Federal Reserve Bank of New York
Staff Reports

Credit Quantity and Credit Quality: Bank Competition and Capital Accumulation

Nicola Cetorelli
Pietro F. Peretto

Staff Report no. 375
June 2009

Credit Quantity and Credit Quality: Bank Competition and Capital Accumulation
Nicola Cetorelli and Pietro F. Peretto
Federal Reserve Bank of New York Staff Reports, no. 375
June 2009
JEL classification: G1, G2, L1, L2, O1, O4

Abstract

This paper shows that bank competition has an intrinsically ambiguous effect on capital accumulation and economic growth. We further demonstrate that banking market structure can be responsible for the emergence of development traps in economies that would otherwise be characterized by unique steady-state equilibria. These predictions explain the conflicting evidence gathered from recent empirical studies of how bank competition affects the real economy. Our results were obtained by developing a dynamic general-equilibrium model of capital accumulation in which banks operate in a Cournot oligopoly. The presence of more banks leads to a higher quantity of credit available to entrepreneurs, but also to diminished incentives to screen loan applicants and thus to poorer capital allocation. We also show that conditioning on economic parameters describing the quality of the entrepreneurial population resolves the theoretical ambiguity. In economies where the average prospective entrepreneur is of low credit quality and where screening would therefore be especially beneficial, less competition leads to higher capital accumulation. The opposite is true when entrepreneurs are innately of higher credit quality.

Key words: bank competition, economic growth, oligopoly, lending

---

Cetorelli: Federal Reserve Bank of New York (e-mail: nicola.cetorelli@ny.frb.org). Peretto: Duke University (e-mail: peretto@econ.duke.edu). This paper substantially revises and expands a previous version that circulated with the title "Oligopoly Banking and Capital Accumulation." The authors thank Xavier Freixas, Helios Herrera, Giuseppe Lopomo, Robert Marquez, Joel Shapiro, and Huseyin Yildirim for very useful comments.

1 Introduction

Recent empirical work has documented multiple, conflicting effects of bank competition on the real economy. Some papers find evidence consistent with the prediction that bank competition leads to more credit availability, more firm entry and more growth (e.g., Black and Strahan, 2001, Beck, Demirguc-Kunt, and Maksimovic 2004, Cetorelli and Gambera, 2001, Cetorelli and Strahan, 2006, Bertrand, Schoar and Thesmar, 2007). Others highlight instead issues concerning credit quality and that in fact credit availability may be higher in less competitive environments (e.g., Petersen and Rajan, 1995, Shaffer, 1999, Cetorelli and Gambera, 2001, Bonaccorsi and Dell'Ariccia, 2004, Zarutskie, 2006). The evidence thus seems to indicate that bank competition definitely matters for capital accumulation and growth, but also that there is fundamental ambiguity about the sign of the relationship.

However, and to the best of our knowledge, we still do not have a theoretical model of economic growth with a fully specified banking sector able to generate the contrasting predictions that the evidence suggests. In this paper, we attempt to do just that. We develop a dynamic general equilibrium model of a production economy with oligopolistic banks. The model produces innovative insights on the role of banking market structure for capital accumulation and economic growth and important refinements to the associated normative prescriptions.

To construct a tractable dynamic general equilibrium model we posit that banks compete in a Cournot fashion in gathering individual savings and in loaning funds to entrepreneurs.[1] As in many models of financial intermediation, banks' role is information production (see, e.g., Leland and Pyle, 1977, Ramakrishnan and Thakor, 1984, Diamond, 1986, Chan, Greenbaum and Thakor, 1986). Specifically, we set up a world where loan applicants are of unknown quality. As in Sharpe (1990) we abstract from asymmetric information issues. Applicants' quality is unknown to both applicants and lenders. However, banks have access to a screening technology that, at a fixed cost per applicant, allows them to identify quality types. Since low quality applicants waste resources and default on loans, whether banks have the proper incentives to activate the screening technology is of paramount importance. However, the information acquired during the screening process may not be fully appropriable. Even if the outcome of the screening test is

[1] The Cournot model has the nice feature that competition and monopoly are the two extremes of a continuum of market structures wherein market power is fully captured by the number of banks.

not observable by third parties, competitor banks could still extract information about screened applicants by simply observing whether a bank extends or denies a loan.[2] In other words, there is an informational externality that generates a free-riding problem.[3]

We show that more competition leads to higher volumes of credit, but it also reduces banks' incentives to produce information on the quality of prospective entrepreneurs.[4] When we take into account the endogenous feedbacks between the banking sector and the rest of the economy we uncover an intrinsically ambiguous effect of bank competition on the path of capital accumulation. However, the model offers indications on how to resolve such ambiguity. More precisely, we show that perfect competition is the banking market structure that maximizes long-run income only when the benefit of collecting information on borrowers is relatively small because on average most of them are of high quality, and therefore sorting the good from the bad is not so crucial. At the opposite end, monopoly maximizes long-run income when the benefit of collecting information is relatively large because on average most borrowers would be destined to fail, and thus identifying the few good ones is critical. In less clear-cut situations, where the average probability of success is neither very low nor very high, the market structure that maximizes long-run development is an oligopoly. The model's main insight, therefore, is that the role of bank competition depends on the informational friction that gives rise to the special role of financial intermediation in the first place. The stronger the friction, the more important is that banks screen, and the further we deviate from the traditional view that competition is beneficial because it reduces the margin of intermediation.

We also find that banking market structure *per se* can give rise to a development trap in environments where the fundamentals would otherwise be consistent with unique, high-income equilibria in which banks screen borrowers and lend efficiently. Interestingly, the market structure that allows

[2] As recognized by Bhattacharya and Thakor (1993), "bank loans are special in that they signal quality in a way that other forms of credit do not" (p. 3).

[3] We are certainly not the first to stress the problem of appropriability in information production. This issue was central, for instance in the model of financial intermediation of Campbell and Kracaw (1980), where the authors concluded that because of this problem, information could only be under–produced in equilibrium. The potential for free riding in banks' information production was also explicitly mentioned in Thakor (1996, p. 303). Although not in a model of ex-ante information acquisition, but rather in one of ex-post information learning by lending, the same problem is also highlighted in Petersen and Rajan (1995).

[4] Fischer (2000) provides empirical evidence that banks' information production in their lending activity is higher in more concentrated banking markets.

the economy to escape from the trap may not be the one that maximizes steady-state income outside of the trap. The possibility for development traps determined by banking market structure is another potential explanation for the ambiguous evidence documented in the literature. It also enhances the appreciation for the complexity of the task faced by regulators in a dynamic environment where the effect of competition on banking practices varies with the economy's level of development.

Two strands of literature on banking have developed in recent years that relate to our study. The first one focuses on the role that banks play in promoting economic growth. This line of research, part of the broader debate on the importance of finance for real economic activity, has contributed to solidifying the consensus view that a more developed, more efficient banking sector has causal, positive effects on the real economy.[5]

The second strand of literature, more in the tradition of banking studies and corporate finance, focuses on the effects of bank competition on the equilibrium of the credit market. Standard industrial organization arguments applied to the banking industry predict that more competition leads to lower lending rates and larger credit quantities. However, more nuanced claims, recognizing that in "producing" loans banks are simultaneously engaged in the resolution of information-based problems, suggest that more competition may not generate the incentives for banks to play this role. Hence, an increase in competition may lead to worse credit practices and perhaps even lower credit availability overall.[6] While these contributions are deep in the analysis of the banking market, they abstract from broader considerations for aggregate economic variables.

Some comments on our modelling approach. Given the declared objective of this project, to link theoretically bank competition with capital accumulation and growth, building a dynamic general equilibrium model strikes us as the natural thing to do. It turns out that virtually all of the insights and predictions of the model do derive from the explicit general equilibrium set up.[7] The fact that banks face a well-defined downward sloping demand

[5]This literature is vast. A very exaustive review is offered, for example, in Levine (2004). A review with a specific focus on the issue of causality is in Cetorelli (2009).

[6]For alternative theoretical arguments see, e.g., Rajan (1992), Petersen and Rajan (1995), Marquez (2002), Dell'Ariccia and Marquez (2004), Hauswald and Marquez (2006) roughly on one side and Boot and Thakor (2000), Boyd and De Nicolo' (2005) on the other).

[7]Guzman (2000) is one contribution we are aware of that analyzes bank competition in a general equilibrium framework. However, his model did not focus on issues related to information production, and bank competition was unambiguously predicted to yield the best outcome for capital accumulation and growth.

for credit — which derives from the equilibrium conditions in the production sector — and an upward sloping supply of deposits — which derives from the workers' intertemporal decision between current and future consumption — leads naturally to feedback effects that enrich our understanding of the role of bank competition. Take for instance the basic result that more competition reduces banks' incentives to produce information. In our approach it is also the case that when more banks compete for deposits, they offer higher rates to savers, which in turn leads to a higher supply of savings and therefore to a higher supply of credit. But a higher credit supply *raises* banks' incentives to screen, since — all else equal — the incidence of the screening cost is lower. At the same time, if this translates into a higher amount of capital that goes into production, the return on capital decreases, and this *weakens* the incentive to screen. Clearly, the overall effect of a change in the number of banks on the aggregate quantity of lending and on the extent of information production is far from being straightforward and the model allows us to sort out the various channels and obtain sharp predictions that can reconcile the theory with the evidence.[8]

Also, as mentioned, we focus on the role of banks as information producers, hence on screening, and we abstract from agency problems associated with asymmetric information. We justify this choice with two arguments. The first is that we think of screening as the activity that banks perform at the outset of their relationship with loan applicants. Understanding if and how competition affects banks' incentives to undertake such activity in the first place is of first-order importance. The second argument is that by ignoring asymmetric information we are able to capture the trade-off between quantity and quality of credit in an extremely parsimonious model with very few free parameters. Our main ingredients are the information externality associated with the screening activity of banks, their oligopolistic rivalry, and the general equilibrium structure of the economy. This is all that is needed to obtain a rich set of results. Adding further detail to include, for example, the contractual issues associated with the resolution of asymmetric information would augment the banking part of the model at the cost of an

[8] The intuition that a general equilibrium approach is important is confirmed in a number of theoretical studies that have focused specifically on bank competition and financial stability. As stated in Allen and Gale (2004a): "In simple partial-equilibrium models, it is possible to generate a negative trade-off between competition and financial stability. However, ... the nature of the trade-off [...] is more complicated than was first thought." Indeed, significant qualifiers to this statements have been presented in general equilibrium models such as, e.g., Allen and Gale (2004b), Boyd, De Nicolo' and Smith (2004).

unnecessary burden that in the end would obscure the models' insights.[9]

Finally, we also take banking structure as exogenous. This assumption is based primarily on the observation that, in contrast to most other industries, where the default is that market structure and competitive conduct evolve endogenously, banking industries have historically been heavily regulated. This is true both in the U.S. and in other countries. Hence, it is plausible to consider banking market structure as exogenously determined when studying its effect on the real economy.[10]

2 The Economy

The economy is populated by overlapping generations living for two periods. Each generation consists of a continuum of mass one of identical individuals. Population is thus constant. Each young agent is endowed with no capital and with one unit of labor. When old, the agent does not work and lives off his savings. We abstract from labor-leisure choice so that young agents supply their entire labor endowment in the market.

