

The Development and Structure of Financial Systems*

SHANKHA CHAKRABORTY[†]

TRIDIP RAY[‡]

Abstract

We introduce monitored bank loans and non-monitored tradeable securities as sources of external finance for firms in a dynamic general equilibrium model. Due to frictions arising from moral hazard, access to credit and each type of financial instrument are determined by the wealth distribution. We study the depth of credit markets (financial development) and conditions under which the financial system relies more on either type of external finance (financial structure). We identify initial inequality, investment size and institutional factors as key determinants of financial development, while an economy's financial structure is shaped by its investment technology and legal and financial institutions. The model's predictions are consistent with historical and recent development experience.

KEYWORDS: Financial Development, Financial Structure, Bank Finance, Market Finance, Credit Frictions.

JEL CLASSIFICATION: E44, G20, G30, O15, O16

*For helpful comments and suggestions we thank Joydeep Bhattacharya, Nancy Chau, Sudipto Dasgupta, Chris Edmond, Claudian Kwok, Raoul Minetti, Dilip Mookherjee, Kunal Sengupta, Yong Wang and seminar participants at various places this paper was presented. All remaining errors are ours.

[†]Department of Economics, University of Oregon, Eugene, OR 97403-1285. Email: shankhac@oregon.uoregon.edu

[‡]Department of Economics, Hong Kong University of Science & Technology, Clear Water Bay, Kowloon, Hong Kong. Email: tridip@ust.hk

1 Introduction

Researchers concerned with institutions and economic development seek to understand the dynamic process by which institutions evolve and interact with the rest of the economy. This paper analyzes the evolution of one such institution, the financial system, with the objective of isolating factors that shape its development and long-run character.

The role of a financial system in development depends on the degree to which industrialization benefits from external finance and is naturally determined by access to credit markets.¹ But in a second-best world infested with credit frictions, access to credit is constrained by wealth levels and internal asset positions of individuals and firms. There is thus an intimate connection between the wealth distribution and financial development.

We construct a dynamic model that incorporates this interdependence and allows for a sufficiently rich financial structure. Manufacturing industries require large-scale investment, an amount that cannot be funded simply from internal assets. Potential entrepreneurs may borrow using monitored bank loans (bank finance) or non-monitored sources like bonds and equities (market finance). Credit frictions arise from an agency problem: owner-managers of manufacturing firms may choose an inferior technology in order to enjoy private benefits. The incentive to do so is greater the lower the personal stake an owner has in her investment project (Holmstrom and Tirole, 1997).

In presence of this moral hazard, wealth thresholds determine who invests, using what type of financial instrument. Poorer individuals do not obtain any funding since they cannot guarantee lenders the required return. They work instead for a wage. Others who obtain loans produce capital using a risky technology. When investment succeeds, these capitalists hire workers to operate the machinery and produce a final good. Among capitalists, wealth thresholds again determine how they borrow. Individuals of medium wealth levels are able to borrow only through a combination of intermediated (bank) and unintermediated (market) finance. Bank finance entails monitoring that partially eliminates the incentive problem; perceiving this, direct lenders are willing to lend.² Capitalists who are wealthy enough, on the other hand, do not need to be monitored and use only market finance.

The wealth distribution thus determines access to credit markets (financial depth) as well as dependence on each type of external finance (financial structure). Those that are

¹An extensive literature exists on the role of financial systems in development. See Gerschenkron (1962), Goldsmith (1969) and Gurley and Shaw (1955) for early work, and Levine (1997) for a survey of more recent contributions.

²Costly monitoring makes bank finance a more expensive, but necessary, alternative to market finance.

rationalized out of the credit market, the wage earners, may ultimately accumulate enough assets to become capitalists. Whether or not they are able to do so depends on how high their income is, that is, on the extent of industrialization. At the same time, today's capitalists may find themselves denied access to credit in the future if they suffer losses on their current investment.

This dynamic interplay of the wealth distribution and borrowing choices determines the path of financial development and structure. We identify initial inequality, investment size and institutional factors as key determinants of financial development, while an economy's financial structure is shaped by its investment technology and legal and financial institutions.

An unequal distribution, specifically a large peasant population relative to capitalists, is inimical to financial development. When few individuals obtain credit, sparse industrialization prevents workers from earning enough to access credit markets in the future. Low to moderate degrees of inequality, on the other hand, see the emergence of developed financial systems. Even then, financial depth is negatively related to inequality, a prediction consistent with evidence from a cross-section of countries. Initial inequality also impacts an economy's historical financial structure – by reducing dependence on monitored finance it results in more market-oriented financial systems, a pattern we observed during industrialization in Western Europe.

The investment technology plays a crucial role. When investment requirements are too large relative to average wealth levels, fewer individuals obtain credit – industrialization and financial development remain low. An implication is that poorer countries that are characterized by high inequality, such as Latin America and sub-Saharan Africa, ought to rely more on small- and medium-scale industries that do not require large setup costs. An emphasis on import-substituting heavy industries, for instance, would be counterproductive in the long-run.

Larger capital requirements of industry also promote a bank-based financial system, at least during the initial stages of financial development. This outcome too matches well Western Europe's historical experience. In particular, the British industrial revolution occurred in industries, textiles for example, that did not call for enormous investments (Landes, 1969). Germany, in contrast, was involved in heavy manufacturing and chemicals, both of which required large injections of external capital. Such technological differences could have played a key role in Britain's historical reliance on market finance and Germany's dependence upon its banking systems (Baliga and Polak, 2002).³

³Using monitored and non-monitored debt in a static model of moral hazard, Baliga and Polak (2002) too

Although agents are risk-neutral in our model investment risk influences financial structure by affecting wealth dynamics. In particular, riskier technologies are more conducive to market-based systems because a larger fraction of capitalists can accumulate enough assets to be able to shed their reliance on bank intermediation. This result provides an interesting counterpoint to the typical portfolio effects analyzed in the corporate finance literature, although there is no consensus whether risk diversification is better handled in bank-based or market-based financial systems (Levine, 1997, Allen and Gale, 2001).

Institutional factors like agency costs shape a financial system in intuitively plausible ways. Bank-based systems result when bank monitoring is particularly efficient in resolving agency problems, although it depends upon monitoring costs as well. Our theoretical results lend credence to recent empirical evidence, LaPorta et al. (1997, 1998) and Levine (2002) in particular, that shows institutional and legal factors to be crucial determinants of financial structure.

This paper contributes to the emerging literature pioneered by Banerjee and Newman (1993) and Galor and Zeira (1993) on the dynamic link between credit frictions and wealth distribution. Recent work in this area include Aghion and Bolton (1997), Piketty (1997), Ghatak and Jiang (2002) and Mookherjee and Ray (2002, 2003). Similar to some of these papers, an important aspect of our analysis lies in the dependence of factor prices on wealth distribution. It is well-known that such factor price dependence may give rise to complicated non-linear dynamics, preventing one from drawing substantive conclusions about the economy. An advantage of our framework is its ability to circumvent this problem and flesh out interesting and empirically plausible predictions regarding financial development.

The key innovation we bring to the above literature is a rich financial structure. Treatment of financial systems in the existing literature has been somewhat incomplete since its primary goal has been to characterize distributional dynamics. What is missing, specifically, is the variety of financing choices that firms typically face. In contrast, our interest lies, first and foremost, in the development and structure of financial systems. A big part of the policy debate on developing and transition country financial institutions has centered on the pros and cons of bank-finance versus market-finance (Allen and Gale, 2000, Holmstrom, 1996, Demirgüç-Kunt and Levine, 2001a). An explicit dynamic model of financial structure is, therefore, a necessary starting point to address such policy issues. While firm financing

cite the influence of investment size on the Anglo-Saxon and German financial systems. Our results differ in one key respect: investment size affects only the initial financial structure and has no persistent effect in an industrialized society. Allen and Gale (2000) document how financial systems seem to be converging in recent decades with even traditionally bank-based economies moving toward financial markets.

choices are at the heart of corporate finance research, the literature there deals primarily with static (and often partial equilibrium) models of developed financial systems. Conclusions about developing societies are hard to infer without an understanding of how financial institutions evolve in the process of development. Our paper bridges the gap between these two strands of the literature.

We draw insights especially from Diamond (1991) and Holmstrom and Tirole (1997), both of which analyze the link between firm financing choices and some form of asset distribution. Diamond (1991) considers how firms switch from expensive (and monitored) bank finance to cheaper forms such as public debt as they develop better reputations.⁴ Holmstrom and Tirole (1997), whose incentive structure we adopt here, observe how incentive problems, and hence access to different types of external finance, depend on a firm's internal assets. However, neither of these papers incorporate the feedback that macro-fundamentals have on financing choices.

The paper is organized as follows. The model is developed in section 2 and optimal financial contracts characterized in section 3. Section 4 looks at the static general equilibrium while section 5 analyzes the dynamics and discusses model implications for financial development and structure. We conclude in section 6 indicating some possible extensions.

2 The Model

Consider a small open economy populated by a continuum of agents of measure one. Time is continuous and successive generations are connected by a bequest motive.

An agent is born with an initial wealth, a , received as bequest from her parent and a labor time endowment of one unit. This labor can be either supplied inelastically to the labor market, or used to oversee an investment project that produces capital. Inheritance is the sole source of heterogeneity among newly borns. We denote the cumulative distribution of agents at instant- t by $G_t(a)$ and assume that the initial distribution G_0 is continuous and differentiable.

Preferences are given by the “warm-glow” utility function:

$$u_t = c_t^\beta b_t^{1-\beta}, \quad \beta \in (0, 1),$$

where c denotes consumption and b denotes bequest left to offspring. Given a realized income

⁴Reputation is, of course, a form of asset.

z , optimal consumption and bequests are linear functions of z :

$$c_t = \beta z_t, \quad b_t = (1 - \beta)z_t. \quad (1)$$

The indirect utility function is then also linear in income, $U_t = \varphi z_t$ with $\varphi \equiv \beta^\beta(1 - \beta)^{1-\beta}$, implying that agents are risk-neutral.

Our formalization of the generational structure follows Banerjee and Newman (1993). Newly born agents become economically active only when they become ‘mature’; time to maturity, T , is distributed exponentially with the density function $h(T) = \eta e^{-\eta T}$, $\eta > 0$, across members of the same cohort. All economic activity occurs at the instant an agent becomes mature: she chooses her occupation and earns income accordingly, gives birth to one offspring, consumes, leaves bequests and dies.⁵ There is thus no population growth and members of a cohort do not all die at the same time. Without loss of generality we set $\eta = 1$ so that agents live for a unit length of time on average.⁶

2.1 Production and Occupation

Whether or not an individual is a worker or a capitalist is determined by access to external finance. Production of capital requires an indivisible investment of size q . Only individuals able to raise the requisite funds (from internal and external sources) become capitalists, the rest join the labor force.

A worker supplies her unit labor endowment to the labor market, earning a wage income w_t . A capitalist’s income, on the other hand, is uncertain. In particular the investment project is risky – a successful project yields capital amounting to θq ($\theta > 1$), while failure yields nothing. Successful capitalists become producers of final goods by hiring workers to operate the capital. This capital depreciates completely upon use.

Markets for the final good and for labor are perfectly competitive. For reasons that will become evident later, we assume an Arrow-Romer type technological spillover in the final goods sector. Specifically, for a successful capitalist j , the *private* technology for producing the unique consumption good is constant returns to scale in private inputs:

$$Y_t^j = (K_t^j)^\alpha (A_t N_t^j)^{1-\alpha}, \quad \alpha \in (0, 1), \quad (2)$$

⁵In other words, successive generations are non-overlapping, a framework quite distinct from continuous time OLG models such as Yaari (1965) or Burke (1996).

⁶Our motivation for using such a non-standard generational structure in continuous time is the same as Banerjee and Newman’s: it avoids complicated dynamics that could arise purely from a discrete time specification and simultaneous demographics. See also footnote 17 in section 5 below.

Here A_t denotes time-dependent labor-efficiency that is common to all final goods producers. Labor efficiency A depends upon capital per worker, k , through a learning-by-doing externality:

$$A_t = \hat{A}k_t. \quad (3)$$

Productivity improvements in any particular firm spills over instantaneously to the rest of the economy, becoming public knowledge. The *social* (intensive-form) production function is thus an Ak type technology $y_t = Ak_t$, where $A \equiv \hat{A}^{1-\alpha}$.

It remains to characterize the investment decision facing a potential capitalist. An individual with assets $a_t < q$ can become a capitalist only if she is able to borrow the shortfall $q - a_t$. To obtain a rich financial structure, we introduce an agency problem similar to that in Holmstrom (1996) and Holmstrom and Tirole (1997).

Specifically, the probability of success of investment depends upon an unobserved action taken by the capitalist – her choice on how to spend q . She can spend it on an efficient technology that yields θq units of capital with probability π_G , but uses up all of q . Alternatively, she can spend it on one of two inefficient technologies. One of these technologies is a low moral hazard project, costing $q - vq$, leaving vq for the capitalist to appropriate. This project too yields θq units of capital when it succeeds, but it succeeds less often, with probability $\pi_B < \pi_G$. The other inefficient technology is a high moral-hazard project, costing $q - Vq$. This leaves Vq in private benefits.

Both inefficient technologies carry the same probability of success, π_B , but since $0 < v < V < 1$ by assumption, the capitalist would prefer the high moral-hazard project over the low moral-hazard one. Only the efficient technology is, however, economically viable.⁷ The table below summarizes these investment choices.

TABLE 1

PROJECT	GOOD	LOW MORAL HAZARD (low pvt. benefit)	HIGH MORAL HAZARD (high pvt. benefit)
Private Benefit	0	vq	Vq
Success Probability	π_G	π_B	π_B

⁷To ensure this we make the following parametric assumption:

$$\pi_G \alpha A \theta q - r^* q > 0 > \pi_B \alpha A \theta q - r^* q + Vq, \quad (\text{Assumption 1})$$

where r^* is the world return on investment and, additionally, we anticipate that a successful capitalist's return per unit capital produced, ρ_t , equals αA in equilibrium (see equation (11) below).

