Article

Financial Deepening

Arthaniti: Journal of Economic Theory and Practice
1–27

SAGE Publications
sagepub.in/home.nav
DOI: 10.1177/0976747918814031
http://journals.sagepub.com/home/ath

Shankha Chakraborty[1]

Abstract

This article proposes a tractable model of the evolution of financial structure. Firms invest out of internal assets and by borrowing from banks and the financial market. In the presence of moral hazard, whereby owner–managers may intentionally reduce profitability of investment to appropriate resources, banks can monitor firms and partially alleviate agency problems. Under the optimal financial contract, banks monitor and outside investors lend to firms only if they borrow from banks too. The model is broadly consistent with financial development facts. Capital accumulation is facilitated by an increasing reliance on both types of external finance. Initially firms rely more heavily on expensive bank finance. With further development, banks eliminate much of the agency problem and firms substitute in favour of cheaper market finance. The short- and long-run effects of financial sector reforms are considered.

Keywords

Financial structure, external finance, intermediation, banks, financial development

JEL: E44, G20, O16

I. Introduction

Financial deepening, as first articulated by Gurley and Shaw (1955, 1967), denotes a wide array of changes in financial structure accompanying economic development. These changes include loosening credit constraints, more intensive use of external finance, fewer distortions in the credit market and a general increase in financial activity. This article studies one aspect of this transformation: the

[1] Department of Economics, University of Oregon, Eugene, USA.

Corresponding author:
Shankha Chakraborty, Department of Economics, University of Oregon, Eugene, Oregon, OR 97 403 1285, USA.
E-mail: shankhac@uoregon.edu

increasing role of external finance and the concomitant rise of banks and securities markets.

Available cross-country evidence points to a positive correlation of financial sector activity with income levels: financial intermediaries tend to get larger (as measured by total assets or liabilities relative to GDP) as one moves from poorer to richer countries and similarly for markets in tradeable securities like equities and bonds (Demirgüç-Kunt & Levine, 1996). The evolution of bank-finance, however, differs from market-finance in an important respect. In countries where financial markets are not particularly developed, firms tend to rely more heavily on bank debt rather than equity and it is only in sufficiently developed markets that equity finance substitutes for bank finance (Demirgüç-Kunt & Maksimovic, 1996). Using more recent data, Demirgüç-Kunt, Feyen, and Levine (2013) find that while both banks and securities markets get larger relative to the size of the economy as countries get richer, 'the association between an increase in economic output and an increase in bank development becomes smaller' while that 'between an increase in economic output and an increase in securities market development becomes larger'.

The goal of this article is to formalise the following aspects of financial deepening: how banks and markets come into play as an economy increases its reliance on external finance and why, initially, economies rely more heavily on bank debt instead of securities like stocks and bonds and later, on securities. A by-product of our analysis is a tractable general equilibrium model that may be used to study financial sector reforms.

The model proposed further distinguishes between bank finance and market finance based on their information content. Bank finance comes with the lender's involvement (intermediated) while market finance is arm's-length lending (unintermediated). Banks possess a technology to monitor firms and align their incentives more in line with the investors'. Arm's-length investors (bond and equity holders) do not possess this technology or are too dispersed to effectively exercise it.

In the absence of bank finance and monitoring, direct/outside investors are unwilling to lend because they expect borrowing firms to grossly misappropriate (misallocate) funds. When a firm finances part of its investment through banks, the market expects the firm to be monitored and the agency problem to be contained; it becomes willing to lend. This delegated monitoring role played by banks (Diamond, 1984) effectively makes bank and market-finance complementary: market finance becomes available only if a firm simultaneously borrows from banks and, given its internal equity, a firm is able to invest more through increased market borrowing only if it also borrows more heavily from banks.

It is precisely this mechanism that drives the evolution of banks and markets in our model. Along the path of capital accumulation, as the demand for capital rises and firms invest greater amounts, they face tighter incentive constraints. As they can access cheaper market finance only by borrowing more from banks, the increasing reliance on external finance is accompanied by an expansion of both banks and financial markets. Banks, especially, assume a greater role by resolving incentive problems, monitoring firms more intensively and funding a rising proportion of all investment. As a result, bank finance may initially rise faster than market finance. With sufficient economic development, banks eliminate agency problems to the fullest extent possible. Thereafter, it is market finance that grows faster than bank finance.

With this framework in place, we analyse the effect of policy interventions in the financial sector. Since the Washington Consensus, there has been a push towards market-finance in developed as well as developing countries. Yet, even if one were to believe in the efficacy of markets, it is far from obvious that such a policy slant is helpful especially in poorer countries that lack adequate financial structure. Facing limited resources, should governments look for ways to reform their banking sector or their financial sector? As bank-finance enables financial markets to be active and develop in the model, not surprisingly, comparative dynamics results show that policies that improve the effectiveness of the banking sector (lower cost of bank finance) serve better in the long-run compared to those that lower the cost of market finance. Conversely an ill-functioning banking system delivers little information content to the financial market, leaving both stunted.

This article is not the first, in a relatively thin literature in macro-development theory, to study different sources of external finance. Most closely related are Bose and Neumayer (2015), Chakraborty and Ray (2006, 2007) and Boyd and Smith (1998). In the former, loan contracts are either debt issue or equity issue depending on firm type due to an adverse selection problem. The authors illustrate how the mix of financing changes with development and, interestingly, found non-unique financing outcomes in intermediate stages of development. The equilibrium in our model, in contrast, is unique and well defined, given initial conditions. Chakraborty and Ray (2006, 2007) use a similar incentive problem as the current article without the richness that varying monitoring intensity provides. While the first article studies the emergence of dichotomous bank-based or market-based financial systems, the second is interested in the effects of technology and income distribution on financial structure. Firms borrow using a mix of bank and market finance in Boyd and Smith's (1998) model of costly state verification. The gradual switch to market finance is, however, driven by a narrow assumption: verification costs are incurred in the form of final goods while firms produce capital. Hence as capital becomes relatively cheaper with development, bank finance gets more expensive and firms switch towards observable-return projects where they have less trouble raising equity finance.

This article is also related to works on financial deepening that models either bank or market borrowing. Financial intermediaries arise endogenously in Greenwood and Jovanovic (1990) with costly investment in organisational capital: capital deepening raises returns on investment via non-diminishing returns and enables economies to utilise more efficient and costlier financial structures. In Khan (2001), economic development raises collateral that firms offer which, then, reduces the cost of bank finance and enables firms to finance larger investment. In more recent work, Greenwood, Sanchez, and Wang (2010) augment the standard costly state verification model with stochastic monitoring outcomes and heterogeneous firm stochastic returns to show how financial intermediation becomes more efficient over time.

Among articles that study market finance, both Levine (1991) and Bencivenga, Smith and Starr (1995) emphasise the ability of equity markets to provide liquidity and promote investment that requires a longer-term commitment of capital but yield higher returns. The presence of fixed costs of market formation in Greenwood and Smith (1997) imply that economic growth loosens the ability of poorer countries to open markets and increase financial market activity that then feed into faster growth through specialisation.

The rest of the article proceeds as follows: We describe the model in the following section, consider returns on savings with and without monitoring in section III, and study financial intermediation and firm investment decisions in section IV. Section V analyses the general equilibrium. In section VI we discuss implications of the model before concluding in section VII.

II. Structure of the Economy

We analyse the evolving nature of external finance in a two-period overlapping-generations economy. The standard model is modified in two ways: (a) 'firms' borrow in the presence of agency problems and (b) internal funds play a role in resolving incentive problems.

Our economy comprises of three sectors: a capital goods sector, a final goods sector, and a financial sector. Following four types of agents participate in these sectors: (a) *households* supply labour to the final goods sector and invest their savings with banks and on the financial market, (b) capital goods producers, called *entrepreneurs*, borrow and produce capital that they rent out to the final goods sector, (c) *final goods producers* manufacture the unique consumption good, utilising labour and capital goods and (d) financial intermediaries, or *banks*, use household deposits as inputs to produce loans for entrepreneurs.

Final Goods Producers

The final good (numeraire) is produced using raw labour (h) and various types of intermediate capital goods (K^j) using a constant-returns technology:

$$Y_t = A_t h_t^{1-\alpha} \left[\int_{j=0}^{1} \left(K_t^j \right)^{\alpha} dK_t^j \right]. \tag{1}$$

Here $j \in [0, 1]$ indexes intermediate capital goods that are distributed uniformly on the unit interval, and A_t denotes aggregate productivity that, without loss of generality, is normalised $A_t = 1 \; \forall t$. In addition, because the aggregate labor supply is equal to one (see further), we set $h_t = 1 \; \forall t$.