2.1 The primitives: technology and preferences

There exists a competitive firm producing a homogeneous final good with a standard neoclassical production function that satisfies the Inada conditions,

$$Y_t = F(K_t, L_t) = K_t^{\gamma} L_t^{1-\gamma}, \quad 0 < \gamma < 1 \tag{1}$$

where Y, K and L are, respectively, output, capital and labor. Since labor supply is inelastic, in equilibrium $L_t = 1$. Therefore, in our analysis we can work with the intensive-form version of (1), which we denote $f(\cdot) = K_t^{\gamma}$. All of our results obtain with a generic, neoclassical production function.[11] We shall work with the Cobb-Douglas specification to streamline the exposition.

[9] The standard approach in models of asymmetric information in banking is to ignore general equilibrium considerations and rely instead on simplifying assumptions such as production projects of fixed size that pay fixed return and, more generally, a perfectly elastic supply of funds from savers.

[10] For instance, before the process of deregulation initiated in the mid 1970s, the U.S. banking industry had been effectively partitioned, since the nineteenth century, within state boundaries, and even within states there were significant restrictions to bank expansion: at the beginning of the 1970's, 38 states prohibited bank branching within a state (unit banking states) or imposed significant limitations to branching. At the same time, banks were completely prohibited from acquiring banks outside the state in which they were headquartered (see, e.g., Jayratne and Strahan, 1995).

[11] See, e.g., Barro and Sala-i-Martin (2005) for a list of its properties.

The competitive final producer's profit maximization problem yields the following demand schedules for capital and labor:

$$R_t^K = f'(K_t) = \gamma K_t^{\gamma-1}; \quad (2)$$
$$W_t = f(K_t) - K_t f_K(K_t) = (1-\gamma) K_t^{\gamma}, \quad (3)$$

where R^K is the rental rate on capital and W is the wage rate.

Let c_t and c_{t+1} be consumption at time t and $t+1$ for a representative member of generation t. The agent maximizes[12]

$$U(c_t, c_{t+1}) = u(c_t) + u(c_{t+1}) = c_t^{\alpha} + c_{t+1}^{\alpha}, \quad \alpha < 1 \quad (4)$$

subject to:

$$c_t = W_t - s_t;$$
$$c_{t+1} = s_t r_{t+1},$$

where s_t is the amount of saving (bank deposits) at time t and r_{t+1} is the rate of return on saving.[13] As for the production function, all of our results obtain with a generic utility function but we shall work with the power function form to streamline the exposition.

Substitution of the two constraints into (4) yields directly that the solution to the maximization problem is the saving supply schedule

$$r_{t+1} = h(S_t; W_t) = \left[\frac{S_t}{W_t - S_t}\right]^{\frac{1-\alpha}{\alpha}}, \quad (5)$$

where we use the assumption that there is a mass one of identical young agents to write the function in terms of aggregate savings, S_t.

2.2 Capital accumulation

In modeling investment we wish to stay as close as possible to standard capital theory. Thus, we think of investment as the usual linear transformation

[12] We set the discount factor equal to one because it plays no essential role in our analysis.

[13] In this model banks make positive profits. In order to account for these profits, we assume that banks are owned by young agents. More precisely we assume that young agents save by both depositing and purchasing equity shares of banks. Formally, $s_t = d_t + e_t$, where d_t is deposits and e_t is equity capital. Banks in turn use both deposits and equity capital to supply credit. A standard arbitrage argument requires that the rate of return to deposit be equal to the rate of return to equity. r_t is this rate of return. Banks' profits are thus part of the resources that old agents use to finance consumption.

of final output into capital. To assign a significant role to financial intermediation we need to introduce a friction. The simplest way to do so is to imagine that each young agent, in addition to working for a wage, can obtain credit from banks to invest in the production of capital. For concreteness, we call a young agent engaged in this activity an entrepreneur. In a frictionless world all credit becomes capital. We posit, instead, agent-specific uncertainty. Entrepreneurs belong to two types: H, who always succeed in transforming credit into capital, and L, who always fail. Let $\theta \in [0, 1]$ be the time-invariant proportion of type H entrepreneurs. As in Sharpe (1990), agents do not know their type, they only know the distribution of types. This implies that there is an information problem to be solved, providing the rationale for banks' special function. We elaborate on this central feature of the model in the next subsection. In the remainder of this subsection we discuss how this structure modifies the traditional characterization of capital accumulation.

If successful, the entrepreneur rents capital services to the final producer at the competitive rental rate; if not successful, he defaults on the loan and the borrowed resources are lost. Consequently, the amount of credit that becomes capital is only a fraction of the total credit issued by banks. Banks play a crucial role because they possess a screening technology that allows them to learn an entrepreneur's type by spending a fixed amount β of final output. Thus, in principle, banks can discriminate between good and bad entrepreneurs and lend only to the H types, thereby eliminating the potential losses. However, banks do not always have incentives to screen because once a bank learns the entrepreneur's type it might not be possible to prevent other banks from acquiring the information. Our model studies how the market structure of the banking sector regulates the incentives for banks to undertake screening even though there is the potential of free riding by the competition.

Before turning to that crucial component of the model, it is useful to summarize the timing of events. At time t old agents of generation $t-1$, who have saved resources to finance time t consumption, supply their savings to banks. Entrepreneurs borrow from banks. They either succeed or fail in transforming credit into capital. The successful entrepreneurs add to the aggregate capital stock, which is then used to produce the final good. Given total output Y_t, a fraction represents the compensation for the successful entrepreneurs, which is used to pay back bank loans. Banks pay savers who consume the payment at time $t+1$. A fraction of output Y_t is the labor income of young agents of generation t who, according to their preferences, decide how much to consume and how much to save. Their savings are then intermediated by banks to generate credit supply for entrepreneurs of generation $t + 1$.

2.3 The free-riding problem

Banks gather savings from old agents and lend to entrepreneurs. As all other agents in the economy, banks do not know the quality of individual entrepreneurs, they only know the distribution of types. Each bank, however, has a screening technology that at a cost of β units of final output allows it to learn the type of the entrepreneur who is applying for credit. This cost of screening does not vary with the scale of the entrepreneur's investment. If the entrepreneur is of high quality, the bank extends a loan at conditions determined by market equilibrium. If the entrepreneur is of low quality the bank rejects the loan application. The bank can choose not to perform screening and lend indiscriminately to capture the proportion θ of type H entrepreneurs.

Because entrepreneurs are young agents with no record of past performance, the screening cost should be interpreted as an investment by the bank to learn specific characteristics of the entrepreneur (attitude, potential expertise, etc.). As such, screening always has a fixed, per-project component to it, and this is captured by the parameter β. The cost could also be interpreted as a direct investment by the bank offering its own expertise in an initial set-up stage, which combined with the entrepreneur's characteristics can either guarantee success or expose failure.[14]

Screening produces valuable information on entrepreneurs. However, if such information is not appropriable, a free-riding problem arises. Suppose, for example, that the results of the screening test performed by a bank on an entrepreneur were public knowledge. A competitor bank could extend a safe loan to this tested, high quality entrepreneur without bearing the screening cost. In fact, to have free riding it is not necessary to assume that the outcome of the screening test is observable. It is just sufficient to assume that the very decision by a bank to extend or deny a loan is observable. The free riding problem then arises because should a bank routinely screen all its clients, competitor banks would be able to infer each entrepreneur's type, since those who receive a loan must be of high quality and those who do not must be of low quality.

The main conclusion of this discussion is that under plausible conditions

[14] Put it differently, this form of investment can be interpreted as a type of informed, relationship-based lending. The alternative, no screening and indiscriminate lending, can instead be viewed as a type of uninformed, transaction-based activity.

there exists an *informational externality* associated with banks' screening activity that has important implications for the functioning of the credit market and the dynamics of capital accumulation.[15] The existence of such externality may prevent, or limit, the extent to which banks engage in information-based lending.

3 Lending strategies

We now study the optimal lending strategy of banks, given the existence of competitors and the non appropriability of information. Because of free riding, always screening may not be an optimal strategy because a bank could save the per-entrepreneur screening cost in the event an entrepreneur were also screened with certainty by at least another bank. At the same time, never screening when no other bank is screening may also not be optimal, if the benefit of screening from better allocation were larger than the marginal screening cost. We therefore posit, and subsequently prove, that the optimal lending strategy of banks entails random screening, that is each time an entrepreneur approaches a bank, the bank performs screening with probability $p \in [0, 1]$.

3.1 The bank's profit with random screening

Let $i \in [0, 1]$ denote a loan applicant from the mass one of entrepreneurs, and let $j = 1, ...N$ denote one of the banks in operation. When i applies for a loan, he applies to all banks. From the perspective of bank j, the following three cases are possible:

1. with probability p_j the bank screens the entrepreneur and makes a safe loan l_{ij}^{safe};
2. with probability $1 - p_j$ the bank does not screen the entrepreneur and two outcomes are possible:
    (a) with probability $\Pi_{q \neq j} (1 - p_q)$ none of the other banks screens the entrepreneur and bank j makes a risky, unscreened loan, l_{ij}^{risky};

[15] This externality could be ruled out by assuming that all information related to banks' screening activity and lending decisions is private. This scenario seems unrealistic, since an entrepreneur that has received a loan from a screening bank could not be prevented from making such decision public.

(b) but with probability $1 - \Pi_{q \neq j}(1 - p_q)$ at least one of the other banks screens the entrepreneur and bank j can make a safe loan l_{ij}^{safe} by free riding.

Let R denote the interest rate on loans and recall that r denotes the interest rate on deposits. Then, the bank's expected profit from loan i is

$$\begin{aligned}\pi_{ji} &= p_j \left[(R-r)\, l_{ji}^{safe} - \beta \right] \\ &\quad + (1-p_j)\, \Pi_{q \neq j}(1-p_q)\,(\theta R - r)\, l_{ji}^{risky} \\ &\quad + (1-p_j)\left[1 - \Pi_{q \neq j}(1-p_q)\right](R-r)\, l_{ji}^{safe},\end{aligned}$$

where β is the cost of screening an entrepreneur. As said, the third term captures free riding, i.e., the bank can make a safe loan without paying the screening cost. Since all entrepreneurs apply to all banks, in equilibrium we can think of all loans as syndicated loans of which each bank gets a share $1/N$ (this is similar to Thakor, 1996). Aggregating over the mass of applicants, the bank's total profit is

$$\begin{aligned}\pi_j &= \int_0^1 p_j \left[(R-r)\, l_{ji}^{safe} - \beta \right] di \\ &\quad + \int_0^1 (1-p_j)\, \Pi_{q \neq j}(1-p_q)\,(\theta R - r)\, l_{ji}^{risky} di \\ &\quad + \int_0^1 (1-p_j)\left[1 - \Pi_{q \neq j}(1-p_q)\right](R-r)\, l_{ji}^{safe} di\end{aligned}$$

Collecting terms we can then write the bank's total profit as

$$\begin{aligned}\pi_j &= (R-r)\int_0^1 \left[1 - \Pi_q(1-p_q)\right] l_{ji}^{safe} di - \int_0^1 p_j \beta di \\ &\quad + (\theta R - r)\int_0^1 \Pi_q(1-p_q)\, l_{ji}^{risky} di.\end{aligned}$$

We can simplify this expression if we observe that

$$\int_0^1 \left[1 - \Pi_q(1-p_q)\right] l_{ji}^{safe} di = x_j^{safe}$$

is the total amount loaned by the bank to screened entrepreneurs, while

$$\int_0^1 \Pi_q(1-p_q)\, l_{ji}^{risky} di = x_j^{risky}$$

is the total amount loaned to unscreened entrepreneurs. Accordingly, we can write

$$x_j^{safe} = [1 - \Pi_q (1 - p_q)] \, x_j$$

and

$$x_j^{risky} = \Pi_q (1 - p_q) \, x_j,$$

where $x_j = x_j^{safe} + x_j^{risky}$ is the total amount of credit extended by the bank. We can then collect terms and write

$$\pi_j = \{R [1 - \Pi_q (1 - p_q)] + \theta R \Pi_q (1 - p_q) - r\} \, x_j - p_j \beta$$

and therefore

$$\pi_j = \{R [1 - (1 - \theta) \Pi_q (1 - p_q)] - r\} \, x_j - p_j \beta. \tag{6}$$

This expression allows us to concentrate on two choice variables only: the total amount of lending x_j that the bank does and the probability p_j with which it screens individual entrepreneurs.

