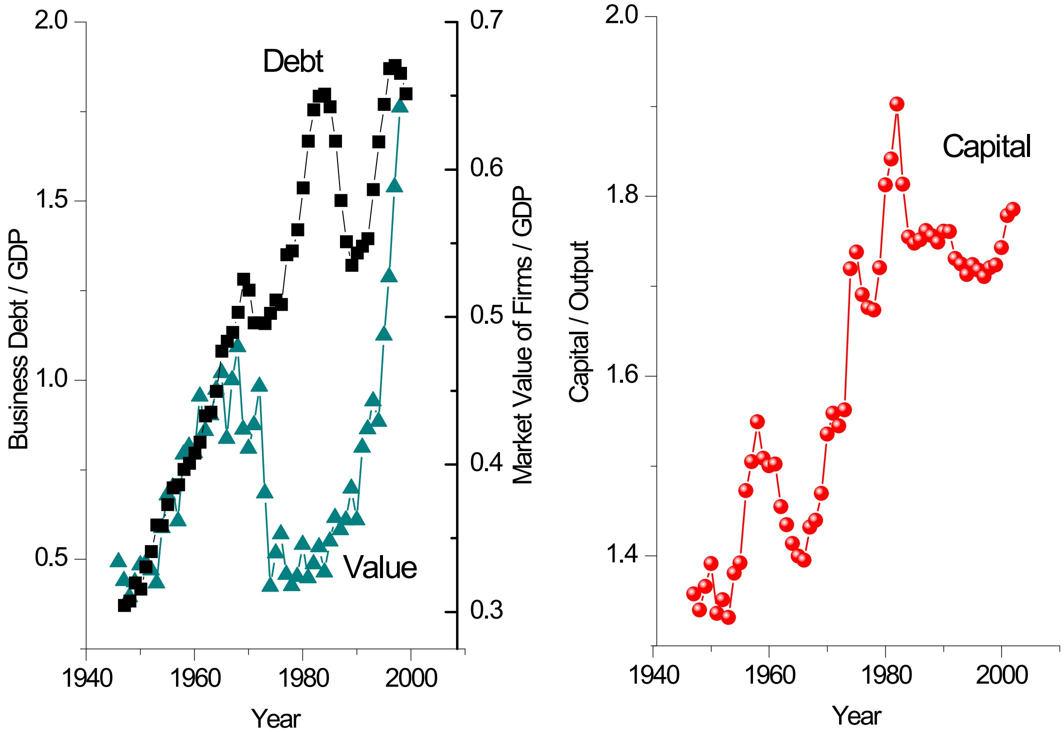

Figure 1: Trends in financial intermediation

observe little change in the rate of return earned by savers over time, because aggregate savings will adjust in equilibrium so that this return reflects savers' rates of time preference.

If the wedge between the internal rate of return earned by firms and the rate of return received by savers falls, due to more effective intermediation, then the capital stock in the economy should rise. Indeed, there is some evidence that this may be the case. Figure 1 plots the capital-to-GDP ratio for the business sector of the economy, where the capital stock includes intangible investments such as firms' research and development, following the lead of Corrado et al. (2006). It has increased significantly over the postwar period. Of course, an economy's capital-to-GDP ratio may rise on other accounts, too. For example, lower taxes on capital income should increase it, as should declines in the price of capital goods due to investment-specific technological advance. To the extent that capital stock measures exclude intangibles, or fail to incorporate improvements in quality due to embodied technological progress, more effective intermediation will be picked up as increases in productivity, instead.

There is evidence of the above phenomena in the cross-country data, too. Figure 2 plots the relationship between interest-rate spreads and output for a sample of 49 countries. As can be seen, there is a negative association. Additionally, capital-to-output ratios and output are positively related, in a sample of 48 countries. Hence, there is evidence of capital deepening both in the US time-series data and in the cross-country data.

While the above facts are stylized, to be sure, it will be noted that empirical researchers have used increasingly sophisticated methods to tease out the relationship between financial intermediation and growth. This literature is surveyed masterfully by Levine (2005). The upshot is that financial development has a causal effect on economic development; specifically, financial development leads to higher rates of growth in income and productivity.

A general equilibrium model of firm finance, with competitive intermediation, is presented to address the impact that financial intermediation has on economic development. At the heart of the framework developed here is a costly-state verification paradigm that has its roots in classic work by Robert M. Townsend (1979) and Stephen D. Williamson (1986). The model here has two novel ingredients, though. First, in the standard costly-state verifi-

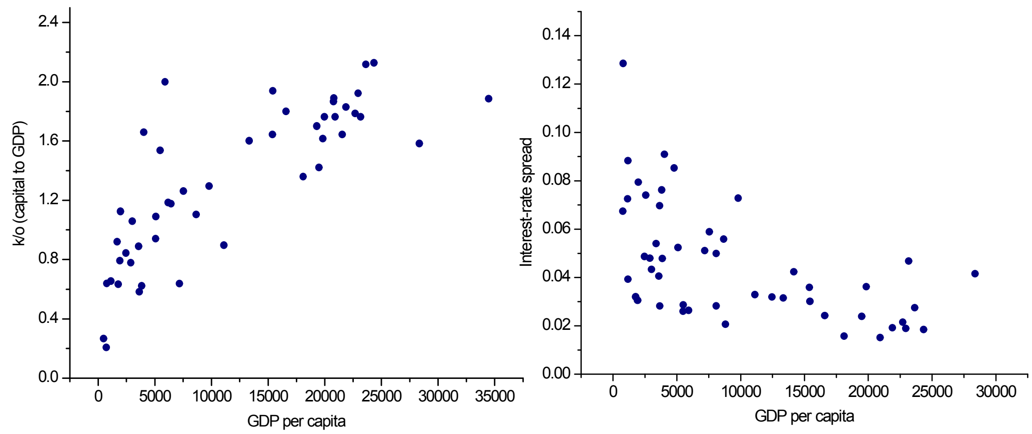

Figure 2: The cross-country relationship between per-capita GDP, on the one hand, and interest-rate spreads and capital-output ratios on the other

cation paradigm, the realized return on a firm's investment activity is private information. This return can be monitored, but the *outcome* of this auditing process is deterministic: once monitoring takes place, the true state of the world is revealed with certainty. This is true whether or not a deterministic *decision rule* for monitoring is employed, as in Townsend (1979) or Williamson (1986, 1987), or a stochastic one is used, as in Ben Bernanke and Mark Gertler (1989) or John H. Boyd and Bruce D. Smith (1994)—see also Townsend (1979, Section 4). By contrast, in the current setup the outcome of monitoring is random. Specifically, the probability of detecting malfeasance depends upon the amount of resources devoted to policing the returns on a project and the efficiency of the monitoring technology. In the model a project's funding and level of monitoring will be jointly determined. The information asymmetry between the firm and the intermediary gives the firm an opportunity to exploit its private knowledge about the realization of the investment return. In particular, it can extract rents from the intermediary, and hence savers.

Second, a firm's production technology is subject to idiosyncratic randomness. This is true in the standard costly-state verification framework as well. Here, though, there is a distribution across firms over the distribution of these returns. In particular, some firms may have investment projects that offer low expected returns with little variance, while others may have projects that yield high expected returns with a large variance.

Two key features in the analysis follow from these ingredients. First, the setup yields a finance-based theory of the equilibrium size distribution for firms. This theory derives from the facts that: (a) investment opportunities differ in the expected returns and levels of risk that they offer and (b) producing information about these returns is costly. A simple threshold rule for funding results. All firms with an expected return at least as great as the cost of raising capital are funded. Funding is increasing in a firm's expected return, and is decreasing in its risk. Loan size is determinate for a given type of project because the costs of financing a project are increasing and convex in the level of the monitoring activity. Thus, high mean projects receive limited funding because the costs of monitoring will rise disproportionately with loan size. The riskier a project is, the bigger is the difference between the returns in good and bad states. This increases the incentive for a firm to low-ball its earnings in good states. Hence, more diligent monitoring will be required, which increases its financing cost. In an abstract sense, one could think that the diminishing returns in information production modeled here provide a microfoundation for Robert E. Lucas's (1978) span of control model.

Second, the framework provides a link between the efficiency of the financial system and the level of economic activity. Such a tie was developed earlier in the models of Valerie R. Bencivenga and Smith (1991), Greenwood and Boyan Jovanovic (1990), Levine (1991), and Albert Marcet and Ramon Marimon (1992). The analysis here provides a crystal clear delineation of the Goldsmith (1969) mechanism, however. It stresses the connection between the state of technological development in the financial sector, on the one hand, and capital accumulation, both along the extensive and intensive margins, on the other. If technological improvement in the financial sector occurs at a faster pace than in the rest of the economy, then financial intermediation becomes more efficient. Loans are monitored more diligently and the rents earned by firms shrink. Additionally, lending activity will change along both extensive and intensive margins. Projects with high (low) expected returns will now receive more (less) funds. Those investments with the lowest expected returns will be cut. At high levels of efficiency in the financial sector the economy approaches the first-best equilibrium achieved in a world without informational frictions. This reallocation effect distinguishes the analysis from earlier research by Shankha Chakraborty and Amartya Lahiri (2007) and Aubhik Khan (2001) that also embed the standard costly-state verification framework into a growth model. In these frameworks all firms receive the same amount of capital.

In order to assess the importance of financial intermediation for the level of economic activity, the developed model is solved numerically. The model is calibrated to match some stylized facts for the U.S. economy. It is found that the model can replicate the increase in the U.S. capital/output ratios, shown in Figure 1. It does a very good job matching the firm-size distribution for the U.S. economy for the year 1972. The improvement in financial sector productivity required to generate the observed rise in the U.S. capital/output ratio also appears to be reasonable. In the baseline model, improvements in financial intermediation account for 1/3 of U.S. growth.

The calibrated model is then taken to the cross-country data. Others have tried to quantify the importance of factors such as limited investor protection [Castro et al. (2009)] for economic development. The emphasis here is on taking a micro-foundation finance model to the data in order to quantify the importance of information production for economic development. Townsend and Ueda (2006) estimate a version of the Greenwood and Jovanovic (1991) model using Thai data. Their model emphasizes the role that financial intermediaries play in producing ex ante information about the state of the economy at the aggregate or sectorial levels. Financial intermediaries offer savers higher and safer returns. In the current analysis, intermediaries produce ex post information about firm-level returns. The return earned by savers is fixed (and therefore safe) in the analysis. The gain to the economy from improved intermediation derives from squeezing, via information production, the rents appropriated by firms on capital investment. This leads to a reduction in borrowing rates and more efficient capital allocation.

The model's imputed measures for the efficiency of financial intermediation in various countries match up very well with independent measures. It also does a reasonable job predicting the differences in cross-country interest-rate spreads or capital/output ratios shown in Figure 2. Financial intermediation turns out to be important quantitatively. For example, in the baseline model Sri Lanka would increase its GDP by 70 to 76 percent if it could somehow adopt Switzerland's financial system. World output would rise by 28 to 48 percent if all countries adopted Switzerland's financial practice. Still, the bulk (or 81 to 87 percent) of cross-country variation in GDP cannot be accounted for by variation in financial systems.

2 The Environment

Imagine a world resting in a steady state that is made up of three types of agents: consumer/workers, firms and financial intermediaries. In a nutshell, firms produce output using capital and labor. The consumer/worker supplies the labor, and intermediaries, the capital. All funding for capital must be raised outside of the firm. Financial intermediaries raise the funds for capital from consumer/workers. They also use labor in their lending activity. Output is used for consumption by consumer/workers and for investment in capital by intermediaries. The behavior of firms and intermediaries will now be described in more detail. The consumer/worker plays a more passive role in the analysis, which is relegated into the background by assuming that he supplies one unit of labor and saves at some fixed interest rate, $\widehat{r}$.[2]

3 Firms

Firms produce output, o, in line with the production function

$$o = \theta k^{\alpha} l^{1-\alpha},$$

[2] Think about a representative consumer with time separable preferences over consumption. The steady state interest rate will then be given by $\widehat{r} = 1/\beta - 1$, where β is his discount factor.

where k and l represent the inputs of capital and labor used in production. The variable θ gives the productivity level of the firm's production process. Productivity is a random variable drawn from a two-point vector $\tau \equiv (\theta_1, \theta_2)$ with $\theta_1 < \theta_2$. Let $\Pr(\theta = \theta_1) = \pi_1$ and $\Pr(\theta = \theta_2) = \pi_2 = 1 - \pi_1$. The mean and variance of θ are given by $\pi_1\theta_1 + \pi_2\theta_2$ and $\pi_1\pi_2(\theta_1 - \theta_2)^2$, respectively.[3] Thus, for a given set of probabilities these statistics differ in accordance with the values specified for θ_1 and θ_2. The realized value of θ is the firm's private information.

Now, the productivity vector, τ, differs across firms. In particular, suppose that firms in the economy are distributed over productivities in line with the distribution function $F : \mathcal{T} \rightarrow [0, 1]$, where $\mathcal{T} \subseteq R^2_+$ and

$$F(x, y) = \Pr(\theta_1 \leq x, \theta_2 \leq y).$$

Think of this distribution as somehow specifying a trade-off between the mean and variance of project returns. Due to technological progress in the production sector of the economy, this distribution will evolve over time. Figure 3 plots the density function for F in mean/variance space that is used in the quantitative analysis.

The firm borrows capital, k, from the intermediary *before* it observes the technology shock, θ. It does this with both parties knowing its type, τ. It can employ labor, at the wage rate w, *after* it sees the realization for θ. In order to finance its use of capital the firm must enter into a contract with a financial intermediary. Last, note that a firm's production is governed by constant returns to scale. In the absence of financial market frictions no rents would be earned on production. Additionally, in a frictionless world only firms offering the highest expected return would be funded. With financial market frictions, deserving projects are underfunded, while undeserving projects are simultaneously over funded.

[3] Observe that $\pi_1\theta_1^2 + (1-\pi_1)\theta_2^2 - [\pi_1\theta_1 + (1-\pi_1)\theta_2]^2 = \theta_1^2\pi_1 - \theta_1^2\pi_1^2 - 2\theta_1\theta_2\pi_1 + 2\theta_1\theta_2\pi_1^2 + \theta_2^2\pi_1 - \theta_2^2\pi_1^2$ $= (1-\pi_1)\pi_1\theta_1^2 + (1-\pi_1)\pi_1\theta_2^2 - 2\theta_1\theta_2\pi_1(1-\pi_1) = (1-\pi_1)\pi_1[\theta_1^2 + \theta_2^2 - 2\theta_1\theta_2] = (1-\pi_1)\pi_1(\theta_1 - \theta_2)^2$.

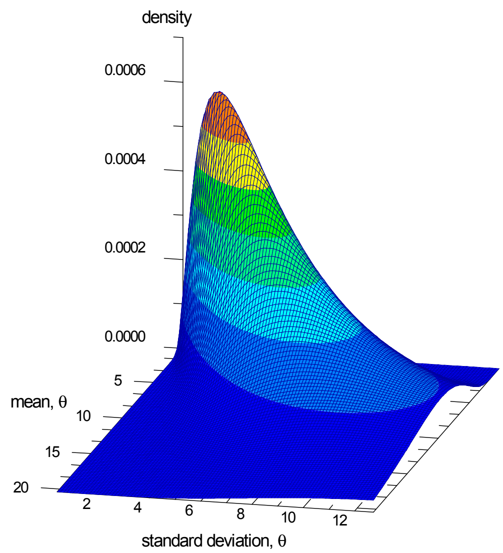

Figure 3: The distribution for F in mean/variance space that is used in the baseline quantitative model

3.1 Profit Maximization by Firms

Consider the problem faced by a firm that receives a loan in terms of capital in the amount k. The firm hires labor after it sees the realization of its technology shock, θ. It will do this in a manner so as to maximize its profits. In other words, the firm will solve the maximization problem shown below.

$$R(\theta, w)k \equiv \max_{l}\{\theta k^{\alpha} l^{1-\alpha} - wl\}. \tag{P1}$$

The first-order condition associated with this maximization is

$$(1-\alpha)\theta k^{\alpha} l^{-\alpha} = w,$$

which gives

$$l = \left[\theta \frac{(1-\alpha)}{w}\right]^{1/\alpha} k. \tag{1}$$

Substituting the solution for l into the maximand and solving yields the unit return function, $R(\theta, w)$, or

$$r = R(\theta, w) = \alpha(1-\alpha)^{(1-\alpha)/\alpha} w^{-(1-\alpha)/\alpha} \theta^{1/\alpha} > 0. \tag{2}$$

Think about $r_i = R(\theta_i, w)$ as giving the gross rate of return on a unit of capital invested in the firm given that state θ_i occurs. The expected gross rate of return will be $\pi_1 r_1 + \pi_2 r_2$, while the variance reads $\pi_1 \pi_2 (r_1 - r_2)^2$.