3 The Financial Sector

Capital is perfectly mobile across borders so that this small open economy has free access to the international capital market. The time-invariant (gross) world rate of return on investment, r^* , is taken as given.

Supply of loans in the domestic financial sector comes from two sources: through financial intermediaries or banks, and directly from workers and international investors. Workers are indifferent between bank deposits, lending directly to capitalists and investing on the international capital market as long as all three yield a return of r^* . In other words, r^* is the return that banks promise their depositors as also the guaranteed return on direct lending.

On the demand side, loans are obtained by individuals who invest in the production of capital; they invest their entire wealth, borrowing the remainder from the domestic financial sector. Credit-constrained agents work for the capitalists. They deposit their wealth with banks or lend directly to domestic capitalists or the international capital market.

Capitalists face a perfectly elastic supply curve of loanable funds since they can freely access the international capital market. Availability of domestic investable resources is thus not an issue that concerns us. What is crucial is *where* capitalists obtain their loans from. Direct borrowing from domestic (workers) and foreign investors will be referred to as *direct* (or *market*) finance, and should be thought of as occurring through the purchase of one-period corporate bonds and equities. Borrowings intermediated by the banking sector will be called *indirect* (or *bank*) finance.

Bank finance plays a specific role. Banks are assumed to have a monitoring technology that allows them to inspect a borrowing capitalist's activities and ensure that she conforms to the terms agreed upon in the financial contract (Hellwig, 1991; Holmstrom and Tirole, 1997). Direct lenders (workers and foreign investors) do not possess this technology. Thus, as in Diamond (1984, 1991), banks are the delegated monitors.

Bank monitoring partially resolves the moral hazard problem and reduces a capitalist's opportunity cost of being diligent. By choosing to monitor borrowers, banks are able to eliminate the high moral-hazard project but not the low moral-hazard one (see Holmstrom, 1996 and Holmstrom and Tirole, 1997). For instance, a bank could stipulate conditions that prevent the firm from implementing the high moral-hazard project when it negotiates a loan contract with the bank. But such monitoring is privately costly for the bank and requires it to spend a nonverifiable amount γ per unit invested by the capitalist. Evidently, bank monitoring will be an optimal arrangement only if the gains from resolving the incentive problem is commensurate with monitoring costs.

The distinction between the two types of external finance is being kept relatively simple for reasons of tractability. The literature on corporate finance and financial intermediation discusses several aspects of bank- and market-finance (see Hellwig, 1991, Levine, 1997, Allen and Gale, 2000 and the references therein). But the key issue for us is that banks (or bank-like institutions) exist primarily to perform certain services that dispersed investors on the financial market do not or cannot provide. The Holmstrom-Tirole framework is a convenient way to capture this, without worrying too much about the specifics of the two types of external finance.

3.1 Optimal Contracts

Faced with the incentive problem outlined above, a capitalist will behave diligently to the extent that she receives an incentive-compatible expected payoff and whether or not she is monitored. Consider the financing options a prospective capitalist faces in borrowing from banks or from the market.

Since banks monitor firms while outside investors do not, we shall find it convenient to refer to the former as *informed* investors. We consider optimal contracts that induce investment in the good project.

Direct Finance

An optimal contract between a capitalist and direct financiers has a simple structure. Capitalist- i invests her entire wealth, a_t^i , on her own project since that yields a strictly higher return than she would otherwise obtain. Direct lenders provide the remaining, $q - a_t^i$. Neither party is paid anything if the investment fails. When the project succeeds, the capitalist earns an amount $x_t^C > 0$ while uninformed investors are paid $x_t^U > 0$. Denote a successful capitalist's rate of return per unit capital produced by ρ_t . Since a successful project produces θq units of capital, we have $x_t^C + x_t^U = \rho_t \theta q$.⁸

In order to invest in the good project, capitalist- i must earn an incentive compatible expected income. Moreover, the contract should satisfy each lender's participation constraint,

⁸Since project returns are observable and verifiable, optimal contracts between direct financiers and capitalists may be interpreted either as debt or as outside equity. For an equity contract, the capitalist sells a share s_t of her project return, $x_t^U = s_t (\rho_t \theta q)$. For a debt contract, the capitalist borrows $q - a_t$, promising to repay a return of r^* in case of success. The implicit return on equity has to be r^* for both assets to be held simultaneously, that is, $s_t (\rho_t \theta q) = r^* (q - a_t)$. Again, what matters is that neither of these is monitored lending.

that is, lenders should be guaranteed at least as much as they would earn on the international capital market. Combining these two constraints, we obtain (refer to Appendix A.1 for details) that only capitalists with wealth exceeding \bar{a}_t would be able to obtain direct finance, where

$$\bar{a}_t \equiv \frac{q}{r^*} \left[\frac{\pi_G}{\pi_G - \pi_B} V - \{\pi_G \rho_t \theta - r^*\} \right]. \quad (4)$$

Indirect Finance

Indirect or *intermediated* finance entails three parties to the contract: the bank, besides uninformed investors and the capitalist. An optimal contract here too stipulates that no one earns anything when the project fails. In case of success, total returns, $\rho_t \theta q$, are divided up as $x_t^C + x_t^U + x_t^B = \rho_t \theta q$, with x_t^B denoting the bank's returns.

Besides the incentive compatibility constraint of the capitalist and the participation constraint of the uninformed investors, we have to take into account an additional incentive compatibility constraint, that for bank monitoring. At the same time the loan size has to be chosen optimally so as to maximize bank profits subject to the capitalist's incentive constraint and the bank's incentive and resource constraints. Moreover, in a competitive equilibrium, the banking sector earns zero profits. Together these have the following implications (see Appendix A.1): (i) bank finance is relatively more expensive than direct finance (due to monitoring costs), that is, the (gross) loan rate charged by the bank, r_t^L , is greater than r^*

$$r_t^L = \left(\frac{\pi_G}{\pi_B} \right) r^* > r^*, \quad (5)$$

and (ii) capitalists with wealth $\bar{a}_t > a_t^i \geq \underline{a}_t$, where

$$\underline{a}_t \equiv \frac{q}{r^*} \left[\frac{\pi_G v}{\pi_G - \pi_B} - \{\pi_G \rho_t \theta - (1 + \gamma) r^*\} \right], \quad (6)$$

are able to convince uninformed investors to supply the remaining funds for the investment project only after the bank lends an amount (and agrees to monitor)⁹

$$l_t^i = \gamma \left(\frac{\pi_B}{\pi_G - \pi_B} \right) q. \quad (7)$$

⁹In order that the loan size does not exceed investment size, that is $l_t^i \leq q$, monitoring costs should not be so high as to make it impossible for bank intermediation to resolve agency problems. Hence, we restrict monitoring cost such that:

$$\gamma \leq (\pi_G - \pi_B) / \pi_B. \quad (\text{Assumption 2})$$

Also, it is natural to assume that $\bar{a}_t > \underline{a}_t$, or else there will be no demand for intermediation since monitoring would be too costly to be socially useful. This is true as long as the expected gain from monitoring exceeds

Capitalists with wealth level below \underline{a}_t cannot obtain any external finance, direct or indirect.

3.2 Occupational Incomes

Denote individual- i 's income by z_t^i . Consider the case where i does not obtain any external finance since her wealth is too small, $a_t^i \leq \underline{a}_t$. Her income consists of labor earnings and returns on investment in the domestic and/or international capital market

$$z_t^i = w_t + r^* a_t^i.$$

Next consider those who borrow both from banks and the market, that is, using *mixed* finance. For these capitalists with $a_t^i \in [\underline{a}_t, \bar{a}_t)$, equations (5) and (7) imply that income from a successful project would be

$$z_t^i = \rho_t \theta q - r_t^L l_t^i - r^* [q - l_t^i - a_t^i] = [\rho_t \theta - (1 + \gamma)r^*] q + r^* a_t^i.$$

Failure gives them zero returns.

Finally, capitalists with adequate wealth, $a_t^i \geq \bar{a}_t$, borrow only from the market and earn

$$z_t^i = \rho_t \theta q - r^* [q - a_t^i] = (\rho_t \theta - r^*) q + r^* a_t^i,$$

from a successful project. Of course, we are assuming that the rate of return from the project ($\rho_t \theta$) is high enough for the capitalist's participation constraint to be satisfied. That is, we require that the capitalist's expected income $\pi_G z_t^i$ is greater than $r^* a_t^i$, what she could earn for sure by investing her entire wealth on the domestic and international capital markets. This will be true under appropriate restrictions on θ and the final goods technology (α, A).

Earnings for each type of economic agent are thus given by

$$z_t^i(a_t^i) = \begin{cases} w_t + r^* a_t^i, & \text{for } a_t^i \in [0, \underline{a}_t) \\ \left. \begin{array}{l} [\rho_t \theta - (1 + \gamma)r^*] q + r^* a_t^i, \text{ with prob. } \pi_G \\ 0, \text{ otherwise} \end{array} \right\} & \text{for } a_t^i \in [\underline{a}_t, \bar{a}_t) \\ \left. \begin{array}{l} (\rho_t \theta - r^*) q + r^* a_t^i, \text{ with prob. } \pi_G \\ 0, \text{ otherwise} \end{array} \right\} & \text{for } a_t^i \in [\bar{a}_t, \infty) \end{cases} \quad (8)$$

its cost:

$$\left[\frac{\pi_G}{\pi_G - \pi_B} \right] (V - v) \geq \gamma r^*. \quad (\text{Assumption 3})$$

To summarize properties of optimal loan contracts and external financing choices, we note that given $q > a_t^i$ and the wealth distribution G_t ,

- (i) individuals with $a_t^i < \underline{a}_t$ are unable to obtain any external finance and work as laborers;
- (ii) individuals with $a_t^i \in [\underline{a}_t, \bar{a}_t)$ obtain external finance from banks as well as households: they borrow an amount l_t^i from banks at the loan rate r_t^L , given by (5) and (7) above, agree to being monitored, and raise the remaining $(q - l_t^i - a_t^i)$ directly from investors at the rate r^* ; optimal contracts guarantee these capitalists incentive compatible payments such that they behave diligently;
- (iii) individuals with $a_t^i \geq \bar{a}_t$ borrow only from investors, paying them a return of r^* ; here too, incentive compatible payments to these capitalists ensure that investments occur in the good project; and
- (iv) income in each case is given by (8) above.

4 Static Equilibrium

Given the wealth distribution G_t , occupational choices determine incomes according to (8). An agent's occupational choice depends upon her wealth relative to the cutoffs \underline{a}_t and \bar{a}_t . Individuals sort themselves into three categories: those who work, those who become capitalists by borrowing through intermediated and unintermediated loans (*mixed-finance* capitalists), and those who become capitalists by borrowing solely from the market (*market-finance* capitalists).

Parametric assumptions we make below ensure that workers earn a strictly lower income than either type of capitalist. Moreover, market-finance capitalists earn a higher expected income than mixed-finance capitalists (see equation (8)). Which of the three occupations an individual “chooses” is thus solely determined by her wealth. If wealth were not a constraint, all agents would want to become market-finance capitalists.

Consider the static equilibrium at time t . Denote the fractions of the three types of agents by (f_{1t}, f_{2t}, f_{3t}) , where

$$f_{1t} = G_t(\underline{a}_t), f_{2t} = G_t(\bar{a}_t) - G_t(\underline{a}_t), f_{3t} = 1 - G_t(\bar{a}_t).$$

Clearly $\sum_{\ell} f_{\ell t} = 1$, and at any instant t , there are f_{1t} workers and $1 - f_{1t}$ capitalists. Given the law of large numbers, π_G proportion of these capitalists succeed in producing capital,

amounting to $K_t^j = \theta q$ each. The aggregate capital stock is then $K_t = \pi_G \theta q (1 - f_{1t})$ and the workforce $N_t = f_{1t}$. Capital per worker is, thus,

$$k_t = \pi_G \theta q \left[\frac{1 - f_{1t}}{f_{1t}} \right]. \quad (9)$$

Since all successful capitalists produce the same amount of capital, given w_t , they hire the same number of workers

$$N_t^j = \frac{f_{1t}}{\pi_G (1 - f_{1t})}.$$

Note the private technology (2). In equilibrium, substituting for the labor augmenting technological progress ((3) and (9)) into this production function gives output produced by a successful capitalist as $Y_t^j = A \theta q$.

Under competitive markets, the equilibrium wage rate is given by the private marginal product of capital,

$$w_t = (1 - \alpha) A k_t = (1 - \alpha) \pi_G A \theta q \left[\frac{1 - f_{1t}}{f_{1t}} \right]. \quad (10)$$

A successful capitalist then earns the income $\widehat{Y}_t^j = \alpha A \theta q$, net of wage payments $w_t N_t^j$, from her capital θq . The (gross) rate of return on capital, which we defined as ρ_t above, is clearly equal to αA , the private marginal product of capital, that is,

$$\rho_t = \alpha A. \quad (11)$$

Due to overall constant returns to capital, this return is time-invariant. Since all successful capitalists earn the same return on their capital, we assume, without loss of generality, that they produce final goods using only their own capital.

