Oligopoly Banking and Capital Accumulation*

Nicola Cetorelli[†]
*Federal Reserve Bank
of Chicago*

Pietro F. Peretto
*Duke University
Department of Economics*

December 2000

Abstract

We develop a dynamic general equilibrium model of capital accumulation where credit is intermediated by banks operating in a Cournot oligopoly. The number of banks affects capital accumulation through two channels. First, it affects the quantity of credit available to entrepreneurs. Second, it affects banks' decisions to collect costly information about entrepreneurs, and thus determines the efficiency of the credit market. We show that under plausible conditions, the market structure that maximizes the economy's steady-state income per capita is neither a monopoly nor competition, but an intermediate oligopoly. Moreover, the credit market splits in two segments: one in which loans are screened and only high quality entrepreneurs obtain credit, and one in which banks extend credit indiscriminately to all entrepreneurs. The relative size of the two segments depends on the market power of banks and evolves endogenously along the path of capital accumulation. We thus obtain the prediction that the banking sector becomes more sophisticated as the economy develops.

JEL code: G1, G2, L1, L2, O1, O4

[†]Corresponding author. Mailing address: Federal Reserve Bank of Chicago, 11th Floor, 230 South La Salle St., Chicago, IL 60604, USA. E-mail: ncetorelli@frbchi.org

Abstract

We develop a dynamic general equilibrium model of capital accumulation where credit is intermediated by banks operating in a Cournot oligopoly. The number of banks affects capital accumulation through two channels. First, it affects the quantity of credit available to entrepreneurs. Second, it affects banks' decisions to collect costly information about entrepreneurs, and thus determines the efficiency of the credit market. We show that under plausible conditions, the market structure that maximizes the economy's steady-state income per capita is neither a monopoly nor competition, but an intermediate oligopoly. Moreover, the credit market splits in two segments: one in which loans are screened and only high quality entrepreneurs obtain credit, and one in which banks extend credit indiscriminately to all entrepreneurs. The relative size of the two segments depends on the market power of banks and evolves endogenously along the path of capital accumulation. We thus obtain the prediction that the banking sector becomes more sophisticated as the economy develops.

1   Introduction

Banks have been shown to promote economic growth by providing liquidity and risk pooling, by screening entrepreneurs, and by reducing or eliminating the agency problems associated with adverse selection and moral hazard.[1] The literature, however, has paid much less attention to the implications for growth of the market structure of the banking industry. Does it matter whether banks operate in a highly competitive market or whether they have monopoly power? A standard argument suggests that any departure from competition can only be detrimental for economic growth because banks with market power restrain the supply of loanable funds in order to charge higher interests rates. Recent empirical studies, however, sketch a more complicated picture and suggest the existence of multiple channels working in opposite directions through which banking market structure affects growth.

Shaffer [25] analyzes U.S. cross-sectional data and finds that household income grows faster in markets with a higher number of banks. Similarly,

---

[1]A representative subset of the numerous contributions to this literature are those of Greenwood and Jovanovic [12] and Bencivenga and Smith [1] among the theoretical ones, and those of King and Levine [15], Rajan and Zingales [21] and Levine and Zervos [18] on the empirical side. A survey of the literature is in Levine [17].

Black and Strahan [3], analyze cross-state U.S. data and find that the creation of new firms is higher in less concentrated banking markets. In contrast, Bonaccorsi and Dell'Ariccia [4] analyze cross-industry, cross-markets Italian data and find an overall positive effect of bank concentration on new firms growth, and even an amplification of such effect for more informationally opaque industrial sectors. Petersen and Rajan [20], look at small business firms in the U.S. and find that such firms are less credit constrained, and younger ones are charged lower loan rates, if they are located in more concentrated banking markets. Finally, Cetorelli and Gambera [8] analyze cross-country, cross-industry data and find that a more concentrated banking industry imposes a deadweight loss in the credit market, resulting in a reduction in the total quantity of loanable funds, but that the effect is *heterogenous* across industrial sectors. Specifically, firms in industries more dependent on external finance actually benefit from being in a country with a concentrated banking sector.

The evidence thus suggests the existence of multiple effects of banking market structure for economic growth, and that the relationship may be non-monotonic. The main goal of this paper is to provide theoretical foundations for this relationship, identifying opposite effects of banking concentration.[2] To this end, we develop a general equilibrium model of capital accumulation[3] in which banks compete in a Cournot fashion.[4]

The literature has long stressed that one of banks' crucial roles is to produce information about entrepreneurs in order to evaluate their credit worthiness (e.g., Leland and Pyle [16], Ramakrishnan and Thakor [23]). We construct a model where information production is the main role played by banks. However, we also highlight the free-riding problem associated with information production (see, e.g., Campbell and Cracaw [5]). We then study how the market power of banks affects their incentives to spend on information production, given that the information they produce is not fully appropriable.

---

[2]There exists a recent literature on the potential shortcomings of banking competition, e.g., Cao and Shi [6], Dell'Ariccia [9], Manove, Padilla and Pagano [19], Petersen and Rajan [20], Riordan [24], Shaffer [25]. These contributions do not focus on growth.

[3]We build, in part, on Cetorelli [7] who uses a similar approach but provides only a comparison of perfect competition and monopoly. A related contribution is Guzman [13], who studies a general equilibrium model of capital accumulation with a focus on banking market structure. His results, however, highlight only a negative effect of banks' market power on growth.

[4]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 firms – banks in our case. This provides us with a very tractable model.

Banks compete with each other in gathering individual savings and in loaning funds to entrepreneurs. We let banks have access to a screening technology that, at a cost, allows them to discriminate between high and low quality entrepreneurs. While the outcome of the screening test may not be observable by third parties, competitor banks can still extract information about the screened entrepreneurs by simply observing whether the bank extends or denies the loan.[5] In other words, there is an informational externality that generates a free-riding problem. This weakens banks' incentives to incur the cost of screening and to carry out an information-based (efficient) lending strategy. We show that the bank's optimal strategy entails screening entrepreneurs only with some probability, thereby extending both "safe" (screened) and "risky" (unscreened) loans. The credit market is thus endogenously segmented in two parts whose relative size evolves along, and has feedbacks with, the path of economic development. Within this theoretical framework, we identify two major effects of banking market structure on economic growth. On one hand, the fewer the number of banks, the smaller the total quantity of credit available to entrepreneurs. On the other hand, the fewer the number of banks, the higher the incentive to produce information,[6] and therefore the larger the proportion of funds allocated to screened, high quality entrepreneurs. Therefore, the number of banks governs a trade-off between the overall size of the credit market and its efficiency. The size and efficiency of the credit market, in turn, determine the return to capital accumulation and therefore to saving.