4 Financial Intermediaries

There is a competitive intermediation sector that borrows funds from consumers and lends capital to firms. While the intermediary knows a firm's type it cannot observe the state of a firm's business either costlessly or perfectly.[4] That is, the intermediary cannot costlessly observe θ, o and l. The firm will make a report to the intermediary about its business situation. The intermediary can devote some resources in order to assess the veracity of this report. The payments, p, from a firm to the intermediary will be conditioned both upon the report made by former, and the outcome of any monitoring activity done by latter. By channelling funds through financial intermediaries consumers avoid a costly duplication of monitoring effort that would occur in an equilibrium with direct lending between them and firms—see Diamond (1984) and Williamson (1986) for more detail. Likewise, in the environment under study, it is optimal for a firm to borrow from only one intermediary at a time.

Suppose a firm reports that the productivity on its project in a given period is θ_j, which may differ from the true state θ_i. The intermediary can devote resources, m_j, to verify this claim. The probability of detecting fraud is increasing in the amount of resources devoted to this activity. In particular, let $P_{ij}(m_j/k)$ denote the probability that the firm is caught cheating conditional on the following: (1) the true realization of productivity is θ_i; (2) the firm makes a report of θ_j; (3) the intermediary spends m_j in monitoring; (4) the total

[4] Recall that the intermediary knows the firm's type, τ. One could think about this as representing the activity, industry or sector that a firm operates within. For instance, Castro et al (forthcoming, Figure 3) present data suggesting that the capital goods sector is riskier than the consumption goods one. It would be possible to have a screening stage where the intermediary verifies the initial type of a firm. The easiest way to do this would be to have them pay a fixed cost to discover τ. If the firm's type can't be undercovered perfectly, as in the classic work of Boyd and Prescott (1986), then it may be possible to design the contract to reveal it.

amount of borrowing is k (which represents the size of the project). The function $P_{ij}(m_j/k)$ is assumed to be monotonically increasing in m_j/k. Additionally, let $P_{ij}(m_j/k) = 0$ if the firm truthfully reports that its type is θ_i. Any lender to the firm must monitor the *whole* project to detect cheating, because his claim to profits will depend on the total level of receipts vis à vis the total amount of disbursements paid out to others. Borrowing through a single intermediary then avoids a costly duplication of monitoring effort.

A convenient formulation for $P_{ij}(m_j/k)$ is

$$P_{ij}(m_j/k) = \begin{cases} 1 - \frac{1}{(\epsilon m_j/k)^{\psi}} < 1, \text{ with } 0 < \psi < 1, \\ \quad \text{for a report } \theta_j \neq \theta_i \text{ and } m_j/k > 1/\epsilon, \\ 0, \\ \quad \text{for a report } \theta_j = \theta_i \text{ or } m_j/k \leq 1/\epsilon. \end{cases}$$

To guarantee that $P_{ij}(m_j/k) \geq 0$, this specification requires that some threshold level of monitoring, $m_j > k/\epsilon$, must be exceeded to detect cheating. Note that this threshold level of monitoring can be made arbitrarily small by picking a large enough value for ε.[5] Also, an arbitrarily large value for ε can be chosen so that the threshold level of monitoring is very small. Figure 4 makes this clear, while illustrating the function $P_{ij}(m_j/k)$.

Monitoring is a produced good, measured in units of consumption. The production of monitoring is project specific. Monitoring produced for detecting fraud in one project cannot be used in a different one. Let monitoring be produced in line with the production function

$$m = z l_m^{1/\gamma}, \text{ with } 0 \leq 1/\gamma \leq 1,$$

where l_m represents the amount of labor employed in this activity. The cost function, $C(m/z; w)$, associated with monitoring is given by

$$C(m/z; w) = w(m/z)^{\gamma}.$$

Costs are linear in wages, w. With diminishing returns to scale in production ($1/\gamma < 1$), the cost function is increasing and convex in the amount of monitoring, m, and decreasing and

[5] The choice of ε can be thought of as normalization relative to the level of productivity in the production of monitoring services–see footnote 7.

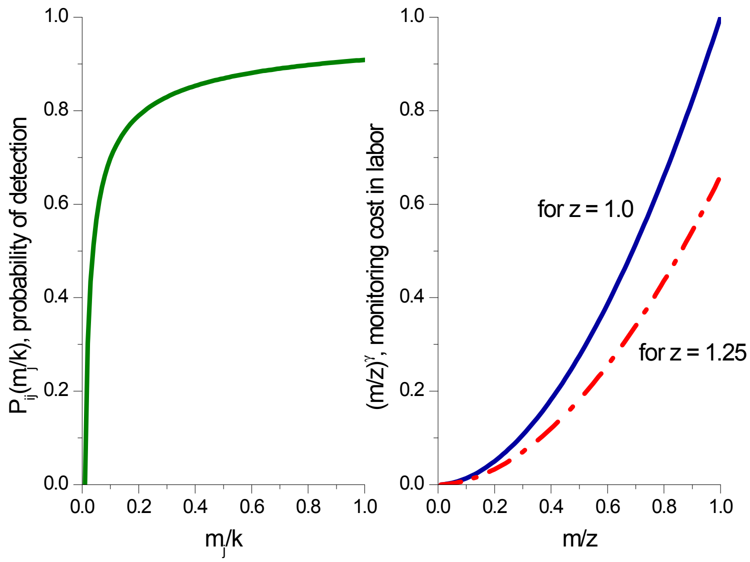

Figure 4: The functions $P_{ij}(m_j/k)$ and $C(m/z;w)/w$, for $\varepsilon = 100$, $\psi = 0.52$, and $\gamma = 1.86$ (the values used in the quantitative analysis)

convex in the state of the monitoring production technology, z. Figure 4 portrays the cost of monitoring in terms of labor, or plots $C(m/z;w)/w$.

Now, exactly which firms are funded depends on three things: (1) the firm's type, τ; (2) the state of the monitoring production technology in the financial intermediation sector , z; (3) the expense of monitoring effort as reflected by the wage, w. As will be seen, when the variance of a firm's project becomes larger, the informational problems associated with contracting become more severe. Therefore, high variance projects are less likely to get funded, ceteris paribus.

5 The Financial Contract

A contract between a firm and an intermediary is summarized by the quadruple $\{k, p_j, p_{ij}, m_j\}$. Here k represents the amount of capital lent by the intermediary to the firm, p_j is the firm's payment to the intermediary if it reports θ_j and is not found cheating, p_{ij} is payment to the bank if the borrower reports θ_j and monitoring reveals that productivity is $\theta_i \neq \theta_j$, and m_j

is the intermediary's monitoring effort when θ_j is reported. Denote the value of the firm's outside option by v.

The intermediary chooses the details of the financial contract, $\{k, p_j, p_{ij}, m_j\}$, to maximize its profits. The contract is designed to have two features: (1) it entices truthful reporting by firms; (2) it offers firms an expected return of v. The optimization problem is

$$I(\tau, v) \equiv \max_{p_1,p_2,p_{12},p_{21},m_1,m_2,k} \{\pi_1 p_1 + \pi_2 p_2 - \widetilde{r}k - \pi_1 w(m_1/z)^\gamma - \pi_2 w(m_2/z)^\gamma\}, \tag{P2}$$

subject to

$$p_1 \leq r_1 k, \tag{3}$$

$$p_2 \leq r_2 k, \tag{4}$$

$$p_{12} \leq r_1 k, \tag{5}$$

$$p_{21} \leq r_2 k, \tag{6}$$

$$[1 - P_{12}(m_2/k)](r_1 k - p_2) + P_{12}(m_2/k)(r_1 k - p_{12}) \leq r_1 k - p_1, \tag{7}$$

$$[1 - P_{21}(m_1/k)](r_2 k - p_1) + P_{21}(m_1/k)(r_2 k - p_{21}) \leq r_2 k - p_2, \tag{8}$$

and

$$\pi_1(r_1 k - p_1) + \pi_2(r_2 k - p_2) = v. \tag{9}$$

Note that the cost of capital, $\widetilde{r}$, is given by $\widetilde{r} = \widehat{r} + \delta$; i.e., the interest paid to investors plus the depreciation on capital. The first four constraints just say the intermediary cannot demand more than the firm earns; that is, the firm has limited liability. Equations (7) and (8) are the incentive-compatibility constraints. Take (8). This simply states that the expected return to the firm from reporting state one when it actually is in state two, as given by the left-hand side, must be less than telling the truth, as represented by the right-hand side. Observe that the constraint set is not convex due to the way that m_1 enters (8). Therefore, the second-order conditions for the maximization problem are important to consider. The last constraint (9) specifies that the contract must offer the firm an expected return equal to v, its option value outside. A firm's outside option is the expected return that it could earn on a loan from another intermediary. This will be determined in equilibrium. Finally, note the solution for $\{p_j, p_{ij}, m_j, k\}$ is contingent upon the firm's type, $\tau = (\theta_1, \theta_2)$. To conserve on notation, this dependence is generally suppressed.

The lemma below characterizes the solution to the above optimization problem.

Lemma 1 *(Terms of the Contract) The solution to problem (P2) is described by:*

1. *The size of the loan from the intermediary to the firm, k, is*
$$k = \frac{v}{\pi_2(r_2 - r_1)[1 - P_{21}(m_1/k)]}. \tag{10}$$

2. *The amount of monitoring per unit of capital when the firm reports a bad state, m_1/k, solves the problem*
$$I(\tau, v) \equiv \max_{m_1/k}\{(\pi_1 r_1 + \pi_2 r_2 - \widetilde{r})k - \frac{\pi_1 w}{z^\gamma}k^\gamma(\frac{m_1}{k})^\gamma - v\}, \tag{P3}$$
where k can be eliminated using (10) above.
   (a) *Monitoring in the bad state is simply given by*
$$m_1 = (m_1/k)k,$$
   where m_1/k solves (P3).
   (b) *The intermediary does not monitor when the firm reports a good state so that*
$$m_2 = 0. \tag{11}$$

3. *The payment schedule is*
$$p_1 = r_1 k, \tag{12}$$
$$p_2 = r_2 k - v/\pi_2, \tag{13}$$
$$p_{12} = r_1 k, \tag{14}$$
$$p_{21} = r_2 k. \tag{15}$$

Proof. See Appendix 14.1. ■

It is intuitive that there are no benefits to the firm from claiming a better outcome than it actually realizes, since it will only have to pay the intermediary more. The intermediary would like to reduce the firm's incentive to report being in the low state. So, suppose the firm reports a low state. If cheating is not detected, then the firm pays all of its revenue (minus labor cost) that would be realized in the low state—see (12). If the firm is caught cheating, then it must surrender all of the revenue (sans labor cost) that it earns in the high state—see (15). Note that due to the incentive-compatibility constraints a false report will never occur so that the payments shown by (14) and (15) do not occur in equilibrium.

The contract specifies that the intermediary should only monitor the firm when it reports a bad outcome (state 1) on its project—see (11). Monitoring in the low state is done to maximize the intermediary's profits, subject to the incentive-compatibility cum promise-keeping constraint (10), as problem (P3) dictates. Note that the higher is the value of the firm, v, the bigger must be the loan, k, to satisfy the incentive-compatibility cum promise-keeping constraint (10). This constraint (10) ensures that the contract provides the firm an expected return equal to what it would earn if it misrepresented the outcome in the good state, $\pi_2(r_2-r_1)[1-P_{21}(m_1/k)]k$. Furthermore, this expected return is set equal to the firm's outside option, v. The size of the loan, k, is increasing in the amount of monitoring that occurs in the low state, m_1/k. This happens because the probability of the firm not getting caught from misrepresenting its revenues, $1-P_{21}(m_1/k)$, is decreasing in the intermediary's monitoring activity.

The theory of intermediation presented here is at an abstract level, as in Boyd and Prescott (1986). It is not intended to explain the presence of real world forms of intermediation, such as the existence of banks, bond markets, or stock markets. Levine (2005) notes that, empirically speaking, it is of secondary importance whether the source of the development in financial systems arises from improvements in banks, stock markets or bond markets.

Now, a financial contract will be offered by an intermediary to a firm only if it yields the former nonnegative profits, $I(\tau, v) \geq 0$. Suppose that $r_1 < \widetilde{r}$. A necessary condition for a contract to yield nonnegative profits is for the intermediary to devote more than the minimal level of resources per unit of funds lent, $1/\epsilon$, to monitoring a report of a bad state. If this is not done, the firm will always claim that it is in the low state, and the intermediary can only earn a loss on the contract. Briefly consider the case where $r_1 \geq \widetilde{r}$. Here the firm's return on capital in its worst state of nature is at least as large as the cost of capital, $\widetilde{r}$. Firms would desire to borrow an infinite amount of capital. An equilibrium will not exist.

Assumption: *$r_1 = R(\theta_1, w) < \widetilde{r} = \widehat{r} + \delta$, for all firm types.*

Later, a lower bound on the level of productivity in the financial sector, ζ, will be imposed that guarantees that this assumption holds whenever $z > \zeta$. This lower bound ensures that the equilibrium wage, w, is high enough so that the assumption will always hold–see (2).

Lemma 2 *(Interior solution for monitoring) $m_1/k > 1/\epsilon$, for all $v > 0$.*

Proof. The argument is presented in Appendix 14.1. ■

6 Competitive Financial Intermediation

In the economy there is perfect competition in the financial sector. Consequently, an intermediary must offer a contract that maximizes a firm's value, subject to the restriction that the former does not incur a loss. If an intermediary failed to do so it would be undercut by others. The upshot is that intermediaries will make zero profits on each type of loan. Furthermore, for a firm to produce, it must make non-negative profits too. It is intuitive that for all this to happen a project must offer some potential for a surplus or that $\pi_1 r_1 + \pi_2 r_2 \geq \widetilde{r}$; i.e., the expected return on capital must exceed its cost. Assume this. The value of a firm, v, is then determined by the condition

$$v = V(\tau) \equiv \arg\max_{x}\{x : I(\tau, x) = 0\}. \tag{16}$$

The implications of perfect competition will now be analyzed. Key questions are: (ii) What will be the loan size? (ii) Which firms will get funded?

The size of a loan for a project, k, can now be determined. To do this, substitute (10) in Problem (P3) and solve for the optimal level of monitoring, m_1/k. Plug this solution for monitoring in the objective function for (P3) to obtain a formula for $I(\tau, v)$.[6] Next, solve for v using condition (16). Plugging the obtained formulae for m_1/k and v into (10) yields

$$\begin{aligned} k &= (\psi\gamma + \gamma - \psi)^{\gamma/(\psi\gamma-\psi)}(\frac{1}{\psi})^{-\psi/(\psi\gamma-\psi)}(\frac{1}{\psi\gamma+\gamma})^{(\gamma+\psi)/(\psi\gamma-\psi)} \\ &\quad \times(\pi_1 r_1 + \pi_2 r_2 - \widetilde{r})^{(\gamma+\psi)/(\psi\gamma-\psi)}(\pi_1 w)^{-\psi/(\psi\gamma-\psi)}[\frac{(\epsilon z)^{\psi}}{\pi_2(r_2 - r_1)}]^{\gamma/(\psi\gamma-\psi)}. \end{aligned} \quad (17)$$

Equation (17) gives a determinate loan size for each type of funded project. Furthermore, funding is increasing in a project's expected return and is decreasing in its volatility.[7]

Lemma 3 *(Loan size) The level of investment in a firm, k, is increasing in its expected net return, $\pi_1 r_1 + \pi_2 r_2 - \widetilde{r}$, and the state of technology in the financial sector, z, and is decreasing in the variance of the return, $r_2 - r_1$ (holding the wage rate, w, fixed).*

Proof. See Appendix 14.1. ■

Attention will now be directed toward determining which projects will be funded. Consider the set of firms, $\mathcal{A}(w)$, defined by

$$\mathcal{A}(w) \equiv \{\tau : \pi_1 r_1 + \pi_2 r_2 - \widetilde{r} > 0\}. \quad (18)$$

Intuitively, one might expect that this set of projects will be funded in equilibrium because they offer an expected return on capital, $\pi_1 r_1 + \pi_2 r_2$, that is greater than its user cost, $\widetilde{r}$. This turns out to be true. The contracting problem (P2) implies that the firm will never make negative profits, given (3) to (6). The construction of equation (17) suggests that the intermediary will be able to make a loan in this situation, and not incur a loss.