Final goods producers face perfectly competitive factor and goods markets and the equilibrium prices of capital and labour equal their respective (value of) marginal products.

Households

A continuum of two-period lived overlapping generations of working households populate the unit interval. Each such household is endowed with 1 unit of labour time in youth which it supplies inelastically to the final goods sector. A generation-t worker household's lifetime utility depends only upon its old-age consumption:

$$U_t^H = u(c_{t+1}^H) = c_{t+1}^H, \tag{2}$$

which has the convenient implication that all labour income, w_t, is saved in the first period of life. Each worker household—referred to as simply household from hereon—begets another at the end of the first period.

Savings may be held in bank deposits and/or one-period corporate securities (bonds and equities). Let R^D denote the (gross) return on bank deposits. In the equilibrium studied further, households invest in both assets and R^D also denotes the return on bonds and equities.[1]

Households (and entrepreneurs) also have access to a storage technology that allows them to store their savings for a period. Gross return on this is $\sigma \geq 1$. Storage is never used in the financial arrangement analysed further but implicitly sustains the financial equilibrium.

Entrepreneurs

Overlapping generations of entrepreneurs live for two periods and their measure is normalised to unity. As these agents have the ability to produce various types of capital goods, they are the 'capitalists' of this economy.

Each entrepreneur is born with one unit of time that he/she can spend overseeing his/her capital goods production. She cares only about old-age consumption and is risk neutral:

$$U_t^E = c_{t+1}^E. \tag{3}$$

We incorporate a role for internal funds by endowing each entrepreneur with b amount of goods in the first period of her life.[2]

Entrepreneur-j is the monopoly supplier and owner–manager of a firm producing the capital good, K^j. Production of this capital requires resources invested one period in advance, together with entrepreneur-specific skills.[3] Entrepreneurs rent out their capital, repay their lenders and consume out of the net returns. Hence, a generation-t entrepreneur-j invests an amount I_t^j to maximise her income net of loan repayments, x_{t+1}^E.

Let R_t^j denote the rental on K_t^j. We assume all capital goods depreciate fully upon use. Demand for the jth capital good is obtained from the static maximisation problem faced by final goods producers:

$$\max_{\{K_t^j\}} \left[\int_{j=0}^{1} \left(K_t^j\right)^{\alpha} dK_t^j \right] - \int_{j=0}^{1} R_t^j K_t^j dK_t^j.$$

Table 1. A Typical Entrepreneur's Project Choices

Investment Project	Good	v-project	V-project
Probability of success	1	π	π
Private benefit to entrepreneur	0	vI	VI

Source: The author.

Separability of capital goods in the production function implies separable demand for these goods with firm-j facing the inverse demand curve:

$$R_t^j = \alpha \left(K_t^j\right)^{\alpha-1}, \tag{4}$$

$1/(1-\alpha)$ being the constant price elasticity.

An entrepreneur's total investment consists of his/her internal participation b,[4] and the amount of external finance he/she obtains. We focus on an equilibrium where both types of external finance—bank borrowing and household borrowing—are used. Following Gurley and Shaw (1955) and Holmstrom and Tirole (1997), we distinguish between bank borrowing and household borrowing as *indirect* (or intermediated) finance versus *direct* (or unintermediated) finance.

Now consider this entrepreneur's investment decision where new investment in period t generates a verifiable amount of capital in period $(t+1)$. But, when proper incentives are lacking or when entrepreneurs are not suitably monitored, they may enjoy a private benefit and, in the process, reduce the success probability of their investment (Holmstrom, 1996; Holmstrom & Tirole, 1997).

Take an entrepreneur who has raised investment resources amounting to $I_t^j > b$. Suppose that investment outcomes depend on the entrepreneur's investment choice and whether or not he/she appropriates part of the investment funds. The entrepreneur is diligent when he/she gives full attention to his/her project and invests his/her entire funds I_t^j. In this case, the investment succeeds for sure and capital produced equals the investment he/she undertook:

$$K_{t+1}^j = I_t^j. \tag{5}$$

When the entrepreneur shirks, he/she devotes less time to overseeing his/her project and spends that time appropriating part of the investment funds.[5] In this case, the project may go bankrupt as a result of insufficient supervision.

We formalise this by endowing the entrepreneur with the three types of investment choices illustrated in Table 1. The best investment is the 'good' one given in Equation (5). The worst outcome occurs for the 'V-project' where the entrepreneur steals VI_t^j and invests the remainder to produce capital worth I_t^j (i.e., it costs less) with probability π. Expected capital produced is thus $K_{t+1}^j = \pi I_t^j$.

In contrast to the worst outcome, v-projects result in better outcomes even with the entrepreneur shirking. For these projects, the entrepreneur enjoys a private benefit proportional to his/her investment with the proportion v varying. The remaining $(1-v)I^j$ funds are invested which produces capital as in Equation (5) with probability π, zero otherwise. The success probability of a v-project is as same as the V-project, so an entrepreneur clearly prefers the latter to the former. To develop a rich monitoring structure, we allow the v-project to be a set of project choices over the interval $[\underline{v}, \overline{v}]$ with $\underline{v} < \overline{v} < V$, all with the same probability of success, π.[6]

Consider an equilibrium where banks monitor entrepreneurs when lending to them. External monitoring has two consequences: banks can completely eliminate the V-project, narrowing down the entrepreneur's choice to those between the best investment project and the set of v-projects. But monitoring can do more. Depending on how intensively the entrepreneur is monitored, it can also eliminate some of the v-projects. In particular, monitoring at intensity γ eliminates all v-projects yielding a private benefit higher than $v(\gamma) \in [\underline{v}, \overline{v}]$. That is, an entrepreneur monitored at intensity γ enjoys a private benefit of $v(\gamma)I$ at best. Under monitoring, private benefits may therefore be characterised by a decreasing function $v : [0, \widehat{\gamma}] \to [\underline{v}, \overline{v}]$, $v' < 0$, and where $\widehat{\gamma} \leq 1$ is the monitoring intensity required to eliminate anything worse than the $\underline{v}$-project.

III. Monitoring and the Return on Savings

Financial intermediaries, or banks, are endowed with a monitoring technology that households do not have, or alternatively, households being too dispersed cannot effectively exercise. Bank monitoring can take a variety of forms including inspection of the firm's cash flows and balance sheets, and keeping tabs on the manager's activities (Hellwig, 1991). These activities ensure that the manager's incentives do not stray too far from those of the bank's and outside investors'.

Consider how monitoring alters returns on savings. Suppose entrepreneur-j wants to invest $I^j > b$ so that he/she has to borrow an amount $I^j - b$. Let R^E denote the implicit return that the entrepreneur earns on his/her funds. In the absence of monitoring, the entrepreneur behaves diligently only if,

$$R^E b \geq \pi R^E b + V I^j \tag{6}$$

$$\Rightarrow R^E b \geq \frac{V}{1-\pi} I^j, \tag{7}$$

where the left-hand side of Equation (6) denotes his/her earnings when diligent, and the right-hand side denotes earnings when he/she shirks and appropriates the maximum amount possible without detection. Households, requiring a return of $R^S \geq \sigma$ on their investments, earn

$$R^S (I^j - b) = \left[R^j - \frac{V}{1-\pi} \right] I^j \tag{8}$$

at most.

Had these households, instead, stored away their savings, they would have earned a (gross) return σ. Whether or not they are willing to lend directly to firms depends upon how severe the agency problem is. Suppose $V = 1$. This means that entrepreneurs can steal the entire investment funds, storing them away for future consumption without any detection. As nothing is left over to invest, the project produces zero capital and households get back nothing. In this case, because $\sigma \geq 1$ by assumption, households are strictly better off not lending directly to firms when they are the sole external financiers. In other words, no one is willing to lend to firms. Even for $V < 1$, as long as V is high, direct finance would yield lower returns than storage. Without loss of generality, we set $V = 1$ henceforth.

Next consider what happens when an entrepreneur is monitored at intensity γ, so that the worst investment choice he/she can make is the $v(\gamma)$ project. The entrepreneur behaves diligently as long as

$$R^E b \geq \frac{v(\gamma)}{1-\pi} I^j, \tag{9}$$

leaving at most

$$R^S(I^j - b) = \left[R^j - \frac{v(\gamma)}{1-\pi} \right] I^j, \tag{10}$$

to be divided up between banks and households, and on monitoring expenses. If monitoring were costless, as $v(\gamma) \leq \overline{v} < V$, returns under monitoring would exceed those under no monitoring and, if σ (or $\underline{v}$) is low enough, also that from storage.