As a result of this tension, the model shows that the relationship between banking market structure and steady-state income per capita may be non-monotonic. In other words, the market structure that maximizes economic development is neither a monopoly nor perfect competition, but an oligopoly. Moreover, the transitional dynamics exhibit an endogenous regime switch whereby the qualitative properties of the financial sector evolve with its quantitative growth: for given market structure, as the economy grows and the financial sector expands, the share of screened lending rises and the banking sector becomes more sophisticated, in the sense that credit intermediation is more and more based on the production of information prior to the extension of loans. The predictions associated with these theoretical results, all stemming from the non-monotonic relationship between banking market structure and capital accumulation, are consistent with the empirical

---

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

[6]Fischer [11] provides evidence that concentration enhances banks' information production in their lending activity.

evidence surveyed above.

2 The Economy

The economy is populated by overlapping cohorts living for two periods. Each cohort is a continuum of mass one and population is constant. Each young agent is a potential entrepreneur, endowed with no capital and with one unit of labor. When old, agents do not work and do not operate a productive technology.

2.1 The primitives: technology and preferences

The assumption that there is a mass one of workers implies that there is no distinction between aggregate and per worker variables. There exists a competitive firm producing a final good with technology

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

where Y_t and K_t are, respectively, output and capital at time t. The function $f(\cdot)$ is a standard neoclassical production function that satisfies the Inada conditions.[7] Most of our results are obtained in full generality and do not require us to specialize the production function to the Cobb-Douglas case. When needed, we impose restrictions on the parameter γ of the Cobb-Douglas specification. Factors' demand schedules are well-behaved functions:

$$
\begin{aligned}
R_t &= f_K(K_t) = \gamma K_t^{\gamma-1}; \tag{2}\\
W_t &= f(K_t) - K_t f_K(K_t) = (1-\gamma) K_t^{\gamma}, \tag{3}
\end{aligned}
$$

where R_t is the rate of return on capital and W_t is the wage rate.

Agents have preferences over consumption in both periods. Let c_t and c_{t+1} be consumption at time t and $t+1$ for a representative member of generation t. Agents maximize the utility function[8]

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

---

[7]These conditions are: $f(0) = 0$; $f_K(K) > 0$; $f_{KK}(K) < 0$; $\lim_{K \to 0} f_K(K) = \infty$; $\lim_{K \to \infty} f_K(K) = 0$. In words, the marginal product of capital is positive and decreasing and convex in K.

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

subject to:

$$c_t = W_t - s_t;$$
$$c_{t+1} = s_t r^s_{t+1},$$

where s_t is the amount of saving at time t and r^s_{t+1} is the rate of return on saving.[9] The function $u(\cdot)$ is a standard utility function that satisfies the Inada conditions (see above). As for the production function, for most of our results we do not need to specialize it to the Cobb-Douglas case. When needed, we highlight which results require restrictions on the parameter α. Substituting the two constraints into (4), the solution to the maximization problem is the upward sloping saving supply schedule

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

where we have used the assumption that there is a mass one of workers to write the function in terms of aggregate savings, S_t. The saving function exhibits the following properties: $0 \leq S < W$; $h(0; W) = 0$; $h_S(S; W) > 0$; $\lim_{S \to W} h(S; W) = \lim_{W \to S} h(S; W) = \infty$; $h_W(S; W) < 0$; $\lim_{W \to \infty} h(S; W) = 0$.[10]

2.2 Capital accumulation

Investment is a two-stage process. In a first stage, entrepreneurs borrow from banks to finance production of capital goods. Production of capital goods is subject to idiosyncratic, entrepreneur-specific uncertainty. If it is successful, in a second stage the entrepreneur rents capital services to the final producer at the competitive rental rate. If it is not successful, the borrowed resources are lost and the entrepreneur defaults on the loan.

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,

---

[9] In this model banks make positive profits. In order to account for these profits, we assume that banks are institutions 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^s_t is this rate of return. Banks' profits are thus part of the resources that old agents use to finance consumption.

[10] The following additional properties hold: $h_{SS}(S; W) > 0$; $h_{SW}(S; W) < 0$; $\lim_{W \to \infty} h_S(S; W) = 0$. In words, the return to saving is increasing and convex in S. Moreover, the marginal return to saving is decreasing in W and vanishes for W very large.

supply their savings to banks. Entrepreneurs borrow from banks. They either succeed or fail in the investment stage. From the successful investment projects the aggregate capital stock is obtained, 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 young entrepreneurs of generation $t+1$.

2.3 The free-riding problem

There are two types of entrepreneurs: a type H always succeeds in the investment stage; a type L fails with probability one. Let $\theta \in [0,1]$ be the time-invariant proportion of type H entrepreneurs. As in Sharpe [26], entrepreneurs 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 the function that banks perform.

Banks gather savings from old agents and lend to entrepreneurs. As all other agents in the economy, banks do not have ex-ante information regarding the quality of individual entrepreneurs but they know the distribution of types. Banks, however, are endowed with a screening technology according to which they can learn the entrepreneurs' quality type by spending an amount β per entrepreneur. If an entrepreneur is recognized of high quality, the bank extends a line of credit at conditions determined by market equilibrium. If an entrepreneur is recognized of low quality the bank rejects the loan application. If a bank does not perform screening, it lends indiscriminately to all entrepreneurs relying on the law of large numbers to capture the proportion θ of type H entrepreneurs.[11]

Screening produces valuable information concerning entrepreneurs. If such information is not appropriable, a free-riding problem arises. Assume, for example, that the results of the screening tests performed by a bank on its clients were public knowledge. A competitor bank could extend safe

---

[11]We treat the number of banks as exogenous because the banking industry is typically subject to regulations and controls that determine the number of banks in relation to the objective function of the regulator, as opposed to entry/exit dictated by profitability (see Rajan and Zingales [22] for an extensive discussion of "political economy" arguments that rationalize this observation.) The available empirical evidence supports this view. Looking at a cross-section of countries, Cetorelli and Gambera [8] cannot reject the hypothesis that banking concentration is exogenous.

loans to tested, high quality entrepreneurs without bearing the screening cost. Given this incentive, no bank would undertake screening and credit would be allocated inefficiently. Notice that to have free riding it is not necessary to impose that the results of the tests are public knowledge – an overly restrictive assumption – but it is sufficient to assume that a bank's decision to extend or deny a loan is observable. This assumption is more realistic because it requires that in order to certify to another bank that he is of high quality, an entrepreneur simply declares that he has been extended a loan (e.g., by showing an offer letter by the screening bank). The free riding problem then arises because should a bank screen *all* its clients, even if the results of the tests are private information, competitor banks would still infer the entrepreneur's type, since those who receive a loan would be necessarily of high quality.