Using the equilibrium return on capital from (11) the cutoff wealth levels defined by (6) and (4) now do not depend upon time:

$$\underline{a} = \delta_1 q, \quad \bar{a} = \delta_2 q, \quad (12)$$

where¹⁰

$$\begin{aligned} \delta_1 &\equiv [v \pi_G / (\pi_G - \pi_B) - \{\pi_G \alpha \theta A - (1 + \gamma) r^*\}] / r^*, \\ \delta_2 &\equiv [V \pi_G / (\pi_G - \pi_B) - \{\pi_G \alpha \theta A - r^*\}] / r^*. \end{aligned}$$

It remains to check whether or not a worker earns lower income than a capitalist. This is by no means guaranteed. For instance, when there are “too few” workers, the marginal

¹⁰Assumptions 1 and 3 ensure that $\delta_1 < \delta_2 < 1$.

product of labor may be so high that even individuals who could have obtained external finance choose to work. It turns out that this happens when the proportion of credit-constrained agents falls below \tilde{f}_1 , where \tilde{f}_1 satisfies

$$(1 - \alpha)\pi_G\alpha A \left[\frac{1 - \tilde{f}_1}{\tilde{f}_1} \right] + r^*\underline{a} = \pi_G [\{\alpha A\theta - (1 + \gamma)r^*\}q + r^*\underline{a}].$$

We restrict ourselves to empirically plausible distributions, those that are positively skewed. We assume hence that G_0 satisfies $f_{10} > \tilde{f}_1$. This ensures that occupational “choice” is stable over time, and we can simply focus on the proportions of the three types of agents without having to worry about the effect of income on occupational dynamics.

5 Dynamics of Financial Development and Structure

The financial system, by which we mean the degree to which an economy relies upon external finance in general, and bank and market-finance in particular, is determined by access to credit markets. Our purpose is to trace the evolution of the financial system and isolate factors that influence its dynamics and character. Drawing upon the instantaneous equilibrium from the previous section, we now turn to this analysis.

Given an initial wealth distribution G_0 , wealth thresholds \underline{a} and \bar{a} determine the proportion of individuals able to invest and the relative composition of bank- and market-finance in aggregate investment. These investment choices lead to income realizations that determine the subsequent distribution through bequests. The process continues recursively, with financial development tracking the wealth distribution through time.

Substituting labor and capital’s equilibrium returns into (8), and using optimal bequests (1), we obtain the intergenerational law of motion:

$$b_t = \left\{ \begin{array}{ll} (1 - \beta) \left[r^*a_t + (1 - \alpha)\pi_G A\theta q \left(\frac{1 - f_{1t}}{f_{1t}} \right) \right], & \text{for } a_t \in [0, \underline{a}) \\ (1 - \beta) [r^*a_t + \{\alpha A\theta - (1 + \gamma)r^*\}q], \text{ with prob. } \pi_G \\ 0, \text{ otherwise} & \left. \vphantom{\begin{array}{l} (1 - \beta) [r^*a_t + \{\alpha A\theta - (1 + \gamma)r^*\}q], \text{ with prob. } \pi_G \\ 0, \text{ otherwise} \end{array}} \right\} \text{ for } a_t \in [\underline{a}, \bar{a}) \quad (13) \\ \\ (1 - \beta) [r^*a_t + (\alpha A\theta - r^*)q], \text{ with prob. } \pi_G \\ 0, \text{ otherwise} & \left. \vphantom{\begin{array}{l} (1 - \beta) [r^*a_t + (\alpha A\theta - r^*)q], \text{ with prob. } \pi_G \\ 0, \text{ otherwise} \end{array}} \right\} \text{ for } a_t \in [\bar{a}, \infty) \end{array}$$

Figure 1 depicts this wealth dynamics for various possibilities (see below); the dotted lines represent expected income from investment.

To ensure that this mapping is convergent we need a particular assumption. All three regimes of (13) are piecewise linear, with slopes equal to $(1 - \beta)r^*$. In order to rule out a dynasty from becoming arbitrary rich over time by simply reinvesting its wealth, we require that

$$(1 - \beta)r^* < 1. \quad (\text{Assumption 4})$$

We would also like to rule out a dynasty from being able to self-finance its entire investment. When investment succeeds, the fixed-point of the mapping for $a_t \in [\bar{a}, \infty)$ is given by $a^U = (1 - \beta)(\alpha A\theta - r^*)q/[1 - (1 - \beta)r^*]$. For this to be less than q , we assume that

$$(1 - \beta)\alpha A\theta < 1. \quad (\text{Assumption 4'})$$

Investment is undertaken only if the return from it, $\alpha A\theta$, is greater than the return to be paid to lenders, r^* . Assumption 4' is sufficient to ensure that Assumption 4 is satisfied. We maintain it henceforth and, without loss of generality, restrict ourselves to distributions on the domain $[0, a^U]$.

In tracing the evolution of the financial system we first note that distributional dynamics are essentially nonlinear. The current wealth distribution and the threshold \underline{a} determine the size of the working class (f_{1t}) which then determines equilibrium wages through (10). This endogeneity of the wage rate gives rise to nonlinear dynamics since the future wealth distribution depends upon wages via optimal bequests.

It is well-known that there are no general mathematical methods for analyzing such nonlinear dynamics and the dynamic behavior of such systems can be quite complex. But in tracking the financial system our objective is somewhat modest. We do not need to completely characterize wealth dynamics, simply tracking the evolution of (f_{1t}, f_{2t}, f_{3t}) is sufficient for our purpose.¹¹ Besides, a couple of specific features of our model greatly simplifies the analysis. In the first place, there is no feedback from the wealth distribution to \underline{a} and \bar{a} , which are independent of time.¹² This eases the task of analyzing the dynamics of (f_{1t}, f_{2t}, f_{3t}) . Secondly, constant returns to capital at the aggregate level ($\rho_t = \alpha A$) guarantees that recursion dynamics for wealth levels exceeding \underline{a} is independent of time.¹³

¹¹We are interested in two features of a financial system, its depth and structure. Financial depth is captured by $(1 - f_{1t})$, the proportion of unconstrained borrowers, while the financial structure is characterized by the relative measure of capitalists relying on bank-finance (f_{2t}) and market-finance (f_{3t}).

¹²This is a feature our paper shares with Banerjee and Newman (1993) and Galor and Zeira (1993). See also Ghatak and Jiang (2002).

¹³See Aghion and Bolton (1997) and Piketty (1997) for models where returns to capital depend upon the wealth distribution.

Specifically, wealth dynamics for the two upper categories are not affected by the endogeneity of the wage rate which impacts only working-class dynamics. By exploiting the investment technology, the feature that investment failure yields zero returns, here too we are able to precisely characterize the dynamics.

5.1 Financial Development

The degree of credit rationing among potential capitalists determines the depth of a financial system. The simplest measure of financial development hence comes from observing the movement of f_1 .

We begin by reconsidering Figure 1.¹⁴ Suppose that a λ_t fraction of the f_{1t} working dynasties leave bequests exceeding \underline{a} . This implies that the offsprings of these $\lambda_t f_{1t}$ workers are able to borrow and become capitalists, once they are economically active. Figures 1(a)-(c) differ only in the position of the lowest regime relative to \underline{a} , and hence, in λ_t .

In Figure 1(a), the wealth recursion line for $[0, \underline{a}]$ lies entirely above \underline{a} so that $\lambda_t = 1$. This happens when the wage rate is high enough, that is, when there are fewer workers:

$$f_1 \leq \underline{f}_1 \equiv \left[1 + \frac{\underline{a}}{(1-\beta)(1-\alpha)\pi_G A \theta q} \right]^{-1}.$$

We characterize dynamics on the two-dimensional unit simplex in (f_1, f_3) . Since $\sum_{\ell} f_{\ell} = 1$, this is sufficient to determine the time-path of f_{2t} . Suppressing time subscripts, transition dynamics when $f_1 \leq \underline{f}_1$ is given by the pair of differential equations

$$\begin{aligned} \dot{f}_1 &= (1 - \pi_G)(f_2 + f_3) - f_1 = (1 - \pi_G) - (2 - \pi_G)f_1, \\ \dot{f}_3 &= \pi_G f_2 - (1 - \pi_G)f_3 = \pi_G(1 - f_1) - f_3. \end{aligned}$$

The first equation follows from noting that the outflow from the stock of workers is f_1 whereas the inflow comes from the fraction $(1 - \pi_G)$ of capitalists who suffer losses on their investment and lose their entire wealth. The mass of workers, f_1 , decreases or increases over time according to whether f_1 exceeds $(1 - \pi_G)(f_2 + f_3)$. The second differential equation is obtained similarly: the stock of market-finance capitalists increases as long as mixed-finance capitalists moving up, $\pi_G f_2$, are more numerous than market-finance capitalists suffering losses, $(1 - \pi_G)f_3$. Now turn to Figure 2 for the phase-plane: when $f_1 \leq \underline{f}_1$, the $\dot{f}_1 = 0$

¹⁴We assume, for now, that returns from successful investment are high enough; specifically, successful mixed-finance capitalists become wealthy enough so that their offsprings are able to borrow using market finance only.

locus is given by the equation $f_1 = (1 - \pi_G)/(2 - \pi_G)$ while the $\dot{f}_3 = 0$ locus is given by $f_3 = \pi_G(1 - f_1)$.

Figure 1(b) looks at another possibility, where the lowest regime of the transition mapping lies entirely below \underline{a} . None of the working dynasties leave bequests exceeding \underline{a} here which means $\lambda_t = 0$. This happens when the wage rate is low enough, that is, workers are more numerous:

$$f_1 > \bar{f}_1 \equiv \left[1 + \frac{\underline{a}[1 - (1 - \beta)r^*]}{(1 - \beta)(1 - \alpha)\pi_G A \theta q} \right]^{-1}.$$

The corresponding transition dynamics is given by:

$$\begin{aligned} \dot{f}_1 &= (1 - \pi_G)(f_2 + f_3) = (1 - \pi_G)(1 - f_1), \\ \dot{f}_3 &= \pi_G f_2 - (1 - \pi_G)f_3 = \pi_G(1 - f_1) - f_3. \end{aligned}$$

In Figures 2(a) and (b), the $\dot{f}_1 = 0$ locus is given by $f_1 = 1$ and the $\dot{f}_3 = 0$ locus by $f_3 = \pi_G(1 - f_1)$, when $f_1 > \bar{f}_1$.¹⁵

A third possibility arises when the wealth recursion line on $[0, \underline{a})$ lies neither fully above nor fully below \underline{a} . This occurs for $\underline{f}_1 < f_1 < \bar{f}_1$. In Figure 1(c), working dynasties distributed on $[\tilde{a}_t, \underline{a})$ leave bequests exceeding \underline{a} , those on $[0, \tilde{a}_t)$ do not. For a scenario like this, λ_t would depend upon the exact distribution on $[0, \underline{a})$ in general. But a moment's reflection shows we do not need details about the distribution on this interval; information about f_{1t} alone is sufficient to determine the dynamics.

To see this, we establish first that λ_t is a monotonically decreasing function of f_{1t} . An increase in f_{1t} lowers the wage rate by increasing the supply of labor; this raises \tilde{a}_t and, *ceteris paribus*, lowers λ_t . Obviously how an increase in f_{1t} gets distributed on $[0, \underline{a})$ matters, which is why detailed information about G_t may be necessary. But recall that investment failure yields zero income, which means all *new* workers start out with zero wealth. The pool of workers increases when the influx of capitalists whose investments have failed exceeds the outflux of workers who have accumulated wealth beyond \underline{a} . This means an increase in f_{1t} results in a bulging of the distribution at zero; hence such an increase further reduces λ_t .¹⁶

In addition, λ_t is a continuous function of f_{1t} . The continuous time demographic structure¹⁷ and a continuous initial distribution imply that changes in G_t and f_{1t} (and hence in

¹⁵ $\bar{f}_1 > \underline{f}_1$ since $(1 - \beta)r^* < 1$.

¹⁶Note the crucial role played by the investment technology. If failure resulted in low, but positive, returns, we would need more information about the distribution to determine how λ_t responds to G_t .

¹⁷A discrete-time framework would have led to several complications. For example, optimal trajectories of f_{1t} could be oscillatory even if convergent. Since such behavior would arise solely from a discrete time specification, we have adopted a continuous-time formalization instead.

\tilde{a}_t) occur in a continuous fashion. Thus λ_t , defined by $1 - G_t(\tilde{a}_t)/f_{1t}$, also moves continuously with f_{1t} .

We can therefore specify the dynamics corresponding to Figure 1(c) by the differential equations:

$$\begin{aligned}\dot{f}_1 &= (1 - \pi_G)(1 - f_1) - \lambda(f_1)f_1, \\ \dot{f}_3 &= \pi_G(1 - f_1) - f_3.\end{aligned}$$

Appendix A.2 demonstrates the existence of the $\dot{f}_1 = 0$ locus for a continuous $\lambda(f_{1t})$. Multiple such loci are possible but, generically, there will be an odd number of these.

Figures 2(a) and (b) illustrate dynamics under one and three such loci respectively. In both cases, when $f_{10} \leq \underline{f}_1$, point D represents a locally stable stationary distribution, while point L is a locally stable stationary distribution for $f_{10} > \overline{f}_1$. Point D , in fact, represents a well-developed financial system and is given by $(f_1^*, f_2^*, f_3^*) = \left(\frac{1-\pi_G}{2-\pi_G}, \frac{1-\pi_G}{2-\pi_G}, \frac{\pi_G}{2-\pi_G}\right)$. Point L likewise represents a less-developed financial system. Indeed, there we have $(f_1^{**}, f_2^{**}, f_3^{**}) = (1, 0, 0)$, that is, a complete collapse of the financial structure.

In Figure 2(a), \hat{f}_1 acts as a threshold. For values of f_{10} below \hat{f}_1 , the economy converges to the developed financial system, whereas for values above \hat{f}_1 , the long-run outcome is the primitive system. For three loci (f_1^a, f_1^b, f_1^c) , as in Figure 2(b), the intermediate one acts as a local attractor for $f_{10} \in (f_1^a, f_1^c)$. In addition to the developed and underdeveloped financial systems, we now have a third kind, a moderately developed financial structure, at point M .

The complete collapse of the financial system at point L is an unattractive outcome, a consequence of there being no way out of the working class when $f_1 > \overline{f}_1$. Following Banerjee and Newman (1993), perhaps it makes sense to get rid of this extreme result by perturbing the dynamics slightly. We do so by allowing a very small probability (ξ) of moving up from the working- to the middle-class. This probability corresponds to winning a lottery or some other form of windfall gain not captured by the model.