Since we are interested in characterizing the banks' strategic interaction in a Cournot model, we allow the individual bank to take into account the effect of its own actions on three aggregates: the amount of deposits X raised from the households and lent to entrepreneurs, the probability $P \equiv 1 - \Pi_q (1 - p_q)$ that an entrepreneur is screened, and consequently the amount of capital K supplied to the production firms. To see these contributions, observe that:

$$\begin{aligned} K &= X^s + \theta X^u \\ &= \sum_j x_j^s + \theta \sum_j x_j^u \\ &= \sum_j [1 - \Pi_q (1 - p_q)] \, x_j + \theta \sum_j \Pi_q (1 - p_q) \, x_j \\ &= \sum_j [1 - (1 - \theta) \Pi_q (1 - p_q)] \, x_j \end{aligned}$$

We then denote

$$m \equiv 1 - (1 - \theta) \Pi_q (1 - p_q) \tag{7}$$

and since

$$X = \sum_j x_j.$$

we can write

$$K = mX, \tag{8}$$

where m is a reduced-form measure of the efficiency with which the banking sector transforms credit into capital. Therefore, we can think of m as the *endogenous* productivity of the banking sector. If $p_q = 1\ \forall q$, then $m = 1$ and capital is allocated with no waste. If $p_q = 0\ \forall q$, then $m = \theta$ and capital allocation is the most inefficient. According to this expression, bank j contributes to capital formation through its contribution to aggregate credit, X, and through its contribution to aggregate screening, m.

We are almost ready to derive the bank's optimal choice. One more step is required to show how the interest rate on loans R that appears in the profit function depends on aggregate capital K. Recall that the aggregate demand for capital is $R^K = f'(K) = \gamma K^{\gamma-1}$ where R^K is the rental rate on capital. Let l be the size of the loan demanded by an entrepreneur. With no screening, the project yields $k = \theta l$. What determines l? The entrepreneur wants to maximize the profit from the project, i.e., he solves $max_l \left(R^K k - Rl\right)$, where R is the interest rate on the loan. Since the entrepreneur is atomistic, he takes R^K and R as given, and therefore this problem becomes $max_l \left(R^K\theta - R\right) l$, which yields that the entrepreneur demands $l = \infty$ for $R^K\theta > R$, $l = 0$ for $R^K\theta < R$ and is indifferent about the size of l for $R^K\theta = R$. If, as is the case in our model, the entrepreneur knows that he is screened with probability p and that banks offer loans of the same size to screened and unscreened entrepreneurs, he expects the project to yield $k = p\theta l + (1-p)\theta l = \theta l$, exactly as before because as long as the loan size is the same random screening makes no difference. What if screened and unscreened loans are different? In this case, the yield is $k = p\theta l^{safe} + (1-p)\theta l^{risky}$ and the profit is

$$R^K\left[p\theta l^{safe} + (1-p)\theta l^{risky}\right] - R^{safe} p l^{safe} + R^{risky}(1-p)\, l^{risky},$$

where R^{safe} is the interest rate on a screened loan and R^{risky} on an unscreened loan. Maximization with respect to l^{safe} and l^{risky} yields the indifference condition $R^K\theta = R^{safe} = R^{risky}$ for both l^{safe} and l^{risky}. Therefore under random screening as well we have that entrepreneurs approach banks with the given reservation rate

$$R(K) = \theta\gamma K^{\gamma-1} \tag{9}$$

and are indifferent to loan size.

3.2 The bank's optimal choice of x_j and p_j

Equations (6), (7) and (8) allow us to rewrite the bank's problem as:

$$\max_{x_j,p_j} \left[mR\left(mX\right)-r\left(X\right)\right]x_j-\beta p_j \quad s.t.\ (5)\ and\ (9).$$

The first-order condition with respect to x_j is

$$mR-r+\left(m^2\frac{\partial R}{\partial X}-\frac{\partial r}{\partial X}\right)x_i=0. \tag{10}$$

The first-order condition with respect to p_j is

$$p_j=\begin{cases} 0 \quad for \quad x_j\left[R+m\frac{dR}{dK}X\right]\frac{dm}{dp_j}<\beta \\ ? \quad for \quad x_j\left[R+m\frac{dR}{dK}X\right]\frac{dm}{dp_j}=\beta \\ 1 \quad for \quad x_j\left[R+m\frac{dR}{dK}X\right]\frac{dm}{dp_j}>\beta \end{cases}. \tag{11}$$

It is useful to discuss these conditions separately.

Equation (10) describes the behavior of a bank with market power in the output (oligopolist) and input (oligopsonist) markets. To highlight what this implies, we rewrite it as

$$\frac{R}{r}=\underbrace{\frac{1}{m}}_{\substack{\text{inverse of}\\ \text{credit efficiency}\\ \text{due to screening}}}\cdot\underbrace{\frac{1+\frac{x_j}{X}\frac{1}{\varepsilon_r}}{1-\frac{x_j}{X}\frac{1}{\varepsilon_R}}}_{\substack{\text{exercise of}\\ \text{market power}}}, \tag{12}$$

where

$$\varepsilon_r\equiv\frac{\partial X}{\partial r}\frac{r}{X}=\frac{\alpha}{1-\alpha}\frac{W-X}{W} \tag{13}$$

and

$$\varepsilon_R\equiv-\frac{\partial X}{\partial R}\frac{R}{X}=\frac{1}{1-\gamma}. \tag{14}$$

are, respectively, the elasticity of saving supply and credit demand derived from (5) and (9). These elasticities capture the property that our banks internalize the effects of their individual quantity decisions on the total quantity of credit, which in turn affects the interest rates on loans and deposits. If we impose symmetry and write $\frac{x_j}{X}=\frac{1}{N}$ the equation captures the traditional view holding that the differential between the interest rate on loans and the interest rate on deposits is decreasing in the number of competing banks, N.

Therefore, the main benefit of competition is that the total volume of credit is larger while entrepreneurs obtain credit at lower rates. A novel feature of (12) is the presence of the inverse of the credit efficiency term m, which says that the more banks screen, the smaller is the spread between interest rates on loans and deposits. The reason is that when less credit is wasted, the economy accumulates more capital, and the corresponding lower marginal product of capital results into entrepreneurs' lower willingness to pay for loans.

Equation (11) states that when the marginal benefit of screening is smaller than the marginal cost the bank does no screening and therefore no information production. Conversely, if the marginal benefit of screening exceeds the marginal cost it is optimal to screen every entrepreneur, irrespective of free riding considerations. The middle line says that when the marginal benefit of screening equals the cost, the bank wants $p_j > 0$ but is indifferent to the specific value of p_j, which in equilibrium is determined by the simultaneous solution of the two first-order conditions. Observe that

$$m\frac{dR}{dK}X = m\frac{dR}{dX}\frac{dX}{dK}X = m\frac{dR}{dX}\frac{1}{m}X = \frac{dR}{dX}\frac{X}{R}R = -\frac{R}{\varepsilon_R}.$$

Using this result and (9), we can rewrite the indifference condition as

$$\underbrace{x_j}_{\substack{\text{total}\\ \text{credit}\\ \text{issued}}} \cdot \underbrace{\theta\gamma K^{\gamma-1}}_{\substack{\text{interest}\\ \text{rate}\\ \text{on loans}}} \cdot \underbrace{\left(1-\frac{1}{\epsilon_R}\right)}_{\substack{\text{contribution to}\\ \text{capital via}\\ \text{contribution to}\\ \text{credit efficiency}}} \cdot \underbrace{(1-\theta)\,\Pi_{q\neq j}\,(1-p_q)}_{\substack{\text{contribution to}\\ \text{aggregate}\\ \text{prob of screening}}} = \beta. \qquad (15)$$

Note that the bank's marginal benefit from screening is decreasing in p_q for all $q \neq j$, once again capturing the role of free riding. Note also that the condition can hold iff $0 < \theta < 1$ and $\beta > 0$. Intuitively, if either $\theta = 0$ or $\theta = 1$ screening does not matter and it is optimal to set $p_j = 0$. Similarly, if $\beta = 0$ the benefit is always larger than the cost and it is optimal to set $p_j = 1$ regardless of what the other banks do. Finally, using symmetry to write $\frac{x_j}{X} = \frac{1}{N}$ we have that, given X, as $N \to \infty$ the left-hand side vanishes and the bank does no screening.

4 The banking sector's symmetric equilibrium

To characterize the symmetric equilibrium of the banking sector, we solve simultaneously (12) and (15). We first rewrite them as:

$$\frac{\theta\gamma\left(mX\right)^{\gamma-1}}{\left(\frac{X}{W-X}\right)^{\frac{1-\alpha}{\alpha}}}=\frac{1}{m}\frac{N+\frac{1-\alpha}{\alpha}\frac{W}{W-X}}{N-1+\gamma};$$

$$\frac{X}{N}\theta\gamma^{2}\left(mX\right)^{\gamma-1}\left(1-\theta\right)\left(1-p\right)^{N-1}=\beta.$$

Then use (7) to obtain:

$$1-\left(1-\theta\right)\left(1-p\right)^{N}=\left[\frac{X^{1-\gamma}\left(\frac{X}{W-X}\right)^{\frac{1-\alpha}{\alpha}}}{\theta\gamma}\frac{N+\frac{1-\alpha}{\alpha}\frac{W}{W-X}}{N-1+\gamma}\right]^{1/\gamma}; \tag{16}$$

$$X=\left[\frac{\beta N}{\gamma^{2}}\frac{\left[1-\left(1-\theta\right)\left(1-p\right)^{N}\right]^{1-\gamma}}{\theta\left(1-\theta\right)\left(1-p\right)^{N-1}}\right]^{1/\gamma}. \tag{17}$$

We graph these two functions in (X,p) space in Figure 1. To fix ideas, we refer to (16) as the "lending curve" since it yields the optimal lending volume given the banks' screening probability. Similarly, we refer to (17) as the "screening curve" since it yields the optimal screening probability given the banks' lending volume. The equilibrium is the intersection of the two curves. The following two lemmas state formally the properties of the two curves. Since we are interested in the role of competition, we highlight the role of three structural parameters: N, because it is our measure of competition in banking; θ, because it regulates crucially the role of N; β, because it determines whether screening is profitable in the first place.