The main conclusion of this discussion is that under plausible conditions 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.[12] A solution to the free-riding problem is for banks to inject noise by screening entrepreneurs only with some probability p. In the following section we show that this is indeed the strategy that banks play in equilibrium.

3 Lending strategies

In this section, we study banks' lending strategies. First, we show that banks randomize and screen with some probability. Next, we develop the implications of this strategy for the relation between credit and capital.

3.1 Endogenous credit market segmentation

Consider one entrepreneur wishing to finance a project. The entrepreneur applies for credit. If he obtains credit and the project is successful, it generates positive revenues for the bank who finances it. If the project is unsuccessful the borrowed resources are lost and revenues are zero. Consider now a bank's choices. If the bank decides to screen the entrepreneur, it sustains the cost of screening and makes a safe loan regardless of what the other banks do. If the bank decides not to screen, two outcomes are possible. First, at

---

[12]This externality could be ruled out by assuming that all information related to banks' screening activity and lending decisions is private. This scenario is as unrealistic as the one in which all information is public.

least one of the other banks screens the entrepreneur, the bank learns the entrepreneur's type and makes a safe loan without sustaining the screening cost. Second, no other bank screens the entrepreneur, and the bank makes a risky loan whose expected payoff depends on the unconditional distribution of types.

To formalize these ideas, assume that there are N banks, where N is an exogenous number, and consider a two-stage game where in stage one the bank decides whether to screen (strategy YS) or not (strategy NS). In stage two, Cournot competition determines the bank's payoff given its stage-one choice. The following matrix describes stage-two payoffs. These depend on what the other banks do. The first column gives the number of other banks that play YS. The other two columns give the payoff to the typical bank, given its stage-one choice and how many other banks play YS.

	YS	NS
$N-1$	$\pi_{YS} - \beta$	π_{YS}
$N-2$	$\pi_{YS} - \beta$	π_{YS}
.	.	.
.	.	.
0	$\pi_{YS} - \beta$	π_{NS}

For the purposes of this analysis, we do not need the exact expressions for the payoffs in stage two of the game. All we need is their ordering. We calculate the payoffs in the next subsection, where we solve the stage-two Cournot game. Notice, however, that our assumption that the screening cost does not depend on the amount of credit yields that the gross profit per bank, in the case when at least one bank screens the entrepreneur, is the same regardless of how many other banks play YS (we show this formally in subsection 4.2). Thus, if the bank chooses YS, then in stage two it earns $\pi_{YS} - \beta$ regardless of the other banks' choices. If the bank chooses NS, then in stage two it earns π_{YS} if at least one other bank chooses YS, while it earns π_{NS} if no other bank chooses YS.

We now characterize the equilibria of this game.

Proposition 1 *There are two cases:*

- *if $\pi_{NS} > \pi_{YS} - \beta$ there is one pure-strategy Nash equilibrium where banks do not screen;*

- *if $\pi_{NS} \leq \pi_{YS} - \beta$ there is one mixed-strategy Nash equilibrium where*

banks screen the entrepreneur with probability

$$p = 1 - \left[\frac{\beta}{\pi_{YS} - \pi_{NS}}\right]^{\frac{1}{N-1}}. \tag{6}$$

Proof. See the Appendix.

The proposition states that banks choose to screen if it is profitable, but randomization is necessary in order to eliminate the free-riding problem. The game discussed above posited one entrepreneur. Aggregating over the continuum of entrepreneurs, the equilibrium behavior in mixed strategy implies the existence of two separate segments of the credit market: one in which banks screen, accept high-quality entrepreneurs, and reject low-quality applications. The other segment is one where banks do not screen and lend regardless of quality. We refer to them as, respectively, the "efficient" and the "inefficient" segments of the credit market.

3.2   From credit to capital: the role of banks

We now construct the relationship between credit and capital. Let X_t be the total volume of credit issued by banks at time t. (This results from aggregation of individual quantities, x_{it}, that solve profit maximization problems. In the next subsection we provide a full characterization of individual and aggregate quantities.) Let X_t^S and X_t^U denote, respectively, credit allocated to screened and unscreened entrepreneurs. The entire quantity X_t^S becomes productive capital, while only a fraction θX_t^U does. Therefore, capital is

$$K_t = X_t^S + \theta X_t^U.$$

Depending on the information flow across banks, we have two cases.

Consider a regime where banks can observe whether or not another bank has extended a loan, but they cannot observe the terms of the loan contract. This implies that screened and unscreened loans can differ in size without revealing the result of the screening test. Let x_t^S be the size of the loan offered to a screened entrepreneur found to be of high quality. Let x_t^U be the loan offered to an unscreened entrepreneur. Aggregation across the two groups yields $X_t^S = p_t \theta x_t^S$ and $X_t^U = (1 - p_t) x_t^U$. (Given our assumption that there is a mass 1 of entrepreneurs, p_t is the fraction of entrepreneurs that are screened.) Since all entrepreneurs are ex-ante identical, they all expect to receive the same amount of credit, x_t, at the market interest rate on loans. With a mass 1 of entrepreneurs, it must be $x_t = X_t$. We refer to

10

x_t as the unconditional loan.[13] We then notice that an entrepreneur applies for credit knowing that if he is not screened nothing changes for him and he receives a loan of size x_t, while if he is screened he receives a loan of size x_t^S with probability θ and a loan of size 0 with probability $1 - \theta$. The definition of unconditional loan then implies the following relationship

$$x_t = (1 - p_t) x_t^U + p_t \theta x_t^S + p_t (1 - \theta) 0.$$

The important step in our calculation is to notice that $x_t = x_t^U$ because an unscreened entrepreneur is identical to an ex-ante entrepreneur. Substituting, we obtain $x_t = \theta x_t^S$, which yields $x_t^S = \frac{1}{\theta} x_t$. Therefore, under the assumption that other banks cannot observe the terms of the loan contract, banks are able to offer larger loans to screened, high quality entrepreneurs by allocating to them credit from screened and rejected ones. We can now calculate

$$K_t = p_t \theta x_t^S + \theta (1 - p_t) x_t^U = [p_t + \theta (1 - p_t)] X_t. \qquad (7)$$

The term in brackets measures the efficiency of the credit market.