External monitoring is, however, privately costly: banks have to spend resources in monitoring the borrowers. We assume that these costs are directly proportional to the volume of investment undertaken by a borrowing intermediate goods firm. Let c be the unit cost of monitoring a borrower at full intensity. When a bank monitors at an intensity γ, its *effective* unit monitoring cost is then $c\gamma$.

Whether or not households earn higher returns with monitored finance than with only non-monitored finance depends on how large monitoring costs are. As long as they are not too high, the returns given by Equation (10) are higher than those given by Equation (8). Once again, the assumption that $V = 1$ ensures that this is indeed the case.[7]

Thus, households are better off with storage than with purely direct lending. However, when firms borrow from banks and are monitored, households earn a return not only higher than that under purely direct finance but also higher than storage. Whereas these households would not lend to firms when firms do not simultaneously borrow from banks, under bank finance, they would.[8]

Note here that how bank-finance transmits a signal to direct lenders. When a firm and a bank enter into a loan agreement, outside investors assume, correctly, that the firm would be monitored. Knowing that the agency problem will be reduced due to this, households recognise that they earn a strictly greater return from lending to firms now than they would if they were the sole financiers. In other words, households do not need to know exactly how much resources are spent on monitoring firms: the very fact that banks lend to firms ensures that the entrepreneur's incentive to shirk drops from V to at least $\overline{v}$.

Not surprisingly, costly monitoring drives a wedge between the prices of direct and indirect finance. Bank finance is the more expensive alternative because of its information content. While entrepreneurs cannot do without bank finance, at the same time they will economise on it by borrowing from banks only the minimum amount necessary to access the cheaper market finance. The following section looks at this choice.

IV. Investment Choice and Loan Contracts

We solve for an equilibrium where banks and firms enter into loan contracts that allow for monitoring and, at the same time, guarantee entrepreneurs enough returns to induce them to behave diligently (incentive constraint). In a symmetric equilibrium, each entrepreneur will produce as much capital as invested, similar to the standard neoclassical model:

$$K^j_{t+1} = I^j_t = I_t \text{ for all } j \in [0, 1] \text{ and } t.$$

The Bank's Problem

Consider the optimisation problem facing a bank that has lent L^j to the jth firm. As all intermediate goods firms are identical, we assume they each borrow the same amount from banks, $L^j = L$ for all j.

Banking profits in $(t + 1)$ are given by

$$\Pi^B_{t+1} = R^L_{t+1} L_t - R^D_{t+1} D_t, \tag{11}$$

where R^L_{t+1} is the (gross) loan rate charged to borrowers and D_t denotes the flow of deposits into the banking sector. Banks take as given deposits D_t, as well as the price vector (R^D_{t+1}, R^L_{t+1}).

A bank's cost of monitoring a firm at intensity γ, given unit monitoring cost c, is proportional to the firm's total investment: $c\gamma I$. These costs are paid out of current deposits, so that banks face the resource constraint:

$$L_t \leq D_t - c\gamma_t I_t. \tag{12}$$

For bank monitoring to be an equilibrium, banks have to earn at least as much as they would be without monitoring. As banks monitor firms in period t, but realise returns on their loans in period-$t + 1$, they discount monitoring costs at their opportunity cost, that is, the deposit rate. The banking participation constraint then requires net returns under monitoring to be at least as great as those under no monitoring:

$$\begin{aligned} R^L_{t+1} L_t - c\gamma_t R^D_{t+1} I_t \;&\geq\; \pi R^L_{t+1} L_t \\ \Rightarrow R^L_{t+1} L_t &\geq \frac{c\gamma_t}{1-\pi} R^D_{t+1} I_t. \end{aligned} \tag{13}$$

Likewise, in order for bank monitoring to be an equilibrium outcome, entrepreneurs have to be willing to accept it. This means each entrepreneur must earn at least as much as he/she would without monitoring. Therefore,

$$\begin{aligned} R^E_{t+1} b &\geq \pi R^E_{t+1} b + v_t I_t \\ &\Rightarrow R^E_{t+1} b \geq \frac{v_t}{1-\pi} I_t, \end{aligned} \tag{14}$$

where $R^E_{t+1} b \equiv x^E_{t+1}$ and $v_t \equiv v(\gamma_t)$. The optimisation problem facing the banking sector is then

$$\begin{gathered} \text{Max}_{\{L_t\}}\ \Pi^B_{t+1} \\ \text{subject to: Equations (12), (13) and (14),} \end{gathered} \tag{15}$$

given deposits, D_t, and the price vector $(R^D_{t+1},\ R^L_{t+1})$. Note that banks do not directly choose monitoring intensity. That is determined in equilibrium through active competition for bank finance: entrepreneurs who would like to invest more, agree to be monitored more intensively and, in their willingness to accept higher γ, push their participation constraint Equation (14) to equality.

From Equation (15), it follows that in a monitoring equilibrium, banks lend out the maximum amount, that is,

$$L_t = D_t - c\gamma_t I_t.$$

Moreover, Equation (13) implies that banks are willing to monitor as long as they are allowed to fund a minimum fraction, ϕ_t, of a firm's total investment, where

$$\phi_t \geq \frac{L_t}{I_t} \left[\frac{c\gamma_t}{1-\pi} \right] \left[\frac{R^D_{t+1}}{R^L_{t+1}} \right]. \tag{16}$$

As bank finance is relatively more expensive than market-finance (see further), entrepreneurs accept only the minimum amount necessary, so that Equation (16) holds as an equality in equilibrium.

The Entrepreneur's Problem

Now turn to the investment decision. The entrepreneur has to decide how much to invest and how much to borrow from banks versus the securities market. Given an investment of size I and bank loan L, denoted by $M = I - L - b$ the amount borrowed directly from the market.

For households to be willing to lend to firms directly, they have to be guaranteed a return of at least R^D. But that is not the only cost firms face in using direct finance. In particular, we assume that there are transactions costs in the financial market: for every unit firms are required to pay back to households on their bond and stock holdings, they incur a transactions cost τ. These costs are deadweight losses and capture, in a simple way, the ease of placing and trading in corporate securities, the liquidity of the financial market and the cost of enforcing contracts between direct lenders and firms.[9]

Entrepreneurs, thus face the cost $(1+\tau)R^D$ on direct finance versus R^L on bank loans. In order for bank finance to be relatively more expensive than market finance, it is sufficient to assume that

$$\tau < \frac{1-\pi}{\pi}.$$

This condition guarantees that $R^L > (1+\tau)R^D$ in equilibrium (see Equation [23]).

Under this assumption, bank finance is relatively more expensive and each entrepreneur borrows from banks only the minimum amount necessary. In other words, Equation (16) holds as an equality, with bank-finance constituting ϕ_t fraction of I_t. This leaves an amount $M_t = (1-\phi_t)I_t - b$ to be raised through direct finance.[10] Income earned by the entrepreneur in $(t+1)$, net of loan repayments is then:

$$x^E_{t+1} = R_{t+1}I_t - R^L_{t+1}(\phi_t I_t) - (1+\tau)R^D_{t+1}\left[(1-\phi_t)I_t - b\right]. \tag{17}$$

As the entrepreneur's lifetime utility is proportional to this income, he/she chooses I_t to maximise Equation (17), given Equations (4) and (5), or,

$$\max_{K_{t+1}} \left\{\alpha K^\alpha_{t+1} - [\phi_t R^L_{t+1} + (1-\phi_t)(1+\tau)R^D_{t+1}]K_{t+1} + (1+\tau)R^D_{t+1}b\right\}.$$

The first-order condition gives optimal investment as:

$$K_{t+1} = \left[\frac{\alpha^2}{\mu_{t+1}}\right]^{1/(1-\alpha)}, \tag{18}$$

where

$$\mu_{t+1} \equiv \phi_t R^L_{t+1} + (1-\phi_t)(1+\tau)R^D_{t+1}, \tag{19}$$

is the effective unit borrowing cost using Equations (23) and (24). Substituting the optimal choice of investment from Equation (18) into Equation (17), maximal entrepreneurial income is:

$$\begin{aligned}\widehat{x}^E_{t+1} &= (R_{t+1} - \mu_{t+1})K_{t+1} + (1+\tau)R^D_{t+1}b \\ &= \left(\frac{1-\alpha}{\alpha}\right)\mu_{t+1}K_{t+1} + (1+\tau)R^D_{t+1}b. \end{aligned} \tag{20}$$

This income is strictly greater than $R^D_{t+1}b$, what the entrepreneur would have earned had he/she invested his/her saving with banks or lent directly to other firms. Income in excess of the opportunity cost of funds consists of monopoly rents from capital goods production as well as compensation for transactions costs incurred in using market finance. Hence, entrepreneurs put up their entire endowment b as internal equity on their projects.