Lemma 1 *Denote the right-hand side of (16) as $F\left(X;W,N,\theta\right)$. The lending curve in (X,p) space is a function*

$$p^{L}(X;W,N,\theta)=\begin{cases}0 & 0\leq X\leq X_{0}^{L}\left(W,N,\theta\right)\\ 1-\left[\frac{F\left(X;W,N,\theta\right)}{1-\theta}\right]^{\frac{1}{N}} & X_{0}^{L}\left(W,N,\theta\right)<X<X_{1}^{L}\left(W,N,\theta\right)\\ 1 & X\geq X_{1}^{L}\left(W,N,\theta\right)\end{cases},$$

where:

- $X_0^L(W,N,\theta) = \arg\text{solve}\{\theta = F(X;W,N,\theta)\}$;
- $X_1^L(W,N,\theta) = \arg\text{solve}\{1 = F(X;W,N,\theta)\}$;
- in the region $X_0^L(W,N,\theta) < X < X_1^L(W,N,\theta)$ the following holds:
  - $p_X^L(X;W,N,\theta) > 0$;
  - $p_{XX}^L(X;W,N,\theta) < 0$;
  - $p^L(X;W,N,\theta)$ increasing in W, N, θ.

Proof The non-negativity constraint $p \geq 0$ binds whenever

$$p = 1 - \left[\frac{F(X;W,N,\theta)}{1-\theta}\right]^{\frac{1}{N}} < 0.$$

Therefore, there exists a value $X_0^L(W,N,\theta)$ defined by

$$1 - \left[\frac{F(X;W,N,\theta)}{1-\theta}\right]^{\frac{1}{N}} = 0 \Rightarrow F(X;W,N,\theta) = \theta$$

and such that $p = 0$ for $X \leq X_0^L(W,N,\theta)$. The constraint $p \leq 1$ binds whenever

$$p = 1 - \left[\frac{F(X;W,N,\theta)}{1-\theta}\right]^{\frac{1}{N}} > 1.$$

Therefore, there exists a value $X_1^L(W,N,\theta)$ defined by

$$1 - \left[\frac{F(X;W,N,\theta)}{1-\theta}\right]^{\frac{1}{N}} = 0 \Rightarrow F(X;W,N,\theta) = 0$$

and such that $p = 1$ for $X \geq X_1^L(W,N,\theta)$. The other properties follow directly from differentiating $F(X;W,N,\theta)$ with respect to X, W, N, θ.□

As said, equation (16) describes the aggregate amount of credit X that banks wish to issue given their screening strategy p. If banks choose to do no screening, they only make risky loans and therefore they make the smallest amount of credit. The curve is increasing because as banks do more screening, they wish to lend more since lending becomes more efficient and it generates a higher rate of return. As banks approach the extreme where they always screen, and make only safe loans, the aggregate amount of credit reaches its maximum. The convexity of the curve reflects the fact that as banks want to extend more credit, because they screen more, they must also pay a higher interest rate on deposits. An important property of this curve is that when the wage rises (because the previous period capital stock is larger), savers demand a lower interest rate on deposits and the banks' profit margin rises. Accordingly, they lend more for any given screening strategy p.

Lemma 2 *Denote the right-hand side of (17) as $\Omega(p; N, \theta, \beta)$. The screening curve in (X, p) space is a function*

$$p^S(X; N, \theta, \beta) = \begin{cases} 0 & 0 \leq X \leq X_0^S(N, \theta, \beta) \\ \Omega^{-1}(p; N, \theta, \beta) & X > X_0^S(N, \theta, \beta) \end{cases},$$

where:

- $X_0^S(N, \theta, \beta) = \left[\frac{\beta N}{\gamma^2} \frac{\theta - \gamma}{1 - \theta}\right]^{1/\gamma}$;
- *in the region $X > X_0^S(N, \theta, \beta)$ the following holds:*
  - $p_X^S(X; N, \theta, \beta) > 0$;
  - $p_{XX}^S(X; N, \theta, \beta) < 0$;
  - $\lim_{X \to \infty} p^S(X; N, \theta, \beta) = 1$;
  - $p^S(X; N, \theta, \beta)$ *increasing in N, β, U-shaped in θ.*

Proof Study (17) in (p, X) space and then plot its inverse in (X, p) space. Direct differentiation establishes how the curve shifts with N, β, θ.□

Equation (17) describes the screening strategy p that banks wish to adopt given the aggregate amount of credit X. The first property to note is that if aggregate credit is too low, the amount of credit x of the individual bank is too small and the bank cannot cover the fixed cost of screening. Accordingly, banks choose $p = 0$. Only when aggregate credit is sufficiently large banks start screening. This property, of course, stems from the assumption that screening entails a fixed cost per entrepreneur. In the region where $p > 0$, the curve is monotonically increasing because as aggregate credit gets larger banks spread the fixed cost of screening on larger loans. If $N > 1$, it converges asymptotically to $p = 1$. The concavity of the function reflects the fact that the bank's benefit from screening depends on its contribution to credit efficiency m and through it to K. Since the interest rate on loans decreases with K, an increase in credit X that induces an increase in p is subject to diminishing returns. Note that because the marginal benefit of screening depends only on the interest rate on loans, the screening curve does not depend on the wage W.

The equilibrium of the banking sector is the intersection of the lending curve (16) and the screening curve (17). Consider Figure 1. The curves intersect only once for positive p. This intersection is stable in that a deviation with higher X for the same p leads banks to reduce X and thus return to the intersection point. The point $X_0^S(N, \theta, \beta)$ is also a possible equilibrium, since it yields the optimal amount of credit given $p = 0$, but it is unstable and thus can be ignored.

Inspection of Figure 1 reveals that there exists a critical value of the wage $W_0(N, \theta, \beta)$ such that for $W \leq W_0(N, \theta, \beta)$ the equilibrium is $\left(X_0^L, 0\right)$ since the two curves do no intersect for $p > 0$. The value of $W_0(N, \theta, \beta)$ follows from solving

$$X_0^L(W, N, \theta) = X_0^S(N, \theta, \beta).$$

This is the first aggregate implication of the model and highlights the role of the fixed cost of screening β.

Remark 1 *For $W \leq W_0(N, \theta, \beta)$, the banking sector is unable to afford information production. Only if $W > W_0(N, \theta, \beta)$ screening and the associated information production is profitable. In this case the equilibrium has the property that both the quantity and the quality of credit increase in W because increases in W shift the lending curve (16) up and yield a movement along the screening curve (17). As $W \to \infty$, $X \to \infty$ while $p \to 1$.*

The following proposition summarizes these insights.

Proposition 3 *The equilibrium of the credit market is represented by two functions*

$$X(W; N, \theta, \beta) = \begin{cases} X_0^L(W, N, \theta) & 0 \leq W \leq W_0(N, \theta, \beta) \\ X^*(W; N, \theta, \beta) & W > W_0(N, \theta, \beta) \end{cases}$$

and

$$p(W; N, \theta, \beta) = \begin{cases} 0 & 0 \leq W \leq W_0(N, \theta, \beta) \\ p^*(W; N, \theta, \beta) & W > W_0(N, \theta, \beta) \end{cases}$$

with the following properties:

- $W_0(N,\theta,\beta)$ *is increasing in* N, θ, β *with*
  - $\lim_{N\to\infty} W_0(N,\theta,\beta) = \infty$,
  - $\lim_{\theta\to 1} W_0(N,\theta,\beta) = \infty$,
  - $\lim_{\beta\to\infty} W_0(N,\theta,\beta) = \infty$.
- $X_0^L(W,N,\theta)$ *is increasing in* W *with*

$$X_0^L(W_0(N,\theta,\beta);N,\theta) = X^*(W_0(N,\theta,\beta);N,\theta,\beta).$$

- $X_W^*(W;N,\theta,\beta) > 0$ *with* $\lim_{W\to\infty} X^*(W;N,\theta,\beta) = \infty$.
- $p_W^*(W;N,\theta,\beta) > 0$ *with* $\lim_{W\to\infty} p^*(W;N,\theta,\beta) = 1$.

The next three remarks highlight equilibrium properties related to bank competition.

Remark 2 The number of banks N has an ambiguous effect on the equilibrium values p^* and X^*.

Inspection of Figure 1 provides the intuition for this property. When N rises and the market becomes more competitive, profit margins shrink and banks increase the volume of credit they issue, given the screening strategy p. This effect is captured by the downward shift of the lending curve (16). On the other hand, banks also wish to do less screening since the marginal benefit from screening shrinks. This effect is captured by the downward shift of the screening curve (17). As one can see, these shifts yield an ambiguous effect of the number of banks on both the individual probability of screening p and total lending X.

Remark 3 *Since* $X_0^L(W;N,\theta)$ *is increasing in* N*, a larger number of banks delays the onset of screening in the sense that the more competitive is the banking sector, the higher is the volume of saving (due to a higher* W*) that triggers screening.*

The reason is that with lower profit margins, the indifference condition (15) holds only if the overall market is larger so that the individual bank lends more. When W crosses the threshold $W_0(N,\theta,\beta)$ we start seeing the trade-off that our model identifies: competition reduces margins and tends to raise total credit, but it also reduces the incentive to screen, the efficiency of credit, and therefore the banks' willingness to lend.

Remark 5 *It is possible to have two intersections, in which case the rightmost one is stable and the other one unstable. This pattern yields multiple equilibria because the point $X_0^L(W; N, \theta)$ is feasible and locally stable.*

The most interesting consequence of this configuration is that as the wage grows, the lending curve (16) shifts down and eventually becomes tangent to the screening curve (17). It is then possible to have a discontinuous jump to the interior stable equilibrium with $p > 0$. In the analysis below we focus on the case of a unique equilibrium because it is simpler and captures fully the model's main insight.

The analysis in the previous section has characterized how banks' individual decisions about screening drive the aggregate efficiency term

$$m = 1 - (1 - \theta)(1 - p)^N .$$

This term plays a critical role in determining capital accumulation since

$$K = mX.$$

It is therefore useful to restate Proposition 3 in terms of these two variables in order to highlight how they depend on the wage and the model's parameters. The main advantage of this exercise is that it yields directly the equation that governs the aggregate dynamics of the model.

To this end, we rewrite (16)-(17) as:

$$m = \left[\frac{\left(\frac{K}{m}\right)^{1-\gamma} \left(\frac{\frac{K}{m}}{W - \frac{K}{m}}\right)^{\frac{1-\alpha}{\alpha}}}{\theta\gamma} \frac{N + \frac{1-\alpha}{\alpha}\frac{W}{W - \frac{K}{m}}}{N - 1 + \gamma} \right]^{1/\gamma} ; \tag{18}$$

$$K = \left[\frac{\beta N}{\gamma^2} \frac{m}{\theta (1-\theta)^{\frac{1}{N}} (1-m)^{\frac{N-1}{N}}} \right]^{1/\gamma} . \tag{19}$$

These two loci have properties that are isomorphic to those of the (16)-(17) curves studied in Figure 1. We refer to them as the "efficiency curve" and the "accumulation curve", respectively. We can then construct Figure 2 and obtain the following result.