Consider now a regime where banks can observe the terms of the contract that another bank offers to a client. If another bank can observe the size of the loan, it can infer the quality of the entrepreneur, since there are only two types of loans and an entrepreneur offered the larger loan must be a high-quality one who passed the screening test. This implies that no bank can discriminate between screened and unscreened loans without revealing the result of the screening test. Since loans cannot differ in size, banks set $x_t^S = x_t^U$ and reallocate the credit denied to screened entrepreneurs who failed the test among *all* other entrepreneurs. The total mass of entrepreneurs who obtain credit is $p_t \theta + 1 - p_t$. Given the volume of credit X_t, each loan is of size

$$x_t^S = x_t^U = \frac{X_t}{p_t \theta + 1 - p_t}.$$

This loan is larger than the unconditional loan, x_t. To see this, notice that

$$x_t = (1 - p_t) x_t^U + p_t \theta x_t^S + p_t (1 - \theta) 0 \Rightarrow x_t = X_t < x_t^S$$

since $p_t \theta + 1 - p_t < 1$. The relationship between credit and capital is then

$$K_t = p_t \theta x_t^S + \theta (1 - p_t) x_t^U = \frac{\theta}{p_t \theta + 1 - p_t} X_t. \qquad (8)$$

---

[13]The important point here is that heterogeneity across entrepreneurs comes into play only *after* the screening test, not before. This means that when they approach the banks, all entrepreneurs have the same expectations about the size of the loan that they obtain, the interest rate at which they obtain it, and their probability of success in turning the loan into productive capital.

The quantity X_t that enters this equation is of course different from the one that enters equation (7).

It is informative to compare these two cases. An important aspect of the model is that lending to unscreened entrepreneurs is risky. In the first case, the amount of funds that banks lend to unscreened entrepreneurs is lower than the amount they lend to screened ones. Hence, all entrepreneurs borrow at the same rate but unscreened entrepreneurs are "rationed" relatively to screened entrepreneurs. This is because banks discriminate in order to take into account risk. In the second case, banks cannot discriminate because offering loans at different conditions reveals the result of the test and thus makes randomization ineffective. In order to solve the free-riding problem, banks must offer the same conditions to all entrepreneurs — whether they screen them or not. But there is a cost associated with reallocating credit from screened, low-quality entrepreneurs to everybody else because some of this credit goes to unscreened entrepreneurs who are of low quality with probability $1 - \theta$. In contrast, in the first case all reallocated credit goes to high-quality entrepreneurs who succeed with probability 1. This means that given the volume of credit and the probability of screening that arise in equilibrium in each of the two cases, the efficiency of the credit market is higher in the first one. Despite this difference, the two cases give rise to aggregate dynamics that are qualitatively similar. We thus concentrate the rest of the paper on the first case, while deferring the analysis of the second one to the appendix.[14]

4 Equilibrium of the banking sector

In this section, we solve for the equilibrium of the banking sector and characterize the total volume of credit and its composition in credit to screened entrepreneurs and credit to unscreened entrepreneurs. These steps are necessary to characterize the probability p as a function of variables exogenous

---

[14]We think that the model with unobservable contracts is more realistic because the revelation of information due to the observability of the terms of the loan contracts stems from the fact that there are only two loan sizes. In reality, entrepreneurs are heterogeneous and thus apply for loans of different size. In the presence of two distributions of entrepreneurs, one of high quality and one of low quality, observing the size of a loan contract would not be sufficient to determine with certainty what distribution an entrepreneur belongs to. One could also argue that with observable contracts banks do not need to give identical loans but could just randomize and give the large loan with probability q and the small loan with probability $1 - q$. We believe that introducing this additional layer of complication would not produce new insights.

to the banking sector and structural parameters. Characterizing p is necessary to characterize the relationship between credit and capital, in turn necessary to characterize the general equilibrium path of the economy.

4.1 Banks' quantities and profits with credit market segmentation

Recall that the demand for capital is a function $R_t = f_K(K_t)$, where R_t is the capital rental rate and K_t is the capital stock. Since the market for capital services is competitive, we obtain the following demand for credit

$$R_t = [p + (1-p)\theta] f_K([p + (1-p)\theta] X_t), \tag{9}$$

where R_t is the interest rate on loans charged by banks.[15]

The probability p_t is determined in stage one of the game and is taken as given in stage two. The saving function (5) can be written as $r_t^s = h(X_t, W_{t-1})$, where we use the fact that in equilibrium total credit must equal aggregate saving, $X_t = S_{t-1}$. This relationship gives the unit cost of lending for banks. Since in this part of the analysis we do not need to keep track of the dating of variables, for the remaining of this section we drop the time subscript and denote $W_{t-1} = W$ with the understanding that W is the wage in the previous period.

Let x_i be the total credit issued by bank i. The bank's *gross* profit can be written $\pi = (R - r) x_i$, where to simplify the notation we let R denote the interest rate on loans and r the interest rate on deposits. These rates are given by, respectively, (9) and (5).

The bank maximizes the total *net* profit from lending in both segments of the credit market,

$$\underset{x_i}{\text{Max}} \quad (R - r) x_i - \beta p,$$

where $X = \sum_i x_i$. The first order condition is

$$R - r + \left(\frac{\partial R}{\partial X} - \frac{\partial r}{\partial X}\right) x_i = 0,$$

---

[15] To see this, note that entrepreneurs who supply capital services solve

$$\max_{K_t} [R_t K_t - r_t^x X_t] \quad \text{s.t.} \quad K_t = X_t [\theta + p(1-\theta)],$$

where they take as given R_t, r_t^x, and p. This problem is linear in K_t and yields the equilibrium condition $R_t = r_t^x$, which implies (9).

which can be rewritten

$$R = r\frac{1 + \frac{1}{N\varepsilon_r}}{1 - \frac{1}{N\varepsilon_R}}, \tag{10}$$

where

$$\varepsilon_r \equiv \frac{\partial X}{\partial r}\frac{r}{X} = \frac{\alpha}{1 - \alpha}\frac{W - X}{W}, \tag{11}$$

$$\varepsilon_R \equiv -\frac{\partial X}{\partial R}\frac{R}{X} = \frac{1}{1 - \gamma}. \tag{12}$$

ε_r and ε_R are, respectively, the elasticity of saving supply and credit demand derived from (9) and (5). Equation (10) describes the behavior of a Cournot oligopolist with market power in the output and input markets. Symmetry allows us to drop the subscript i and write $x = \frac{X}{N}$. This equation captures the traditional view of the positive role of competition in banking, whereby 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. This supports a larger volume of investment and the economy accumulates more capital, thereby enjoying all the beneficial effects that follow in terms of higher income per capita.