Optimal Loan Contract

Banks offer loan contracts that are accepted or rejected by entrepreneurs. Under free entry and exit into the banking sector, banking profits are zero in equilibrium.

Likewise, active competition for bank finance means entrepreneurs' participation and incentive constraints will be binding.

Proposition 1. *The optimal loan contract between a bank and a borrowing firm is a triple* $\chi = (L_t,\ R^L_{t+1},\ \gamma_t) \in R_+ \times [1, R^j] \times [0, \hat{\gamma}]$ *that*
(i) solves the bank's optimisation problem Equation (15) *and*
(ii) maximises entrepreneurial income, x^E_{t+1}*, subject to the entrepreneur's participation constraint Equation* (14).

To solve for this contract, using Equations (12) and (13), we first obtain

$$L_t = D_t - (1-\pi)\frac{R^L_{t+1}}{R^D_{t+1}}L_t$$
$$\Rightarrow \left[R^D_{t+1} + (1-\pi)R^L_{t+1}\right]L_t = R^D_{t+1}D_t. \tag{21}$$

Substituting this expression into the bank's profit function Equation (11) gives maximal profits as:

$$\widehat{\Pi}^B_{t+1} = \left[\frac{R^L_{t+1}}{R^D_{t+1} + (1-\pi)R^L_{t+1}} - 1\right]R^D_{t+1}D_t. \tag{22}$$

With free entry and exit into the banking sector, maximal profits can only be zero in equilibrium. Imposing this zero-profit condition on Equation (22) implies that banks charge a constant markup over the deposit rate:

$$R^L_{t+1} = \frac{R^D_{t+1}}{\pi} \tag{23}$$

reflecting the risk of project failure.[11] Now we can substitute Equation (23) into Equations (16) and (21), to obtain the following equilibrium relations:

$$\phi_t = c\gamma_t\left(\frac{\pi}{1-\pi}\right), \quad L_t = \pi D_t. \tag{24}$$

The loan amount is independent of monitoring costs, but banks participate more intensively when they monitor more intensively.

Next we have to determine how intensively firms are monitored in equilibrium. For that, using Equation (24) first simplify Equation (19) to

$$\mu_{t+1} = \left[\phi_t/\pi + (1-\phi_t)(1+\tau)\right]R^D_{t+1} = (1+\delta_t)R^D_{t+1}, \tag{25}$$

where $\delta_t \equiv c\gamma_t + \tau[1 - c\pi\gamma_t/(1-\pi)]$. Entrepreneurs actively compete for bank finance which ensures that, given I_t and $\hat{x}^E_{t+1}$,

$$\hat{x}^E_{t+1} = \frac{v(\gamma_t)}{1-\pi}I_t.$$

Finally, substituting for optimal investment choice and maximal entrepreneurial income implicitly determines γ_t via

$$\left(\frac{1-\alpha}{\alpha}\right)\mu_{t+1}K_{t+1} + (1+\tau)R^D_{t+1}b = \varphi(\gamma_t)K_{t+1}, \tag{26}$$

where $\varphi(\gamma) \equiv \upsilon(\gamma)/(1-\pi)$.

V. General Equilibrium

Equilibrium Prices of Labour and Capital Goods

As all intermediate goods producers face similar prices and constraints, a natural equilibrium to consider is the symmetric one where, $K^j_{t+1} = K_{t+1}$, $\forall j$. Having normalised the size of the entrepreneurial class to one, this has the convenient implication that each entrepreneur's capital stock is also the economy-wide capital stock. Moreover, recognizing that labour supply is also equal to one, it is straightforward to normalise all variables by the size of the labour force. Variables expressed as 'per worker' are denoted in lower case. The competitive equilibrium prices for labour and capital goods are hence:

$$\begin{aligned} w(k_t) &= (1-\alpha)k_t^{\alpha}, \\ R^j(k_t) &= \alpha k_t^{\alpha-1} \text{ for all } j. \end{aligned}$$

General equilibrium in this economy, as in the standard overlapping generations model, can now be fully characterised by a first-order difference equation in capital per worker, k.

Equilibrium in the Loanable Funds Market

The financial sector is a conduit for transforming household savings into capital goods. Part of these savings flows through banks, the remainder flows directly from households to entrepreneurs through the purchase of corporate debt and equity.

Investment undertaken by each entrepreneur, when both monitored and non-monitored finance are used, is

$$\begin{aligned} I_t &= b + L_t + M_t \\ &= b + \pi D_t + (w_t - D_t) \\ &= b + w_t - (1-\pi)D_t \\ &= b + w_t - \left(\frac{1-\pi}{\pi}\right)\phi_t I_t \\ &= b + w_t - c\gamma_t I_t, \end{aligned}$$

where the last step is obtained using Equation (24). As all entrepreneurs behave diligently, given the terms of the optimal loan contract, each of them supplies capital goods amounting to

$$K_{t+1} = k_{t+1} = I_t = \frac{b + w_t}{1 + c\gamma_t}. \tag{27}$$

Using the aforementioned equilibrium wage rate, we then have

$$k_{t+1} = \frac{G(k_t)}{1 + c\gamma_t}, \tag{28}$$

where $G(k) \equiv b + w(k)$ is an increasing, concave function of capital per worker. We now determine how γ depends on economy-wide aggregates.

Monitoring Intensity

Using Equations (18), (26) and (28), lead us to the relationship

$$\left[\frac{1 + \delta_t}{1 + c\gamma_t}\right]\left[\frac{\varphi(\gamma_t)}{(1 + c\gamma_t)^{1-\alpha}} - \frac{\alpha(1-\alpha)}{G(k_t)^{1-\alpha}}\right] = \frac{\alpha^2(1+\tau)b}{G(k_t)^{2-\alpha}} \tag{29}$$

that implicitly determines γ_t as a function of the current capital stock, k_t. In particular, the left-hand side of Equation (29) is a continuous function of γ_t and k_t, decreasing in γ but increasing in k. The right-hand side is a continuous and increasing function of k_t. Hence, Equation (29) defines the optimal monitoring intensity as an increasing function of capital per worker: $\gamma_t = \gamma(k_t)$ with $\gamma' > 0$ (see Appendix A.1).

Intuitively, given their internal funds b, as entrepreneurs invest more over time, they face higher incentives to shirk. To align their incentives with those of the banks', banks monitor them more intensively and hence, supply a higher proportion of their investment funds. Recall, however, that γ is bounded above by $\widehat{\gamma}$ at which point the entrepreneur's worst project choice is $\underline{v}$. Monitoring at an intensity greater than $\hat{\gamma}$ does not allow banks to attenuate the agency problem any further. Hence, for all values of capital exceeding a threshold level $\widehat{k}$ corresponding to $\widehat{\gamma}$, monitoring intensity stays at $\widehat{\gamma}$. The following lemma summarises these results:

Lemma 1. *The optimal monitoring intensity is weakly increasing in capital per worker and is given by a function*

$$\gamma_t = \Gamma(k_t) \equiv \begin{cases} \gamma(k_t), & \text{if } k_t \le \widehat{k} \\ \widehat{\gamma}, & \text{if } k_t > \widehat{k} \end{cases}$$

where $\Gamma(0) = 0$, $\gamma' > 0$ *and* $\widehat{k}$ *solves* $\gamma(\widehat{k}) = \widehat{\gamma}$.

This monitoring function, as well as the threshold capital stock $\widehat{k}$, depends upon the underlying cost parameters, c and τ. For our discussions later, Lemma 2 presents (without proof), how (c, τ) affect γ and $\hat{k}$:

Lemma 2. *(i) The optimal monitoring intensity, $\gamma = \Gamma(k; c, \tau)$, decreases with an increase in the unit cost of monitoring, c, and increases with a rise in the transactions cost, τ, that is, $\partial\gamma/\partial c < 0$, $\partial\gamma/\partial\tau > 0$;*
(ii) the threshold capital stock, $\widehat{k}$, increases when c goes up and decreases with an increase in τ, that is, $\partial\widehat{k}/\partial c > 0$ and $\partial\widehat{k}/\partial\tau < 0$.

Results from these two lemmas are illustrated in Figure 1.