Proposition 4 *The equilibrium level of credit market efficiency and the equilibrium amount of capital that the economy builds within each period are represented by two functions*

$$K(W;N,\theta,\beta)=\begin{cases} K_0^A(W;N,\theta) & 0\le W\le W_0(N,\theta,\beta)\\ K^*(W;N,\theta,\beta) & W>W_0(N,\theta,\beta)\end{cases}$$

and

$$m(W;N,\theta,\beta)=\begin{cases} \theta & 0\le W\le W_0(N,\theta,\beta)\\ m^*(W;N,\theta,\beta) & W>W_0(N,\theta,\beta)\end{cases}$$

with the following properties:

- $W_0(N,\theta,\beta)$ *is increasing in* N, θ, β *with*
  - $\lim_{N\to\infty} W_0(N,\theta,\beta)=\infty$,
  - $\lim_{\theta\to 1} W_0(N,\theta,\beta)=\infty$,
  - $\lim_{\beta\to\infty} W_0(N,\theta,\beta)=\infty$.
- $K_0^A(W;N,\theta)$ *is increasing in* W.
- $K_0^A(W_0(N,\theta,\beta);N,\theta)=K^*(W_0(N,\theta,\beta);N,\theta,\beta)$.
- $K_W^*(W;N,\theta,\beta)>0$ *with* $\lim_{W\to\infty} X^*(W;N,\theta,\beta)=\infty$.
- $m_W^*(W;N,\theta,\beta)>0$ *with* $\lim_{W\to\infty} m^*(W;N,\theta,\beta)=1$.

Proof The graph in the lower panel of Figure 2 shows the construction of the efficiency curve in the upper panel. Denote the right-hand side of (18) as $\Upsilon(m;K,W,N,\theta)$. This is a monotonically decreasing function of m with a vertical asymptote at

$$m=\frac{K}{W}.$$

This function is also increasing in K. Thus as K rises, $\Upsilon(m;K,W,N,\theta)$ shifts up and traces the 45^0 line, thereby generating a function $m^A(K;W,N,\theta)$ increasing in K. The constraint $m\ge\theta$ (i.e., $p\ge 0$) binds whenever $K\le K_0^A(W,N,\theta)$ defined by

$$\Upsilon\left(\theta;K_0^A,W,N,\theta\right)=\theta.$$

Similarly, The constraint $m\le 1$ (i.e., $p\le 1$) binds whenever $K\ge K_1^A(W,N,\theta)$ defined by

$$\Upsilon\left(1;K_1^A,W,N,\theta\right)=1.$$

The function $m^A(K; W, N, \theta)$ is decreasing in W since $\Upsilon(m; K, W, N, \theta)$ is decreasing in W.

Constructing the efficiency curve $m^E(K; N, \theta, \beta)$ simply requires studying the function (19) in (K, m) space and then plotting it in (m, K) space. Note that there exists a value $K_0^A(N, \theta, \beta)$ such that for $K \leq K_0^A(N, \theta, \beta)$ we have $m = \theta$.

The equilibrium is the intersection of the two curves. The effect of the wage follows from the downward shift of the accumulation curve. If the wage is too low, that is, if $W \leq W_0$, where W_0 is defined by $K_0^A(W, N, \theta) = K_0^E(N, \theta, \beta)$, the equilibrium is $m = \theta$ and $K_0^A(W, N, \theta)$. If instead $W > W_0$, the interior equilibrium generates two functions $K^*(W; N, \theta, \beta)$ and $m^*(W; N, \theta, \beta)$ both increasing in W since as W rises the accumulation curve shifts down and yields a movement along the efficiency curve. As $W \to \infty$ we have that $K \to \infty$ and $m \to 1.\square$

As said, Proposition 4 yields directly the equation governing the general equilibrium path of the economy. Note that the effect of the number of banks N on credit efficiency m is in principle ambiguous, but this follows directly from Remark 2. However, referring to Figure 2, we see that if the shift of the efficiency curve (19) dominates over the shift of the accumulation curve (18) the effect of an increase in N is to reduce m. Inspection of the two equations suggests that, in fact, this property is likely to hold. The reason is that the efficiency curve (19) shifts down without bound, while the accumulation curve (18) shifts with the term

$$\frac{N + \frac{1-\alpha}{\alpha}\frac{mW}{mW-K}}{N-1+\gamma},$$

which is bounded above. Thus, while we cannot rule out that starting from small values of N the initial effect of increasing N is to raise m, we can fully expect that as N grows very large its effect on m becomes negative.

An interesting way of interpreting this property is to think of $1-m$ as the *losses-to-loans ratio*. Then, the prediction of the model is that (too much) competition raises the losses-to-loans ratio. This prediction is consistent with the empirical evidence presented by Shaffer [29], who documents a negative relationship between the number of banks operating in a market and the losses-to-loans ratio.

5 Aggregate implications

5.1 Dynamics

Recall that the wage is an increasing function of the lagged capital stock, $W_t = W(K_{t-1})$; see equation (3). Accordingly, there is a value $K_{t-1} = K_0(N, \theta, \beta)$ such that $W(K_0) = W_0(N, \theta, \beta)$. Recall that $W_0(N, \theta, \beta)$ is increasing in N, θ, β. Therefore, $K_0(N, \theta, \beta)$ is increasing in N, θ, β. Proposition 4 then yields

$$K_{t+1} = \begin{cases} K_0^A(W(K_t); N, \theta) & 0 \leq K_t \leq K_0(N, \theta, \beta) \\ K^*(W(K_t); N, \theta, \beta) & K_t > K_0(N, \theta, \beta) \end{cases},$$

This implies that there are two regions of the state-space wherein banking is, respectively, fully inefficient and only partially inefficient. Once K_t passes the threshold $K_0(N, \theta, \beta)$, the economy moves to a higher capital accumulation trajectory because banks reach the minimum scale necessary to make screening profitable. The following proposition states these properties formally.

Proposition 5 *The economy's general equilibrium is described by the first-order difference equation*

$$K_{t+1} = \Phi(K_t; N, \theta, \beta), \tag{20}$$

where

$$\Phi = \begin{cases} K_0^A(W(K_t); N, \theta) & 0 \leq K \leq K_0(N, \theta, \beta) \\ K^*(W(K_t); N, \theta, \beta) & K > K_0(N, \theta, \beta) \end{cases}.$$

The function $\Phi(K; \cdot)$ is continuous, differentiable everywhere except at the point $K = K_0(N, \theta, \beta)$, and exhibits the following properties which ensure that there exists at least one non-trivial steady state $K_{ss} > 0$:

- $\Phi(0) = 0$;
- $\Phi_K(\cdot) > 0$;
- $\lim_{K \to 0} \Phi_K(K) = \infty$;
- $\lim_{K \to \infty} \Phi_K(K) = 0$.

The trajectory marked in bold in Figure 3 illustrates the dynamics of the economy. Note that because of the threshold $K_0(N, \theta, \beta)$ multiple steady states may emerge. This is an important result since it is exactly the number of banks N that determines whether this happens.

Remark 6 *Even with standard primitives that would normally guarantee a well-behaved dynamic transition to a unique steady state with screening, the market structure of the banking sector can yield a development trap with no screening.*

Inspection of Figure 3 shows that conditional on being in the lower steady state with no information production, a change in N can remove the steady state with no screening by shifting the curve $\Phi(K_t; N, \theta, \beta)$ above the 45^0 line for all $K_t \leq K_0(N, \theta, \beta)$ and thereby putting the economy on a path that leads to the steady state with screening. Yet, conditional on being on the upper trajectory with screening, the number of banks that maximizes steady-state income can be different from the one that allows escaping from the trap.

It is useful to be specific. Removing the trap requires that the number of banks satisfies

$$K_0^A(W(K_0); N, \theta) > K_0(N, \theta, \beta). \tag{21}$$

This condition says nothing about the level of N that makes the curve $K^*(W(K_t); N, \theta, \beta)$ as high as possible, which is what is required to maximize the steady-state level of capital produced by equilibria with screening. In fact, the change in N required to remove the trap can very well shift the $K^*(W(K_t); N, \theta, \beta)$ curve *down*, thereby reducing steady-state capital.

Note also that since both sides of (21) are increasing in N, the sign of the change in N required to satisfy it, starting from some arbitrary level of N, is ambiguous. Thus, it is quite possible to have that escaping the trap requires an increase in competition while maximizing steady-state income requires a decrease. If so, the model's implication is that the optimal number of banks is contingent on K_t. A specific example makes this discussion more concrete.

Suppose there exists a regulator who sets the number of banks. Suppose also that (21) yields a threshold N_{trap} such that for $N > N_{trap}$ (21) holds while for $N \leq N_{trap}$ it does not. Suppose, finally, that there exists a number of banks $N_{\max}$ that maximizes the level of capital in the steady state with $p > 0$ and such that for $N_{\max}$ the no-trap condition (21) fails, i.e., $N_{\max} < N_{trap}$. Then the regulator should set $N > N_{trap}$ to remove the trap and keep it there until the economy has accumulated $K_t > K_0(N_{\max}, \theta, \beta)$. Once the economy passes this threshold, the regulator can set $N = N_{\max}$ because the economy is out of the basin of attraction of the underdevelopment trap associated to $N_{\max}$.

It is possible to construct several examples with similar features. The general insight is that regulating the banking sector through direct control of the number of banks is a complex exercise that requires detailed information. The analysis of the steady state with screening that we undertake next underscores this point.

5.2 The steady state

For simplicity we focus on the case where the function $\Phi(K_t; N, \theta, \beta)$ has a unique steady state with screening. The steady state value of the capital stock, $K_{ss}(N, \theta, \beta)$, is the solution of the system:

$$m = \left[\frac{\left(\frac{K}{m}\right)^{1-\gamma} \left(\frac{\frac{K}{m}}{(1-\gamma)K^{\gamma} - \frac{K}{m}} \right)^{\frac{1-\alpha}{\alpha}}}{\theta\gamma} \frac{N + \frac{1-\alpha}{\alpha} \frac{(1-\gamma)K^{\gamma}}{(1-\gamma)K^{\gamma} - \frac{K}{m}}}{N - 1 + \gamma} \right]^{1/\gamma} ; \tag{22}$$

$$K = \left[\frac{\beta N}{\gamma^2} \frac{m}{\theta (1-\theta)^{\frac{1}{N}} (1-m)^{\frac{N-1}{N}}} \right]^{1/\gamma} . \tag{23}$$

The graph in the lower panel of Figure 4 shows the construction of the accumulation curve that we use in the upper panel. Denote the right-hand side of (22) as $\Psi(m; K, N, \theta)$. This is a monotonically decreasing function of m with a vertical asymptote at

$$m = \frac{K^{1-\gamma}}{1-\gamma}.$$

The function is also increasing in K so that as K rises, $\Psi(m; K, N, \theta)$ shifts up and traces the 45^0 line, thereby generating a function $m_{ss}^{A}(K; N, \theta)$ increasing in K. The constraint $m \geq \theta$ (i.e., $p \geq 0$) binds whenever $K \leq K_0^L(N, \theta)$ defined by

$$\Psi\left(\theta; K_0^A, N, \theta\right) = \theta.$$

Similarly, The constraint $m \leq 1$ (i.e., $p \leq 1$) binds whenever $K \geq K_1^A(N, \theta)$ defined by

$$\Psi\left(1; K_1^A, N, \theta\right) = 1.$$

We then obtain the kinked accumulation curve in the figure. The screening curve is the same as in Figure 2.