The phase diagram for one such perturbation is Figure 2(c). When $f_1 \leq \underline{f}_1$, this perturbation does not alter the wealth dynamics since $\lambda_t = 1$. When $f_1 > \overline{f}_1$, perturbed wealth dynamics is given by

$$\begin{aligned}\dot{f}_1 &= (1 - \pi_G)(f_2 + f_3) - \xi f_1 = (1 - \pi_G) - (1 + \xi - \pi_G)f_1, \\ \dot{f}_3 &= \pi_G(1 - f_1) - f_3.\end{aligned}$$

The perturbed locus $\dot{f}_1 = 0$ (when $f_1 > \overline{f}_1$) lies to the left of the original one while the $\dot{f}_3 = 0$ locus remains unchanged. The stationary distributions are now represented by D

and L' , both of which are locally stable. L' still represents a very under-developed financial structure, but we now have both f_2^{**} and $f_3^{**} > 0$.

Thus developed, underdeveloped and even moderately-developed financial systems may emerge depending upon the initial measure of credit-constrained individuals, f_{10} . High values of f_{10} are particularly inimical to financial development. For low to moderate values, a developed financial system results in the long-run, but even here, the degree to which it develops may depend upon initial conditions.

Recall that $f_{10} \equiv G_0(\underline{a})$. Clearly, f_{10} depends on the initial distribution (G_0) and on the factors that determine \underline{a} – investment size (q) and institutional parameters (v and γ).¹⁸ We consider these determinants one by one.

Inequality and Financial Development

Wealth distribution is a key predictor of financial development. Suppose a high value of f_{10} results from a highly unequal initial distribution. Figure 2 shows how a developed financial system (D) results for values of f_{10} less than \underline{f}_1 , that is, for a relatively equal distribution. Under a relatively unequal distribution, individual wealth levels will be more commonly below \underline{a} . When $f_{10} > \overline{f}_1$, these inequities hamper development leading to a ‘collapse’ of credit transactions (Figures 2(a) and 2(b)), or more plausibly, an underdeveloped financial system (Figure 2(c)). For moderate degrees of inequality such that $\underline{f}_1 < f_{10} < \overline{f}_1$, outcomes depend upon the existence of additional stationary distributions. Persistent financial underdevelopment results in Figure 2(a) for values of f_{10} above \widehat{f}_1 , whereas the economy converges to a developed financial regime for values below \widehat{f}_1 . In Figures 2(b) or 2(c), a moderately developed financial system arises when $f_1^a < f_{10} < f_1^c$. Not as many individuals become capitalists in this case as they would when $f_{10} < f_1^a$, but credit-constrained borrowers are less numerous than at L' , and credit markets thicker.

Our model thus predicts that concentrated ownership of assets like land and natural resources, which directly or indirectly determine access to investment opportunities, would hamper financial development and industrialization. Initial inequities persist when lower capital accumulation significantly depresses income for the working classes, the potential entrepreneurs of the future.

This negative association between inequality and financial development finds support with available evidence. Using data on income Gini from Deininger and Squire (1996) and indices of financial depth from Levine (2002), we obtained a negative correlation of -0.49

¹⁸An inspection of (12) shows that \underline{a} depends positively upon investment size, q , and on the agency problems parameters v and γ .

between inequality and financial development.¹⁹ Indeed, some of the financially least developed countries in Levine (2002) (mostly from Latin America and Africa) are, at the same time, characterized by severe distributional problems.

This is all the more evident when we contrast middle-income countries in East Asia and Latin America. East Asia's better asset and income distribution has received considerable attention in the development literature;²⁰ our model relates how this difference may have been vital for their financial development. Latin American nations like Argentina, Brazil, Chile, Peru and Venezuela are by and large financially underdeveloped while Hong Kong, Malaysia, South Korea, Taiwan, Thailand and Singapore's financial systems are comparable to those in Western Europe and North America.²¹

Investment Size

Consider next the effect of investment size (q). An increase in q raises the cutoff \underline{a} , given the wealth distribution. This could lead to financial underdevelopment if the economy is pushed over \hat{f}_1 or f_1^c . An immediate implication is that poorer countries which are characterized by high inequality, such as Latin America and sub-Saharan Africa, ought to rely more on small- and medium-scale industries for their development. An emphasis on import-substituting heavy industries, for instance, would be counterproductive in the long-run.

Institutional Factors

Individuals do not differ in terms of their innate abilities in our model and access to credit markets is limited solely by informational asymmetries and costs. A relevant question is how better institutions mitigate these asymmetries and what that implies for financial structure. A simple way to interpret institutions here is through the parameters γ , v and V . These parameters affect the depth and structure of a financial system through the nature and magnitude of agency problems and costs of controlling it.

As noted earlier, the degree of credit-rationing, f_{10} , depends positively upon the institutional parameters γ and v through \underline{a} . When legal and financial institutions are too inefficient

¹⁹For financial development, we use Levine's (2002) "Finance Aggregate" measure (Column 4, Table 2), constructed using indicators of financial activity, size and efficiency over the period 1980-95. For Levine's sample of 48 countries this index ranges from -2.2 to 1.88 . For this sample of 48 countries, income Gini ranges from 24.9 to 62.5 . We use income Gini for the year 1980 (as close as possible, permitted by availability) as a proxy for initial inequality. Details available upon request.

²⁰Income Gini was 34.6 in East Asia and 53 in Latin America during the 1960s; land Ginis were 44.8 and 82 respectively (Deininger and Squire, 1998, Tables 1 and 2).

²¹See Table 2 in Levine (2002) and Table 3.12 in Demirgüç-Kunt and Levine (2001b).

($f_{10} > \widehat{f}_1$ or $f_{10} > f_1^c$), the financial system remains underdeveloped in the long run; efficient institutions lead to financial development in the long-run. This prediction is along the lines of recent studies of the ‘legal-based’ view of financial development in LaPorta et al. (1997, 1998), where the quality and nature of legal rules and law enforcements protecting shareholders and creditors are seen as fundamental to financial activity. Systematic empirical support for this view, covering a wide range of countries, is offered by Levine (2002).

Inefficient institutions are more widespread and informational problems more acute in poorer countries presumably because better institutions are costly to implement. In our model, one way to get around such institutional bottlenecks is through a temporary income redistribution that relaxes credit constraints for a sizeable number of potential entrepreneurs. The benefits of such a redistribution will be persistent if it pushes f_{10} below the relevant threshold, and in fact, could be politically more palatable than permanent distributive policies such as land reforms.

5.2 Financial Structure

Financial structure refers to the combination of financial instruments, markets and institutions operating in an economy (Goldsmith, 1969). A change in that structure towards the ideal mode assumed in competitive general equilibrium theory is an essential feature of financial development.

Turning to the model’s implications for financial structure, we consider the relative importance of market- and bank-finance, $\psi_t \equiv f_{3t}/f_{2t}$, as an index of structure and study its evolution.

In the analysis so far we have implicitly allowed high investment returns to ensure that successful mixed-finance capitalists move up to the next wealth category. What if that were not the case? We illustrate such a scenario in Figure 4 which depicts wealth recursion dynamics for $f_1 \leq \underline{f}_1$ (the other two cases would parallel Figures 1(b) and (c)). It turns out that the nature of the transition dynamics does not change here although the composition of the stationary distribution does.

Transition dynamics corresponding to Figure 4 is given by:

when $f_1 \leq \underline{f}_1$,

$$\begin{aligned}\dot{f}_1 &= (1 - \pi_G) - (2 - \pi_G)f_1, \\ \dot{f}_3 &= -(1 - \pi_G)f_3;\end{aligned}$$

when $f_1 > \overline{f_1}$,

$$\begin{aligned}\dot{f}_1 &= (1 - \pi_G)(1 - f_1), \\ \dot{f}_3 &= -(1 - \pi_G)f_3;\end{aligned}$$

and when $\underline{f_1} < f_1 < \overline{f_1}$,

$$\begin{aligned}\dot{f}_1 &= (1 - \pi_G)(1 - f_1) - \lambda(f_1)f_1, \\ \dot{f}_3 &= -(1 - \pi_G)f_3.\end{aligned}$$

In Figure 5(a), when $f_{10} \leq \underline{f_1}$, D represents a locally stable stationary distribution with a ‘developed’ financial system, where $(f_1^*, f_2^*, f_3^*) = \left(\frac{1-\pi_G}{2-\pi_G}, \frac{1}{2-\pi_G}, 0\right)$. As before, when $f_{10} > \overline{f_1}$, point L represents a locally stable stationary distribution representing a ‘less-developed’ financial system with $(f_1^{**}, f_2^{**}, f_3^{**}) = (1, 0, 0)$, and when $f_1^a < f_{10} < f_1^c$, point M represents a moderately developed financial system. Not surprisingly, the long-run distribution has no capitalist relying purely on market finance since middle-class capitalists cannot move up.

As before, such an outcome can be avoided with perturbations that allow workers to move up to the middle-class with a small probability (ξ), and middle-class capitalists to move up with a similar probability (ε). Under perturbation, the stationary distributions in Figure 5(b) are represented by D' when $f_{10} \leq \underline{f_1}$, L' when $f_{10} > \overline{f_1}$ and M' when $f_1^a < f_{10} < f_1^c$. All three points are locally stable.

Compare now the financial structure of D (or D') in Figures 2 and 5 under different rates of return. When investment returns are low, all (or a very large proportion of) eligible capitalists go through bank intermediation in the long-run. That is, bank-finance is relatively more important when investment returns are lower (this would be true even for a moderately developed financial system). The long-run financial structure is, hence, more market-based for a configuration like Figure 1 and more bank-oriented for a situation like Figure 4.

A market-based system (Figure 1) occurs when the height of C exceeds \bar{a} , that is when,

$$(1 - \beta) r^* \left[\frac{\pi_G v}{\pi_G - \pi_B} + (1 - \pi_G) \alpha A \theta \right] \geq \frac{\pi_G V}{\pi_G - \pi_B} - (\alpha \pi_G A \theta - r^*). \quad (14)$$

This is more likely to occur when V is relatively low and v relatively high, or when θ is relatively high but π_G low (holding $\pi_G \theta$ constant).²² On the other hand a bank-based financial system (Figure 4) results when the height of H is less than \bar{a} . This happens if

$$(1 - \beta) r^* [(1 - \pi_G) \alpha A \theta - \gamma r^*] < \left(\frac{\pi_G}{\pi_G - \pi_B} \right) [1 - (1 - \beta) r^*] V - \pi_G \alpha A \theta + r^*. \quad (15)$$

²²We allow for a proportionate decrease in π_B as well.

This inequality is more likely when V and γ are relatively high, or, holding $\pi_G\theta$ constant, when θ is relatively low and π_G high.²³

Thus, the financial structure of an economy is determined by the investment technology (π_G, θ) and institutional factors $(\gamma, v$ and $V)$. We consider these issues next.

Investment Risk

Although banks and individuals are risk-neutral in our model, investment risk affects financial structure through its effect on wealth dynamics. We begin by contrasting two types of investment that yield the same expected return, $\pi_G\theta$: type-I projects yield a high θ but are riskier since π_G is low, while type-II projects succeed more often but realize low θ .²⁴

Suppose now that the two project types differ significantly in their riskiness so that Figure 1 depicts wealth dynamics for type-I projects while that for type-II projects is given by Figure 4. From (14) and (15) we know that Figure 1 is more likely to occur when θ is relatively high but π_G low (holding $\pi_G\theta$ constant) while Figure 4 is more likely to occur for the opposite case.

Figures 1 and 4 lead to dynamics shown by Figures 2 and 5 respectively. We draw two conclusions on the role of investment risk. First, lower π_G leads to higher f_1^* so that credit rationing is more widespread in the long-run. Secondly, when investment is less risky, all or a large proportion of eligible capitalists go through bank intermediation in the long-run. In other words, bank-finance is relatively more important for safer technologies, whereas market finance gains importance for riskier ones.

This dependence of financial structure on risk is quite distinct from, but complementary to, the ones commonly analyzed in the finance literature. Specifically, since agents are risk-neutral our analysis misses the typical portfolio effect discussed in the literature.²⁵ At the same time it brings to the analysis the macroeconomic feedback that investment risk has on asset positions and financing dynamics, an effect entirely absent from the existing literature.

Institutional Quality

Institutional parameters also affect long run financial structure in the model. Recall that a market-based system is more likely to occur (that is, (14) holds) when V is relatively low and v relatively high. On the other hand, a bank-based system (condition (15)) is more likely

²³Again, we change π_B by the same proportion as π_G .

²⁴We allow for a proportionate change in π_B between the two project types.

²⁵Risk averse agents would clearly make the analysis much less tractable. It is to be noted, though, that the literature is not unequivocal about whether banks or markets diversify risks better, suggesting only that both are important. Levine (1997) and Allen and Gale (2001) discuss these issues.

to occur when V and γ are relatively high. It is quite intuitive that a bank-based system is more likely when the residual moral hazard under bank monitoring (v) is low relative to what incentives would be in the absence of monitoring (V). This conclusion is consistent with Demirguc-Kunt and Levine’s (2001b) finding that countries with strong shareholder rights relative to creditor rights and strong accounting systems (that is, low V relative to v) tend to have more market-based systems.

It may seem surprising that higher monitoring costs, γ , lead to a more bank-oriented systems even though these costs are borne by the banking sector. This is easy to understand once we recognize that wealth and financing dynamics depend upon investment earnings. A higher cost of monitoring means that banks need to inject a larger amount of their own resources into the investment project. This forces mixed-finance capitalists to rely more heavily on expensive intermediated finance; consequently less of them are able to move up to become market-finance capitalists.

5.3 Historical Implications

Several of the model’s implications shed light on the historical development of financial systems during the Industrial Revolution.