---

[6] The solution obtained for $I(\tau, v)$ is

$$\begin{aligned} I(\tau, v) = (\psi\gamma + \gamma - \psi)(\frac{1}{\psi})^{-\psi/(\psi\gamma+\gamma-\psi)}(\frac{1}{\psi\gamma+\gamma})^{(\psi\gamma+\gamma)/(\psi\gamma+\gamma-\psi)}(\pi_1 r_1 + \pi_2 r_2 - \widetilde{r})^{(\psi\gamma+\gamma)/(\psi\gamma+\gamma-\psi)} \\ \times(\pi_1 w)^{-\psi/(\psi\gamma+\gamma-\psi)}[\frac{(\epsilon z)^{\psi}}{\pi_2(r_2 - r_1)}]^{\gamma/(\psi\gamma+\gamma-\psi)} v^{\gamma/(\psi\gamma+\gamma-\psi)} - v. \end{aligned}$$

This solution presumes that $r_1 < \widetilde{r}$, so there is an interior solution for monitoring, and that $\pi_1 r_1 + \pi_2 r_2 \geq \widetilde{r}$. It is easy to see that the intermediary's profit function, $I(\tau, v)$, is $\cap$ shaped with the following properties: (i) $I(\tau, 0) = 0$; (ii) $\lim_{v\to 0} \partial I(\tau, v)/\partial v = \infty$; (iii) $\partial^2 I(\tau, v)/\partial v^2 < 0$; (iv) $\lim_{v\to\infty} I(\tau, v) = -\infty$. Therefore, there is only one $v > 0$ that solves (16).

[7] Observe that loan size is a function of ϵz. In this sense, the choice of the constant ϵ in the odds of detection function can be thought of as normalization relative to z.

Lemma 4 *(The set of funded firms) A necessary and sufficient condition for a type-τ firm to be active or funded, or for $k(\tau) > 0, I(\tau, V(\tau)) = 0$ and $V(\tau) > 0$, is that $\tau \in \mathcal{A}(w)$.*

Proof. Those interested should go to Appendix 14.1. ■

Therefore, as was mentioned in the introduction, a simple threshold rule exists for funding, as characterized by (18). Call $\mathcal{A}(w)$ the set of active firms. From (2) it is easy to see that

$$\pi_1 r_1 + \pi_2 r_2 - \widetilde{r} = \alpha(1-\alpha)^{(1-\alpha)/\alpha} w^{-(1-\alpha)/\alpha}[\pi_1 \theta_1^{1/\alpha} + \pi_2 \theta_2^{1/\alpha}] - \widetilde{r} \geq 0.$$

Observe that the firm's profits are decreasing in wages. A type-τ firm will operate when $w \leq \overline{W}(\tau)$, and will not otherwise, where the cutoff wage, $\overline{W}(\tau)$, is specified by

$$\overline{W}(\tau) \equiv \alpha^{\alpha/(1-\alpha)}(1-\alpha)(\frac{\pi_1 \theta_1^{1/\alpha} + \pi_2 \theta_2^{1/\alpha}}{\widetilde{r}})^{\alpha/(1-\alpha)}. \tag{19}$$

So, the set of active projects $\mathcal{A}(w)$ can be expressed equivalently as

$$\mathcal{A}(w) = \{\tau : w < \overline{W}(\tau)\}. \tag{20}$$

The active set depends on the wage because $r_i = R(\theta_i, w)$. It contracts (expands) with a rise (decrease) in the real wage, since $R(\theta_i, w)$ is decreasing in w. In equilibrium the wage rate, w, turns out to be increasing function of the state of technology in the financial sector, z. Hence, the active set will shrink with technological improvement in the financial sector, or a rise in z. This will become clearer in Section 8.

Figure 5 summarizes the discussion on funding. As the expected return on a project rises more capital is allocated to it, as is illustrated in the first panel. (Note that the direction of the x and y axes is specific to each panel.) The increase in funding (or the scale of the firm) is associated with higher monitoring costs given the increasing, convex form of the cost function (fourth panel). The amount of monitoring done per unit of capital is then economized on (third panel). As a consequence, the odds of detecting fraud drop (second panel). In response the expected rents earned by the firm rise (fifth panel).

As the risk associated with a project rises its funding is slashed (first panel). When the difference between the good and bad state widens there is more incentive for the firm to falsify its earning. The intermediary therefore monitors more per unit of capital lent (third panel). Total monitoring costs fall with risk, because the size of the loan is smaller (fourth panel). The probability of detecting malfeasance therefore moves up since monitoring per unit of capital is now higher (second panel). A firm's rents will fall with a rise in risk (fifth panel), because it receives a smaller loan and faces more vigilant policing.

7 Stationary Equilibrium

The focus of the analysis is on stationary equilibria. First, the labor-market-clearing condition for the model will be presented. Second, a definition for a stationary equilibrium will be given. Third, it will be demonstrated that a stationary equilibrium for the model exists.

On the demand side for labor, only firms with $\tau \in \mathcal{A}(w)$ will be producing output. On the supply side, recall that the economy has one unit of labor in aggregate. The labor-market-clearing condition will then appear as

$$\int_{\mathcal{A}(w)} [\pi_1 l_1(\theta_1, \theta_2) + \pi_2 l_2(\theta_1, \theta_2) + \pi_1 l_{m1}(\theta_1, \theta_2)] dF(\theta_1, \theta_2) = 1. \tag{21}$$

It is now time to take stock of the situation so far by presenting a definition of the equilibrium under study. It will be assumed that the economy rests in a stationary state where the cost of capital is $\widetilde{r} = \widehat{r} + \delta$.

Definition 1 *Set the steady-state cost of capital at* $\widetilde{r}$. *A stationary competitive equilibrium is described by a set of labor allocations,* l *and* l_m, *a financial contract,* $\{p_1, p_2, p_{12}, p_{21}, k, m_1, m_2\}$, *a set of active monitored firms,* $\mathcal{A}(w)$, *firm values* v, *and a wage rate,* w, *such that:*

1. *The financial intermediary offers a contract,* $\{p_1, p_2, p_{12}, p_{21}, k, m_1, m_2\}$, *which maximizes its profits,* I, *in accordance with (P2), given the cost of capital and wages,* $\widetilde{r}$ *and* w, *and the value of firms,* v. *The intermediary hires labor for monitoring in the amount* $l_m = (m/z)^\gamma$.

2. *The financial contract offered by the intermediary maximizes the value of a firm,* v, *in line with (16), given the prices* $\widetilde{r}$ *and* w.

3. *A firm is offered a contract if and only if it lies in the active set,* $\mathcal{A}(w)$, *as defined by (18), given* $\widetilde{r}$ *and* w. *It hires labor,* l, *so as to maximize its profits in accordance with (P1), given,* w, *and the size of the loan,* k, *offered by the intermediary.*

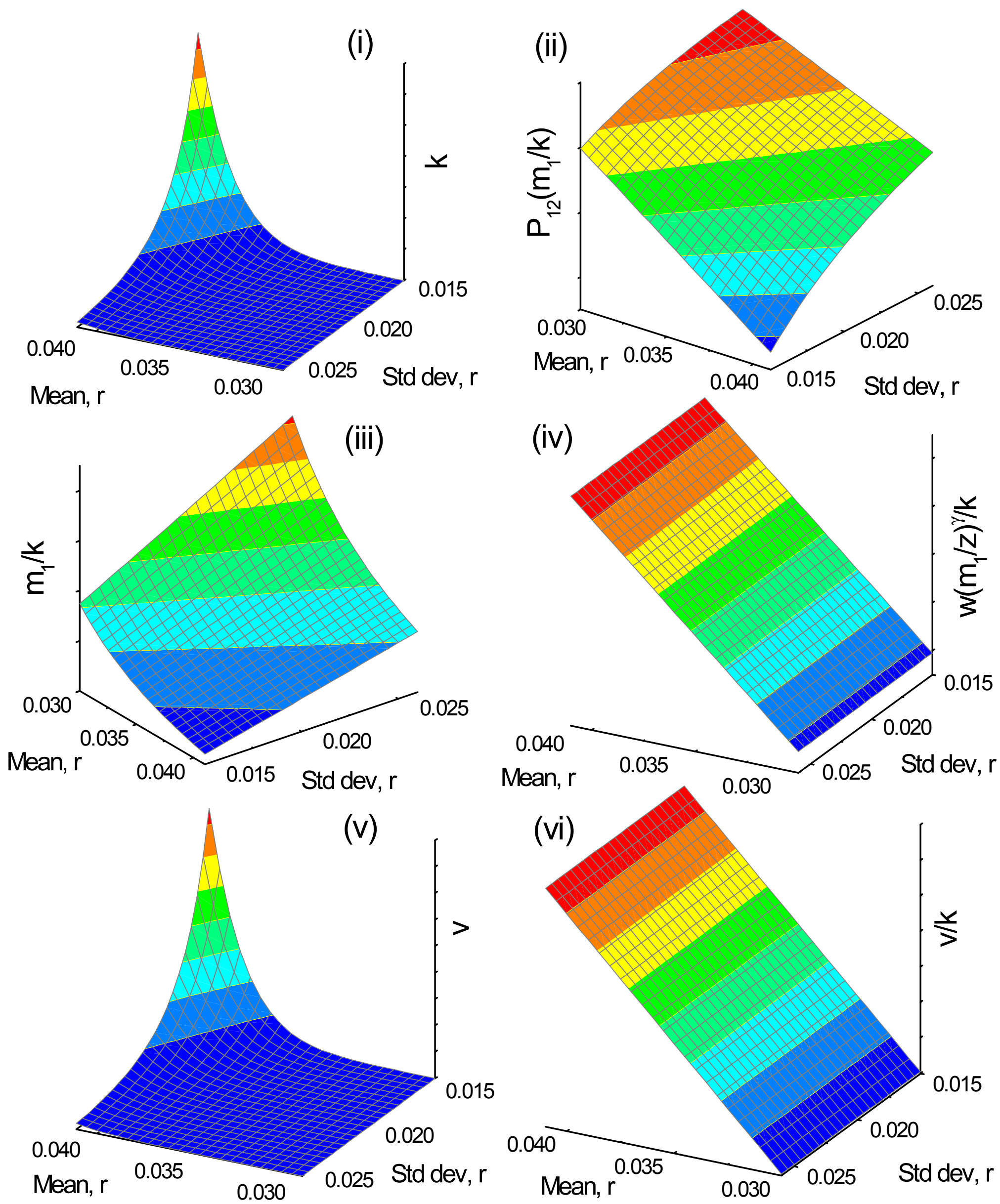

Figure 5: The determination of firm size as a function of the mean and standard deviation of $r = R(\theta, w)$

4. The wage rate, w, is determined so that the labor market clears, in accordance with (21).

When will an equilibrium exist for the economy under study? To address this question, let $\overline{\theta}_1 \equiv \max\{\theta_1 : (\theta_1, \theta_2) \in \mathcal{T}$, over all $\theta_2\}$. Next, define the constant ω by the equation

$$R(\overline{\theta}_1, \omega) = \widetilde{r}. \tag{22}$$

The constant ω specifies a lower bound on the feasible equilibrium wage rate.[8] When $w = \omega$ for a type-$(\overline{\theta}_1, \theta_2)$ project it will happen that $r_1 = \widetilde{r}$. In this situation the intermediary could simply ask for a payment of $r_1 k$ in both states of the world and engage in no monitoring. It would earn zero profits. The firm's profits would be $\pi_2(r_2 - r_1)k$. It would desire a loan of infinite size. Therefore, as the equilibrium wage, w, approaches ω from above the equilibrium defined above will eventually become tenuous.

The situation is portrayed in Figure 6, which graphs the demand and supply for labor. The demand for labor is portrayed by the solid line labeled L. The properties of this demand schedule are established during the course of the proof for Lemma 5. Demand is downward sloping in w. The question is whether or not it will cross the vertical supply schedule for labor. Now, at a given wage rate, loan size increases with more efficient intermediation. This leads to more labor being demanded when intermediation improves. In other words, it can be shown that the demand for labor schedule shifts rightward with an upward movement in z. Now, define ζ as the level of z such that the demand curve intersects the supply curve at the point $(1, \omega)$.

Lemma 5 *(Existence of an equilibrium) There is a constant ζ such that for all $z > \zeta$ there exists a stationary equilibrium for the economy.*

Proof. See Appendix 14.1. ■

[8] The function $R(\theta, w)$ is continuous and strictly decreasing in w, with $\lim_{w \to 0} R(\theta, w) = \infty$, and $\lim_{w \to \infty} R(\theta, w) = 0$; hence, ω is well defined.

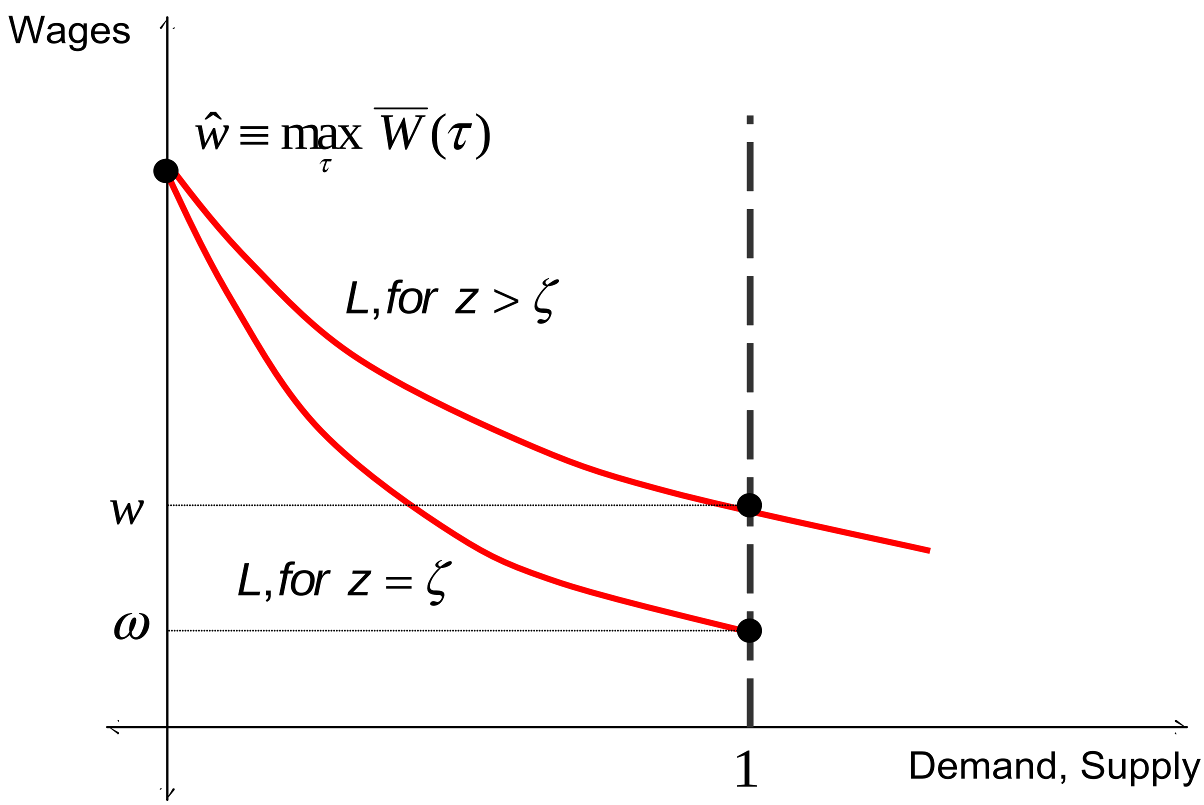

Figure 6: Existence, demand and supply for labor

8 The Impact of Technological Progress on the Economy

The primary goal of the analysis is to understand how technological advance in the financial sector affects the economy. To this end, the impact that technological progress, in either the financial or production sector, has on the portfolio of funded projects will be characterized. To develop some intuition for the economy under study, some special cases will be examined.

8.1 Balanced Growth

In the first special case, technological progress in the financial sector proceeds in balance with the rest of the economy. Specifically, assume that the economy is moving along a balanced growth path where the $\theta_i^{1/\alpha}$'s grow at the common rate $g^{1/\alpha}$ and z grows at rate $g^{1/(1-\alpha)}$.[9]

Therefore, $F_{t+1}(\theta_1, \theta_2) = F_t(\theta_1/g, \theta_2/g)$. The salient features of this case are summarized by the proposition below.