We have two types of solutions. If (23) is above (22) for all values of $m \geq \theta$, the equilibrium is given by (22) evaluated at $m = \theta$. The more interesting case is when (23) and (22) intersect for $m \in [0,1]$, that is for $p > 0$.

5.3 Effect of the model's parameters on the steady state

As mentioned earlier, the main parameters of interest are N, θ and β. A change in the cost of screening β has an obvious and unambiguous effect on the steady state levels of both m and K. A decrease in β shifts up the efficiency curve but it does not affect the accumulation curve, thus leading to both higher m and K.

Since our main focus is the role of competition, we want to study how this equilibrium changes with the number of banks. From what gathered so far, the role of N is intrinsically ambiguous. When N increases, both (23) and (22) shift down, capturing the fact that more competition reduces the interest rate spread and generates more credit, given banks' choice of p, while it reduces bank's incentives to screen, given the size of the credit market X.

However, this ambiguity can be resolved if we investigate further the role of N *in conjunction* with the role of θ. This parameter measures the average quality of entrepreneurs and thus is a good indicator of agents' need to tackle informational frictions in the model.

In order to clarify this relationship we begin first by looking at the steady state at the extreme cases $N = 1$ and $N \to \infty$. Consider first (22)-(23) when $N = 1$:

$$m = \Psi\left(m; K, 1, \theta\right) \equiv \left[\frac{\left(\frac{K}{m}\right)^{1-\gamma}\left(\frac{\frac{K}{m}}{(1-\gamma)K^{\gamma}-\frac{K}{m}}\right)^{\frac{1-\alpha}{\alpha}}}{\theta\gamma}\frac{1+\frac{1-\alpha}{\alpha}\frac{(1-\gamma)K^{\gamma}}{(1-\gamma)K^{\gamma}-\frac{K}{m}}}{\gamma}\right]^{1/\gamma}; \tag{24}$$

$$K = \left[\frac{\beta}{\gamma^2}\frac{m}{\theta\left(1-\theta\right)}\right]^{1/\gamma}. \tag{25}$$

Note that the efficiency curve no longer converges asymptotically to $m = 1$ for $K \to \infty$ but instead admits a finite value of K that yields $m = 1$. The reason is that the monopoly bank does not face the free riding problem that dampens the incentives to screen faced by oligopolistic banks. In particular, the monopoly bank converges to a steady state with $m = 1$ if the accumulation curve (24) cuts the $m = 1$ line to the right of the point where the screening curve (25) cuts it. This requires that the parameters satisfy:

$$K_{ss}(1, \theta) > \left[\frac{\beta}{\gamma^2} \frac{1}{\theta(1-\theta)} \right]^{1/\gamma},$$

where

$$K_{ss}(1, \theta) = \arg\text{solve}\{1 = \Psi(1; K, 1, \theta)\}. \tag{26}$$

In the following, it is useful (albeit not necessary) to assume that this condition holds.

Consider now (22)-(23) when $N \to \infty$:

$$m = \Psi(m; K, \infty, \theta) \equiv \left[\frac{\left(\frac{K}{m}\right)^{1-\gamma} \left(\frac{\frac{K}{m}}{(1-\gamma)K^{\gamma} - \frac{K}{m}} \right)^{\frac{1-\alpha}{\alpha}}}{\theta\gamma} \right]^{1/\gamma}; \tag{27}$$

$$m = \theta \quad \forall K. \tag{28}$$

The second line captures the property that when the number of banks is too large, banks choose $p = 0$. In this case then, the equilibrium is at the intersection of the accumulation curve (28) with the line $m = \theta$, that is,

$$K_{ss}(\infty, \theta) = \arg\text{solve}\{\theta = \Psi(\theta; K, \infty, \theta)\}. \tag{29}$$

Now we are ready to establish the connection with the parameter θ, which, as said, governs the importance of screening. Observe first that for $\theta = 1$ equations (29) and (26) yield $K_{ss}(\infty, 1) > K_{ss}(1, 1)$. That is, when we shut down the model's friction, and thus make screening unimportant, the equilibrium with the monopolistic distortion of the quantity of credit is inferior to the one without it. In fact, since for $\theta = 1$ we have $m = 1$ regardless of the number of banks, it is clear that to maximize long-run output we should let $N \to \infty$. Consider now $\theta \to 1$, that is, the fraction of good entrepreneurs is not exactly 1 but still so high that screening is almost irrelevant. Again, we obtain that competition raises steady-state capital because it eliminates the dead-weight losses associated with banks' market power. In other words, the relation between N and K_{ss} is monotonically increasing and the number of banks that maximizes steady-state capital is $N^* \to \infty$.

Things change drastically when we move to the opposite end of the spectrum. The case $\theta = 0$ is trivial since it implies shutting down the credit market altogether. We thus focus on $\theta \to 0$, which means that there are so few good entrepreneurs that screening is crucial because the losses from inefficient lending are too large and outweigh the benefits of eliminating market power. In this case, K_{ss} is monotonically decreasing in N and the number of banks that maximizes steady-state capital is $N^* = 1$.

The intuition driving the extreme cases, which we illustrate in Figure 5, suggests that for intermediate values of θ, a hump-shaped relation should emerge yielding that the number of banks that maximizes steady-state capital is a finite value $N^* \in (1, \infty)$. In other words, oligopoly banking strikes the best possible balance between the deadweight losses from banks' market power and the benefits of efficient lending.

6 Conclusion

We have presented a dynamic general equilibrium model of capital accumulation in which oligopolistic banks serve as financial intermediaries between savers and entrepreneurs. The assumption that entrepreneurs are of unknown quality provides the informational friction that assigns a central role to banks. Specifically, entrepreneurs do not know their type but banks can administer costly screening to find out. The question then is under what conditions banks choose to sustain the screening cost given that its outcome is not appropriable. Intuitively, this depends on the degree of competition since it determines both the margin of intermediation and the likelihood that a bank administering the test is subject to free-riding by its competitors. The model allows us to study in detail these issues and to characterize how competition affects the general equilibrium path of the economy.

With this contribution we fill a gap in the theoretical literature on finance and growth, a literature that has recognized the importance of banks in fostering economic growth but has not explored in depth the role played by the market structure of the banking industry. Perhaps the reason is that conventional wisdom suggests that perfect competition — price taking behavior due to a large number of banks — should be the optimal market structure. However, the available empirical evidence, and existing models of financial intermediation, paint a more nuanced picture, suggesting the existence of multiple channels through which banking market structure affects growth, thus ultimately providing an unclear picture regarding the sign of the relationship.

A particularly valuable feature of our model is that it is extremely parsimonious: we do not make special assumptions and have only two free parameters that affect the role of the number of banks — the fraction of good entrepreneurs and the cost of screening. Therefore, our main ingredients are simply the information externality associated with the screening activity of banks and their oligopolistic rivalry. We also stress the importance of a dynamic, general equilibrium approach, which at the cost of added complexity allows us to obtain a rich set of new results.

First, we show that bank competition has an intrinsically ambiguous impact on aggregate economic variables. Without additional conditioning on other covariates it is just not clear whether a more competitive banking sector leads to better outcomes. This theoretical insight would then explain the apparent lack of consistent results from the empirical evidence.

Second, we show how to resolve the ambiguity. Namely, conditioning on variables that proxy for the characteristics of the entrepreneurial population yields sharper predictions. In environments where entrepreneurs are on average of relatively low quality, hence where informational frictions are especially severe, we predict higher capital accumulation under less competitive banking market structures. The opposite is true in environments where the population of entrepreneurs is inherently of high quality. In intermediate cases, where the average probability of success is neither very low nor very high, the market structure that maximizes long-run development is an oligopoly. If we are willing to assume that population characteristics may evolve along the development path, then these results imply a further dynamic dimension to the determination of optimal banking market structure that the literature has ignored.

Third, we show that development traps may emerge just because of banking market structure. This is a very important result for the following reasons: a) It means that we could have otherwise identical economies, including the same banking market structure, exhibiting significantly different levels of income and/or growth trajectories. This is another explanation for the inconclusive empirical results emerging in the literature; b) The normative implication for a banking regulator is that the optimal market structure to escape from the trap is not the optimal market structure to achieve the highest levels of economic development. This is another insight pointing at a necessarily dynamic nature of banking regulation; c) It implies a severe criticism of policies regarding emerging markets, where the traditional prescription is that in order to achieve income convergence it is necessary to "simply" adopt the same institutional and regulatory environments prevalent *today* in developed countries. In fact, such policies, proposed without deeper qualifications, are ineffective at best and very damaging at worst.

References

[1] Allen Franklin and Douglass Gale, (2004a), "Competition and Financial Stability," Journal of Money, Credit and Banking, 36, 3, 453-80.

[2] Allen Franklin and Douglass Gale, (2004b), "Financial Intermediaries and Markets," Econometrica, 72, 4, 1023-1061.

[3] Barro, Robert and Xavier Sala-i-Martin, (2005), *Economic Growth*, MIT Press.

[4] Beck, Thorsten, Asli Demirguc-Kunt, and Vojislav Maksimovic (2004), "Bank competition and access to finance: International evidence", Journal of Money, Credit, and Banking, 36, 627–648.

[5] Bertrand, Marianne, Antoinette Schoar and David Thesmar (2007), "Banking Deregulation and Industry Structure: Evidence from the French Banking Reforms of 1985", The Journal of Finance, 62(2), 597-628.

[6] Bhattacharya, Sudipto and Anjan V. Thakor (1993), "Contemporary Banking Theory", Journal of Financial Intermediation, 3, 2-50.

[7] Black, Sandra E. and Philip E. Strahan (2002) "Entrepreneurship and Bank Credit Availability", Journal of Finance, 57, 2807-2833.

[8] Bonaccorsi di Patti, Emilia and Giovanni Dell'Ariccia, (2004) "Bank Competition and Firm Creation," Journal of Money, Credit and Banking, Blackwell Publishing, 36, 2, 225-51.

[9] Arnoud W. A. Boot and Anjan V. Thakor (2000), "Can Relationship Banking Survive Competition?," Journal of Finance, vol. 55, 2, 679-713.

[10] Boyd, John,Gianni De Nicolo' and Bruce Smith, (2004), "Crises in Competitive Versus Monopolistic Bankings Systems", Journal of Money, Credit and Banking, 36, 2, 487-506.

[11] Boyd, John and Gianni De Nicolo', (2005), "The Theory of Bank Risk-Taking and Competition Revisited", Journal of Finance, 60, 3, 1329-1343.

[12] Campbell, Tim S. and William A. Kracaw (1980), "Information Production, Market Signalling, and the Theory of Financial Intermediation", Journal of Finance, 35, 4, 863-882.

[13] Cetorelli, Nicola (2009), "Banking and Real Economic Activity", in *The Oxford Handbook of Banking*, Allen N. Berger, Phillip Molyneux and John Wilson, Eds., Series : Oxford Handbooks in Finance, Oxford University Press.

[14] Cetorelli, Nicola and Michele Gambera (2001), "Banking Market Structure, Financial Dependence and Growth: International Evidence from Industry Data", Journal of Finance, 56, 2, 617-48.