Inequality and Financial Structure

Consider, for example, the effect of initial inequality on financing choices. For convenience, assume that the initial distribution is lognormal with mean μ_0 and variance σ_0^2 , where $\underline{a} < \mu_0 < \bar{a}$. In Appendix A.3 we establish that an increase in σ_0^2 tends to raise f_{10} , lower f_{20} but increase f_{30} . Higher inequality leads to thinner capital markets since $1 - f_{10}$ is lower. But among those who obtain loans, there is a shift toward market finance and away from bank finance, increasing the ratio ψ_0 . Historical reliance upon the two types of finance may, in other words, depend upon inequality. This prediction seems to be corroborated by what we know about England and Germany during the industrial revolution.

The Anglo-Saxon financial system, with its creditors pursuing more of a “hands-off” lending, was more market-oriented. Banks were mostly concerned with liquidity and did not engage in long-term lending so that British industries primarily depended upon internal finance and the London Stock Exchange for their financing needs (Collins 1995, Allen and Gale, 2000). German industries, in comparison, relied more on bank finance; German bankers kept a continuous watch over the development of companies and were often represented on the company boards (Allen and Gale, 2000, Baliga and Polak, 2002).

At the same time, substantial evidence suggests England had a more unequal land distribution than Germany (and France) (Clapham, 1936, Soltow, 1968). Landes (1969) also notes that a large number of British industrialists were “men of substance”, having accumulated significant wealth from merchant activities. This distributional difference between the two regions could partly explain why the Anglo-Saxon and German financial systems have historically differed. Indeed, this also explains why other societies with better distributions than England, for instance France and Japan (or the newly industrializing East Asian countries), have traditionally relied more on bank-finance (see Allen and Gale, 2000). While a systematic analysis covering a broad sample of countries is clearly required, the model indicates the kind of macro-fundamentals that could prove useful in explaining paths of financial development.²⁶

Investment Size and Financial Structure

Investment size has an impact on initial or historical financial structure. With a higher q , fewer individuals are able to obtain loans either from markets or from banks. At the same time, the shortfall $q - a^i$ that has to be raised through external finance is higher for those who do invest. The Holmstrom-Tirole incentive structure has a straightforward implication: due to limited liability, a borrower’s incentive to be diligent is weaker the less her personal stake in the project, that is, the more she needs to borrow. The only way to attenuate this is through increased monitoring.

Larger scale investments would hence push an economy towards bank finance. But whether or not this happens depends also on the wealth distribution. A higher q raises the importance of bank-finance under two conditions: when the initial wealth distribution among capitalists is more equitable, that is when f_{30}/f_{20} is low; and when banks are particularly effective at resolving incentive problems ($\delta_1 \ll \delta_2$ in (12)) so that the measure of individuals above \underline{a} is sizeable.²⁷

Historical evidence, once again, provides some support in favor of this story. Similar to their financial systems, a distinction is often made between England and Germany’s industrialization patterns. The British process of industrialization mainly relied upon small- and medium-scale industries, textile manufacturing being a prime example. Germany, on

²⁶At the same time, as long as two societies converge to developed financial systems (D), inequality does not have a permanent effect on financial structure. Interestingly, Allen and Gale (2000) point out how even traditionally bank-oriented societies such as France, Germany or Japan have been moving toward market finance since the 1980s. One could interpret this as a convergence in industrialized country financial systems, though, policy shifts would have clearly played a key role.

²⁷This follows from noting that $\partial\psi_0/\partial q < 0$ whenever $\delta_2(1 + f_{20}/f_{30})g_0(\bar{a}) > \delta_1g_0(\underline{a})$.

the other hand, largely utilized heavy manufacturing and chemical industries for its development, both requiring far greater investment than in the case of England (see Landes, 1969, and related references in Baliga and Polak, 2002). Consistent with the evidence, our model suggests that this technological difference would be reflected in a greater German reliance on intermediated finance especially if banks are efficient intermediaries and the wealth distribution is more equitable.

Using a static model of monitored and non-monitored debt, Baliga and Polak (2002) too stress the role of investment size for Western Europe's financial structure. Importantly though, our dynamic model suggests that long-run differences in financial structure are neutral with respect to investment size. As long as $f_{10} < \widehat{f}_1$ (or less than f_1^a), differences in f_{10} do not translate into differences in the stationary distribution, that is, point D . This is because q affects threshold wealth requirements (\underline{a} and \bar{a}) as well as expected income from investment. In the short-run, a higher q could increase reliance upon bank-finance but it also enables successful capitalists to earn more. When more of the middle-class capitalists move above \bar{a} , they do not need to be monitored so that reliance upon bank finance declines.

6 Conclusion

This paper has analyzed the evolution of a financial system and identified factors determining its development and structure. We introduced monitored bank loans and non-monitored tradeable securities in a dynamic general equilibrium model and showed how the path to financial development exhibits non-ergodic behavior – underdeveloped financial systems persist in highly unequal societies or in societies with inefficient legal and financial institutions, whereas societies with more equitable wealth distribution or better financial institutions experience financial development. We examined how the investment technology (size and riskiness) and institutional qualities influence the structure of financial systems. Finally, we showed that some of the model's implications are consistent with the historical development of financial systems during the Industrial Revolution.

This paper continues in the tradition of several works like Banerjee and Newman (1993), Galor and Zeira (1993), Aghion and Bolton (1997), Piketty (1997) and Mookherjee and Ray (2002, 2003) in studying the dynamic interaction between credit markets and the wealth distribution. Existing research has mainly studied the nature of distributional dynamics, credit market imperfections, occupational choice and possibilities of poverty traps. In contrast, our goal has been to obtain a clearer understanding of what drives the development

and structure of financial systems. Thus, an important contribution of this paper lies in extending the literature by incorporating a richer financial structure, one that may be used to analyze the design and effectiveness of financial systems in developing societies.

We conclude by considering several extensions for future work. Throughout the paper we focused on the demand side of financial systems. We did this primarily for tractability given the nonlinear nature of the dynamics. A natural extension would be to consider how important the supply of loanable funds is to financial structure, for example, by looking at a closed economy version of the model. A more challenging extension would be to capture the emergence of different financial institutions (institutions like banks and markets are taken as given in our story). For instance, if there are fixed costs to setting up an intermediary (as in Jovanovic and Greenwood, 1990), the extent to which banks emerge and monitor lending will depend upon, and affect, the pattern of wealth accumulation. It will be also interesting to endogenize the interest rate, allowing the wealth distribution to affect financial structure through returns to bank and market-finance.

Another institutional aspect we have ignored, and one likely to be important in development, is the quality of bank monitoring. In particular, banks do not face any incentive problems vis-a-vis depositors in our model. Extensions incorporating agency problems within the banking sector may be used to examine how the “quality” of bank-finance itself changes over time and with respect to macro-fundamentals.

References

- [1] Aghion, Philippe and Patrick Bolton (1997), “A Theory of Trickle-Down Growth and Development”, *Review of Economic Studies*, vol. 64, pp. 151-172.
- [2] Allen, Franklin and Douglas Gale (2000), *Comparing Financial Systems*, Cambridge, MA, MIT Press.
- [3] Allen, Franklin and Douglas Gale (2001), “Comparative Financial Systems: A Survey”, forthcoming in Bhattacharya, Sudipto, Boot, Arnoud and Anjan Thakor (eds.) *Financial Intermediation*, Oxford University Press.
- [4] Baliga, Sandeep and Ben Polak (2002), “The Emergence and Persistence of the Anglo-Saxon and German Financial Systems,” forthcoming, *Review of Financial Studies*.
- [5] Banerjee, Abhijit and Andrew Newman (1993), “Occupational Choice and the Process of Development”, *Journal of Political Economy*, vol. 101, pp. 274-298.
- [6] Burke, Jonathan L. (1996), “Equilibrium for Overlapping Generations in Continuous Time”, *Journal of Economic Theory*, vol. 70, no. 2, pp. 364-90.
- [7] Clapham, John Harrold (1936), *The Economic Development of France and Germany, 1815-1914*, Cambridge University Press.
- [8] Collins, Michael (1995), *Banks and industrial finance in Britain, 1800-1939*, Cambridge University Press.
- [9] Deininger, Klaus and Lyn Squire (1996), “Measuring Income Inequality: A New Database,” *World Bank Economic Review*, vol. 10, no. 3 (September), pp. 565-91.
- [10] Deininger, Klaus, and Lyn Squire (1998), “New Ways of Looking at Old Issues: Inequality and Growth,” *Journal of Development Economics*, vol. 57, no. 2, pp. 259-287.
- [11] Demirgüç-Kunt, Asli and Ross Levine eds. (2001a) *Financial Structure and Economic Growth*, MIT Press, Cambridge, MA.
- [12] Demirgüç-Kunt, Asli and Ross Levine (2001b), “Bank-based and Market-based Financial Systems: Cross-country Comparisons”, in Asli Demirgüç-Kunt and Ross Levine (eds.) *Financial Structure and Economic Growth*, MIT Press, Cambridge, MA.

- [13] Diamond, Douglas (1984), “Financial Intermediation and Delegated Monitoring,” *Review of Economic Studies*, vol. 51, pp. 393-414.
- [14] Diamond, Douglas (1991), “Monitoring and Reputation: The Choice between Bank Loans and Directly Placed Debt”, *Journal of Political Economy*, vol. 99, pp. 689-721.
- [15] Galor, Oded and Joseph Zeira (1993), “Income Distribution and Macroeconomics,” *Review of Economic Studies*, vol. 60, pp. 35-52.
- [16] Gerschenkron, Alexander (1962), *Economic Backwardness in Historical Perspective*, Cambridge, MA: Harvard University Press.
- [17] Ghatak, Maitreesh and Neville Nien-Huei Jiang (2002), “A Simple Model of Inequality, Occupational Choice and Development”, *Journal of Development Economics*, vol. 69, pp. 205-226.
- [18] Goldsmith, Raymond (1969), *Financial Structure and Development*, Yale University Press, New Haven, CT.
- [19] Greenwood, Jeremy and Boyan Jovanovic (1990), “Financial Development, Growth and the Distribution of Income”, *Journal of Political Economy*, vol. 98, no. 5, pp. 1076-107.
- [20] Gurley, John and E. Shaw (1955), “Financial Aspects of Economic Development,” *American Economic Review*, vol. 45, pp. 515-38.
- [21] Hellwig, Martin (1991), “Banking, Financial Intermediation and Corporate Finance”, in *European Financial Integration*, (eds.) Alberto Giovannini and Colin Mayer, Cambridge University Press, U.K.
- [22] Holmstrom, Bengt (1996), “Financing of Investment in Eastern Europe,” *Industrial and Corporate Change*, vol. 5, pp. 205-37.
- [23] Holmstrom, Bengt and Jean Tirole (1997), “Financial Intermediation, Loanable Funds, and the Real Sector,” *Quarterly Journal of Economics*, vol. 112, pp. 663-91.
- [24] La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert Vishny (1997), “Legal Determinants of External Finance,” *Journal of Finance*, vol. 52, pp. 1131-50.
- [25] La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer and Robert Vishny (1998), “Law and Finance,” *Journal of Political Economy*, vol. 106, pp. 1113-55.

- [26] Landes, D. (1969), *The Unbound Prometheus, Technological Change and Industrial Development in Western Europe from 1750 to the Present*, Cambridge, Cambridge University Press.
- [27] Levine, Ross (1997), “Financial Development and Economic Growth: Views and Agenda”, *Journal of Economic Literature*, vol. 35, pp. 688-726.
- [28] Levine, Ross (2002), “Bank-based or Market-Based Financial Systems: Which is Better?”, *Journal of Financial Intermediation*, vol. 11, pp. 1-30.
- [29] Mookherjee, Dilip and Debraj Ray (2002), “Contractual Structure and Wealth Accumulation”, *American Economic Review*, vol. 92, no. 4, pp. 818-849.
- [30] Mookherjee, Dilip and Debraj Ray (2003), “Persistent Inequality”, *Review of Economic Studies*, vol. 70, pp. 369-93.
- [31] Piketty, Thomas (1997), “The Dynamics of the Wealth Distribution and the Interest Rate with Credit Rationing”, *Review of Economic Studies*, vol. 64, pp. 173-189.
- [32] Soltow, Lee (1968), “Long-Run Changes in British Income Inequality”, *Economic History Review*, Second Series, 21, 17-29.
- [33] Yaari, Menahem E. (1965), “Uncertain Lifetime, Life Insurance, and the Theory of the Consumer”, *Review of Economic Studies* vol. 32, pp. 137-50.

Appendix

A.1. Optimal Contracts

Direct Finance

We have $x_t^C + x_t^U = \rho_t \theta q$.

Capitalist- i 's incentive compatibility constraint (choosing the good project) is given by

$$\pi_G x_t^C + (1 - \pi_G) \cdot 0 \geq \pi_B x_t^C + (1 - \pi_B) \cdot 0 + Vq.$$

Direct lender's participation constraint is

$$\pi_G x_t^U + (1 - \pi_G) \cdot 0 \geq r^* (q - a_t^i).$$

Capitalist's incentive compatibility constraint implies

$$x_t^C \geq \frac{Vq}{\pi_G - \pi_B} \Rightarrow x_t^U = \rho_t \theta q - x_t^C \leq \rho_t \theta q - \frac{Vq}{\pi_G - \pi_B}.$$

Using this, the lender's participation constraint gives

$$\begin{aligned} r^* (q - a_t^i) &\leq \pi_G x_t^U \leq \pi_G \left[\rho_t \theta q - \frac{Vq}{\pi_G - \pi_B} \right] \\ \Rightarrow a_t^i &\geq \bar{a}_t \equiv \frac{q}{r^*} \left[\frac{\pi_G}{\pi_G - \pi_B} V - \{ \pi_G \rho_t \theta - r^* \} \right]. \end{aligned}$$

Indirect Finance

Under indirect finance we have $x_t^C + x_t^U + x_t^B = \rho_t \theta q$. Here the optimal contracts need to satisfy the following three constraints:

- (i) capitalist- i 's incentive compatibility constraint (choosing the good project):

$$\pi_G x_t^C + (1 - \pi_G) \cdot 0 \geq \pi_B x_t^C + (1 - \pi_B) \cdot 0 + vq,$$

- (ii) bank's incentive constraint (for monitoring):

$$\pi_G x_t^B + (1 - \pi_G) \cdot 0 - r^* \gamma q \geq \pi_B x_t^B + (1 - \pi_B) \cdot 0,$$

assuming that banks discount monitoring costs at their opportunity cost, r^* ,

(iii) participation constraint of the uninformed investors:

$$\pi_G x_t^U + (1 - \pi_G) \cdot 0 \geq r^* (q - l_t^i - a_t^i),$$

where l_t^i is the amount that the bank lends to capitalist- i .