We now establish some important partial equilibrium results for total credit, for the scale of activity of the individual bank and for bank profits, necessary to obtain a full characterization of banks' equilibrium lending strategy, summarized by the function p. In characterizing p we identify a second effect of bank competition, namely that screening is decreasing in the degree of bank competition.

Our procedure for analyzing the comparative statics of the partial equilibrium of the banking sector is based on Dixit [10]. Define the function

$$\mu(X; W, \theta, p) \equiv [p + (1 - p)\theta] f_K([p + (1 - p)\theta]X) - h(X; W),$$

which describes the interest rates differential, $R - r$, and let λ be a member of the vector (W, θ, p). Holding X constant, we have $\mu_\lambda > 0$ because θ and p improve the demand conditions that banks face, while W improves the cost conditions. We impose the following assumption.

Assumption 1 $\mu_\lambda + \frac{X}{N}\mu_{X\lambda} > 0$ for $\lambda = W, \theta, p$.

This assumption simply requires that, in case their signs differ, the first-order effects of W, θ, and p dominate their second-order effects. It is surely

14

satisfied in the case of the Cobb-Douglas specifications of the production and utility functions (1) and (4).

Proposition 2 *The total quantity of credit is a function $X(W; N, \theta, p)$ with the following properties:*

- $X(0; \cdot) = 0$;

- $X_W(W; \cdot) > 0$;

- $\lim_{W \to \infty} X(W; \cdot) = \infty$;

- *increasing in N, θ, and p.*

Proof. See the Appendix.

Proposition 3 *The quantity of credit supplied by the individual bank is a function $x(W; N, \theta, p)$ with properties similar to those of the function describing total credit, with the exception that x is decreasing in N.*

Proof. See the Appendix.

The properties of the functions $X(W; N, \theta, p)$ and $x(W; N, \theta, p)$ are intuitive in that an increase in θ, p or W improves the demand or cost conditions for all banks while an increase in N raises competition. Specifically, θ and p raise the probability that loans be successful, while W reduces the cost of collecting savings from households. An increase in N raises the aggregate volume of credit because it reduces the interest rate differential. It will however reduce the scale of activity of the individual bank, since each bank commands a lower share of the credit market and this negative market share effect dominates the positive effect of the larger size of the market.

Next, we characterize profits. Improvements in demand or cost conditions for all banks should normally raise the typical bank's profit. In Cournot oligopoly, however, this is not necessarily the case because the uncoordinated actions of the N banks might lead to an expansion of total credit so large that the ensuing fall in R and rise in r actually reduce their average profit. More precisely, we cannot prove that the typical bank's profit is always monotonically increasing in θ, the probability p, or the wage W. We can however, provide a sufficient condition for this to be the case.

Assumption 2 $\dfrac{N + 1 + X \frac{\mu_{XX}}{\mu_X}}{N + X \frac{\mu_{X\lambda}}{\mu_\lambda}} > \dfrac{N-1}{N}$.

This inequality imposes restrictions on the curvature of the function $\mu(X;\cdot)$, which describes the interest rates differential, $R - r$. It requires that the ratio of the first-order to second-order effect of the parameter λ be sufficiently smaller than the ratio of the first-order to second-order effect of total credit, X. In the case of the Cobb-Douglas specifications of (1) and (4), the condition is surely satisfied for $\lambda = \theta, p, W$ *regardless of the particular value of* N. This property will be useful in our analysis of the aggregate implications of the model. Notice that if the banking sector is a monopoly, we have $N = 1$ and Assumption 2 is surely satisfied. On the other hand, if the banking sector is competitive, we have $N \to \infty$ and $R = r$ for all W, which yields $\pi = 0$ so that it is meaningless to look for comparative statics results on the banks' profitability.[16]

Proposition 4 *The gross profit of the typical bank is a function* $\pi(W; N, \theta, p)$ *with the following properties:*

- $\pi(0; \cdot) = 0$;

- $\pi_W(W; \cdot) > 0$;

- $\lim\limits_{W \to \infty} \pi(W; \cdot) = \infty$;

- *increasing in* p *and* θ, *decreasing in* N.

Proof. See the Appendix.

The properties of the function $\pi(W; N, \theta, p)$ are intuitive in that an increase in θ, p or W improves the demand or cost conditions for all banks while an increase in N raises competition. The possible non-monotonicity of the function $\pi(W; N, \theta, p)$ with respect to θ, p and W is interesting from the theoretical viewpoint because it highlights the importance of the strategic interaction of banks in the loans and deposits markets. However, we prefer to work with restrictions that ensure monotonicity because this simplifies the analysis of the aggregate implications of the model.

Before proceeding, a remark is in order. In solving stage one of the game, we made use of the result that the bank's payoffs when it plays YS are independent of how many other banks play YS. Observing the solution

---

[16]Of course, $\pi = 0$ implies that the condition for $p > 0$ specified in Proposition 1 cannot be met because $\pi_{YS} = \pi_{NS} = 0$. Hence, the banking sector is in its pure strategy Nash equilibrium where all banks play NS. As a result, $p = 0$ for all W, θ and β.

of stage two, we see that the quantity X_{YS} does not depend on the number of banks that play strategy YS. The reason is that the screening cost β is not related to the quantity of credit extended to the screened entrepreneur. In other words, it is a fixed cost that does not affect the marginal cost of credit. This implies that the optimal quantity that bank i chooses is not affected by how many other banks decide to play YS. It follows that the *gross* profit to strategy YS is the same in all cases where at least one bank plays YS.

4.2   The function p

To fully characterize the credit market, we now turn to the determination of the probability p, defined in (6).

Proposition 5 *The function $p\left(W; N, \theta, \beta\right)$ has the following properties:*

- *there is a value $W_0\left(N, \theta, \beta\right)$ such that $p = 0$ for $0 \leq W \leq W_0$; p is increasing in W for $W > W_0$ and converges to 1 as W becomes infinity;*

- *the threshold value $W_0\left(N, \theta, \beta\right)$ is increasing in N, θ, β with*

  - $\lim_{N \to \infty} W_0\left(N, \theta, \beta\right) = \infty,$
  - $\lim_{\theta \to 1} W_0\left(N, \theta, \beta\right) = \infty,$
  - $\lim_{\beta \to \infty} W_0\left(N, \theta, \beta\right) = \infty;$

- *the function $p\left(W; N, \theta, \beta\right)$ is decreasing in N, θ and β.*

Proof.   See the Appendix.