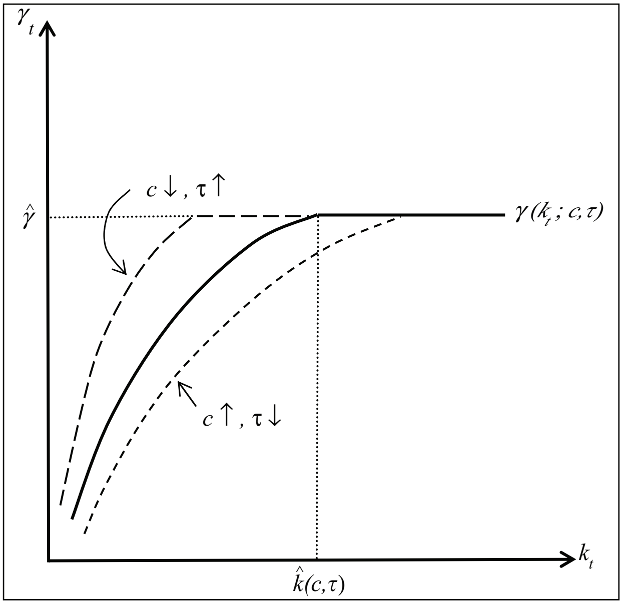

Figure 1. Optimal Monitoring Intensity

Source: The author.

Equilibrium Capital Accumulation

Now that we have the optimal monitoring intensity, the equilibrium law of motion is obtained from Equation (28) as

$$k_{t+1} = \frac{G(k_t)}{1 + c\Gamma(k_t)} \equiv H(k_t). \tag{30}$$

Given an initial stock of capital goods, $k_0 > 0$, owned by old members of the initial entrepreneurial generation, this equation is sufficient to characterise the evolution of the real and financial sectors.

As $G(0) = b$ and $\Gamma(0) = 0$, clearly $H(0) = b$. Thus, zero is not a steady state of this economy. Even without any initial capital, the initial young entrepreneurs are able to convert their goods endowments into capital and jump-start the economy. To look for positive steady states, the H function has to be characterised a bit more. As both G and Γ are increasing functions of k, it is not obvious that Equation (30) describes an increasing phase map. But in Appendix A.1, we show that as long as the unit cost of monitoring, c, is small enough, H is monotonically increasing in capital per worker. This restriction has an intuitive interpretation. Monitoring allows firms to invest more, given their internal funds. However, it also uses up some of the resources that could have been used to finance investment. As long as monitoring is relatively inexpensive, it does not use up too much loanable funds and firms can invest more.

Now define $H_1(k) \equiv G(k)/[1 + c\gamma(k)]$ and $H_2(k) \equiv G(k)/(1 + c\hat{\gamma})$ so that $H_2(0) = b > H_2(0) = b/(1 + c\hat{\gamma})$. As $\gamma(k) \leq \hat{\gamma}$ for all $k \leq \hat{k}$, we also have $H^1(k) \geq H^2(k)$ for $k \leq \hat{k}$. Finally, $\gamma' > 0$ ensures that $H_1' < H_2'$ at any particular point $k < \hat{k}$. The phase map described by Equation (30) is, thus the upper envelope of $\{H_1(k), H_2(k)\}$. As $\lim_{k \to \hat{k}-} H_1' < \lim_{k \to \hat{k}+} H_2'$, $H(k)$ has a kink at $\hat{k}$, the point at which bank monitoring eliminates all investment choices for capital goods producing firms except for the good one and the $\underline{\upsilon}$-project.

Figure 2 depicts the equilibrium law of motion of capital Equation (30) for two possible values of c, which in turn implies two different values for $\hat{k}$. In both cases, the economy converges asymptotically to the unique steady-state capital stock. The dotted line corresponds to the case where $c = 0$, which gives us the usual equilibrium capital accumulation rule, $k_{t+1} = G(k_t)$, when bad investment projects can be costlessly excluded. Depending on how high or low c ($\hat{k}$) is, the economy converges asymptotically to a steady-state $\bar{k}$ that may be lower or higher than $\hat{k}$. In Figure 2(a), relatively high monitoring costs imply that the economy is never able to attenuate agency problems to the fullest possible extent. Here, the steady state solves

$$\bar{k} = \frac{G(\bar{k})}{1 + c\gamma(\bar{k})} < \widehat{k}.$$

Figure 2(b) illustrates, on the other hand, how, with low monitoring costs, per capita income is higher in the long-run because the economy is quickly able to resolve agency problems without have to spend too much resources in the process. The steady state is now the solution to

$$\bar{k} = \frac{G(\bar{k})}{1 + c\widehat{\gamma}} > \widehat{k}.$$

Recall that in section III we implicitly assumed that households earn a higher return under a monitored finance than they would without any borrower monitoring. From an entrepreneur's demand for capital, we have

$$(1 + \delta_t)R^D_{t+1} = \alpha^2 K^{\alpha-1}_{t+1}.$$

In a steady state, $\bar{k}$ ($> \hat{k}$), this gives us the return on savings as

$$\overline{R}^D = \frac{\alpha^2 \overline{k}^{\alpha-1}}{1 + c\widehat{\gamma} + \tau[1 - c\pi\widehat{\gamma}/(1 - \pi)]}.$$

We have been implicitly assuming parametric restrictions such that $\overline{R}^D > \sigma$. This equation shows that the restriction is satisfied if (c, τ, σ) are low.

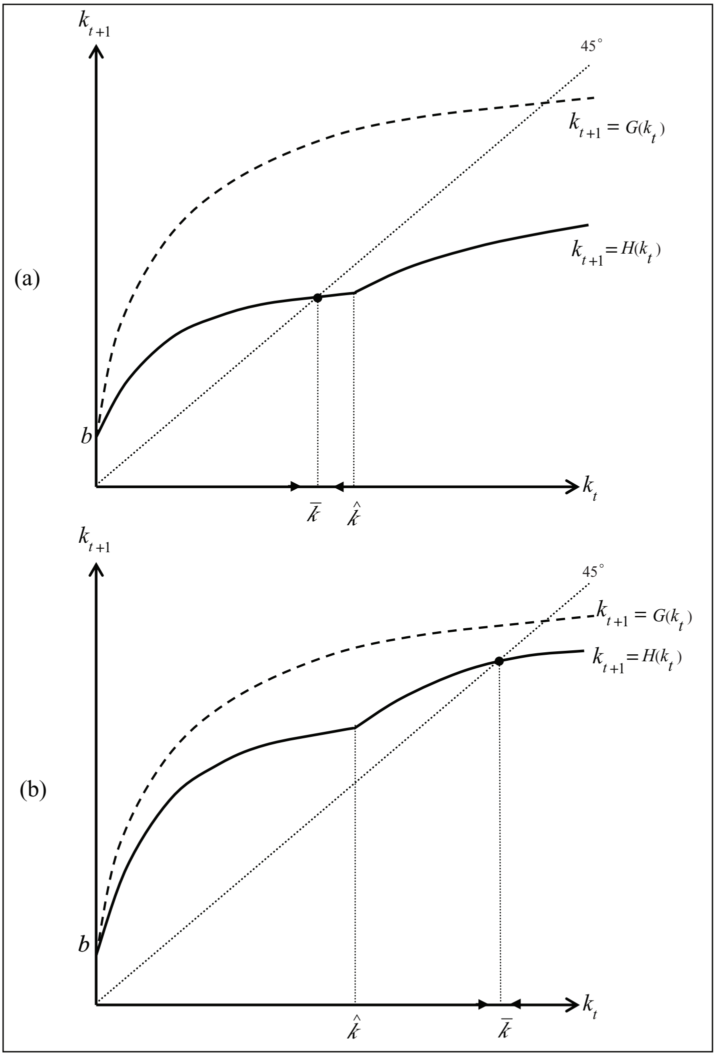

Figure 2. General Equilibrium Dynamics

Source: The author.

VI. Discussion

The model has several implications about the process of financial deepening, in particular, about how external finance and financial structure evolve with economic development.

Financial Deepening: The Composition of Investment

The financial sector here comprises here of both banks and market for bonds and equities. It is instructive to see how each evolves with capital accumulation.

As bank-finance is more expensive than market-finance, firms demand only the minimum amount that they need. Aggregate bank-intermediated investment is $L_t = \phi_t I_t = c\pi\gamma_t I_t/(1-\pi)$, a share ϕ_t of aggregate investment. We have already seen that γ increases with the capital intensity of production—as firms borrow more, banks become more involved in alleviating agency problems. An immediate implication is that banks finance a rising proportion of all investment during the initial stages of economic development, that is ϕ rises monotonically until the economy reaches the threshold $\hat{k}$ at which point all inferior project choices except for the $\underline{v}$-project are eliminated. Thereafter, monitoring intensity remains at $\hat{\gamma}$ and the proportion of bank loans at $\hat{\phi} \equiv c\pi\widehat{\gamma}/(1-\pi)$.