The bank's expected return from lending l_t^i to capitalist- i is

$$\pi_G x_t^B = r_t^L l_t^i, \quad (16)$$

where r_t^L is the (gross) loan rate charged to borrowers.

But bank's incentive constraint implies

$$\pi_G x_t^B \geq \left[\frac{\pi_G}{\pi_G - \pi_B} \right] r^* \gamma q. \quad (17)$$

Since bank finance is relatively more expensive than direct finance (due to monitoring costs), borrowers accept only the minimum amount necessary so that

$$\begin{aligned} r_t^L l_t^i &= \pi_G x_t^B = \left[\frac{\pi_G}{\pi_G - \pi_B} \right] r^* \gamma q \\ \Rightarrow l_t^i(r_t^L) &= \frac{r^* \gamma q \pi_G}{(\pi_G - \pi_B) r_t^L}. \end{aligned} \quad (18)$$

Capitalist's incentive compatibility constraint implies $x_t^C \geq \frac{vq}{\pi_G - \pi_B}$. Then

$$\begin{aligned} x_t^C + x_t^B &\geq \frac{(v + r^* \gamma) q}{\pi_G - \pi_B} \\ \Rightarrow x_t^U &= \rho_t \theta q - (x_t^C + x_t^B) \leq \rho_t \theta q - \frac{(v + r^* \gamma) q}{\pi_G - \pi_B}. \end{aligned}$$

Using this, the uninformed investors' participation constraint gives

$$r^* (q - l_t^i - a_t^i) \leq \pi_G x_t^U \leq \pi_G \left[\rho_t \theta q - \frac{(v + r^* \gamma) q}{\pi_G - \pi_B} \right].$$

It follows that only capitalists with wealth

$$a_t^i \geq \underline{a}_t \equiv q - l_t^i(r_t^L) - \frac{q \pi_G}{r^*} \left[\rho_t \theta - \frac{v + \gamma r^*}{\pi_G - \pi_B} \right] \quad (19)$$

are able to convince uninformed investors to supply enough funds for the investment project.

The Bank's Problem

The aggregate demand for bank loans is

$$L_t = \int_{i \in I_t} l_t^i(r_t^L) dG_t = \left[\frac{\gamma \pi_G r^* q}{(\pi_G - \pi_B) r_t^L} \right] \int_{i \in I_t} dG_t,$$

where I_t denotes the subset of individuals using intermediated finance. The total monitoring cost borne by the banking sector is then

$$\gamma q \int_{i \in I_t} dG_t = \frac{(\pi_G - \pi_B) r_t^L L_t}{\pi_G r^*}.$$

Let D_t denotes the flow of deposits into the banking sector. Then banking profits are

$$\Pi_t^B = r_t^L L_t - r^* D_t. \quad (20)$$

Banks face the resource constraint that total loans cannot exceed total deposits net of monitoring costs:

$$L_t \leq D_t - \gamma q \int_{i \in I_t} dG_t. \quad (21)$$

The banking sector's optimization problem in period t is to choose L_t so as to maximize Π_t^B subject to the capitalist's incentive constraint and the constraints (17) and (21).

Since bank profits are increasing in total loans, (21) holds with equality:

$$L_t = D_t - \gamma q \int_{i \in I_t} dG_t = D_t - \frac{(\pi_G - \pi_B) r_t^L L_t}{\pi_G r^*}. \quad (22)$$

Moreover, in a competitive equilibrium, the banking sector earns zero profits. From (20) we then have

$$r_t^L L_t = r^* D_t, \quad (23)$$

It follows from equations (22) and (23) that

$$L_t = \left(\frac{\pi_B}{\pi_G} \right) D_t,$$

and

$$r_t^L = \left(\frac{\pi_G}{\pi_B} \right) r^*. \quad (24)$$

Hence, using (18), we observe that

$$l_t^i = \gamma \left(\frac{\pi_B}{\pi_G - \pi_B} \right) q. \quad (25)$$

In other words, for all $i \in I_t$, banks finance a fixed proportion of the borrower's investment, irrespective of a_t^i .

Taking into account the optimal loan size (25), the lower wealth cut-off (19) becomes

$$\underline{a}_t = \frac{q}{r^*} \left[\frac{\pi_G v}{\pi_G - \pi_B} - \{\pi_G \rho_t \theta - (1 + \gamma)r^*\} \right]. \quad (26)$$

A.2. Existence of $\dot{f}_1 = 0$ locus when $\underline{f}_1 < f_1 < \overline{f}_1$

In the text we have already established that λ_t is a continuous and monotonically decreasing function of f_{1t} . Now define

$$F(f_1) \equiv (1 - \pi_G)/[1 - \pi_G + \lambda(f_1)] - f_1$$

on the interval $[\underline{f}_1, \overline{f}_1]$. Since $\lambda(f_1)$ is continuous on $[\underline{f}_1, \overline{f}_1]$, F is also continuous on $[\underline{f}_1, \overline{f}_1]$. We have $F(\underline{f}_1) = (1 - \pi_G)/(2 - \pi_G) - \underline{f}_1 < 0$ since $\underline{f}_1 > (1 - \pi_G)/(2 - \pi_G)$ and $F(\overline{f}_1) = 1 - \overline{f}_1 > 0$, since $\overline{f}_1 < 1$. Hence, using the Intermediate Value Theorem, since F is continuous on $[\underline{f}_1, \overline{f}_1]$ and since $0 \in [F(\underline{f}_1), F(\overline{f}_1)]$, we can find an $u \in [\underline{f}_1, \overline{f}_1]$ such that $F(u) = 0$. In other words, $F(f_1) = 0$ for at least one value of $f_1 \in [\underline{f}_1, \overline{f}_1]$. Figure 3(a) illustrates this when the line $(1 - \pi_G)/[1 - \pi_G + \lambda(f_1)]$ intersects f_1 once, at \hat{f}_1 . Multiple such intersections are also possible, but, generically, these have to be in odd numbers. Figure 3(b) depicts three intersections. Figures 2(a) and (b) illustrate dynamics under one and three such intersections respectively.

A.3. Effect of an increase in initial inequality on financing choices

Suppose that the initial distribution G_0 is lognormal with mean μ_0 and variance σ_0^2 . Recall that the lognormal cumulative distribution is given by $\Phi[(\ln x - \mu)/\sigma]$, where Φ is the standard Normal cumulative distribution function. Then:

$$\frac{\partial G_0(\underline{a})}{\partial \sigma_0} = -\phi_0 \left(\frac{\ln \underline{a} - \mu_0}{\sigma_0} \right) \left[\frac{\ln \underline{a} - \mu_0}{\sigma_0^2} \right], \quad \frac{\partial G_0(\overline{a})}{\partial \sigma_0} = -\phi_0 \left(\frac{\ln \overline{a} - \mu_0}{\sigma_0} \right) \left[\frac{\ln \overline{a} - \mu_0}{\sigma_0^2} \right].$$

When $\ln \underline{a} < \mu_0 < \ln \overline{a}$, the first derivative is positive, the second derivative negative. Moreover,

$$\frac{\partial [G_0(\overline{a}) - G_0(\underline{a})]}{\partial \sigma_0} = -\phi_0 \left(\frac{\ln \overline{a} - \mu_0}{\sigma_0} \right) \left[\frac{\ln \overline{a} - \mu_0}{\sigma_0^2} \right] + \phi_0 \left(\frac{\ln \underline{a} - \mu_0}{\sigma_0} \right) \left[\frac{\ln \underline{a} - \mu_0}{\sigma_0^2} \right],$$

is also negative in this case. Thus, as long as $\ln \underline{a} < \mu_0 < \ln \overline{a}$, we have

$$\frac{\partial f_{10}}{\partial \sigma_0} > 0, \quad \frac{\partial f_{20}}{\partial \sigma_0} < 0, \quad \frac{\partial f_{30}}{\partial \sigma_0} > 0.$$

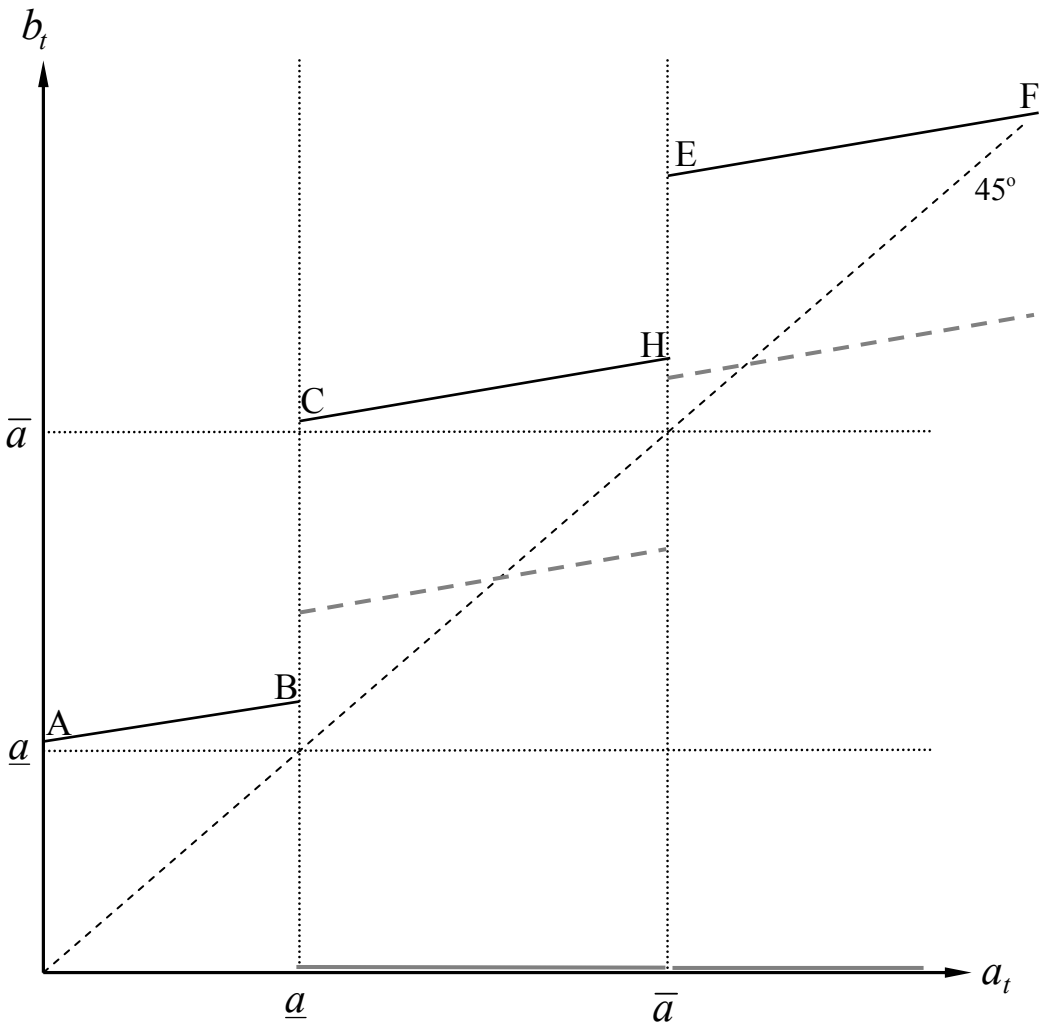


Figure 1(a): Recursion Dynamics of Wealth Accumulation when $f_1 \leq \underline{f}_1$