There is a discontinuity at $\theta = 0$, which yields $p = 0$.[17] Intuitively, it is optimal to do no credit at all, and therefore no screening, when there are no high-quality entrepreneurs. In contrast, when the mass of high-quality entrepreneurs is positive, although very small, it is optimal to screen with probability close to one because this maximizes the chance of finding the few high-quality entrepreneurs. This follows from the fact that investment projects are perfectly divisible and credit can be arbitrarily distributed over many entrepreneurs undertaking small projects, or concentrated on very few entrepreneurs undertaking large projects, the important thing being that the

---

[17]To see this, notice that $\theta = 0$ yields $\pi_{YS} = \pi_{NS} = 0$ and thus $\pi_{YS} - \pi_{NS} = 0 < \beta$.

projects be undertaken by the high-quality entrepreneurs. We also have a discontinuity at $\theta = 1$, which yields $p = 0$.[18] This is again intuitive because it is optimal to do no screening when entrepreneurs succeed with certainty.

An interesting point that emerges from Proposition 5 is that perfect competition, price taking behavior due to an infinite number of banks, is a very special case. For all finite values of the number of banks, the probability of screening is strictly positive. This emphasizes the importance of the Cournot model, which captures in a smooth fashion the degree of competition in the market. Our main result, then, is that higher competition yields a lower equilibrium probability of screening, hence a lower efficiency of the credit market.

5 Aggregate implications

In this section we study the aggregate implications of the model. We first characterize the economy's general equilibrium dynamics and then present the main insights provided by our model and some remarks on its empirical implications.

5.1 General equilibrium dynamics

Capital accumulation is described by (7). Recall that the wage is an increasing function of the lagged capital stock, $W_t = W(K_{t-1})$; see equation (3). The functions $X(W; N, \theta, p)$ and $p(W; N, \theta, \beta)$ constructed in Propositions 2 and 5 can be rewritten

$$X_{t+1} = \tilde{X}(K_t; N, \theta, \beta) \equiv X(W(K_t); N, \theta, p(W(K_t); N, \theta, \beta)),$$

$$p_{t+1} = \tilde{p}(K_t; N, \theta, \beta) \equiv p(W(K_t); N, \theta, \beta)),$$

which are both increasing in K_t. Recall now that Proposition 5 identified a threshold W_0 such that, regardless of N, we have $p_t = 0$ for $0 \leq W_t \leq W_0$. There thus exists a value K_0 such that $W(K_0) = W_0$ and $p_t = 0$ for $0 \leq K_t \leq K_0$. 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, the economy moves to a higher capital accumulation trajectory because banks reach the minimum scale necessary to make screening profitable. Moreover, because p_t is increasing in K_t, the amount of screening increases as capital accumulates, and thereby

---

[18]To see this, notice that $\theta = 1$ yields $\pi_{YS} = \pi_{NS} > 0$ and thus $\pi_{YS} - \pi_{NS} = 0 < \beta$.

generates a positive feedback on capital accumulation. The following proposition makes these points formally.

Proposition 6 *The economy's general equilibrium is described by the first-order difference equation*

$$K_{t+1} = \Phi(K_t; N, \theta, \beta), \tag{13}$$

where

$$\Phi = \begin{cases} \theta \tilde{X}_{NS}(K_t; N, \theta) & \text{for} \quad 0 \leq K \leq K_0 \\ \tilde{X}(K_t; N, \theta, \beta)[\tilde{p}(K_t; N, \theta, \beta)(1 - \theta) + \theta] & \text{for} \quad K > K_0 \end{cases}$$

and

$$\tilde{X}_{NS}(K_t; N, \theta) \equiv X_{NS}(W(K_t); N, \theta).$$

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

- *$\Phi(0) = 0$;*

- *$\Phi_K(\cdot) > 0$;*

- *$\lim_{K \to 0} \Phi_K(K) = \infty$;*

- *$\lim_{K \to \infty} \Phi_K(K) = 0$.*

Proof. See the Appendix.

The trajectory marked in bold in the upper panel of Figure 4 illustrates the dynamics of the economy. Because of the threshold K_0, multiple steady states may emerge and banking market structure determines whether they do, as shown in the bottom panel of the same figure. While we acknowledge that this is potentially an interesting aspect of the model, we think that the issue is not central to the main contribution of the paper and relegate its analysis to a technical appendix which is available upon request.

5.2 Banking market structure and steady-state capital

We begin by identifying some benchmark features of the dynamical system that highlight the intuition behind our results.

Proposition 7 *The function* $\Phi(K_t; N, \theta, \beta)$ *is everywhere below the one corresponding to an economy where infinitely many competitive banks perform screening on all entrepreneurs and the one corresponding to an economy where one monopoly bank does not perform screening. Therefore, the economy converges to a steady state* $K_{ss} \in (\underline{K}, \overline{K})$, *where* $\underline{K}$ *is the steady state achieved by the economy with one monopoly bank that does not screen and* $\overline{K}$ *is the steady state achieved by the economy with competitive banks that screen.*

Proof. A monopoly bank that never screens supplies the lowest equilibrium quantity of credit to all entrepreneurs, thus inflicting the largest deadweight loss and wasting the largest proportion of credit. Competitive banks always screening supply the highest equilibrium quantity of credit only to high quality entrepreneurs, thus inflicting no deadweight loss and wasting no credit. These extreme steady-state values, however, are not attainable as equilibrium outcomes. It follows that an economy with a monopoly bank, or with infinitely many competitive banks, or any market structure in between, converges to a steady-state value of K in the *open* interval $(\underline{K}, \overline{K})$.□

We illustrate these properties in the upper panel of Figure 4, where the thin trajectories are the upper and lower boundaries within which $\Phi(K_t; N, \theta, \beta)$ must be. Note that even for a monopolist there is a value of K that makes the monopoly profit too small to cover the screening cost.[19]

The steady state is the solution of

$$K_{ss} = \Phi(K_{ss}; N, \theta, \beta).$$

We have two types of solutions: $K_{ss} \leq K_0$ and $K_{ss} > K_0$. In the first case we have

$$\Phi_N = \theta \frac{\partial X_{NS}}{\partial N} > 0$$

so that K_{ss} is increasing in N. In the second case, in contrast, we have

$$\Phi_N = \frac{\partial \tilde{X}}{\partial N} \left[\tilde{p} (1 - \theta) + \theta \right] + \tilde{X} \frac{\partial \tilde{p}}{\partial N} (1 - \theta).$$