The more efficacious the banking sector is in enforcing incentive compatible contracts (lower is c), faster does ϕ rise and lower is $\widehat{k}$ (Lemma 2). How rapidly the banking sector grows is closely tied to institutional factors, such as legal systems, accounting procedures and the extent to which firms are required to reveal information to the public, all of which affect c. The implication is that the less distorted systems see a quicker development of their banking sectors.

Moreover, when transactions cost τ in the financial market is low, ϕ is lower at any point in time for $k < \widehat{k}$ (Lemma 2). The cheaper availability of market finance induces firms to substitute away from bank-finance. But the threshold capital stock, $\hat{k}$, is also higher. Not only does ϕ rises more slowly in this case, it takes longer for the economy to eliminate much of the agency problem. The banking sector evolves more gradually in this case.[12]

Consider next the evolution of market finance, $M_t = (1-\phi_t)I_t - b$. As the economy accumulates capital and firms invest more, it tends to raise the demand for market-finance. At the same time, banks fund a rising proportion of investment (for $k < \hat{k}$), which tends to crowd out, market finance. Whether or not bank finance crowds out, market finance depends on how large monitoring costs are. Appendix A.1 lays down a sufficient condition, $c < \hat{c}$, which ensures that bank monitoring does not consume too much resources so as to crowd out market finance along the growth path. It is not hard to see why we need this restriction. Higher volumes of investment need to be monitored more intensively which consumes part of the loanable funds (those with the banking sector). If monitoring is too expensive, not much resources are available for investment after meeting monitoring costs. Although investment would rise, it would not rise as much as is required for firms to increase market borrowing. So long as $c < \hat{c}$, the proportion of investment funded through direct borrowing $m_t \equiv M_t/I_t$, rises over time. Thus, firms utilise both intermediated and unintermediated finance more intensively.

Although the proportion of investment financed by the market rises, it is not obvious how a capital good producing firm's ratio of bank- to market-finance behaves. This ratio is given by $\theta_t \equiv m_t/\phi_t = (1-b/I_t)/\phi_t - 1$. Whether or not it increases depends on how fast m is rising relative to ϕ. Below $\hat{k}$, this ratio will be falling, that is, bank debt rising faster than equity/bond finance if monitoring costs are lower or market transactions cost higher. Above $\widehat{k}$, it is always rising since ϕ stays constant. The implication is that in the earlier stages of economic development, reliance upon bank debt may increase faster than on market finance, but in the latter stages, there is increased substitution towards market finance: the banking sector attains a stable size (relative to investment) while financial markets expand to finance rising percentages of aggregate investment.

The model is thus able to explain Demirgüç-Kunt et al. (2013)'s findings discussed in the Introduction. The rise of the financial market, especially in the latter stages of development ($k > \hat{k}$), is intrinsically tied to the monitoring role of the banking sector without which households would not even lend to firms. In particular, the efficiency with which banks monitor is key: Systems with lower costs of enforcing incentive compatible financial agreements are more effective and see a quicker resolution of agency problems and faster emergence of financial markets.

Policy Trade-offs

Faced with inefficient banking sectors and frequent banking crises, rich and poor countries increasingly moved toward market-based financial systems from the eighties on to the aughts. This shift occurred in Latin America, Eastern Europe as well as developed regions such as France and Japan (Allen & Gale, 2000). A version of this debate played out in India in the last decade. While some commentators rightly emphasised how financial repression—58 per cent of deposits in the Indian banking sector are allocated according to government policy—has hamstrung bank-intermediated allocative efficiency in the country, others saw a way out in developing the corporate bond and equity markets (*The Economist*, 2013; Farrell et al., 2006).

Consider the effect of financial reforms that lower the costs of external finance, for example policies that reduce the cost of bank monitoring, c, and/or transactions costs in non-monitored finance, τ. Faced with limited resources and policy constraints, suppose policymakers have to choose between these two. Contrast, using the phase-portraits, the effect of either type of reform on capital accumulation and long-run living standards.[13]

Lower cost of external finance through either means improves firms' ability to borrow and invest. There is a difference however. When monitoring costs are not particularly high (Figure 3b), a lower c shifts up the phase map. Per capita income is permanently higher at every point in time and, because it becomes cheaper for banks to monitor, they are also able to reduce agency problems more rapidly (lower $\widehat{k}$).[14] A lower τ, on the other hand, shifts only the lower part of the phase map, that below $\hat{k}$ (Figure 4b). At low values of the capital stock, it speeds up capital accumulation as firms substitute toward relatively cheaper direct finance. However, this substitution also means that the banking sector takes longer to contain agency problems. Moreover, a decline in τ does not affect the long-run per capita capital stock as $\bar{k}$ is insensitive to τ.

The case of an inefficient monitoring technology (high c) is illustrated in Figures 3a and 4a. Figure 3a shows the effect of a reduction in monitoring costs. Lower monitoring costs raise the steady-state $\bar{k}$, but reduce $\hat{k}$. For a sufficient decline in these costs, the phase line may be pushed up sufficiently for the new steady state to exceed $\hat{k}$. Not only would the banking sector become more active for a while but financial markets would also evolve, at first gradually and then rapidly as the

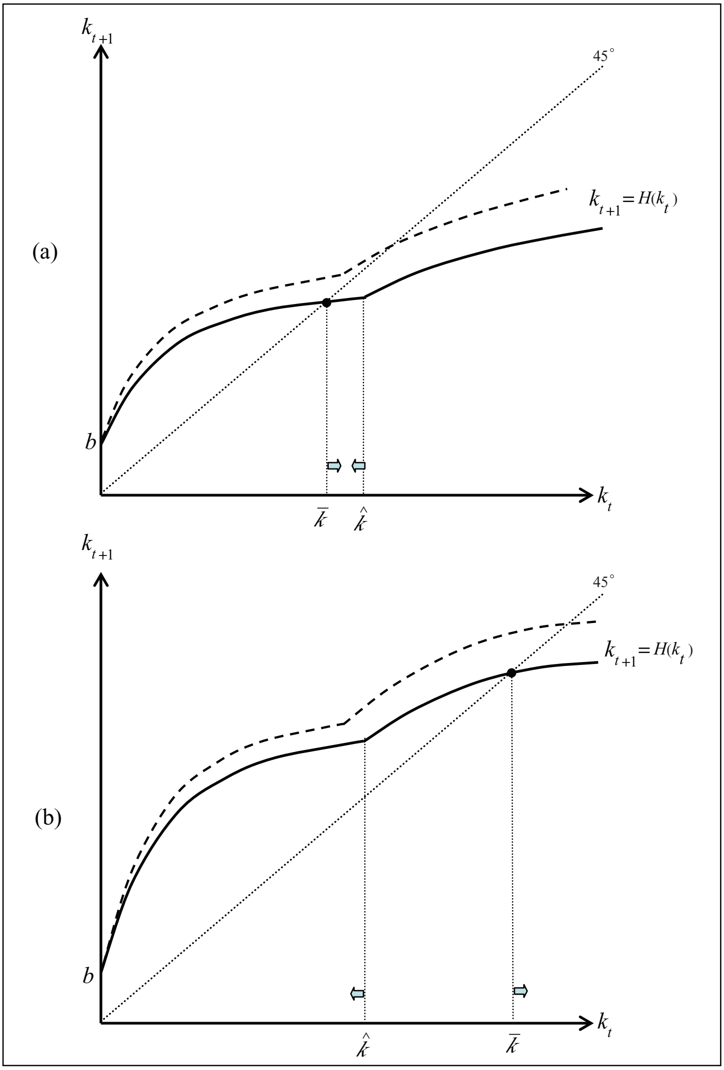

Figure 3. Effect of a Decrease in *c*

Source: The author.

economy crosses $\hat{k}$. Any change that benefits the banking sector, also helps the financial sector develop.