Proposition 1 *(Balanced growth) Let the $\theta_i^{1/\alpha}$'s grow at rate $g^{1/\alpha}$ and z increases at rate $g^{1/(1-\alpha)}$. There exists a balanced growth path where the capital stock, k, wages, w, and rents, v, will all grow at rate $g^{1/(1-\alpha)}$. The amount of resources devoted to monitoring per unit of capital, m_1/k, remains constant.*

Proof. Refer to Appendix 14.1. ■

In this situation the financial sector is not becoming more efficient over time, relative to the rest of the economy. The amount of monitoring done per unit of capital invested remains constant over time. Thus, the probability of a firm getting caught by misrepresenting a high level of earnings, $P_{21}(m_1/k)$, is constant over time too. For any particular project type, the spread between the return on capital (net of labor costs) and the interest earned by investors, $\pi_1 r_1 + \pi_2 r_2 - \widetilde{r}$, is fixed over time. One could easily allow the production of monitoring services to require capital as well labor, say as given by $m = z k_m^{\kappa} l_m^{1/\gamma}$ with $0 < 1/\gamma, \kappa, 1/\gamma + \kappa < 1$. Now, for a balanced growth path to obtain, z would have to grow at rate $g^{\kappa/(1-\alpha)}$. The

[9] In order to get a fixed interest rate assume that the consumer/worker has isoelastic preferences over consumption. Then, in standard fashion along a balanced growth path, $\widehat{r} = g^{\iota/(1-\alpha)}/\beta - 1$, where ι is the coefficient of relative risk aversion.

existence of a balanced growth path results from the fact that the probability of detection, $P_{21}(m_1/k)$, depends on the employment of monitoring services relative to the size of the loan. If loan size did not enter this function, then technological advance in the non-financial sector would lead to a drop in the cost of monitoring as the economy's capital stock rose. That is, there would be a positive feedback loop from the state of development in the non-financial sector to the state of development in the financial one, as is modeled in three different ways in de la Fuente and Marin (1996), Greenwood and Jovanovic (1990), and Harrison, Sussman and Zeira (1999).

8.2 Unbalanced Growth

The above case suggests that for technological progress in the financial sector to have an impact it must outpace advance in the rest of the economy. Suppose that this is the case. Then, one would expect that as monitoring becomes more efficient those projects offering the lowest expected return will be cut.

Proposition 2 *(Technological progress in financial intermediation) Consider z and z' with $z < z'$. Let w and w' be the wage rates associated with z and z', respectively. Then, $\mathcal{A}(w') \subset \mathcal{A}(w)$. Additionally, if $\tau = (\theta_1, \theta_2) \in \mathcal{A}(w) - \mathcal{A}(w')$ and $\tau' = (\theta_1', \theta_2') \in \mathcal{A}(w')$ then $\pi_1(\theta_1)^{1/\alpha} + \pi_2(\theta_2)^{1/\alpha} < \pi_1(\theta_1')^{1/\alpha} + \pi_2(\theta_2')^{1/\alpha}$.*

Proof. See Appendix 14.1. ■

An increase in z makes financial intermediation more efficient. For any given wage rate, w, the aggregate demand for labor will increase for two reasons. First, more capital will be lent to each funded project. Second, more labor will also be hired by the intermediary to monitor the project. Since the demand for labor rises, the wage rate must move up to clear the labor market. This increase in wages causes the set of active projects, $\mathcal{A}(w)$, to shrink, with the projects offering the lowest expected return being culled.

Alternatively, technological advance could occur in the production sector and not the financial one. Here, the lack of development in the financial sector will hinder growth in the rest of the economy. Specifically, technological advance in the production sector of the economy will drive up wages. This leads to the costs of monitoring rising. Therefore, less is done. This lack of scrutiny by intermediaries now allows firms with marginal projects offering low expected returns to receive funding.

Proposition 3 *(Technological progress in production) Suppose all the $\theta_i^{1/\alpha}$'s increase by the factor $g^{1/\alpha}$, holding z fixed. Then, the set of active projects, $\mathcal{A}(w)$, expands with the new projects offering lower expected returns than the old ones.*

Proof. See Appendix 14.1. ■

8.3 Efficient Finance

An extreme example of Proposition 2 would be to assume that z grows forever. Then, the financial sector will become infinitely efficient relative to the rest of the economy. This leads to the fourth special case.

Proposition 4 *(Efficient finance) Suppose $\mathcal{T}$ is a compact and countable subset of R_+^2, with a positive measure of projects for each type, $\tau = (\theta_1, \theta_2)$. Then,*

1. $\lim_{z\to\infty} A(w) = A^* \equiv \arg\max_{\tau=(\theta_1,\theta_2)\in\mathcal{T}}[\, \pi_1(x\theta_1)^{1/\alpha} + \pi_2(x\theta_2)^{1/\alpha}]$,

2. $\lim_{z\to\infty} m_1/z = 0$, *for* $\tau \in A^*$,

3. $\lim_{z\to\infty} m_1/k = \infty$ *and* $\lim_{z\to\infty} P_{12}(m_1/k) = 1$, *for* $\tau \in A^*$,

4. $\lim_{z\to\infty} p_2 = r_2 k$, *for* $\tau \in A^*$,

5. $\lim_{z\to\infty} v = 0$, *for* $\tau \in A^*$,

6.
$$\lim_{z\to\infty} w = w^* \equiv \alpha^{\alpha/(1-\alpha)}(1-\alpha)\{\max_{\tau=(\theta_1,\theta_2)\in\mathcal{T}}[\, \pi_1(\theta_1)^{1/\alpha} + \pi_2(\theta_2)^{1/\alpha}]/\widetilde{r}\}^{\alpha/(1-\alpha)}, \qquad (23)$$

7.
$$\lim_{z\to\infty} \int_{\mathcal{A}(w)} kdF = \mathbf{k}^* \equiv (\frac{\alpha}{\widetilde{r}})^{1/(1-\alpha)}\{\max_{\tau=(\theta_1,\theta_2)\in\mathcal{T}}[\, \pi_1(\theta_1)^{1/\alpha} + \pi_2(\theta_2)^{1/\alpha}]\}^{\alpha/(1-\alpha)}. \qquad (24)$$

Proof. Refer to Appendix 14.1. ■

As the cost of monitoring borrowers drops, the intermediation sector becomes increasingly efficient. The financial intermediary can then perfectly police loan payments without devoting a significant amount of resources in terms of labor to this activity, as points (2) and (3) in the proposition make clear. Since firms are operating constant-returns-to-scale production technologies, no rents will accrue on their activity—see point (5). Firms must pay the full marginal product of capital to the intermediary—point (4). That is, the spread between a firm's internal rate of return (before depreciation), $\pi_1 r_1 + \pi_2 r_2 - \delta$, and the user cost of capital, $\widetilde{r} = \widehat{r} + \delta$, vanishes, where the latter is made of the interest paid to savors, $\widehat{r}$, and rate of depreciation, δ. In this world only projects with the highest return are financed, as point (1) states, even though they may be the most risky. In the aggregate any idiosyncratic project risk washes out. Therefore, in the absence of a contracting problem, only the mean return on investment matters. And, with constant-returns-to-scale technologies everything should be directed to the most profitable opportunity. The wage rate, w^*, and aggregate capital stock, $\mathbf{k}^*$, in the efficient economy are determined in standard fashion by the conditions that the marginal product of capital for the most profitable projects must equal the user cost of capital, $\widetilde{r}$, and the fact that the labor market must clear. These two conditions yield (24) and (23). (By comparison, consider the standard deterministic growth model with the production technology $o = \theta k^{\alpha} l^{1-\alpha}$ and one unit of aggregate labor. Here $w^* \equiv \alpha^{\alpha/(1-\alpha)}(1-\alpha)[\theta^{1/\alpha}/\widetilde{r}]^{\alpha/(1-\alpha)}$ and $\mathbf{k}^* \equiv [\alpha/\widetilde{r}]^{1/(1-\alpha)}(\theta^{1/\alpha})^{\alpha/(1-\alpha)}$. The differences in the formulae are due to two facts that pertain to the current setting: (i) the best projects from a portfolio $\mathcal{T}$ are chosen; (ii) there is uncertainty in θ.)

9 Fitting the Model to the U.S. Economy

9.1 Procedure

The quantitative analysis will now begin. To start with, assume that firms now produce output in line with the production function

$$o = x\theta k^{\alpha} l^{1-\alpha},$$

where x represents a known level of aggregate productivity. To simulate the model, values must be assigned to its parameters. This will be done by calibrating the framework to match some stylized facts for the U.S. economy. Some parameters are standard. They are given conventional values. Capital's share of income, α, is chosen to be 0.33, a very standard number. Likewise, the depreciation rate, δ, is set to 0.07, again a very common number. The chosen value for the discount factor, $\beta = 1/(1+\widehat{r})$, implies that the interest rate earned by savers is 6.3 percent. This is in conformity with Cooley and Prescott's (1995, p. 19) estimate of 6.9 percent for the real return to capital over the postwar period. The concept used for the capital stock is much narrower here, though; i.e., it is just the stock of business capital. Therefore, matching the low observed capital-output ratio will be harder.

Nothing is known about the appropriate choice for parameters governing the intermediary's monitoring technology, or ϵ, ψ, and γ. The selection of a value for ϵ amounts to a normalization (relative to some baseline level of z). Therefore, set $\epsilon = 100$. Similarly, little is known about the distribution of returns facing firms. Let μ_m be the mean *across firms* of expected total factor productivity (TFP); i.e., $\mu_m = \int(\pi_1\theta_1 + \pi_2\theta_2)dF$. Likewise, μ_v will denote the mean *over firms* of the *logarithm* of the volatility of TFP; i.e., $\mu_v = \int \ln[\pi_1\pi_2(\theta_2 - \theta_1)^2]dF$. In a similar vein, ρ will represent the correlation between the means and (ln) volatilities of *firm-level* TFP, while σ_m^2 and σ_v^2 will denote the variance of these firm-level variables. Assume that these means and (ln) volatilities of firm-level TFP are distributed according to a bivariate truncated normal, $N(\mu_m, \mu_v, \sigma_m^2, \sigma_v^2, \rho)$. Thus, values need to be selected for the parameters $\psi, \gamma, \mu_m, \mu_v, \sigma_m^2, \sigma_v^2$, and ρ.

Let $\text{T}\textsc{argets}_j$ represent the j-th component of a n-vector of observations that the model should match. Similarly, $\text{O}\textsc{utput}(\textsc{param})$ denotes the model's prediction for this vector. The model's solution will be a function of the list of calibrated parameters, $\textsc{param}$. These parameters are picked to minimize a weighted sum of the squared deviations between the data targets and the model's output:

$$\min_{\text{PARAM}} \sum_{j=1}^{n} \text{WEIGHTS}_j \left[\text{TARGETS}_j\text{-OUTPUT}_j\left(\text{PARAM}\right)\right]^2, \tag{25}$$

where WEIGHTS_j represents the weight attached to j-th target. The data targets will now be discussed.

The distribution of returns across firms will be integrally related to the distribution of employment across them. Firms with high returns will have high employment, other things equal. Therefore, the size distribution of firms for the year 1972 is chosen as a data target. Eight points on this distribution are picked. As was mentioned in the introduction, to the extent that financial intermediation has become more efficient over time, relative to the nonfinancial sector, one would expect that the capital/output ratio would rise. If a nation's productivities in the financial and non-financial sectors increase in a balanced way then just output will increase. Denote an aggregate quantity in bold. The model provides a mapping between the aggregate level of output (per person), $\mathbf{o}$, and the capital/output ratio, $\mathbf{k}/\mathbf{o}$, on the one hand, and the state of technology in its production and financial sectors, x and z, on the other. Represent this mapping by $(\mathbf{o}, \mathbf{k}/\mathbf{o}) = M(x, z)$–the dependence of this mapping on the model's parameters has been suppressed for notational convenience. Now, while the states of the U.S.'s financial and non-financial technologies are unobservable directly, this mapping can be used to make an inference about (x, z), given an observation on $(\mathbf{o}, \mathbf{k}/\mathbf{o})$, by using the relationship

$$(x, z) = M^{-1}(\mathbf{o}, \mathbf{k}/\mathbf{o}). \tag{26}$$

Therefore, TFP's in the financial and non-financial sectors are picked to match GDP and the capital/output ratio for the years 1972 and 2000. This determines x_{1972}, z_{1972}, x_{2000}, and z_{2000}, as functions of the other estimated parameter values and, of course, the target values for $\mathbf{o}$ and $\mathbf{k}/\mathbf{o}$. The calibrated parameter vector is $\text{PARAM} \equiv (\epsilon, \psi, \gamma, \mu_m, \mu_v, \sigma_m^2, \sigma_v^2, \rho, x_{1972}, z_{1972}, x_{2000}, z_{2000})$. Thus, the calibration procedure is picking 12 parameters to match 12 observations, as best as can be done subject to the restriction (26). Table 1 presents the parameter values for the model.

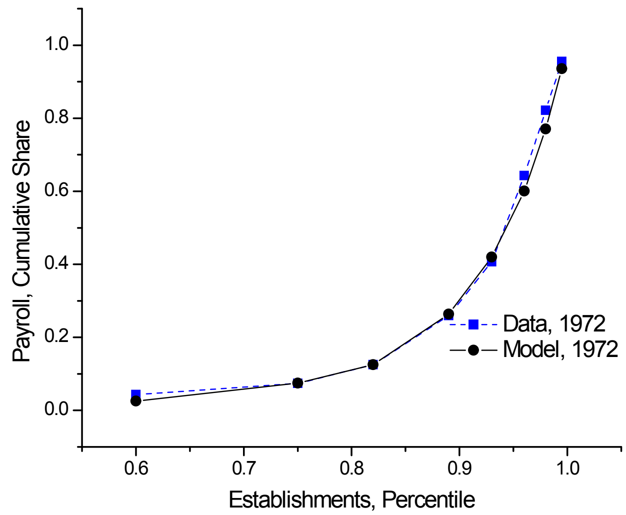

Figure 7: Firm-size distribution, 1972

9.2 Results

The model's prediction for the 1972 firm-size distribution is shown in Figure 7. It fits remarkably well. The firm-size distribution shifts somewhat between 1972 and 2000. Now, monitoring and the provision of financial services are abstract goods, so it hard to know what a reasonable change in z should be. One could think about measuring productivity in the financial sector, as is often done, by $\mathbf{k}/\mathbf{l}_m$, where $\mathbf{k}$ is the aggregate amount of credit extended by financial sector and $\mathbf{l}_m$ is the aggregate labor that it employs. By this traditional measure, productivity in the financial sector rose by 3.8 percent between 1973 and 2000. Berger (2003, Table 5) estimates that productivity in the commercial banking sector increased by 2.2 percent a year over this same period (which includes the troublesome productivity slowdown) and by 3.2 percent from 1982 to 2000. The model's ability to match cross-country differences in the efficiency of financial intermediation and differences in the interest-rate spreads between borrowers and lenders will be addressed in Section 10.

Table 1: Parameter Values

Parameter	*Definition*	*Basis*
$\alpha = 0.33$	Capital's share of income	Standard value
$\delta = 0.07$	Depreciation rate	Standard value
$\beta = 1/(1+\widehat{r}) = 0.94$	Discount factor	Cooley and Prescott (1995)
$\epsilon = 100$	Pr of detection, constant	Normalization
$\psi = 0.52$	Pr of detection, exponent	Calibrated to fit targets
$\gamma = 1.86$	Monitoring cost function	Calibrated to fit targets
$\mu_m = 1.54$, $\mu_v = 0.80$	means	Calibrated to fit targets
$\sigma_m^2 = 0.49$, $\sigma_v^2 = 0.66$, $\rho = 0.88$	variances and correlation	Calibrated to fit targets
$x_{1972} = 60.364, z_{1972} = 1.083e5,$	TFP's	Calibrated to fit targets
$x_{2000} = 88.549, z_{2000} = 6.785e5$		

Table 2 presents results for some other variables of interest. Here, the aggregate value for a variable is again indicated in bold, so that $\mathbf{x} = \int_{\mathcal{A}(w)} xdF$ for $x = m_1$, $w\pi_1(m_1/z)^{\gamma}$, k, etc. The expected value for x is given by $\mathbf{x} \div \int_{\mathcal{A}(w)} dF$. Monitoring becomes less expensive as z rises (relative to x). This results in the amount of monitoring per unit of capital rising, as reflected in the larger values for $\mathbf{m}_1 \div \mathbf{k}$ for 2000 versus 1972. As a consequence, the likelihood of intermediaries detecting fraud increases. The fraction of a firm's output dissipated in pure rents, $\mathbf{v} \div \mathbf{o}$, declines. This can be seen another way. The internal rate of return, i, earned by a firm on its investment is given by $i = \pi_1 r_1 + \pi_2 r_2 - \delta$. The average internal return earned by firms, weighted by their level of investment, will then be defined by $\mathbf{i} \equiv \int_{\mathcal{A}(w)} kidF \div [\mathbf{k} \times \int_{\mathcal{A}(w)} dF]$. Likewise, denote the average rate of return earned by the intermediary on its lending activity by $\widetilde{\mathbf{i}} \equiv \int_{\mathcal{A}(w)} (\pi_1 p_1 + \pi_2 p_2 - \delta k) dF \div [\mathbf{k} \times \int_{\mathcal{A}(w)} dF]$. The gap between these two returns, $\mathbf{e} \equiv \mathbf{i} - \widetilde{\mathbf{i}}$, measures the average excess return earned by firms due to rents. This excess return is squeezed as rents shrink. In similar fashion, the average spread between the rates of return that intermediaries and savers earn, $\mathbf{s} \equiv \widetilde{\mathbf{i}} - \widehat{r}$, reflects the costs of intermediation incurred by the necessity to monitor borrowers. This interest rate spread declines as the costs of intermediation fall due to technological progress in information production. Rousseau (1998, Figure 4) presents evidence suggesting that financial innovation reduced loan-deposit spreads in the U.S. between 1872 to 1929. Li and Sarte (2003, Table 3), using a structural VAR, present evidence suggesting that drops in the cost of financial intermediation account for a significant part of long-run fluctuations in U.S. manufacturing output.