The sign of this derivative is ambiguous because there are two offsetting effects. First, more competition leads to a higher volume of credit. However,

---

[19]More precisely, for $N = 1$ we have $p = 1$ for all $K \geq K_0(1, \theta, \beta)$, for $N > 1$ we have $0 < p < 1$ for all $K > K_0(N, \theta, \beta)$, while for $N \to \infty$ we have $p = 0$ for all K (this is equivalent to saying that $K_0(\infty, \theta, \beta) \to \infty$).

and second, it leads to a fall in screening and less efficient lending. Notice, moreover, that

$$\frac{\partial \tilde{X}}{\partial N} = \frac{\partial X}{\partial N} + \frac{\partial X}{\partial p}\frac{\partial p}{\partial N}.$$

Hence, the negative effect on the probability of screening also reduces the total volume of credit. So, the number of banks has a direct effect on credit, a direct effect on screening, and an indirect effect on credit through its effect on screening. The first effect is positive, the other two are negative. As a result, the overall effect of N on K_{ss} is ambiguous.

Figure 5 provides an intuitive characterization of the general relationship between the number of banks and steady-state capital in the case $K_{ss} > K_0$. As $\theta \to 1$, the fraction of good entrepreneurs is so high that screening becomes irrelevant and we have the traditional result 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 $N^* \to \infty$. To see this formally, notice that

$$\theta \to 1 \Rightarrow \Phi_N \cong \frac{\partial \tilde{X}}{\partial N} > 0 \text{ for all } N.$$

As $\theta \to 0$, in contrast, 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. This means that the relation between N and K_{ss} is monotonically decreasing and the number of banks that maximizes steady-state capital is $N^* = 1$. Formally,

$$\theta \to 0 \Rightarrow \Phi_N \cong \frac{\partial \tilde{X}}{\partial N}\tilde{p} + \tilde{X}\frac{\partial \tilde{p}}{\partial N} = \frac{\partial \tilde{p}}{\partial N}\left(\frac{\partial \tilde{X}}{\partial \tilde{p}}\tilde{p} + \tilde{X}\right) < 0 \text{ for all } N.$$

For intermediate values of θ, a hump-shaped relation emerges, and the number of banks that maximizes steady-state capital is $N^* \in (1, \infty)$. In other words, oligopoly banking strikes the best possible balance between the dead-weight losses from banks' market power and the benefits of efficient lending.

5.3 Banking market structure and growth

As we mentioned in the introduction, the available empirical evidence points at an ambiguous relationship between banking concentration and economic growth. The result highlighted in our theoretical model, i.e., the existence of a non-monotonic relationship between banking market structure and capital

accumulation, yields predictions that are consistent with these findings. To see this, notice that the model yields a growth equation of the form

$$g_t \equiv \frac{Y_{t+1} - Y_t}{Y_t} = \gamma \frac{K_{t+1} - K_t}{K_t} = \gamma \left[\frac{\Phi\left(Y_t^{\frac{1}{\gamma}}; N, \theta, \beta\right)}{Y_t^{\frac{1}{\gamma}}} - 1 \right].$$

As it is well known (see, e.g., Barro and Sala-i-Martin 1995), this equation states that, after controlling for initial income, Y_t, the growth rate is increasing in those factors that raise the function $\Phi(\cdot)$. Hence, for an economy that approaches a steady state with no screening, $K_{ss} \leq K_0$, growth is increasing in θ and N. For an economy that approaches a steady state with screening, $K_{ss} > K_0$, growth is increasing in θ, decreasing in β, while the effect of N is ambiguous. More specifically, the sign of the coefficient of the number of banks in a growth regression depends on the set of factors that determine the relation between the number of banks and steady-state capital that we discussed above.

Cetorelli and Gambera [8] find evidence of two effects going in opposite direction. First, there is an economy-wide negative effect of banking concentration for economic growth. Second, sectors especially dependent on external finance benefit from being in countries with a more concentrated banking sector. Empirically, sectors with higher needs for external finance are also those heavily engaged in Research and Development,[20] hence characterized by a higher degree of uncertainty regarding the final outcome. Translated in the language of our model, these are sectors with a relatively low θ. Cetorelli and Gambera find that such sectors grow faster in countries with a highly concentrated banking sector. Consistent with this, our model predicts that when $\theta \to 0$ the relationship between g and N is monotonically decreasing. The value added from screening, larger with higher market concentration, more than counterbalances the negative effect due to a reduction in quantities. In contrast, when $\theta \to 1$ the relationship between g and N is monotonically increasing. For intermediate values of θ, the relationship is hump-shaped and the effect of N depends on whether the economy has a number of banks that is smaller or larger than the value N^* that maximizes steady-state income per capita.

The model has also an additional prediction that is worth highlighting.

---

[20]In a sample of 35 industrial sectors in OECD countries, the correlation between a measure of sectoral R&D and that of external financial dependence is 0.89.

Define the losses-to-loans ratio as

$$(1 - \theta) \left[1 - \tilde{p} \left(K_t; N, \theta, \beta \right) \right].$$

This ratio is decreasing in $\tilde{p}$. We thus have that the ratio decreases along the transition to the steady state and increases with competition. This prediction is consistent with Shaffer [25], who provides evidence of a negative relationship between the number of banks operating in a market and the losses-to-loans ratio.[21]

6 Conclusion

We have presented a dynamic general equilibrium model of capital accumulation in which credit is intermediated by an oligopolistic banking industry. While the theoretical literature on finance and growth has recognized the importance of banks in fostering economic growth, it has not explored the role played by the market structure of the banking industry. Conventional wisdom would suggest that competition – price taking behavior due to a large number of banks – should be the optimal market structure. However, the available empirical evidence conveys a more ambiguous picture, suggesting the existence of multiple channels through which banking market structure affects growth.

Our model identifies two of such channels and explores their effect on the functioning of the credit market and on the process of capital accumulation. Banking market structure affects both the overall quantity of credit available for investment purposes and the allocative efficiency of the credit market. Specifically, the credit market is endogenously split in two segments, one where banks screen entrepreneurs and lend only to high quality ones, and one where banks lend indiscriminately to all entrepreneurs. The market power of banks determines the relative size of these two segments.