The effect of a decline in τ, however, does affect steady-state incomes now. This is shown in Figure 4a: a decrease in τ has an effect on steady-state monitoring intensity ($\gamma < \hat{\gamma}$), so that $\bar{k}$ rises as does $\hat{k}$. Once again, firms substitute toward cheaper market finance, but now, the banking sector is unable to fully attenuate agency problems any more. This need not be undesirable, per se, especially because monitoring is costly. However, from Figures 3a and 4a, it is arguably possible to conjecture that even in this case (where $\bar{k} < \hat{k}$), a reduction in c is likely to have a

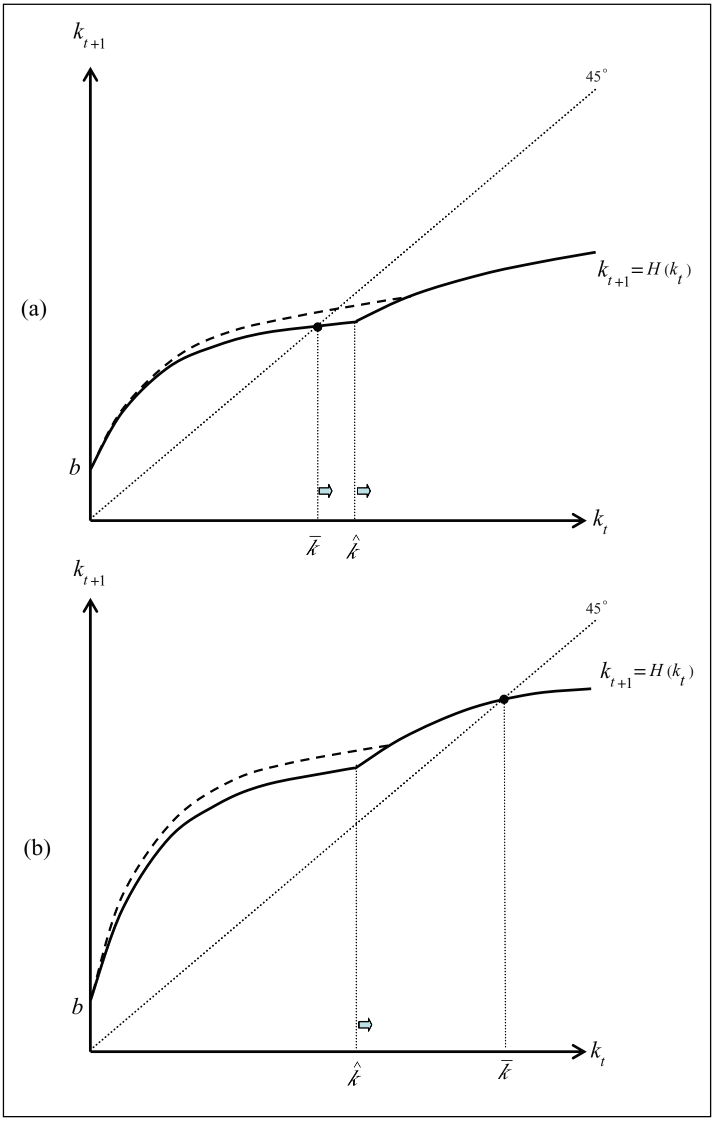

Figure 4. Effect of a Decrease in τ

Source: The author.

bigger long-run impact than a reduction in τ. If it is equally easy to implement a marginal change in c and τ, not only does a lower c speed up capital accumulation in the short-run (as does a lower τ), but it leads to even higher long-run standards of living when $\bar{k}$ is pushed above $\hat{k}$ (which a lower τ cannot).[15]

The lesson to draw is simple: for successful economic development, banking sector reforms that lower the costs of acquiring (and transmitting) information about incentives faced by investing firms are likely to be more effective than reforms that simply make market finance more accessible. The latter helps, of course. But as long as arm's-length finance cannot do much about agency problems within the firm—corporate governance laws are relevant here—they have to rely on banks to do it. An uncompetitive banking sector, that is heavily regulated to direct resources towards less productive sectors or public finance, would be hardly up to the task. Yet that is precisely where reforms need to start: as long as the allocative and information-gathering role that banks have played in financial deepening around the world are stymied, financial sector reforms anywhere else will come up short.

VII. Conclusion

This article has analysed financial deepening by drawing a distinction between two types of external finance within an analytically tractable dynamic framework.

We motivated the existence of financial intermediaries through an agency problem between borrowers and lenders. Owner–managers of borrowing firms may shirk and thereby reduce expected investment returns. Outside investors are too diverse to monitor firms individually so that direct lending requires owner–managers to be paid a high rate of return. Faced with low returns on their investment, households are reluctant to lend to firms when they are the sole suppliers of external finance. They rely on the role of banks as delegated monitors instead. Banks are endowed with a monitoring technology that allows them to partially eliminate bad investment choices. When monitoring costs are modest, households earn higher returns by lending to these firms than they do on alternative assets. In the presence of bank finance, they agree to lend to firms.

The model economy is consistent with financial deepening: (a) rising capital-intensity of production is increasingly financed from external sources, (b) banks finance an increasing percentage of investment, as do financial markets and (c) the ratio of market-to-bank finance rises with development especially in the latter stages of development. Our analysis also suggests that policies that promote an efficient banking sector may have more desirable long-run consequences than those that promote deeper financial markets. Conversely, distorted banking incentives under financial repression leave economies more reliant on more expensive bank finance, hindering the process of financial and economic deepening.

This framework can be the starting point of further studies. One avenue would be to understand better how risky equity finance and less-risky bond finance affect firms' choice between intermediated and unintermediated finance over the course of development. There is scope too to augment the model with technology choice (established versus newer, riskier, ones) or structural change (ability of modern banks to supplant usurious money-lending, facilitate industrialisation) to better understand the role of financial sector reforms in delivering productivity growth and shared prosperity.

Acknowlegement

Thanks to Niloy Bose, Jo Anna Gray and Tridip Ray for helpful conversations and an anonymous referee of this journal for feedback. All errors are mine.

Appendix

A.1 Restrictions on unit monitoring cost *c*

Without loss of generality, let us set $\tau = 0$. This simplifies the algebra without making any qualitative difference to the results.

From $H(k) \equiv G(k)/[1 + c\gamma(k)]$, to have $H' > 0$, we need $(1 + c\gamma)G'/G > c\gamma'$ as long as $k \leq \hat{k}$.

Recall that $\gamma(k)$ is determined from Equation (29):

$$\Phi(\gamma) = \frac{\alpha}{G(k)^{2-\alpha}} \left[(1-\alpha)G(k) + \alpha b\right], \tag{A.1.1}$$

where $\Phi(\gamma) \equiv \varphi(\gamma)/(1 + c\gamma)^{1-\alpha}$. Taking total differentials in Equation (A.1.1), we get

$$\frac{\partial \gamma}{\partial k} = \frac{\alpha}{-\Phi_\gamma} \left[\frac{G'}{G}\right] \frac{1}{G^{2-\alpha}} \left[(1-\alpha)^2 G + \alpha(2-\alpha)b\right]$$

which is clearly positive because $\Phi_\gamma < 0$ (see further). Using Equation (A.1.1) in this, we obtain

$$\frac{\partial \gamma}{\partial k} = \frac{\Phi}{-\Phi_\gamma} \left[\frac{G'}{G}\right] \left[\frac{(1-\alpha)^2 G + \alpha(2-\alpha)b}{(1-\alpha)G + \alpha b}\right], \tag{A.1.2}$$

which is positive. Therefore, to have $H' > 0$, it is necessary that

$$1 + c\gamma > c\left[\frac{\Phi}{-\Phi_\gamma}\right] \left[\frac{(1-\alpha)^2 G + \alpha(2-\alpha)b}{(1-\alpha)G + \alpha b}\right]$$

Note first that the last term on the right-hand side is a decreasing function of k, so that it reaches its maximum at $k = 0$ when it takes the value 1 (because $G(0) = b$).

Thus, for $H' > 0$, it is sufficient to have

$$1 + c\gamma > c\left[\frac{\Phi}{-\Phi_\gamma}\right]. \tag{A.1.3}$$

From the definition of Φ, we obtain

$$\frac{\Phi_\gamma}{\Phi} = \frac{v'}{v} - \frac{(1-\alpha)c}{1 + c\gamma} < 0. \tag{A.1.4}$$

Note here that $-\Phi_\gamma/\Phi$ is increasing in c. From Equation (A.1.2), $\partial\gamma/\partial k$ is therefore decreasing in it; lower monitoring costs lead to faster increase in monitoring intensity (γ), and hence, the proportion of investment financed by banks (ϕ).

Combining Equation (A.1.4) with Equation (A.1.3), and noting that $\gamma \geq 0$, we obtain a sufficient condition for $H' > 0$:

$$c < \frac{1}{\alpha}\left[\frac{-v'}{v}\right]. \tag{A.1.5}$$

In other words, monitoring costs must be sufficiently small for external monitoring to result in higher investment over time.