[15] Cetorelli, Nicola and Philip E. Strahan (2006), "Finance as a Barrier to Entry: Bank Competition and Industry Structure in Local U.S. Markets", Journal of Finance 61(1), 437-61.

[16] Chan, Yuk-Shee, Stuart I.Greenbaum and Anjan V. Thakor, (1986), "Information reusability, competition and bank asset quality," Journal of Banking and Finance, Elsevier, 10, 2, 243-253.

[17] Dell'Ariccia, Giovanni and Robert Marquez, (2004), "Information and Bank Credit Allocation", Journal of Financial Economics, 72, 1, p. 185-214.

[18] Diamond. Douglas (1984), "Financial intermediation and delegated monitoring", Review of Economic Studies, 393-414.

[19] Fischer, Karl H. (2000), "Acquisition of Information in Loan Markets and Bank Market Power. An Empirical Investigation", EFA 0593.

[20] Guzman, Mark, (2000), "Bank Structure, Capital Accumulation, and Growth: A Simple Macroeconomic Model", Economic Theory, 16(2), 421-455.

[21] Hauswald, Robert and Robert Marquez, (2006), "Competition and Strategic Information Acquisition in Credit Markets" Review of Financial Studies, 19, 3, p. 967-1000.

[22] Jayaratne, Jith and Philip E. Strahan (1996), "The Finance-Growth Nexus: Evidence from Bank Branch Deregulation", Quarterly Journal of Economics, 111(3), 639-70.

[23] Leland, Hayne E. and David H. Pyle (1977), "Informational Asymmetries, Financial Structure, and Financial Intermediation ", Journal of Finance, 32, 2, 371-387.

[24] Levine, Ross (2005), "Finance and Growth: Theory and Evidence." in Philippe Aghion and Steven Durlauf, eds. *Handbook of Economic Growth,* The Netherlands: Elsevier Science.

[25] Marquez, Robert, (2002), "Competition, Adverse Selection, and Information Dispersion in the Banking Industry," Review of Financial Studies, 15, 3, p. 901-926.

[26] Rajan, Raghuram G. (1992), "Insiders and Outsiders: The Choice between Informed and Arm's-Length Debt," Journal of Finance, vol. 47(4), p. 1367-400.

[27] Petersen, Mitchell A. and Rajan, Raghuram G. (1995), "The Effect of Credit Market Competition on Lending Relationship", Quarterly Journal of Economics, 110, 407–443.

[28] Ramakrishnan, Ram T.S. and Anjan V. Thakor (1984), "Information Reliability and a Theory of Financial Intermediation", Review of Economic Studies, 51, 415-432.

[29] Shaffer, Sherrill (1998), "The Winner's Curse in Banking", Journal of Financial Intermediation, 7, 4, 359–392.

[30] Sharpe, Steven A. (1990), "Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships", The Journal of Finance, 45, 4, 1069-1087.

[31] Thakor, Anjan V., (1996), "Capital Requirements, Monetary Policy, and Aggregate Bank Lending: Theory and Empirical Evidence", The Journal of Finance, 51, 1, 279-324.

[32] Zarutkie, Rebecca, (2006), "Evidence on the effects of bank competition on firm borrowing and investment," Journal of Financial Economics, 81, 503-537.