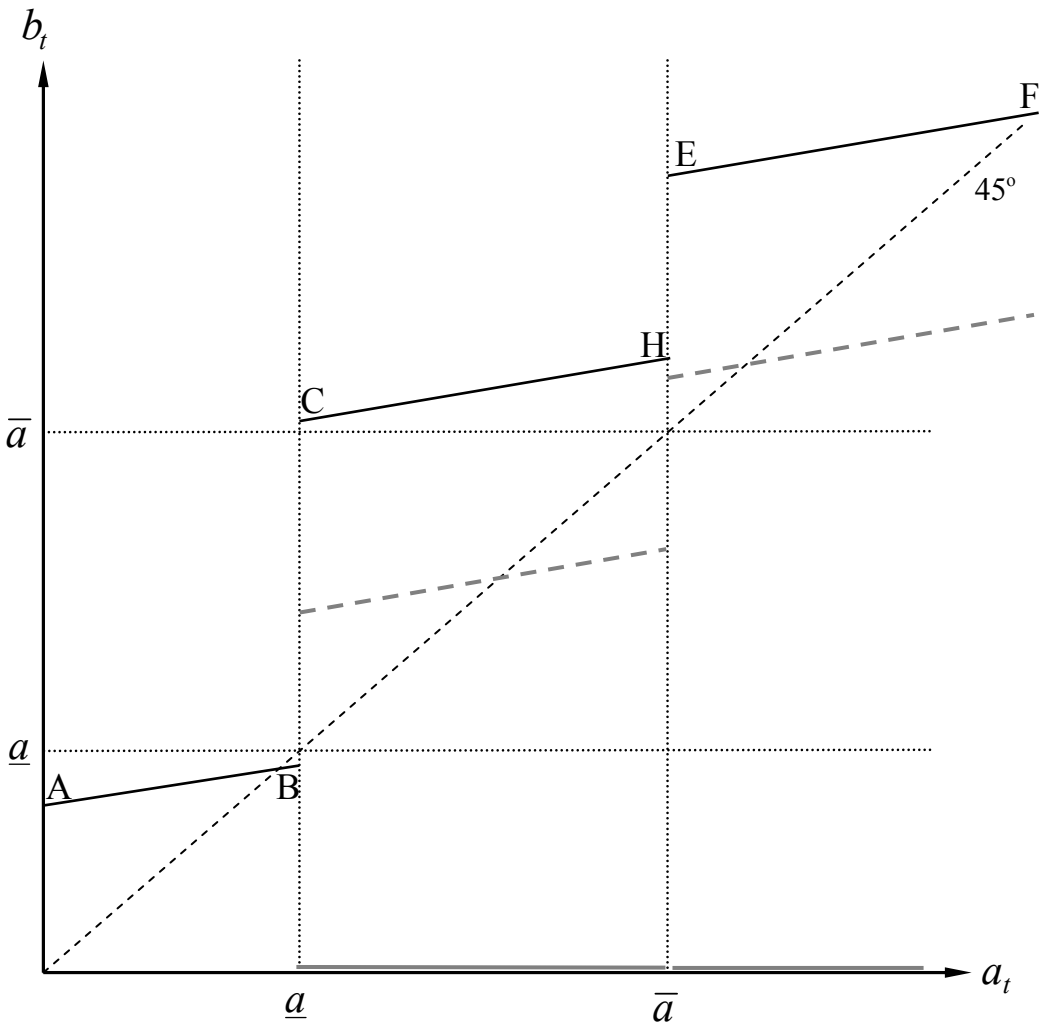


Figure 1(b): Recursion Dynamics of Wealth Accumulation when $f_1 > \bar{f}_1$

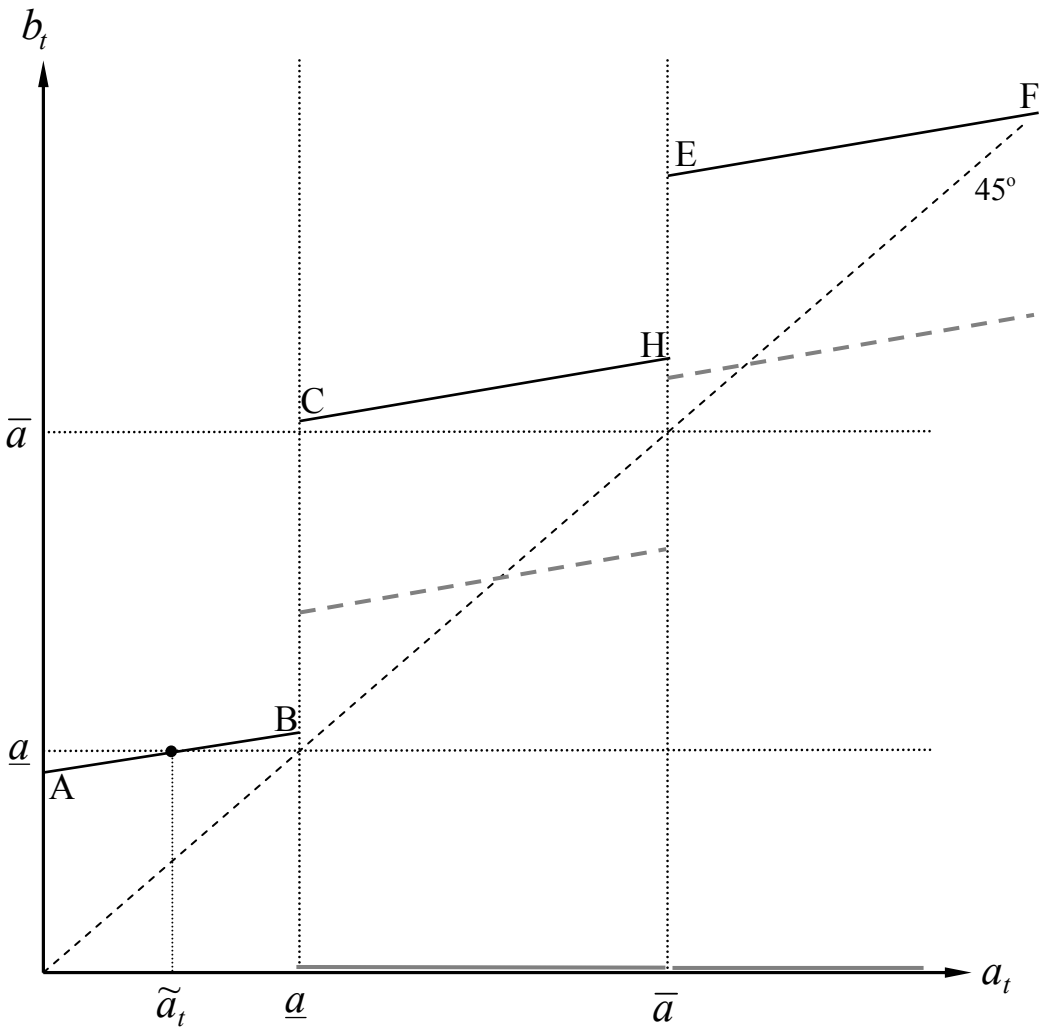


Figure 1(c): Recursion Dynamics of Wealth Accumulation when $\underline{f}_1 < f_1 < \bar{f}_1$

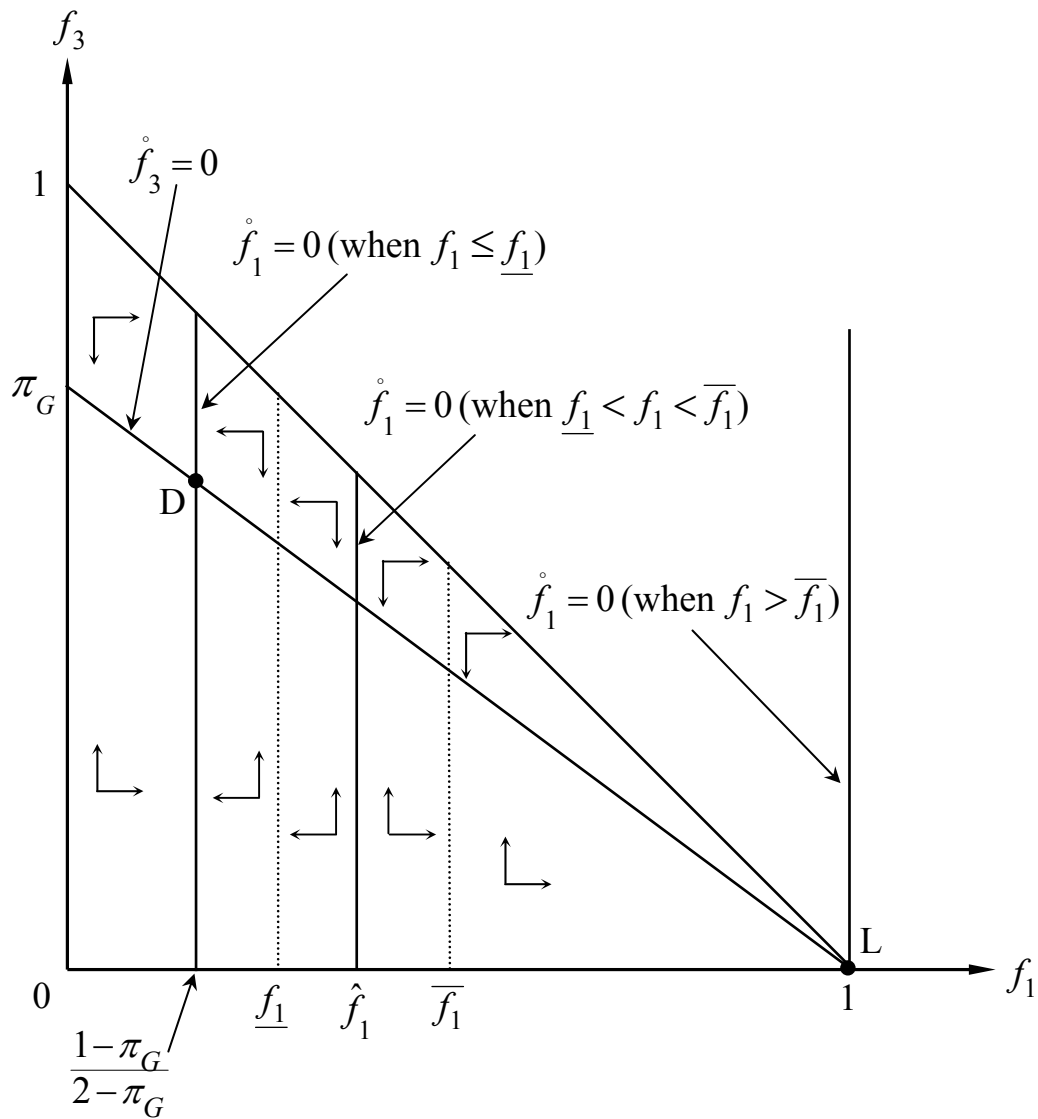


Figure 2(a): Phase Diagram for a single Threshold

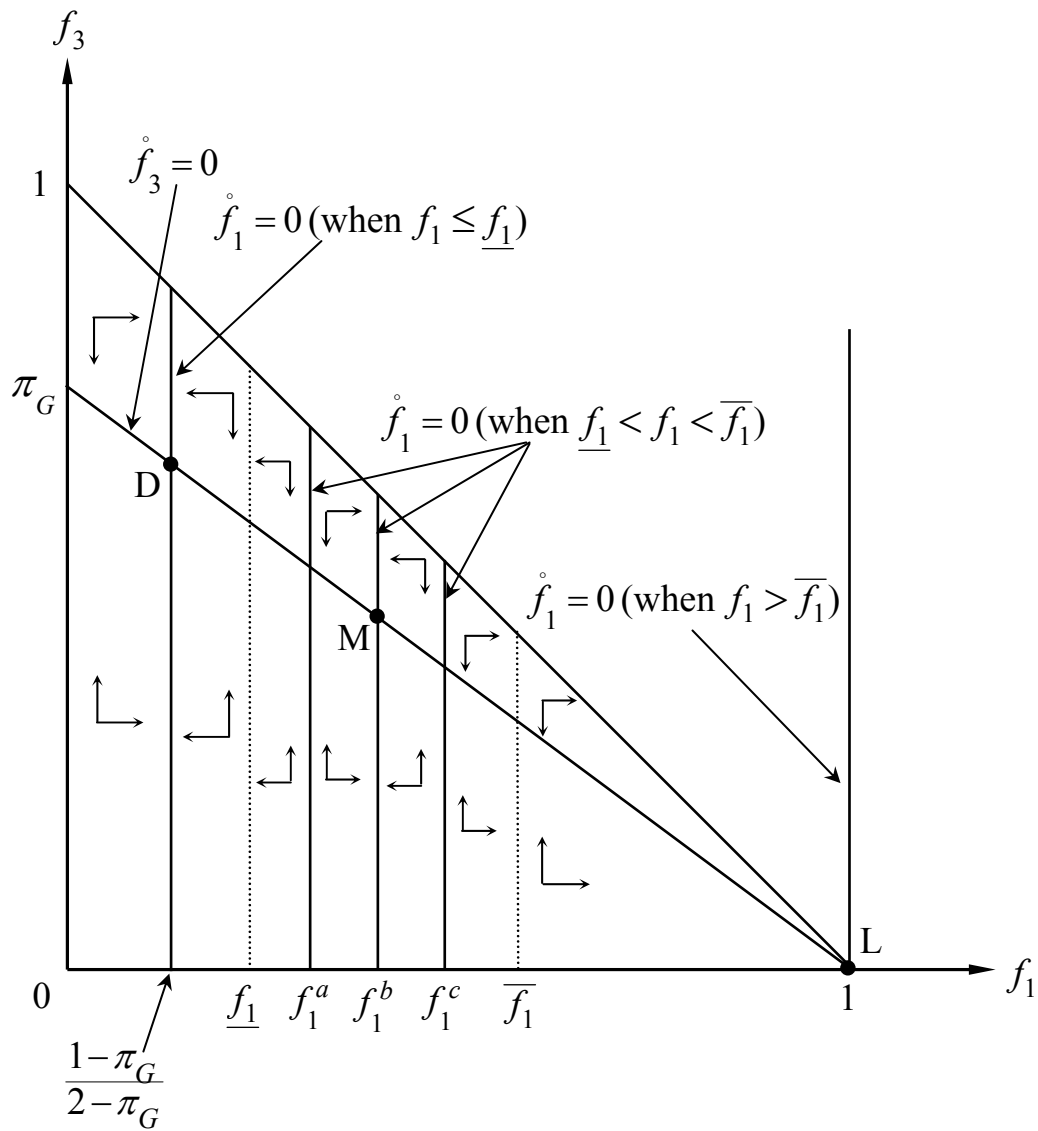


Figure 2(b): Phase Diagram for two Thresholds

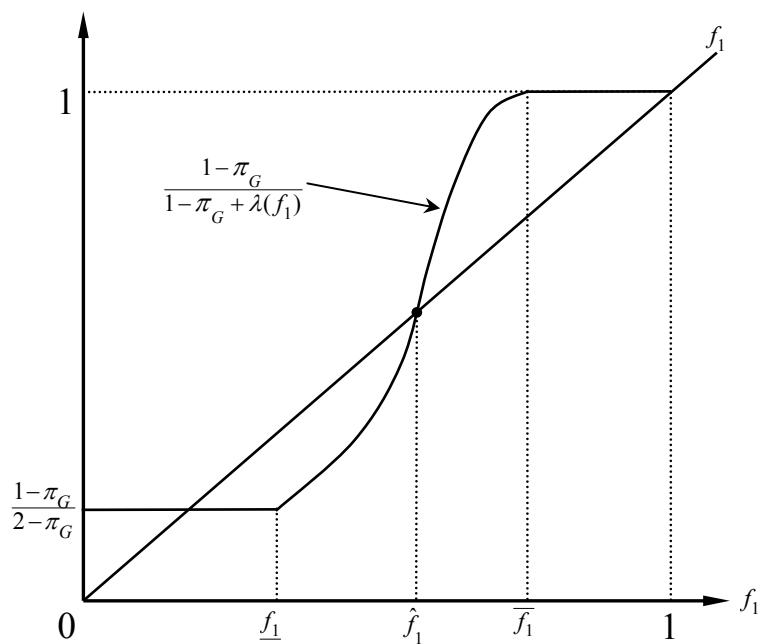


Figure 3(a)