A rise in the probability of detecting fraud relaxes the incentive constraint (8), and makes it easier to lend more capital to firms. This has two effects. First, as borrowing rates decline renting more capital becomes profitable. This results in a higher aggregate amount of capital being invested per unit of output produced, $\mathbf{k} \div \mathbf{o}$. Second, for a given amount of lending, funds are redirected toward those firms offering the highest rate of return. Along with the first effect, this increases GDP, $\mathbf{o}$. Denote the levels of capital and output that would obtain in the first-best economy by $\mathbf{k}^*$ and $\mathbf{o}^*$. As can be seen, capital and output steadily rise, relative to their first-best outcome, as z moves up (relative to x).

Model TFP is quite volatile across plants, as can be seen from Table 2. Hsieh and Klenow (2008, Tables 1 and 2) report (weighted) standard deviations of 0.45 and 0.85 for 1977. The (weighted) number found here lies in the middle of their range. For the U.S., they show little or no increase in this number over time. This is similar to what is found here. Hence, the volatility in plant-level TFP required to match the 1972 U.S. firm size-distribution appears to be reasonable.

Finance is important in the model. This can be gauged by undertaking the following counterfactual question: By how much would GDP have risen between 1972 and 2000 if there had been no technological progress in the financial sector? As can be seen from the third column of Table 2, output would have risen from \$22,097 to \$34,590 or by about 1.6 percent a year (when continuously compounded). This compares with the increase of 2.4 percent (\$22,097 to \$43,268) that occurs when x rises to its 2000 level. Thus, about one third of the increase in growth is due to innovation in the financial sector. Likewise, the model predicts that about 15 percent of TFP growth was due to improvement in financial intermediation. The financial system actually becomes a drag on development when z is not allowed to increase. Wages rise as the rest of the economy develops. This makes monitoring more expensive. Therefore, less will be done. As a consequence, interest rates rise and the economy's capital/output ratio drops. Without an improvement in the financial system, the firm-size distribution actually moves slightly in the wrong direction.

TABLE 2: IMPACT OF TECHNOLOGICAL PROGRESS IN THE FINANCIAL SECTOR

	1972	*2000*	*Counterfactual*
Production Sector, x	60.364	88.549	88.549
Financial Sector, z	1.083e5	6.785e5	1.083e5
Monitoring-to-capital, $\mathbf{m}_1 \div \mathbf{k}$	0.1018	0.1738	0.0809
Pr of detecting fraud, $\int_{\mathcal{A}(w)} P_{12} dF \div \int_{\mathcal{A}(w)} dF$	0.9651	0.9734	0.9608
Financial sector productivity, $\mathbf{k} \div \mathbf{l}_m$	1.029e6	3.004e6	1.431e6
Rents to output, $\mathbf{v} \div \mathbf{o}$	0.1010	0.0768	0.1132
TFP$\equiv x \int_{\mathcal{A}(w)} (\pi_1\theta_1 + \pi_2\theta_2) dF \div \int_{\mathcal{A}(w)} dF$	530.43	807.30	760.86
Std ln(TFP), $\sqrt{\pi_1\pi_2} \int_{\mathcal{A}(w)} k(\ln\theta_2 - \ln\theta_1) dF \div \int_{\mathcal{A}(w)} dF$	0.6953	0.6763	0.7131
Internal return (weighted), $\mathbf{i}$	0.1443	0.1174	0.1611
Lending rate, $\widetilde{\mathbf{i}}$	0.0787	0.0737	0.0818
Excess return, $\mathbf{e} = \mathbf{i} - \widetilde{\mathbf{i}}$	0.0656	0.0436	0.0783
Interest rate spread, $\mathbf{s} = \widetilde{\mathbf{i}} - \widehat{r}$	0.0149	0.0099	0.0180
Return to savers, $\widehat{r} = 1/\beta - 1$	0.0638	0.0638	0.0638
Capital-to-output ratio, $\mathbf{k} \div \mathbf{o}$	1.5400	1.7613	1.4279
Capital relative to first best, $\mathbf{k} \div \mathbf{k}^*$	0.4452	0.5628	0.3903
Output relative to first best, $\mathbf{o} \div \mathbf{o}^*$	0.7128	0.7879	0.6740
GDP, $\mathbf{o}$	\$22,097	\$43,268	\$34,590

10 Fitting the Model to the World Economy

Ever since Goldsmith (1969), economists have been interested in the cross-country relationship between financial structure and economic development. An implication of the current model is that as the state of technology in the intermediation sector advances, the spread between borrowing and lending rates in an economy will shrink, while its capital-to-output ratio and level of aggregate output increases. The cross-country data is suggestive of such a relationship, as Figure 2 shows.

For the cross-country analysis assume that production in a nation is undertaken in the manner described earlier, but let x and z now represent country-specific productivity factors for the non-financial and financial sectors. While the state of a country's productivities in the non-financial and financial sectors is unobservable directly, once again the mapping (26) can be used to make an inference about (x, z), given an observation on $(\mathbf{o}, \mathbf{k}/\mathbf{o})$. This is done for a sample of 40 countries, using the parameter values listed in Table 1. This implies that the distribution of potential projects differs across countries by the factor of proportionality, x, an assumption needed both for discipline and tractability. The results are reported in Table 7 in Appendix 14.2. By construction the model explains all the variation in output and capital/output ratios across countries.[10] Still, one could ask how well the measure of the state of technology in the financial sector that is backed out using the model correlates with independent measures of financial intermediation. Here, take the ratio of private credit by deposit banks and other financial institutions to GDP as a measure of financial intermediation, as reported by Beck et al. (2001). (Other measures produce similar results but reduce the sample size too much.) Additionally, one could examine how well the model explains cross-country differences in interest-rate spreads, $\mathbf{s}$.

[10] The model predicts a positive association between a country's rate of investment and its GDP. Castro et al. (2009, Figure 1) show that this is true. It is stronger when investment spending is measured at international prices, as opposed to domestic ones. They resolve this difference within the context of a two-sector model where the relative price of capital goods is endogenously determined. In their framework, capital goods are more expensive to produce in poor countries. This happens because this sector is risky, implying that the costs of finance are high in countries with poor investor protection. Thus, this puzzle could be resolved here by adopting aspects of their two-sector analysis.

Table 3 reports the findings. The correlation between the imputed state of technology in the financial sector and the independent measure of financial intermediation is quite high. Thus, it appears reasonable to use the constructed values of z for investigating the relationship between output and financial development. Interestingly, Finland and Peru both have a capital-to-output ratio of about 1.6. The model predicts Finland's z is about 160 percent (continuously compounded) higher than Peru's–the former's $\ln(z)$ is 12.99, compared with 11.40 for the latter; again, see Table 7 in Appendix 14.2. But, recall that the units for $\ln(z)$ are meaningless, since monitoring is abstract good. If one measures productivity in the financial sector by the amount of credit extended relative to the amount of labor employed in the financial sector, as was done earlier, then the analysis suggests that intermediation in Finland is about 145 percent (continuously compounded) more efficient than in Peru. Why? Finland has a much higher level of income per worker and hence TFP than does Peru ($40,603 versus $10,200). Therefore, given the higher wages, monitoring will be more expensive in Finland. To give the same capital/output ratio, efficiency in Finland's financial sector must be higher. As can be seen, the interest-rate spreads predicted by the model are positively associated with those in the data. The correlation is reasonably large. That these two correlations aren't perfect, should be expected. There are other factors, such as the big differences in public policies discussed in Parente and Prescott (2000), which may explain a large part of the cross-country differences in capital/output ratios. Differences in monetary policies across nations may influence cross-country interest rate spreads. Additionally, there is noise in these numbers given the manner of their construction—see Appendix 14.2.

Table 3: Cross-Country Evidence

	interest-rate spread	financial intermediation
Correlation(model, data)	0.37	0.72

11 The Importance of Financial Development for Economic Development

It is now possible to gauge how important efficiency in the financial sector is for economic development, at least in the model. To this end, note that the best financial and industrial practices in the world are given by $\overline{x} = \max\{x_i\}$ and $\overline{z} = \max\{z_i\}$, respectively. Represent country i's output, as a function of the efficiency in its industrial and financial sectors, by $\mathbf{o}_i = \mathbf{O}(x_i, z_i)$–this is really just the first component of the mapping $M(x, z)$. If country i could somehow adopt the best financial practice in the world it would produce $\mathbf{O}(x_i, \overline{z})$. Similarly, if country i used the best practice in both sectors it would attain the output level $\mathbf{O}(\overline{x}, \overline{z})$. The shortfall in output from the inability to attain best practice is $\mathbf{O}(\overline{x}, \overline{z}) - \mathbf{O}(x_i, z_i)$. Luxembourg turns out to have the highest value for x, and Switzerland for z.

The percentage gain in output for country i by moving to best financial practice is given by $100 \times [\ln \mathbf{O}(x_i, \overline{z}) - \ln \mathbf{O}(x_i, z_i)]$. The results for this experiment are plotted in Figure 8. As can be seen, the gains are quite sizeable. On average a country could increase its GDP by 31 percent, and TFP by 10 percent. The country with the worst financial system, Sri Lanka, would experience a 76 percent rise in output. Its TFP would increase by 26 percent. While sizeable, these gains in GDP are small relative to the increase that is needed to move a country onto the frontier for income, $\mathbf{O}(\overline{x}, \overline{z})$. The percentage of the gap that is closed by a movement to best financial practice is measured by $100 \times [\mathbf{O}(x_i, \overline{z}) - \mathbf{O}(x_i, z_i)]/[\mathbf{O}(\overline{x}, \overline{z}) - \mathbf{O}(x_i, z_i)] \equiv 100 \times \mathbf{G}(x_i, z_i)$. Figure 9 plots the reduction in this gap for the countries in the sample.[11] The average reduction is this gap is only 17 percent. For most countries the shortfall in output is accounted for by a low level of total factor productivity in the non-financial sector.

Therefore, the importance of financial intermediation for economic development depends on how you look at it. World output would rise by 28 percent by moving all countries to the best financial practice—see Table 4. This is a sizeable gain. Still, it would only close 12

[11] Luxembourg has been deleted from the graph. The reduction in its gap is 100 percent.

percent of the gap between actual and potential world output. Dispersion in cross-country output would fall by about 17 percentage points from 77 percent to 58 percent.[12] Financial development explains about 28 percent of cross-country dispersion in output by this metric.

TABLE 4: WORLD-WIDE MOVE TO FINANCIAL BEST PRACTICE, $\overline{z}$

Increase in world output (per worker)	28%
Reduction in gap between actual and potential world output	12%
Fall in dispersion of ln(output) across countries	17% ($\simeq$ 77% - 59%)
Fall in (pop-wghtd) mean of (cap-wghtd) distortion	15% ($\simeq$ 17% - 2%)
Fall in (pop-wghtd) mean dispersion of (cap-wghtd) distortion	4.65% ($\simeq$ 5% - 0.35%)

The presence of informational frictions causes the expected marginal product of capital, $\pi_1 r_1 + \pi_2 r_2$, to deviate from its user cost, $\widetilde{r}$. Define the induced distortion in investment by $d = \pi_1 r_1 + \pi_2 r_2 - \widetilde{r}$. For a country such as Sri Lanka these deviations are fairly large. The (capital-weighted) mean level of this distortion is 29 percentage points. It varies across plants a lot, as indicated by a coefficient of variation of 32 percent. This is the type of resource misallocation effect emphasized by Restuccia and Rogerson (2008). Here, the distortion is modelled endogenously. If Sri Lanka adopted the Swiss financial practices the average size of this distortion would drop to 1.5 percentage points. Its standard deviation across plants collapses from 9 percentage points to just 0.3 percentage points. The elimination of this distortion results in capital deepening among the active plants. Average TFP would rise by 26 percent in the model, as inefficient plants are culled. For the world at large, the average size of the distortion is 17 percentage points, with an average coefficient of variation of 28 percent. The mean distortion drops to 2 percentage points with a world-wide movement to financial best practice. The average standard deviation across plants falls from 5 percentage

[12] The impact of financial intermediation on income will be larger if the former is allowed to affect TFP in the production sector more directly. Erosa and Hidalgo-Cabrillana (2008) undertake a theoretical analysis where entrepreneurs produce an intermediate good that is important for the production of final output. They employ a Lucas (1978) style span of control model. A limited ability to enforce financial contracts leads to a poor selection of entrepreneurs in the economy. This channel of effect may be important because Levine (2005) documents that financial development has a causal impact of productivity. A related quantitative model is in Amaral and Quintin (2005). They emphasize the capital accumulation channel.

points to a mere 0.35.

Last, it will be noted that the model could be used to make an inference about productivities in the production and financial sectors, x and z, by using interest rate spreads, $\mathbf{s}$, instead of the capital output ratio, $\mathbf{k}/\mathbf{o}$; i.e., by using the mapping of the form $(x, z) = \widetilde{M}^{-1}(\mathbf{o}, \mathbf{s})$. Erosa (2001) uses interest-rate spreads to quantify the effects of financial intermediation on occupational choice. The correlation between the model's prediction between the state of technology in the financial sector and the independent measure of financial development is again high–Table 5. The model does a reasonable job predicting cross-country capital-to-output ratios. Financial development is now more important, but it still does not explain the bulk of cross-country variation in output–see Table 6. The difference in the quantitative significance obtains from the fact that the percentage variation in cross-country interest rate spreads is much larger than in capital/output ratios. The mean distortion in the world is now 21 percentage points, with an average coefficient of variation of 30 percent.

TABLE 5: CROSS-COUNTRY EVIDENCE

	capital-output ratio	financial intermediation
Correlation(model, data)	0.49	0.70

TABLE 6: WORLD-WIDE MOVE TO BEST FIN. PRAC., $\overline{z}$

[alternative results obtained when using interest-rate spread match, $(x, z) = \widetilde{M}^{-1}(\mathbf{o}, \mathbf{s})$]

Increase in world output (per worker)	48%
Drop in shortfall between actual and potential world output	21%
Fall in dispersion of ln(output) across countries	12% ($\simeq$ 75% - 62%)
Fall in (pop-wghtd) mean of (cap-wghtd) distortion	18% ($\simeq$ 21% - 3%)
Fall in (pop-wghtd) mean dispersion of (cap-wghtd) distortion	5.44% ($\simeq$ 6% - 0.56%)