Actually we impose a tighter restriction on c that will also ensure that the proportion of investment funded through market finance increases over time. For this, define $m_t \equiv M_t/I_t = 1 - \phi_t - b/H(k_t)$. To have $dm_t/dk_t > 0$ for $k_t < \hat{k}$, we need $bH'/H^2 > c\pi\gamma'/(1-\pi)$. Substituting for $H'(k)$ and $\gamma'(k)$ from above, we obtain the restriction as

$$b(1+c\gamma) > c\left[\frac{\Phi}{-\Phi_\gamma}\right]\left[b + \frac{\pi}{1-\pi}G(k)\right]\left[\frac{(1-\alpha)^2 G + \alpha(2-\alpha)b}{(1-\alpha)G + \alpha b}\right].$$

As before, a sufficient condition for this to be satisfied is

$$b(1+c\gamma) > c\left[\frac{\Phi}{-\Phi_\gamma}\right]\left[b + \frac{\pi}{1-\pi}G(k)\right],$$

or that,

$$c\left[\alpha + \frac{\pi}{1-\pi}\frac{G(k)}{b}\right] < \frac{-v'}{v},$$

or that,

$$c < \hat{c} \equiv \frac{-v'/v}{\alpha + \{\pi/(1-\pi)\}G(\hat{k})/b}. \tag{A.1.6}$$

Note that Equation (A.1.5) is satisfied whenever Equation (A.1.6) is satisfied. Henceforth, we assume that monitoring costs are lower than this critical value $\hat{c}$.

A.2 Monitoring cost is increasing in c

For total monitoring costs to be increasing in c, that is to have $\partial(c\gamma)/\partial c > 0$, we need $\varepsilon_{\gamma,c} < 1$. The elasticity of γ with respect to c is defined as, $\varepsilon_{\gamma,c} \equiv -\frac{c}{\gamma}\frac{\partial\gamma}{\partial c}$.

Now, from Equation (A.1.1), we get

$$\frac{\partial\gamma}{\partial c} = -\frac{\Phi_c}{\Phi_\gamma} = \frac{(1-\alpha)\gamma\varphi(\gamma)}{\varphi'.(1+c\gamma) - (1-\alpha)c\varphi}.$$

For $\varepsilon_{\gamma,c} < 1$, we require that $-\varphi' > 0$, which is true by assumption because $\varphi' = v'/(1-\pi) < 0$.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship and/or publication of this article.

Notes

1. Equities and bonds are indistinguishable in this model. In particular, if firms issue shares (claims to capital), these become worthless after one period because we assume that the depreciation rate is 100 per cent. Hence, equities must pay the same return as bonds, R^D.
   Modeling equities this way does not take into account some of the complexities of the stock market. But what is key for the present context is that equity- and bond-finance are both arm's-length lending. In some countries, such as the USA, shareholders often have rights that are aggressive enough to throw out a non-performing management. In most countries, that is not so. In other words, for most financial systems, it is sensible to assume that shareholders cannot effectively monitor firms (see Allen & Gale, 2000).
2. Alternatively imagine each entrepreneur having warm-glow altruism towards her single offspring: $U_t^E = c_{t+1}^E + b \ln z_{t+1}$, where z_{t+1} is parental bequest to the child that serves as the latter's internal equity as an entrepreneur in $t+1$. In equilibrium $z_t = b\ \forall t$.
3. In other words, these skills are hereditary and acquired by entrepreneur-j's offspring. An equivalent assumption is one where the technology for producing K^j is handed down from one generation to the next within the jth entrepreneurial family. Modeling the capital goods sector as monopolistically competitive enables us to derive a downward-sloping demand for external funds.
4. While entrepreneurs could invest their endowment in bank deposits or securities, they choose not to do so in equilibrium (see further).
5. As the entrepreneur consumes only in the second period of his/her life, we assume that he/she can store away these goods for future consumption. Storage is also assumed to be impossible to detect.
6. More generally, higher v_i projects can carry lower success probabilities π_i, as long as $v_i/(1-\pi_i)$ falls with v_i. Same holds for the success probability of the V-project vis-a-vis any v-project.
7. It may still be the case that under a severe agency problem (high $\bar{v}$, $\underline{v}$) monitored finance does not yield a return higher than storage. To rule this out, we assume that storage returns and monitoring costs are both low. See section V.
8. Additionally we need to check that entrepreneurs are better off by using external finance than without, so that there would be demand for external funds. This is always true, as long as, given the initial capital stock k_0, b is 'small enough'.
9. See Bencivenga et al. (1995) for a model of equity markets that takes into account such costs. For our analysis, it does not matter who bears the cost. For instance, we could have assumed that these are deducted from the return households get on their security holdings. Firms would then still face the cost $(1+\tau)R^D$ on direct finance as households seek to arbitrage between bank deposits and direct lending.
10. Equilibrium prices will be such that $M_t \geq 0$.
11. Note here that if the v projects differed in their πs, the markup would vary over time.
12. Recall that we have already assumed modest transactions costs ($\tau < (1-\pi)/\pi$) so that market finance is less expensive than bank finance. There may be economies of scale—at lower volumes of financial market activity (i.e., less developed markets), market finance may plausibly be more expensive than bank finance.

13. Recall that, given R^D_{t+1}, the cost of direct finance is $(1+\tau)R^D_{t+1}$ while that for bank finance is $(1+c\gamma_t)R^D_{t+1}$. Results from Lemma 2 allow us to analyse the impact of policies that reduce c or τ on capital accumulation.
14. Although a fall in c increases γ, total monitoring costs $c\gamma$ decline. See Appendix A.2.
15. One can think of an additional benefit that this model does not capture. If entrepreneurs had heterogenous distribution of internal funds, poorer entrepreneurs could be credit rationed. A lower c would allow some of these rationed entrepreneurs to obtain external finance (as monitoring effectively substitutes for lack of internal funds), while a lower τ would have no effect on them.

References

Allen, F., & Gale, D. (2000). *Comparing financial systems.* Cambridge, MA: MIT Press.

Bencivenga, V. R., Smith, B. D., & Starr, R. M. (1995). Liquidity of secondary capital markets: Allocative efficiency and the maturity composition of the capital stock. *Journal of Economic Theory*, *7*(1), 19–50.

Bose, N., & Neumann, R. (2015). An explanation for the diversity in the evolution of financial structure. *Macroeconomic Dynamics*, *19*(2), 270–287.

Boyd, J. H., & Smith, B. D. (1998). The evolution of debt and equity markets in economic development. *Economic Theory*, *12*(3), 519–560.

Chakraborty, S., & Ray, T. (2006). Bank-based and market-based financial systems: A growth-theoretic analysis. *Journal of Monetary Economics*, *53*(2), 329–350.

———. (2007). The development and structure of financial systems. *Journal of Economic Dynamics and Control*, *31*(9), 2920–2956.

Demirguc-Kunt, A., & Levine, R. (1996). Stock markets, corporate finance and economic growth: An overview. *World Bank Economic Review*, *10*(2), 223–240.

Demirguc-Kunt, A., & Maksimovic, V. (1996). Stock market development and financing choices of firms. *World Bank Economic Review*, *10*(2), 341–370.

Demirguc-Kunt, A., Feyen, E., & Levine, R. (2013). The evolving importance of banks and securities markets. *World Bank Economic Review*, *27*(3), 476–490.

Diamond, D. W. (1984). Financial intermediation and delegated monitoring. *Review of Economic Studies*, *51*(3), 393–414.

Farrell, D., Lund, S., Greenberg, E., Doshi, R., Rosenfeld, J., & Morin, F. (2006, May). *Accelerating India's growth through financial system reform.* San Francisco, CA: McKinsey Global Institute.

Greenwood, J., & Jovanovic, B. (1990). Financial development, growth and the distribution of income. *Journal of Political Economy*, *98*(5), 1076–1107.

Greenwood, J., & Smith, B. (1997). Financial markets in development, and the development of financial markets. *Journal of Economic Dynamics and Control*, *21*, 145–181.

Greenwood, J., Sanchez, J. M., & Wang, C. (2010). Financing development: The role of information costs. *American Economic Review*, *100*(4), 1875–1891.

Gurley, J. G., & Shaw, E. S. (1955). Financial aspects of economic development. *The American Economic Review*, *45*(4), 515–538.

———. (1967). Financial structure and economic development. *Economic Development and Cultural Change*, *15*, 257–268.

Hellwig, M. (1991). Banking, financial intermediation and corporate finance. In A. Giovannini & C. Mayer (Eds.), *European financial integration*. Cambridge: Cambridge University Press.

Holmstrom, B. (1996). Financing of investment in Eastern Europe. *Industrial and Corporate Change*, *5*(2), 205–237.

Holmstrom, B., & Tirole, J. (1997). Financial intermediation, loanable funds, and the real sector. *Quarterly Journal of Economics*, *112*(3), 663–691.

Khan, A. (2001). Financial development and economic growth. *Macroeconomic Dynamics*, *5*(3), 413–433.

Levine, R. (1991). Stock markets, growth and tax policy. *Journal of Finance*, *46*(4), 1445–1465.

The Economist. (2013, November 30). Financial-sector reform in India: Bridging the gulf, pp. 69–70.