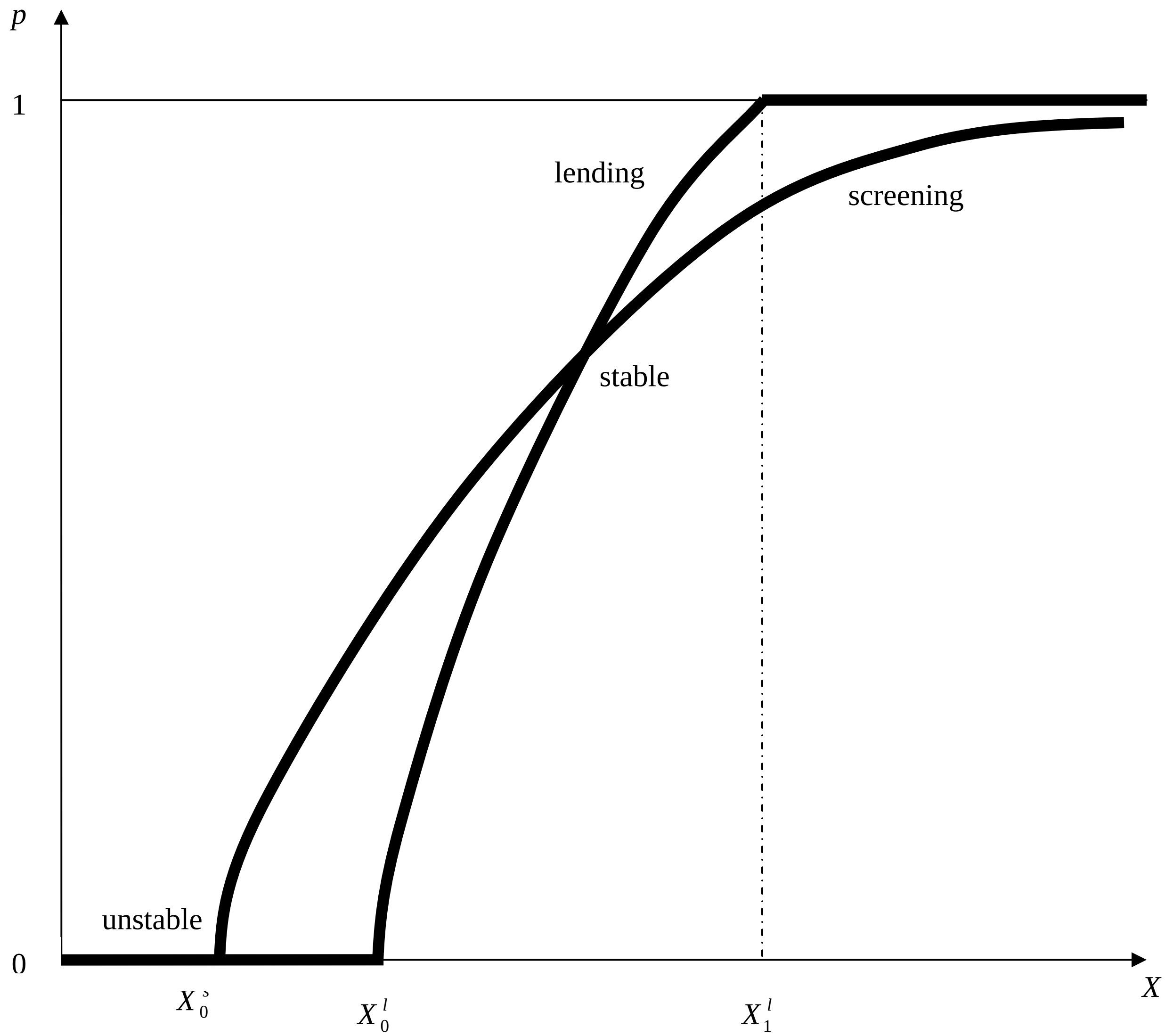

Figure 1: Equilibrium *p* and *X*

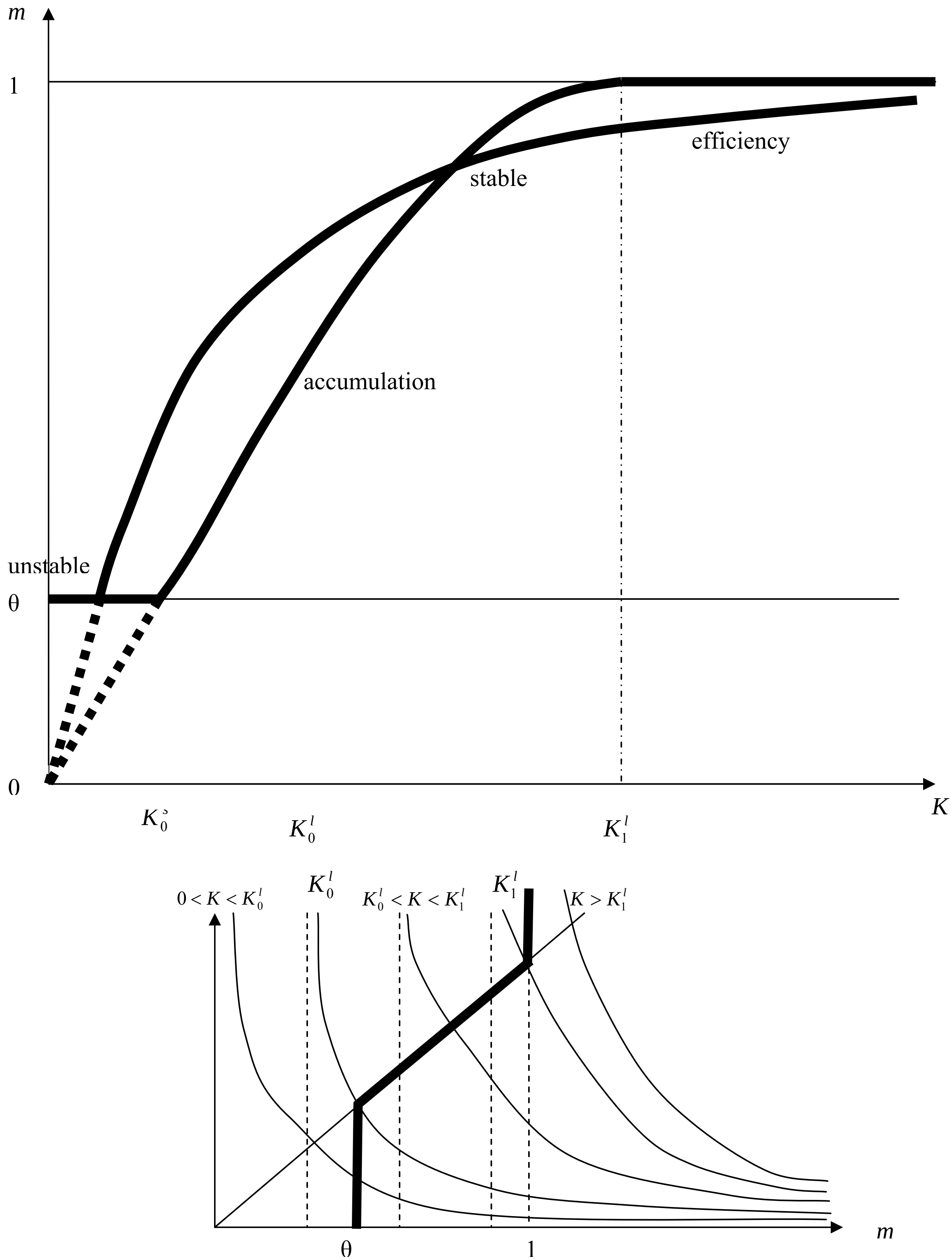

Figure 2: Equilibrium *m* and *K*

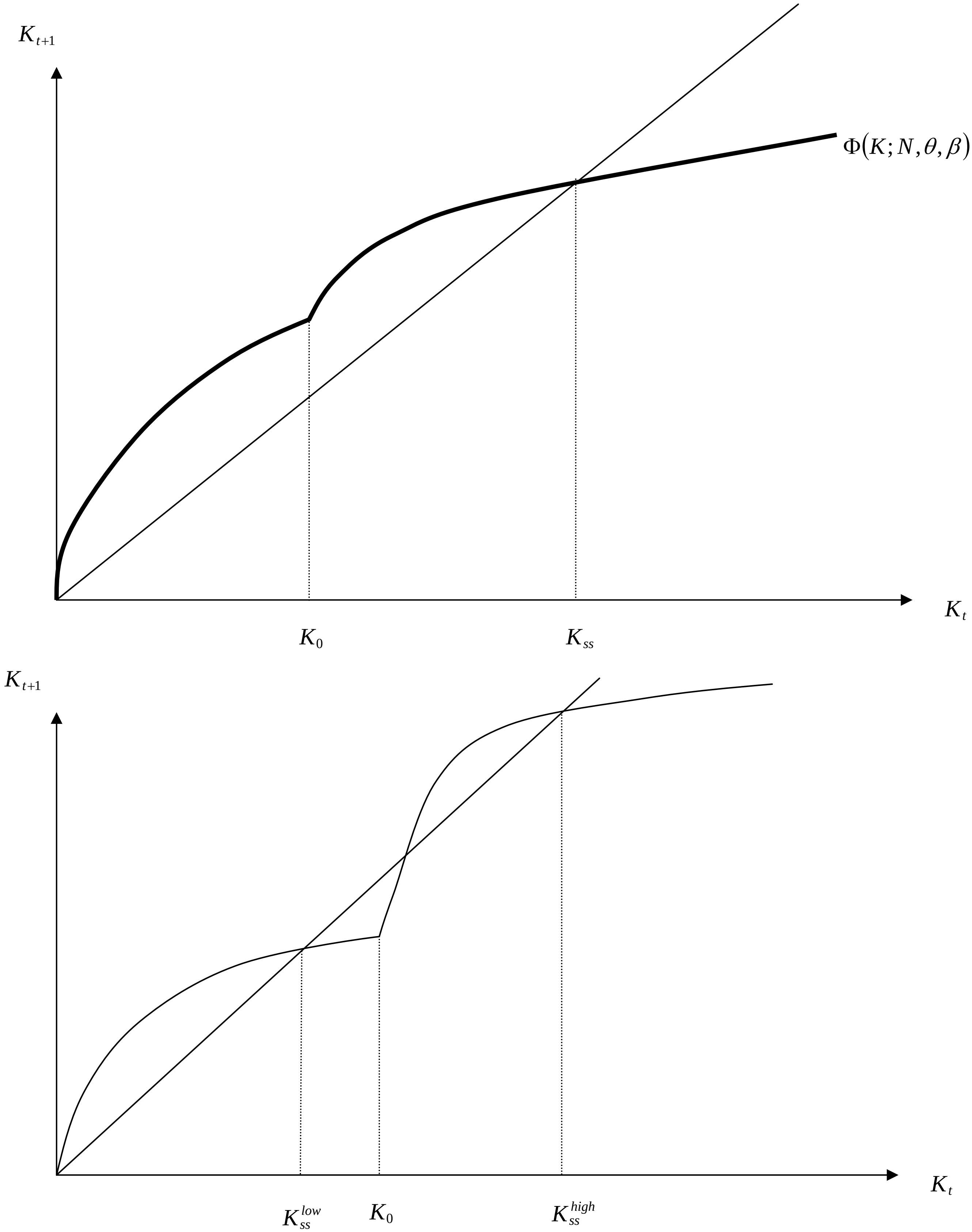

Figure 3: General equilibrium Dynamics

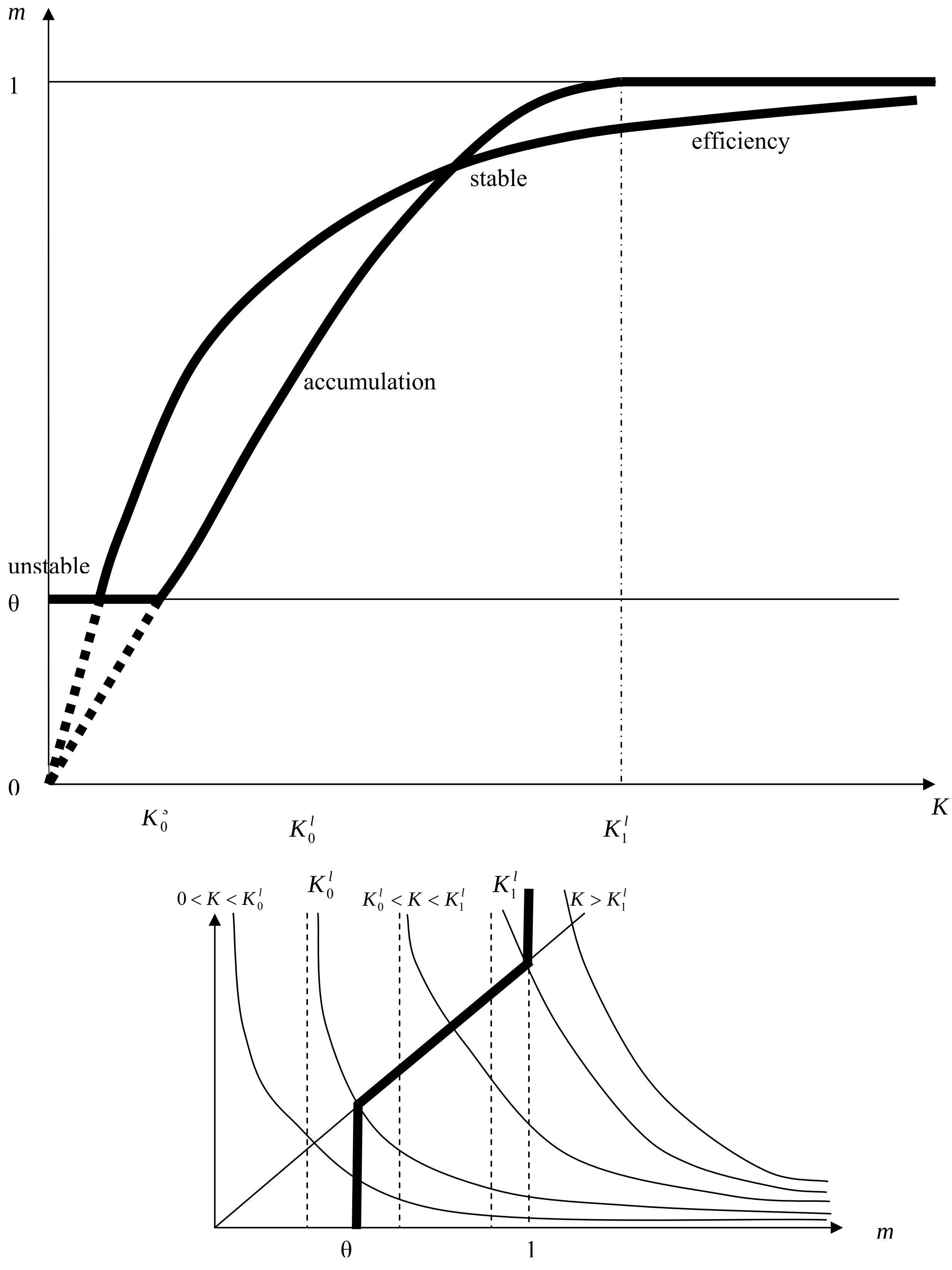

Figure 4: Steady-state equilibrium *m* and *K*

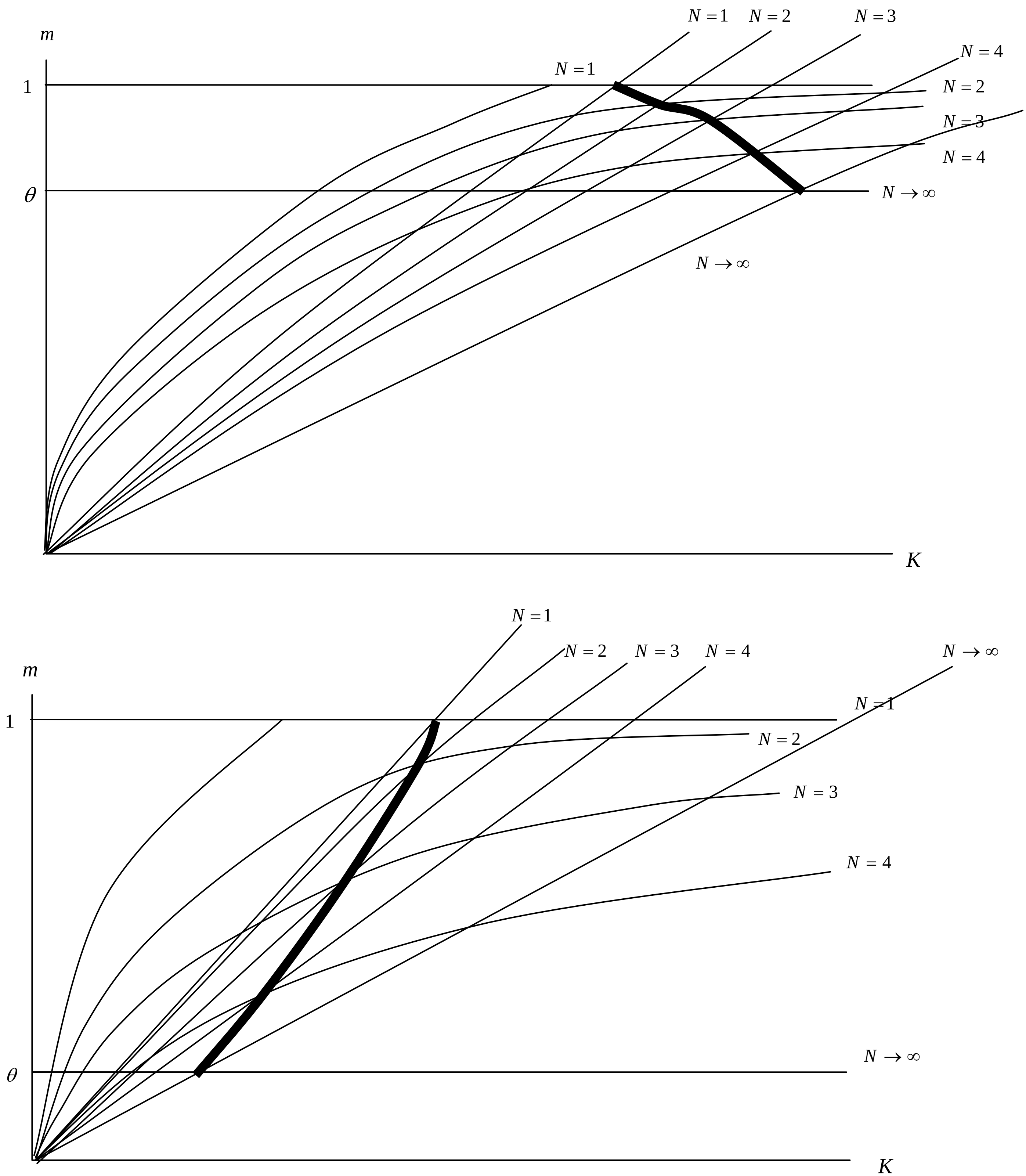

Figure 5: Effect of N when θ is high and low