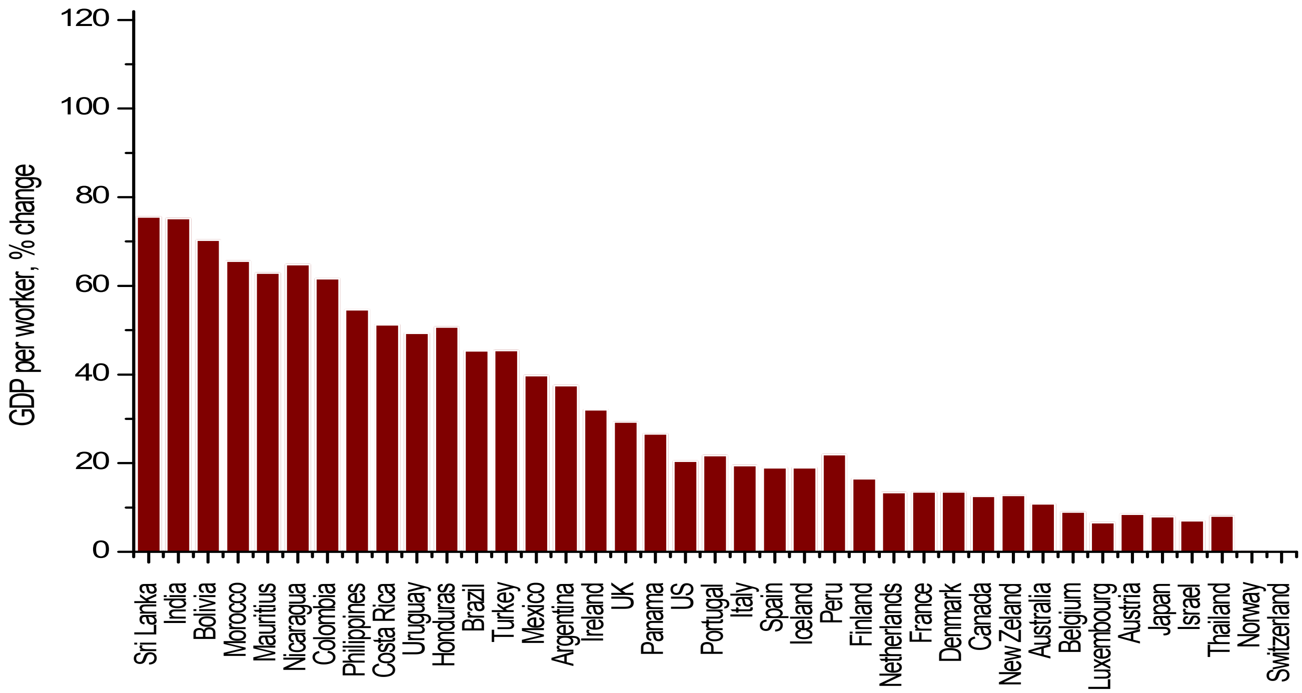

Figure 8: The impact of a move to financial best practice on GDP per worker

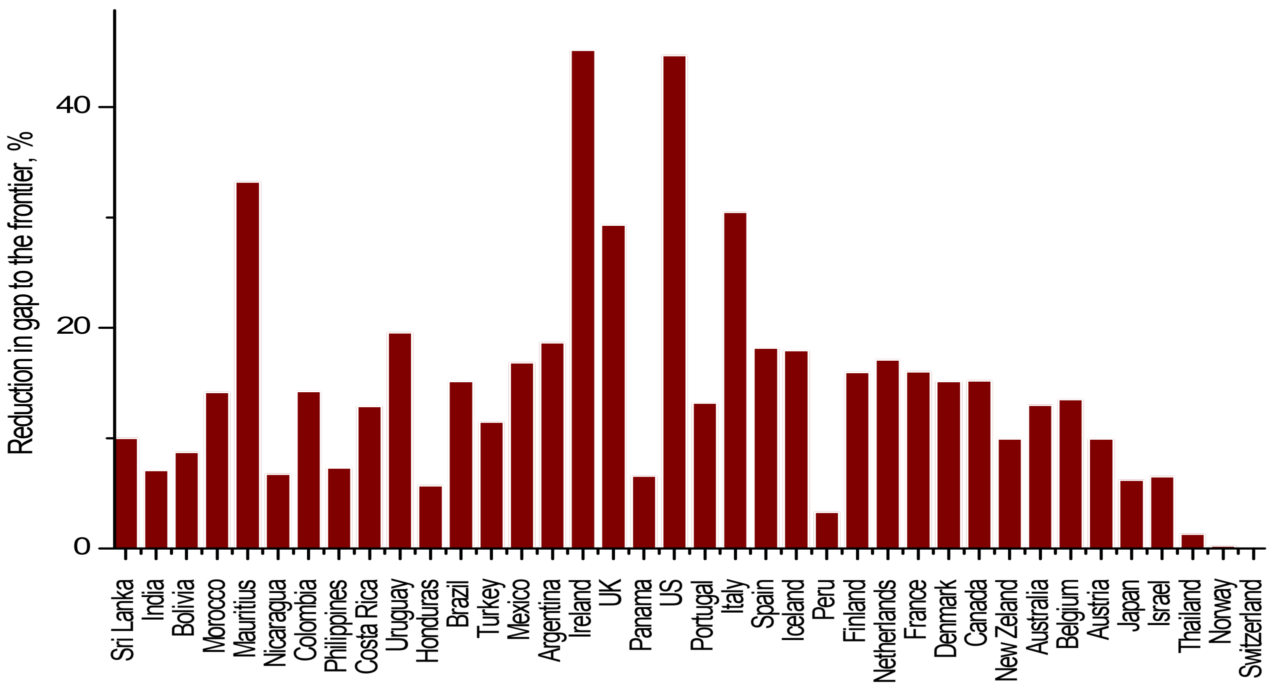

Figure 9: The impact of a move to financial best practice on the gap in GDP per worker

12 Robustness Analysis

12.1 Intangible Investments and Capital's Share of Income

Suppose part of investment spending is undertaken in the form of intangible capital. As a result, measured investment may lie below true investment. This will lead to measured income, GDP, falling short of true output, $\mathbf{o}$. This injects an upward (a downward) bias in the measurement of labor's (capital's) share of income. Specifically, in context of the standard neoclassical model, with a Cobb-Douglas production function, measured labor's share of income, LSI, will appear as

$$\text{LSI} = \frac{\mathbf{o}}{\text{GDP}} \times (1-\alpha) > (1-\alpha).$$

Corrado, Hulten, and Sichel (2007) estimate the amount of intangible investment that was excluded from measured GDP from 1950 to 2003. They show that when output is adjusted to include these unrecognized intangibles, true output, $\mathbf{o}$, is 12 percent higher than measured output, GDP, for the period 2000-2003. As a consequence, it is easy to calculate that

$$\alpha = 1 - \frac{\text{GDP}}{\mathbf{o}} \times \text{LSI} = 1 - \frac{1}{1.12}(1-0.33) = 0.41.$$

How does this larger estimate for capital's share of income affect the analysis?

The calibration procedure described by (25) is redone for the case where $\alpha = 0.41$. The results are in accord with those obtained earlier. The model again fits the U.S. data well. In particular, it matches the firm-size distribution for 1972 very well. With no financial innovation, GDP would have risen by about 1.77 percent a year, compared with its actual rise of 2.4 percent. Hence, financial development accounts for about one quarter of the growth in GDP. About 18 percent of measured TFP growth is due to improvements in financial intermediation.

Financial intermediation is now a little more important for economic development, at least when the model is used to match up GDP and capital/output ratios across countries. World output would increase by 33 percent, as opposed to the 28 percent found earlier, if all countries moved to the best financial practice. When interest-rate spreads are targeted instead of capital/output ratios, financial development is slightly less important than before–world output would rise by 47 percent. Therefore, all in all, the results obtained earlier are quite robust to a change in capital's share of income (when the model is suitably recalibrated).

12.2 Varying the Degree of Substitutability between Capital and Labor

Let output be produced according to a CES production function of the form

$$o = [\alpha k^{\lambda} + (1-\alpha)(x\theta l)^{\lambda}]^{\frac{1}{\lambda}}, \text{ with } \lambda \leq 1.$$

This production function will have have implications for how labor's share of income, LSI, will vary across countries. To see this, think about the one sector growth model. Here labor's share of income can be written as LSI$= (w/l)/(w/l + rk) = 1/[1 + (r/w)(k/l)]$. Therefore, labor's share will rise whenever $(r/w)(k/l)$ falls. With the above production function, $\xi = 1/(1-\lambda)$ represents the elasticity of substitution between capital and labor. Hence, in response to a shock in some exogenous variable, z, it will happen that $d\ln(r/w)/dz = -(1/\xi)d\ln(k/l)/dz$. If the shock induces capital deepening $[d\ln(k/l)/dz > 0]$ then labor's share will rise or fall depending on whether the elasticity of substitution is smaller or bigger than one. In the cross-country data, labor's share either rises slightly or remains constant with per-capita income.[13] This suggests that for the quantitative analysis λ should be restricted so that $1/(1-\lambda) < 1$, which implies $\lambda < 0$; i.e., capital and labor are less substitutable than Cobb-Douglas.

Let $\lambda = -0.38$, roughly in line with Pessoa, Pessoa and Robb (2005). The calibration procedure described above is redone for this value for λ. The CES framework does not fit the firm-size distribution for 1972 as well as the Cobb-Douglas case. It does worse predicting the shift in the 2000 distribution, although the movement is still in the right direction. In

[13] That is, r/w will decrease by more (less) than k/l rises when the elasticity of substitution is smaller (greater) than one.

fact, if one allowed for $\lambda \leq 0$ to be freely chosen in the calibration procedure then a value close to zero (Cobb-Douglas) would be picked. For the U.S. economy, the CES specification predicts a rise in labor's share from 0.73 to 0.75 as the capital stock deepens. The model with a CES production function has a difficult time matching the observed variation in cross-country capital/output ratios. Labor's share varies from 0.70 to 0.75. All in all, both the U.S. and cross-country data prefer the Cobb-Douglas specification. With a CES production structure world output would increase by 18 percent, if all countries move to the best financial practice. This is lower than the Cobb-Douglas case. This occurs because the potential for capital deepening is more limited the higher the degree of complementarity between capital, which is reproducible, and labor, which is fixed, in production.

13 Conclusions

What is the link between the state of financial intermediation and economic development? This question is explored here by embedding a costly-state verification framework into the standard neoclassical growth paradigm. The model has two novel ingredients. As in the standard costly-state verification paradigm, the ex post return on a project is private information and an intermediary can audit the reported return. The first ingredient is that likelihood of a successful audit is increasing and concave in the amount of resources devoted to monitoring. The cost of auditing is increasing and convex in the amount of resources spent on this activity. Second, there is a distribution over firm type, each type offering a different combination of risk and return.

Two key features follow from these ingredients. First, a financial theory of firm size results. All firms are funded that earn an expected return greater than the cost of raising capital from savers. Funding is increasing in a project's expected return and decreasing in its variance. The size of a firm is limited by diminishing returns in information production.

Second, a Goldsmithian (1969) link is created between the state of financial development and economic development. The presence of informational frictions leads to a distortion between the expected marginal product of capital and its user cost, the interest paid to savers plus capital consumption. This distortion is modelled endogenously here. As the efficacy of auditing increases, due to technological progress in the financial sector, the size of this distortion shrinks. The upshot is an increase in the economy's income. Intuitively, the rise in income derives from three effects: (a) as the spread shrinks there is more overall capital accumulation in the economy; (b) capital is redirected toward the most productive investment opportunities in the economy; (c) less labor is required to monitor loans, which frees up resources for the economy.

The developed model is taken to both U.S. and cross-country data. It is calibrated to fit the U.S. firm size distribution for 1972 and the rise in the U.S. capital-to-output ratio between 1972 and 2000. It captures these features of the data well. The model's predictions for the efficiency of financial intermediation in a cross-section of 40 countries matches up well with independent measures. It does a reasonable job mimicking cross-country capital-output ratios and interest-rate spreads. The mechanism outlined above has quantitative significance. The average measured distortion in the world between the expected marginal product of capital and its user cost falls somewhere between 17 and 21 percentage points. The average coefficient of variation in the distortion within a country is 28 to 29 percent. World output could increase by 28 to 48 percent if all countries adopted the best financial practice in the world. Still, this only accounts for 13 to 19 percent of the gap between actual and potential world output. This happens because the bulk of the differences in cross-country GDP are explained by the huge differences in the productivity of the non-financial sector.

There are two natural extensions of the above framework. The first would be to allow for long-term contracts. On this, Smith and Wang (2006) embed a long-term contracting framework into a model of financial intermediation. Clementi and Hopenhayn (2006) and Quadrini (2004) have also examined the properties of dynamic contracting for firm finance in worlds with private information. The use of dynamic contracts could mitigate the informational problems discussed here. How much is an open question. In a competitive world, such contracts may be severely limited by the ability of each party to leave the relationship at any point in time and seek a better partner. Second, firms often use internal funds to finance investment. Incorporating internal finance may affect the importance of financial development. Perhaps accessing external funds involves a fixed cost, and hence is economized on. Then, internal funds may be hoarded in case good investment opportunities come along. An extension along these lines calls for a dynamic theory of the firm–Cooley and Quadrini (2001) or Gomes (2001). In situations where the financial stakeholders of firms delegate its operation to others, however, the returns on any investment activity (whether internally or externally financed) will need to be monitored. Ideally, any predictions from such a model should be matched up with a cross-country data set on the sources of firm finance.

References

Amaral, Pedro, S. and Erwan Quintin. "Financial Intermediation and Economic Development: A Quantitative Assessment." Mimeo, Department of Economics, Southern Methodist University, 2005.

Beck, Thorsten, Demirgürç-Kunt, Asli and Ross Levine. "A New Database on Financial Development and Structure." *World Bank Economic Review*, 14, 3 (September 2000): 597-605.

Beck, Thorsten, Demirgürç-Kunt Asli and Ross Levine. "The financial structure database." *In*: Asli Demirgürç-Kunt and Ross Levine (eds) *Financial Structure and Economic Growth: A Cross-Country Comparison of Banks Markets, and Development.* Cambridge: MIT Press 2001: 17-80.

Bencivenga, Valerie R. and Bruce D. Smith. "Financial Intermediation and Endogenous Growth." *Review of Economic Studies*, 58, 2 (April 1991): 195-209.

Berger, Allen N. "The Economic Effects of Technological Progress: Evidence from the Banking Industry." *Journal of Money, Credit, and Banking*, 35, 2 (April 2003): 141-176.

Bernanke, Ben and Mark Gertler. "Agency Costs, Net Worth and Business Fluctuations." *American Economic Review*, 79, 1 (March 1989): 14-31.

Boyd, John H. and Bruce D. Smith. "How Good are Standard Debt Contracts: Stochastic versus Nonstochastic Monitoring in a Costly State Verification Environment." *Journal of Business*, 67, 4 (October 1994): 539-561.

Boyd, John H. and Edward C. Prescott. "Financial Intermediary-Coalitions." *Journal of Economic Theory*, 38, 2 (April 1986): 211-32.

Castro, Rui, Clementi, Gian Luca and Glenn M. MacDonald. 2009. “Legal Institutions, Sectoral Heterogeneity, and Economic Development,” *Review of Economics Studies*, 76(2): 529-561.

Chakraborty, Shankha and Amartya Lahiri. “Costly Intermediation and the Poverty of Nations.” *International Economic Review*, 48, 1 (February 2007): 155-183.

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

Cooley, Thomas F. and Edward C. Prescott. “Economic Growth and Business Cycles.” *In*: Thomas F. Cooley (ed) *Frontiers of Business Cycle Research*. New Jersey: Princeton University Press, 1995: 1-38.

Cooley, Thomas F. and Vincenzo Quadrini. “Financial Markets and Firm Dynamics.” *American Economic Review*, 91, 5 (December 2001): 1286-1310.

Corrado, Carol A., Hulten, Charles R. and Daniel E. Sichel. “Intangible Capital and Economic Growth.” National Bureau of Economic Research (Cambridge, MA) Working Paper No. 11948, 2006.

Diamond, Douglas W. “Financial Intermediation and Delegated Monitoring.” *The Review of Economic Studies*, 51, 3. (July 1984): 393-414.

Erosa, Andres. “Financial Intermediation and Occupational Choice in Development.” *Review of Economic Dynamics*, 4, 2 (April 2001): 303-334.

Erosa, Andres and Ana Hidalgo Cabrillana. “On Capital Market Imperfections as a Source of Low TFP and Economic Rents.” *International Economic Review*, 49, 2 (May 2008): 437-473.

Foster, Lucia, Haltiwanger, John and Chad Syverson. “Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?” National Bureau of Economic Research (Cambridge, MA) Working Paper No. 11555, 2005.

de la Fuente, Angel and Jose Maria Marin. “Innovation, Bank Monitoring and Endogenous Financial Development." *Journal of Monetary Economics*, 38, 2 (October 1996): 269-301.

Goldsmith, Raymond W. *Financial Structure and Development*. New Haven: Yale University Press, 1969.

Gomes, Joao. “Financing Investment.” *American Economic Review*, 91, 5 (December 2001): 1261-1285.

Greenwood, Jeremy and Boyan Jovanovic. "Financial Development, Growth, and the Distribution of Income." *Journal of Political Economy*, 98, 5 (Part 1, October 1990): 1076-1107.

Harrison, Paul, Sussman, Oren and Joseph Zeira. "Finance and Growth: Theory and New Evidence." Mimeo, Board of Governors of the Federal Reserve, 1999.

Heston, Alan, Summers, Robert and Bettina Aten, Penn World Table, Version 6.1, Center for International Comparisons at the University of Pennsylvania (CICUP), October 2002.

Hobijn, Bart and Boyan Jovanovic. "The Information-Technology Revolution and the Stock Market: Evidence." *American Economic Review*, 91, 5 (December 2001): 1203-1220.

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

Hsieh, Chang-Tai and Peter J. Klenow. "Misallocation and Manufacturing TFP in China and India." Mimeo, Department of Economics, Stanford University, 2008.

Khan, Aubhik. "Financial Development and Economic Growth," *Macroeconomic Dynamics*, 5, 4 (June 2001): 413-433.

Levine, Ross. "Stock Markets, Growth, and Tax Policy." *Journal of Finance*, 46, 4 (September 1991): 1445-65.

Levine, Ross. "Finance and Growth: Theory and Evidence." *In*: Philippe Aghion and Steven N. Durlauf (eds) *Handbook of Economic Growth*, Volume 1A. Amsterdam: Elsevier B.V., 2005: 865-934.