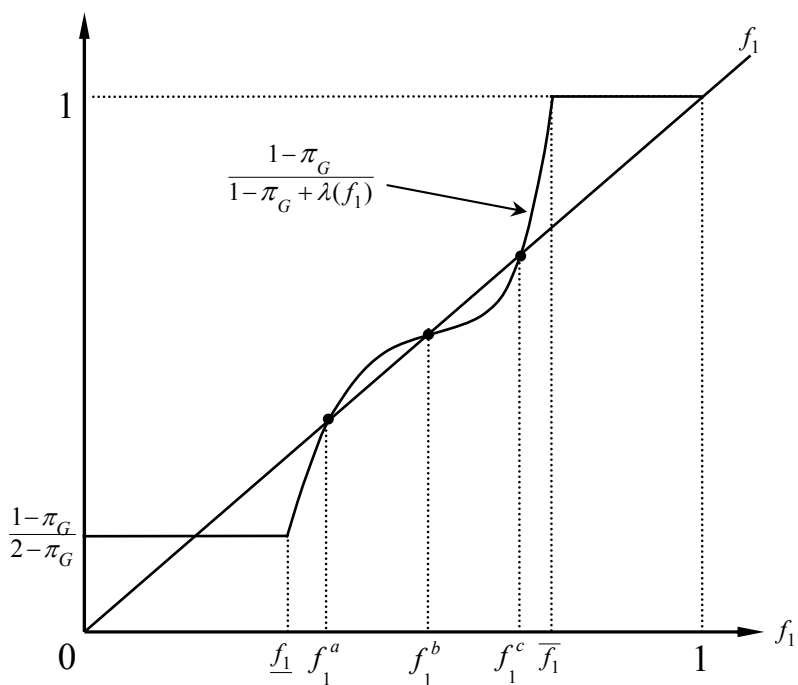


Figure 3(b)

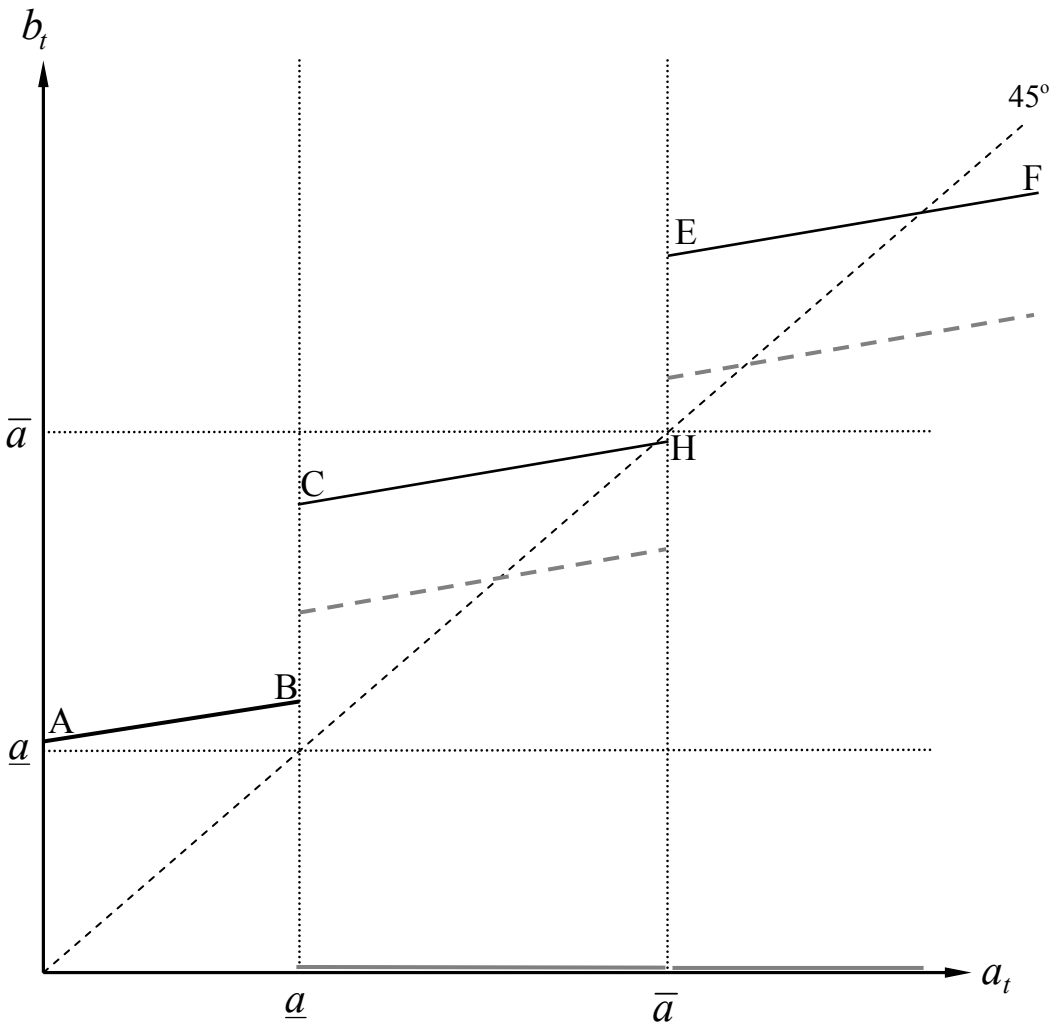


Figure 4: Wealth Dynamics with Low Return on Investment ($f_1 \leq \underline{f}_1$)

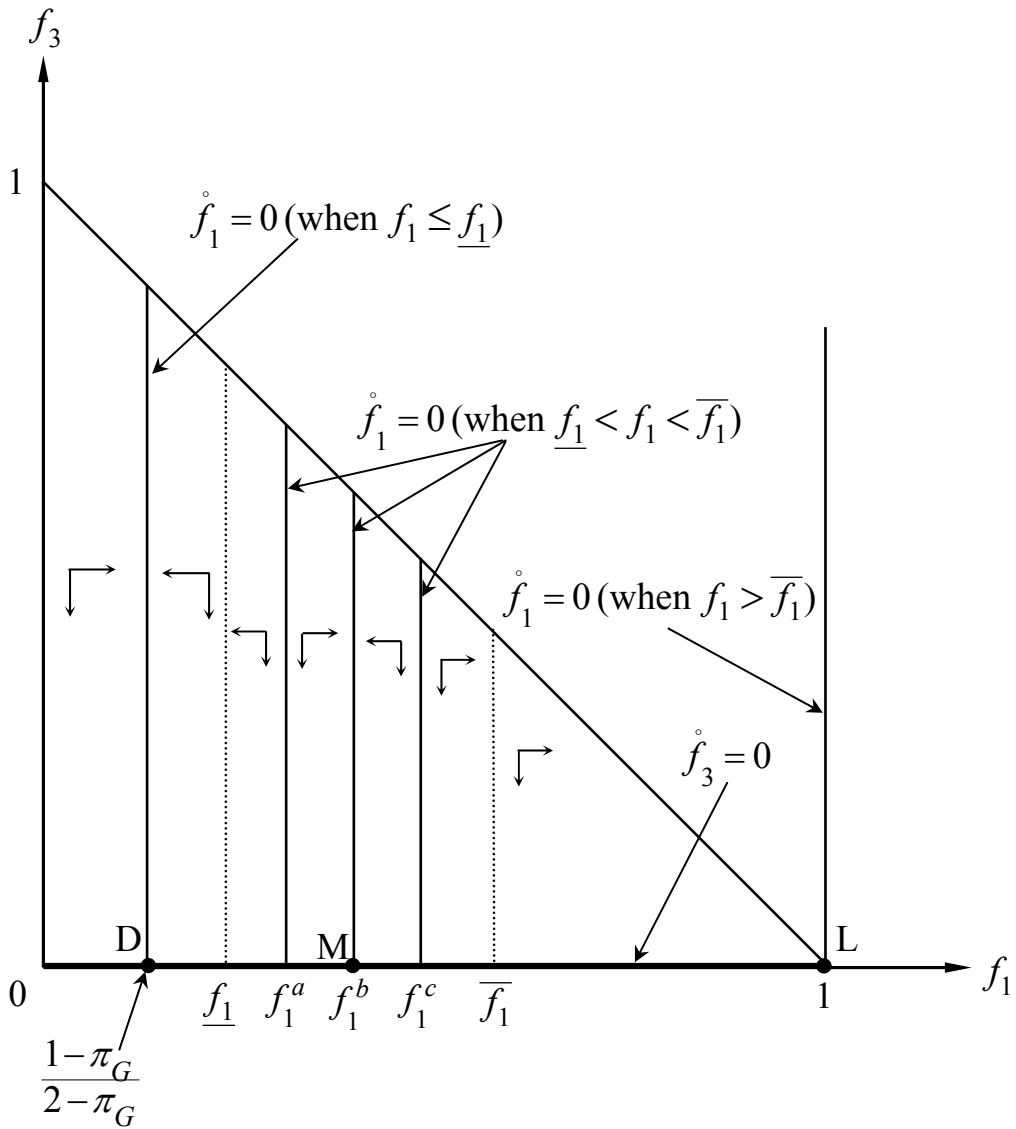


Figure 5(a): Phase Diagram (for two Thresholds) with Low Investment Return

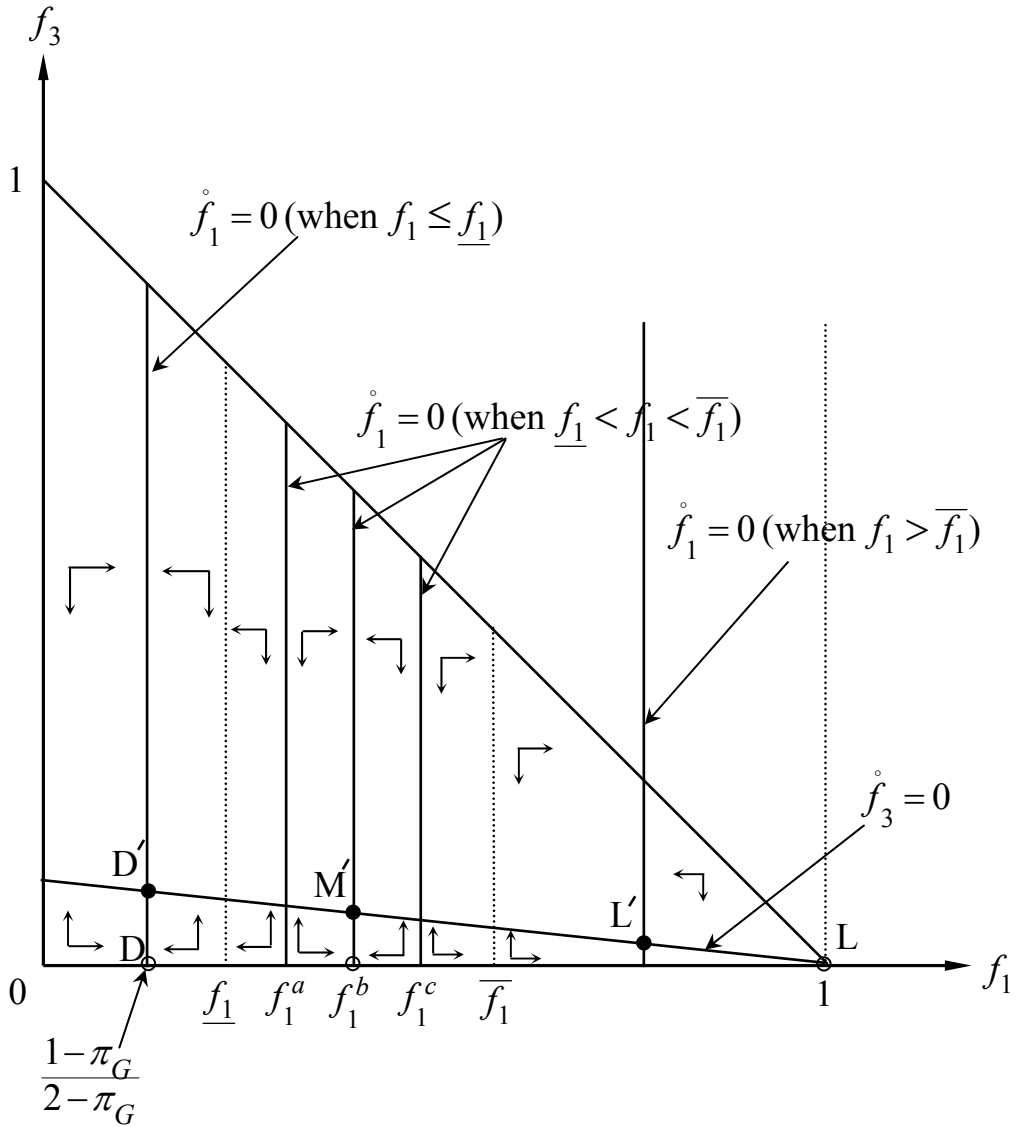


Figure 5(b): Phase Diagram for the Perturbed Wealth Dynamics with Low Investment Return