Li, Wenli and Pierre-Daniel G. Sarte. "Credit market frictions and their Direct Effects on U.S. Manufacturing Fluctuations." *Journal of Economic Dynamics and Control*, 28, 3 (December 2003): 419-443.

Lucas, Robert J., Jr. "On the Size Distribution of Business Firms," *Bell Journal of Economics*, (1978): 508-23.

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

McGrattan, Ellen R. and Edward C. Prescott. "Taxes, Regulations, and the Value of U.S. and U.K. Corporations." *Review of Economic Studies*, 72, 3 (July 2005): 767–796.

Parente, Stephen L. and Edward C. Prescott. *Barriers to Riches*. Cambridge, MA: The MIT Press, 2000.

Pessoa, Samuel de Abreu, Pessoa, Silvia Matos and Rafael Rob. "Elasticity of Substitution Between Capital and Labor and its applications to growth and Development." Pier Working Paper 05-012, 2005.

Quadrini, Vincenzo. "Investment and Liquidation in Renegotiation-Proof Contracts with Moral Hazard." *Journal of Monetary Economics*, 51, 4 (May 2004): 713-751.

Restuccia, Diego and Richard Rogerson. "Policy Distortions and Aggregate Productivity with Heterogeneous Plants." *Review of Economic Dynamics*, 11, 4 (October 2008): 707-720.

Rousseau, Peter L. "The Permanent Effects of Innovation on Financial Depth: Theory and U.S. Historical Evidence from Unobservable Components Models." *Journal of Monetary Economics*, 42, 2 (October 1998): 387-425.

Smith, Anthony A. and Cheng Wang. "Dynamic Credit Relationships in General Equilibrium." *Journal of Monetary Economics* 53, 4 (May 2006): 847-877.

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

Townsend, Robert M. and Kenichi Ueda. "Financial Deepening, Inequality, and Growth: A Model-Based Quantitative Evaluation." *Review of Economic Studies*, 73, 1 (January 2006): 251–293.

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

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

Williamson, Stephen D. "Financial Intermediation, Business Failures, and Real Business Cycles." *Journal of Political Economy*, 95, 6 (December 1987): 1196-1216.

14 Appendix

14.1 Theory

Proof for Lemma 1. First, substitute the promise-keeping constraint (9) into the objective function to rewrite it as

$$(\pi_1 r_1 + \pi_2 r_2 - \widetilde{r})k - \pi_1 w(m_1/z)^\gamma - \pi_2 w(m_2/z)^\gamma - v.$$

Next, it is almost trivial to see that optimality will dictate that $p_{12} = r_1 k$ and $p_{21} = r_2 k$, since this costlessly relaxes the incentive constraints (7) and (8). Next, drop the incentive constraint (7) from problem (P2) to obtain the auxiliary problem now displayed:

$$\widetilde{I}(\tau, v) \equiv \max_{p_1, p_2, m_1, k} \{(\pi_1 r_1 + \pi_2 r_2 - \widetilde{r})k - \pi_1 w(m_1/z)^\gamma - v\}, \tag{P4}$$

subject to

$$p_1 \leq r_1 k, \tag{27}$$

$$p_2 \leq r_2 k, \tag{28}$$

$$[1 - P_{21}(m_1/k)](r_2 k - p_1) \leq r_2 k - p_2, \tag{29}$$

and

$$\pi_1(r_1 k - p_1) + \pi_2(r_2 k - p_2) = v. \tag{30}$$

The strategy will be to solve problem (P4) first. Then, it will be shown that (P2) and (P4) are equivalent. Problem (P4) will now be solved. To this end, note the following points:

1. The incentive constraint (29) is binding. To see why, suppose not. Then, reduce m_1 to increase the objective.

2. The constraint (28) is not binding. Assume, to the contrary, it is. Then, (29) is violated. This happens because the right-hand side is zero. Yet, the left-hand side is positive, given that $p_1 \leq r_1 k < r_2 k$, so that $[1 - P_{21}(m_1/k)](r_2 k - p_1) > 0$.

3. The constraint (27) is binding. Again, suppose not, so that $p_1 < r_1 k$. It will be shown that exists a profitable feasible deviation from any contract where this constraint is slack. Specifically, consider increasing k very slightly by $dk > 0$ while adjusting p_1 and p_2 in the following manner so that (29) and (30) still hold. Also, hold m_1 fixed. The implied perturbations for p_1 and p_2 are given by

$$\begin{bmatrix} dp_1 \\ dp_2 \end{bmatrix} = \begin{bmatrix} -[1 - P_{21}(m_1/k)] & 1 \\ \pi_1 & \pi_2 \end{bmatrix}^{-1} \times \begin{bmatrix} r_2 P_{21}(m_1/k) - (r_2 k - p_1)(m_1/k^2) dP_{21}/d(m_1/k) \\ \pi_1 r_1 + \pi_2 r_2 \end{bmatrix} dk.$$

Note that such an increase in k will raise the objective function.

It will now be demonstrated that the optimization problems (P2) and (P4) are equivalent. First, note that $\widetilde{I}(\tau, v) \geq I(\tau, v)$, because problem (P4) does not impose the constraint (7). It will now be established that $\widetilde{I}(\tau, v) \leq I(\tau, v)$. Consider a solution to problem (P4). It will be shown that this solution is feasible for (P2). On this, note that point 1 implies that

$$r_2k - p_2 = [1 - P_{21}(m_1/k)](r_2k - p_1) \leq r_2k - p_1,$$

so that

$$p_1 \leq p_2.$$

Now, set $m_2 = 0$ in (P2), which is feasible but not necessarily optimal. Then, constraint (7) becomes $p_1 \leq p_2$, which is satisfied by the solution to (P4). Therefore, $\widetilde{I}(\tau, v) \leq I(\tau, v)$.

Last, with the above facts in hand, recast the optimization problem as

$$I(\tau, v) \equiv \max_{p_2, m_1, k} \{(\pi_1 r_1 + \pi_2 r_2 - \widetilde{r})k - \pi_1 w(m_1/z)^\gamma - v\},$$

subject to

$$r_2k - p_2 = (r_2 - r_1)k[1 - P_{21}(m_1/k)], \quad \text{cf (29)},$$

and

$$r_2k - p_2 = v/\pi_2, \text{ cf (30)}.$$

The above two constraints collapse in the single constraint (10), by eliminating $r_2k - p_2$, that involves just m_1 and k. The problem then appears as (P3). ■

Proof for Lemma 2. Suppose that the solution dictates that $m_1/k \leq 1/\epsilon$. Then, from (P3) it is clear that the optimal solution will dictate that $m_1/k = 0$. This happens because $P_{21}(m_1/k) = 0$ for all $m_1/k \leq 1/\epsilon$, yet monitoring costs are positive for all $m_1/k > 0$. Next, by substituting (10) into (P3) it is easy to deduce that the intermediary's profit function can be written as

$$\begin{aligned} I(\tau, v) &= [\frac{\pi_1 r_1 + \pi_2 r_2 - \widetilde{r}}{\pi_2(r_2 - r_1)} - 1]v = \frac{r_1 - \widetilde{r}}{\pi_2(r_2 - r_1)}v \\ &\lesseqqgtr 0 \text{ as } r_1 \lesseqqgtr \widetilde{r}. \end{aligned} \tag{31}$$

Therefore profits are negative if $r_1 < \widetilde{r}$ and $v > 0$. Hence, a contract will not be offered when $m_1/k \leq 1/\epsilon$. ■

Proof for Lemma 3. Clear from equation (17).■

Proof for Lemma 4. *Necessity*: From problem (P3) it is clear that the intermediary will incur a loss when $\pi_1 r_1 + \pi_2 r_2 - \widetilde{r} \leq 0$ and $v > 0$, for any $\tau \in \mathcal{A}(w)$, because $m_1/k > 1/\epsilon$ by Lemma 2.

Sufficiency: Suppose that $\pi_1 r_1 + \pi_2 r_2 - \widetilde{r} > 0$ for some $\tau \in \mathcal{T}$. By equation (17) it is immediate that the intermediary will issue a loan $k > 0$. By construction it will earn zero profits on this loan. Recall that the derivation of (17), discussed in the text, used a solution for v. This solution is

$$v = V(\tau) = (\psi\gamma + \gamma - \psi)^{(\psi\gamma+\gamma-\psi)/(\psi\gamma-\psi)}(\frac{1}{\psi})^{-\psi/(\psi\gamma-\psi)}(\frac{1}{\psi\gamma+\gamma})^{(\psi\gamma+\gamma)/(\psi\gamma-\psi)} \times (\pi_1 r_1 + \pi_2 r_2 - \widetilde{r})^{\frac{\gamma\psi+\gamma}{\gamma\psi-\psi}}[\pi_2(r_2 - r_1)]^{\frac{-\gamma}{\gamma\psi-\psi}}(\pi_1 w)^{\frac{-\psi\gamma}{\gamma\psi-\psi}}(z/\epsilon)^{\frac{\psi\gamma}{\gamma\psi-\psi}}. \tag{32}$$

Therefore, the firm will earn positive rents too. ■

Proof for Lemma 5. To begin with let $k = K(w;\tau,z)$ represent capital stock that will be employed by a type-τ firm when productivity in the financial sector is z and the wage rate is $w > \omega$. Similarly, let $m_1/z = M(w;\tau,z)$ denote the amount of monitoring services, relative to z, that will be devoted to this project. Given this notation, the expected demand for labor by both the firm and intermediary for a project of type τ is

$$L(w;\tau,z) \equiv (\pi_1\theta_1^{1/\alpha} + \pi_2\theta_2^{1/\alpha})[\frac{(1-\alpha)}{w}]^{\frac{1}{\alpha}}K(w;\tau,z) + \pi_1 M(w;\tau,z)^{\gamma}. \tag{33}$$

Note that this demand is only specified for $\tau \in \mathcal{A}(w)$ and $w > \omega$, where ω is defined by (22). In order to characterize $L(w;\tau,z)$, the properties of its components $K(w;\tau,z)$ and $M(w;\tau,z)$ must be developed when $\tau \in \mathcal{A}(w)$ and $w > \omega$. First, take $K(w;\tau,z)$. On this, rewrite equation (17) as

$$\begin{aligned} k &= K(w;\tau,z) \\ &= \Upsilon(\pi_1\Theta_1 + \pi_2\Theta_2 - \widetilde{r}w^{(1-\alpha)/\alpha})^{(\gamma+\psi)/(\psi\gamma-\psi)}[\frac{\epsilon^{\psi}}{\pi_2(\Theta_2-\Theta_1)}]^{\gamma/(\psi\gamma-\psi)}z^{\gamma/(\gamma-1)}w^{-1/[\alpha(\gamma-1)]}, \end{aligned} \tag{34}$$

where $\Theta_i \equiv \alpha(1-\alpha)^{(1-\alpha)/\alpha}\theta_i^{1/\alpha}$ and $\Upsilon \equiv (\psi\gamma+\gamma-\psi)^{\gamma/(\psi\gamma-\psi)}(\frac{1}{\psi})^{-\psi/(\psi\gamma-\psi)}(\frac{1}{\psi\gamma+\gamma})^{(\gamma+\psi)/(\psi\gamma-\psi)}$. Note the following things about this solution for k: (i) The level of investment in a firm, k, is continuous and strictly decreasing in w; (ii) $k \to 0$ as $w \to \overline{W}(\tau) = [(\pi_1\Theta_1+\pi_2\Theta_2)/\widetilde{r}]^{\alpha/(1-\alpha)}$ (when z is finite); (iii) k is continuous and strictly increasing in z.

Now, switch attention to the second term in $L(w;\tau,z)$. A formula for $M(w;\tau,z)$ can be derived in the same manner as the one for $K(w;\tau,z)$. It is

$$m_1/z = M(w;\tau,z) = \Delta(\pi_1\Theta_1+\pi_2\Theta_2-\widetilde{r}w^{(1-\alpha)/\alpha})^{(\psi+1)/(\psi\gamma-\psi)}w^{-1/[\alpha(\gamma-1)]}z^{1/(\gamma-1)}, \tag{35}$$

where $\Delta \equiv \Upsilon[(\psi\gamma+\gamma-\psi)/(\psi\gamma+\gamma)]^{-1/\psi}\{\epsilon^{\psi}/[\pi_2(\Theta_2-\Theta_1)]\}^{1/(\psi\gamma-\psi)}$. Note the following things about this solution for m_1/z: (i) m_1/z is continuous and strictly decreasing in w; (ii) $m_1/z \to 0$ as $w \to \overline{W}(\tau) \equiv [(\pi_1\Theta_1+\pi_2\Theta_2)/\widetilde{r}]^{\alpha/(1-\alpha)}$ (when z is finite); (iii) m_1/z is continuous and strictly increasing in z.

Thus, the demand for labor by a type-τ project has the following properties: (i) $L(w;\tau,z)$ is continuous and strictly decreasing in w; (ii) $\lim_{w\to\overline{W}(\tau)} L(w;\tau,z) = 0$; (iii) $L(w;\tau,z)$ is continuous and strictly increasing in z. Define the function

$$\widetilde{L}(w;\tau,z) = \begin{cases} L(w;\tau,z), & \text{for } w \le \overline{W}(\tau), \\ 0, & \text{for } w > \overline{W}(\tau), \end{cases} \tag{36}$$

for $w > \omega$. The aggregate demand for labor can be expressed as $\int_{\mathcal{A}(w)} L(w;\tau,z)dF(\tau) = \int_{\mathcal{T}} \widetilde{L}(w;\tau,z)dF(\tau)$. Now, determine the constant ζ by $\lim_{w\downarrow\omega}\int_{\mathcal{T}} \widetilde{L}(w;\tau,\zeta)dF(\tau) = 1$,where again the lower bound on wages ω is given by (22). From the simple closed-form solutions for (34) and (35) it is easy to deduce that such a ζ must exist. To summarize, the aggregate demand for labor, $\int_{\mathcal{T}} \widetilde{L}(w;\tau,z)dF(\tau)$, has the following properties:

1. $\int_{\mathcal{T}} \widetilde{L}(w;\tau,z)dF(\tau)$ is continuous and strictly decreasing in w for $w \in (\omega,\infty)$;

2. $\int_{\mathcal{T}} \widetilde{L}(w;\tau,z)dF(\tau)$ is continuous and strictly increasing in z for $z \in (\zeta,\infty)$;

3. $\int_{\mathcal{T}} \widetilde{L}(\widehat{w};\tau,z)dF(\tau) = 0$, where $\widehat{w} = \max_{\tau\in\mathcal{T}} \overline{W}(\tau)$;

4. $\int_{\mathcal{T}} \widetilde{L}(\omega;\tau,\zeta)dF(\tau) = 1$.

Therefore by the intermediate theorem for all $z > \zeta$ there will exist a single value of w that sets labor demand equal to labor supply (or 1)—see Figure 6. ■

Proof for Proposition 1. Express the labor-market-clearing condition as

$$\int_{\mathcal{A}(w)} \{(\pi_1 \theta_1^{1/\alpha} + \pi_2 \theta_2^{1/\alpha})[\frac{(1-\alpha)}{w}]^{1/\alpha} K(w;\tau,z) + \pi_1 M(w;\tau,z)^{\gamma}\} dF = 1. \tag{37}$$

Let the $\theta_i^{1/\alpha}$'s grow at the common rate $g^{1/\alpha}$ and z grow at rate $g^{1/(1-\alpha)}$. Recall that exists a solution to the model without growth, as demonstrated by Lemma 5. It is easy to construct a balanced growth path using this solution. The solution implies that there will be a wage rate that solves (37). Conjecture that along a balanced growth path wages, w, will grow at rate $g^{1/(1-\alpha)}$. From (34) it can be deduced that $K(w;\tau,z)$ will grow at rate $g^{1/(1-\alpha)}$. Therefore, the first term in braces in (37) will be constant. Equation (35) implies that $M(w;\tau,z)$, or the second term, will be constant too. The active set $\mathcal{A}(w)$ will not change—equation (20). Therefore, labor demand remains constant. Hence, the conjectured solution for the rate of growth in wages is true. Using (32) is easy to calculate that v will grow at rate $g^{1/(1-\alpha)}$. Last, since $M(w;\tau,z)$ is constant it must be the case that m_1 is growing at the same rate as z, or $g^{1/(1-\alpha)}$. Therefore, m_1/k will remain unchanged along a balanced growth path. ■

Proof for Proposition 2. First, point 2 in the proof of Lemma 5 established that the aggregate demand for labor is continuous and strictly increasing in z. Therefore, at a given wage rate the demand for labor rises as z moves up. In order for equilibrium in the labor market to be restored, wages must increase, since the demand for labor is decreasing in wages—point 1. Last, recall from (19) that a type-τ project will only be funded when $w < \overline{W}(\tau) = \alpha^{\alpha/(1-\alpha)}(1-\alpha)[(\pi_1\theta_1^{1/\alpha} + \pi_2\theta_2^{1/\alpha})/\widetilde{r}]^{\alpha/(1-\alpha)}$. It's trivial to see that as w rises the set of $\tau \in \mathcal{T}$ satisfying this restriction, or $\mathcal{A}(w)$, shrinks; if $\tau = (\theta_1, \theta_2)$ fulfills the restriction for some wage it will meet it for all lower ones too, yet there will exist a higher wage that will not satisfy it. Furthermore, observe that $\overline{W}(\tau)$ is strictly increasing in $\pi_1\theta_1^{1/\alpha} + \pi_2\theta_2^{1/\alpha}$. Therefore, those τ's offering the lowest expected return will be cut first as w rises, because they have the lowest threshold wage. ■

Proof for Proposition 3. Let the $\theta_i^{1/\alpha}$'s increase by the common factor $g^{1/\alpha} > 1$. Suppose that wages increase in response by the proportion $g^{1/(1-\alpha)}$. Will the labor-market-clearing condition (37) still hold? The answer is no, because the demand for labor will fall. Take the first term behind the integral, which gives the demand for labor by a firm. From (34) it is clear that $K(w;\tau,z)$ will rise by a factor less than $g^{1/(1-\alpha)}$, when z is held fixed. Therefore, the first term in braces in (37) will decline. Turn to the second term. From (35) it is easy to see that $M(w;\tau,z)$ will drop under the conjecture solution. Therefore, wages must rise by less than $g^{1/(1-\alpha)}$, since the demand for labor is decreasing in w (as was established in the proof of Lemma 5). The active set, $\mathcal{A}(w)$, will therefore expand, because $\pi_1 r_1 + \pi_2 r_2$ increases–see (18). ■

Proof for Proposition 4. The set of projects in $\mathcal{T}$ offering the highest expected return is given by

$$\mathcal{A}^* = \underset{\tau=(\theta_1,\theta_2)\in\mathcal{T}}{\arg\max}[\,\pi_1(\theta_1)^{1/\alpha} + \pi_2(\theta_2)^{1/\alpha}].$$

By assumption $\int_{\mathcal{A}^*} dF > 0$. Take *any* equilibrium wage w. From (20) it is immediate that if $\tau \in \mathcal{A}^*$ then $\tau \in \mathcal{A}(w)$, since $\pi_1 r_1 + \pi_2 r_2 - \widetilde{r} = \alpha(1-\alpha)^{(1-\alpha)/\alpha} w^{-(1-\alpha)/\alpha}(\pi_1\theta_2^{1/\alpha} + \pi_2\theta_2^{1/\alpha}) - (\widehat{r}+\delta)$ is increasing in $\pi_1\theta_1^{1/\alpha} + \pi_2\theta_2^{1/\alpha}$. Hence, $\mathcal{A}^* \subseteq \mathcal{A}(w)$ for all w. In equilibrium the wage will be a function of z, so denote this dependence by $w = W(z)$. Now, let $z \to \infty$. It will be shown that $w = W(z) \to w^*$, where

$$w^* \equiv \alpha^{\alpha/(1-\alpha)}(1-\alpha)\{\max_{\tau=(\theta_1,\theta_2)\in\mathcal{T}}[\,\pi_1(\theta_1)^{1/\alpha} + \pi_2(\theta_2)^{1/\alpha}]/\widetilde{r}\}^{\alpha/(1-\alpha)}. \tag{38}$$

To see why, suppose alternatively that $w \to \widetilde{w} \neq w^*$. First, presume that $\widetilde{w} < w^*$. Then, by (19) all projects of type $\tau \in \mathcal{A}^*$ will be funded since their cutoff wage is $\overline{W}(\tau) = w^* > \widetilde{w}$. From equations (33), (34) and (36) it is clear that $\lim_{z\to\infty} \widetilde{L}^d(W(z);\tau,z) = \infty$, for $\tau \in \mathcal{A}^*$. Since, $\int_{\mathcal{A}^*} dF > 0$, this implies that $\lim_{z\to\infty}\int_{\mathcal{T}} \widetilde{L}^d(W(z);\tau,z)dF = \infty$. Therefore, such an equilibrium cannot exist because the demand for labor will exceed its supply. Second, no firm can survive at a wage rate bigger than w^*, by (19). Here, $\lim_{z\to\infty}\int_{\mathcal{T}} \widetilde{L}^d(W(z);\tau,z)dF = 0$. This establishes (23). Last, note that $\mathcal{A}(w^*) = \mathcal{A}^*$.

It is immediate that $\mathcal{A}^* \subseteq \lim_{w \uparrow w^*} \mathcal{A}(w)$, because $\tau \in \mathcal{A}^*$ is viable for *all* wages $w \leq w^* = \overline{W}(\tau)$ by (19). It is also true that $\lim_{w \uparrow w^*} \mathcal{A}(w) \subseteq \mathcal{A}^*$, since from (19) any project $\tau \notin \mathcal{A}^*$ requires an upper bound on wages $\overline{W}(\tau) < w^*$ to survive; that is, for any $\tau \notin \mathcal{A}^*$ there will exist some high enough wage w such that $\overline{W}(\tau) < w < w^*$. Therefore, $\lim_{w \uparrow w^*} \mathcal{A}(w) = \mathcal{A}^* = \mathcal{A}(w^*)$. This establishes point 1 of the Proposition.

To have an equilibrium it must be the case that $m_1/z < \infty$ for $\tau \in \mathcal{A}^*$, otherwise the demand for labor would be infinitely large. From equation (35) this can only happen when $\widetilde{r}w^{(1-\alpha)/\alpha} \to \pi_1\Theta_1 + \pi_2\Theta_2$, or equivalently when $\pi_1 r_1 + \pi_2 r_2 \to \widetilde{r}$. Solve problem (P3) for the optimal level of monitoring, m_1/k, and then use (32) to solve out for v to obtain

$$m_1/k = (\frac{\psi\gamma + \gamma - \psi}{\psi\gamma + \gamma})^{-1/\psi}[\frac{\epsilon^{\psi}(\pi_1 r_1 + \pi_2 r_2 - \widetilde{r})}{\pi_2(r_2 - r_1)}]^{-1/\psi}.$$

It is apparent that $\lim_{z\to\infty} m_1/k = \infty$; because $\pi_1 r_1 + \pi_2 r_2 \to \widetilde{r}$. Consequently, a false report by a firm will be caught with certainty, or $\lim_{z\to\infty} P_{12}(m_1/k) = 1$. The contracting problem (P2) then requires $\lim_{z\to\infty} p_2 = r_2 k$ and $\lim_{z\to\infty} v = 0$. A comparison of (32) and (35) leads to the conclusion that in fact $\lim_{z\to\infty} m_1/z = 0$, when $\lim_{z\to\infty} v = 0$. Using this result and (38), in conjunction with the labor-market-clearing condition, $\int_{\mathcal{T}} \widetilde{L}(w;\tau,z)dF = 1$, then gives (24). ■

14.2 Data

- *Figure 1*: (Left panel) The numbers represent total debt outstanding for businesses (excluding financial ones) relative to gross domestic business value added (excluding gross farm value added). This data derives from the Flow of Funds Accounts of the United States (Table D.3: Debt Outstanding by Sector). The source of the data for the value of firms relative to GDP is Hobijn and Jovanovic (2001, Figure 1, p. 1204). It refers to the total market capitalization of all securities contained in the CRSP data set.

  (Right Panel) A series for the intangible stock of capital is constructed by backing out the implied data series on investment in intangibles that is reported in Corrado, Hulten and Sichel (2006, Figure 1). Specifically, a capital stock series for intangibles is constructed by iterating on the law of motion $k_i' = (1-\delta_i)k_i + i_i$, where k_i is the current stock of intangible capital, i_i is investment in intangibles, and δ_i is the depreciation rate on intangibles. Two issues arise with this procedure. First, what is the depreciation rate on intangible capital? A weighted average of the rates reported in Corrado, Hulten and Sichel (2006, p. 23) suggests that it should be 33 percent. McGrattan and Prescott (2006, p. 782) feel that an upper bound of 11 percent is appropriate. Taking a simple average of these two numbers gives 22 percent, the value used here. Second, what starting value for the intangible stock of capital should be used? Along a balanced growth path the stock of intangible capital is given by $k_i = i_i/(g + \delta_i)$, where g is the growth rate of GDP. This formula is used to construct an initial capital stock for 1947, where g is assigned a value of 0.015. The stock of intangible capital is then simply added to private nonresidential nonfarm fixed assets. The resulting series is divided through by nonfarm business GDP.

- *Figure 7*: The data is for establishments. The horizontal axis orders establishments (from the smallest to highest) by the percentile that they lie in for employment. The vertical axis shows the cumulative contribution of this size of establishment to the total payroll in the U.S. economy. The use of payroll data controls for worker heterogeneity; i.e., measures employment in efficiency units. Note that firms without paid employees (sole-ownership firms and joint proprietorships) are included in 1972 data. The source for the raw data used is Statistics for U.S. Businesses, U.S. Census Bureau.

- *Figure 2* and *Section 10:* The data for the interest-rate spread is taken from Beck, Demirguc-Kunt and Levine (2001). It is defined as the accounting value of banks' net interest as a share of their interest-bearing (total earning) assets averaged over 1990 to 1995. The numbers for the ratio of private credit to GDP (the measure used for financial intermediation) are reported in Beck, Demirguc-Kunt, and Levine (2000) as the sum of deposit money in banks and related institutions, stock market capitalization, and private bond market capitalization. These are averaged over the period 1990 to 1999. The other numbers derive from the Penn World Tables, Version 6.1—see Heston, Summers and Aten (2002). The capital stock for a country, k, is computed for the 1990-2000 sample period using the formula $k = i/(g+\delta)$, where i is gross investment, g is the growth rate in investment, and δ is rate of depreciation. This formula heroically assumes that an economy is on a balanced growth path. The depreciation is taken to be 0.06. The growth rate for investment is calculated from the investment data reported in the tables for the period 1950 to 2000. Investment is recovered by using data on investment's share of GDP and GDP. A country's total factor productivity, TFP, was computed using the formula $TFP = (y/l)/(k/l)^{\alpha}$, where y is GDP, l is aggregate labor, and α is capital's share of income. A value of 0.30 was picked for α. Aggregate labor is backed out using data on per-capita GDP, GDP per worker, and population. The numbers in the analysis are reported in Table 7; the reported values are generally the mean for 1990-2000. Most of the headings in the table are obvious. Here 'Spread' refers to the Beck, Demirguc-Kunt and Levine (2001) interest-rate spread discussed above, while 'Fin. Dev.' represents the measure of financial development from Beck, Demirguc-Kunt and Levine (2000). Recall that $\mathbf{s}$ denotes the average spread in the model between the rates of return that intermediaries and savers. Likewise, $\mathbf{\Delta o}(x, z) \equiv [\mathbf{O}(x, \overline{z}) - \mathbf{O}(x, z)]$ is the increase in a country's income if it moved to financial best practice, while $\mathbf{G}(x, z) \equiv [\mathbf{O}(x, \overline{z}) - \mathbf{O}(x, z)]/[\mathbf{O}(\overline{x}, \overline{z}) - \mathbf{O}(x, z)]$ refers to the fraction of the gap between a country's actual and potential output that would be closed. Last, the mean (capital-weighted) distortion for a country is defined by $\mathbf{d}(x, z) \equiv \int_{\mathcal{A}(w)} (\pi_1 r_1 + \pi_2 r_2 - \widetilde{r}) k dF / \int_{\mathcal{A}(w)} dF$. The standard deviation of this distortion is then given by $\sqrt{\int_{\mathcal{A}(w)} (\pi_1 r_1 + \pi_2 r_2 - \widetilde{r} - \mathbf{d})^2 k dF / \int_{\mathcal{A}(w)} dF}$.

Table 7: Cross-country numbers, data and baseline model

(see Section 14.2 for an explanation of the data and headings)

	Data				**Model** $[\alpha = 0.33, (x, z) = M^{-1}(\mathbf{o}, \mathbf{k}/\mathbf{o})]$						
Country	GDP p.w.	K/GDP	Spread	Fin. Dev.	$\ln(z)$	s	x	$\mathbf{\Delta o}(x, z)$	$\mathbf{G}(x, z)$	$\mathbf{d}(x, z)$	Std Dev $\mathbf{d}(x, z)$
Sri Lanka	7013	0.774	0.051	0.339	7.60	0.054	38.77	0.756	0.100	0.293	0.093
India	5121	0.787	0.030	0.512	6.91	0.053	31.17	0.752	0.071	0.286	0.090
Bolivia	6779	0.839	0.035	0.441	7.60	0.048	36.54	0.703	0.087	0.260	0.080
Morocco	11419	0.884	0.036	0.576	8.29	0.044	50.59	0.656	0.141	0.240	0.073
Mauritius	23705	0.892	0.032	0.683	9.10	0.044	82.19	0.629	0.332	0.236	0.072
Nicaragua	5923	0.915		0.234	7.60	0.042	32.08	0.648	0.067	0.227	0.068
Colombia	12332	0.935	0.064	0.402	8.52	0.041	51.91	0.616	0.142	0.219	0.065
Philippines	7864	1.054	0.042	0.890	8.52	0.033	36.32	0.546	0.073	0.179	0.052
Costa Rica	13913	1.085	0.052	0.158	9.21	0.032	52.50	0.512	0.129	0.170	0.049
Uruguay	20251	1.099	0.056	0.266	9.68	0.031	67.11	0.493	0.196	0.167	0.047
Honduras	6823	1.120	0.069	0.282	8.70	0.030	32.09	0.507	0.057	0.161	0.045
Brazil	18001	1.170	0.120	0.538	9.80	0.027	60.20	0.453	0.151	0.148	0.041
Turkey	14340	1.179	0.094	0.336	9.62	0.027	51.50	0.454	0.115	0.146	0.041
Mexico	22100	1.257	0.053	0.556	10.34	0.024	66.72	0.397	0.168	0.129	0.035
Argentina	25056	1.291	0.082	0.335	10.62	0.023	71.64	0.375	0.186	0.122	0.033
Ireland	46945	1.355	0.016	0.977	11.51	0.020	106.6	0.320	0.452	0.110	0.029
UK	39908	1.416	0.020	2.481	11.63	0.018	93.40	0.293	0.293	0.099	0.026
Panama	15255	1.532	0.020	0.692	11.18	0.015	47.22	0.266	0.065	0.082	0.020
US	57151	1.578	0.039	3.297	12.72	0.014	112.7	0.205	0.447	0.075	0.019
Portugal	30350	1.596	0.035	1.046	12.18	0.014	73.35	0.217	0.132	0.073	0.018
Italy	50569	1.610	0.036	1.104	12.76	0.013	102.8	0.194	0.305	0.071	0.017
Spain	40138	1.639	0.038	1.279	12.68	0.013	87.29	0.189	0.182	0.068	0.016
Iceland	39834	1.639		0.940	12.67	0.013	86.80	0.190	0.179	0.068	0.016
Peru	10200	1.655	0.072	0.312	11.40	0.012	34.69	0.219	0.033	0.066	0.016
Finland	40603	1.695	0.016	1.551	12.99	0.011	86.50	0.165	0.159	0.061	0.014
Netherlands	46929	1.758	0.015	2.430	13.49	0.010	93.60	0.134	0.171	0.054	0.012
France	45317	1.758	0.035	1.592	13.45	0.010	91.43	0..135	0.160	0.054	0.012
Denmark	44024	1.759	0.049	1.741	13.43	0.010	89.65	0.135	0.151	0.054	0.012
Canada	45933	1.781	0.018	1.706	13.60	0.010	91.66	0.125	0.152	0.051	0.012
New Zealand	36422	1.794	0.025	1.363	13.45	0.009	78.18	0.127	0.099	0.050	0.011
Australia	45907	1.824	0.019	1.470	13.86	0.009	90.55	0.108	0.130	0.047	0.011
Belgium	50839	1.862	0.023	1.487	14.21	0.008	95.96	0.090	0.135	0.043	0.010
Luxembourg	80702	1.881	0.007	2.064	14.80	0.008	130.1	0.066	1.000	0.042	0.009
Austria	45560	1.884	0.019	1.143	14.24	0.008	88.65	0.085	0.099	0.041	0.009
Japan	37061	1.917	0.018	3.043	14.26	0.007	76.53	0.079	0.062	0.038	0.008
Israel	40777	1.933	0.033	0.979	14.48	0.007	81.26	0.070	0.065	0.037	0.008
Thailand	11632	1.995	0.030	1.786	13.69	0.006	34.52	0.080	0.013	0.033	0.007
Norway	47845	2.112	0.031	1.403	16.16	0.004	86.58	0.002	0.002	0.022	0.004
Switzerland	45706	2.122	0.016	3.464	16.22	0.004	83.77	0.000	0.000	0.022	0.004