The model yields predictions that are consistent with recent empirical evidence. We have shown that under plausible conditions the market structure of the banking sector that maximizes steady-state income per capita is neither competition nor monopoly, but an intermediate oligopoly. In addition, the transition to the developed state exhibits an endogenous regime switch whereby the qualitative properties of the credit sector evolve with growth. Namely, as the economy grows and the banking sector expands, the share of screened lending rises, and the banking sector becomes more

---

[21]This property has a nice interpretation at business-cycle frequency: in a recession the smaller credit market implies a lower p, which means that the losses-to-loans ratio rises.

sophisticated. The ensuing improvement in lending efficiency accelerates the pace of capital accumulation.

A particularly valuable feature of the model is that it is extremely parsimonious in the sense that we do not make special assumptions and have very few free parameters. Our main ingredients are the information externality associated with the screening activity of banks and their oligopolistic rivalry. This is all it is necessary to obtain a rich set of results and to successfully isolate a fundamental force that is at work in real-world economies, banking market structure, so far neglected in this literature.

7 Appendix

7.1 Proof of Proposition 1

Consider the case $\pi_{NS} > \pi_{YS} - \beta$. Observing the matrix discussed above, we see that YS is a dominated strategy. It follows that NS for all banks is the unique pure-strategy Nash equilibrium. Consider now the case $\pi_{NS} \leq \pi_{YS} - \beta$. Each bank prefers to play NS when at least one other bank plays YS, and to play YS when no one else plays NS. There are multiple possible combinations of banks' actions that yield such an outcome. All of these are pure-strategy Nash equilibria. The mixed-strategy Nash equilibrium obtains when all banks are indifferent between strategy YS and strategy NS. The payoff to strategy YS is always $\pi_{YS} - \beta$. The payoff to strategy NS is π_{YS} if at least one bank plays YS. This event occurs with probability $1 - (1-p)^{N-1}$. If no other bank plays YS, then the bank makes π_{NS}. This event occurs with probability $(1-p)^{N-1}$. The mixed-strategy Nash equilibrium requires

$$\pi_{YS} - \beta = [1 - (1-p)^{N-1}]\pi_{YS} + (1-p)^{N-1}\pi_{NS}.$$

The solution to this equation yields (6). $\square$

7.2 Proof of Proposition 2

Bank i's profit is

$$\pi_i = x_i \mu(X; \cdot) \tag{14}$$

and the first- and second-order conditions are:

$$\mu(X; \cdot) + x_i \mu_X(X; \cdot) = 0;$$

$$2\mu_X(X; \cdot) + x_i \mu_{XX}(X; \cdot) < 0.$$

The function $\mu(X; \cdot)$ has the properties:[22]

- $\lim_{X \to 0} \mu(X; \cdot) = +\infty$ and $\lim_{X \to W} \mu(X; \cdot) = -\infty$;

- $X \mu_X(X; \cdot) < 0$;

- $\lim_{X \to 0} X \mu_X(X; \cdot) = 0$ and $\lim_{X \to W} X \mu_X(X; \cdot) = -\infty$.

---

[22]These properties follow from the properties of the functions $f(K)$ and $h(X; W)$ established in Section 2. Recall, in particular, that $f(K)$ exhibits diminishing returns to scale, which implies that $\lim_{K \to 0} K f_K(K) = 0$.

Symmetry allows us to write

$$\mu(X;\cdot) = -\frac{X\mu_X(X;\cdot)}{N}. \tag{15}$$

Figure 1 illustrates the determination of X as the intersection of two curves representing, respectively, the left-hand and right-hand side of (15).

We now establish the properties of the function $X(W;\cdot)$ with respect to W. $X(0;\cdot) = 0$ follows from the fact that X is a positive variable that must always be less than W. Hence, as W approaches 0, X goes to 0 as well. To prove $X_W(W;\cdot) > 0$, define $y_i \equiv \sum_{j\neq i} x_j$. Totally differentiating the first-order condition, we have

$$a_i dx_i + b_i dy_i + c_i d\lambda = 0,$$

where $a_i \equiv 2\mu_X + x_i\mu_{XX} < 0$, $b_i \equiv \mu_X + x_i\mu_{XX} < 0$, $c_i \equiv -\mu_\lambda - x_i\mu_{X\lambda}$. Notice that $a_i < 0$ follows from the second-order condition, while $b_i < 0$ is the generalized Hahn stability condition that we assume to hold because it requires that the marginal revenue of bank i falls as the credit activity of any of the other banks rises (see Dixit [10], p. 118-119). Observe now that $dy_i = dX - dx_i$. Hence,

$$dx_i = -\frac{b_i}{a_i - b_i}dX - \frac{c_i}{a_i - b_i}d\lambda. \tag{16}$$

A similar relation holds for bank j. Summing across all $j \neq i$, we obtain

$$dy_i = -\sum_{j\neq i}\frac{b_j}{a_j - b_j}dX - \sum_{j\neq i}\frac{c_j}{a_j - b_j}d\lambda,$$

which yields

$$dy_i = -\frac{\sum_{j\neq i}\frac{b_j}{a_j - b_j}}{1 + \sum_{j\neq i}\frac{b_j}{a_j - b_j}}dx_i - \frac{\sum_{j\neq i}\frac{c_j}{a_j - b_j}}{1 + \sum_{j\neq i}\frac{b_j}{a_j - b_j}}d\lambda.$$

The slope of the (collective) reaction function of the other $N - 1$ banks in response to an expansion of activity of bank i is

$$\frac{\partial y_i}{\partial x_i} = -\frac{\sum_{j\neq i}\frac{b_j}{a_j - b_j}}{1 + \sum_{j\neq i}\frac{b_j}{a_j - b_j}}, \tag{17}$$

26

which is negative and less than one in magnitude since $b_j < 0$ and $a_j - b_j = \mu_X < 0$ for all $j \neq i$. Consider now (16) and sum across all banks. This yields

$$dX = -\sum_j \frac{b_j}{a_j - b_j} dX - \sum_j \frac{c_j}{a_j - b_j} d\lambda,$$

which can be solved for

$$\frac{dX}{d\lambda} = -\frac{\sum_j \frac{c_j}{a_j - b_j}}{1 + \sum_j \frac{b_j}{a_j - b_j}}.$$

Exploiting the symmetry of the model, we can now write

$$\frac{dX}{d\lambda} = -\frac{Nc}{a + b(N - 1)}. \tag{18}$$

Stability of the Nash equilibrium requires $a + b(N - 1) < 0$ (see Dixit [10], p. 117-119). Hence, the sign of the derivative depends on the sign of c. It is immediate to see that Assumption 1 yields

$$\lambda = W \implies c = -\frac{\partial r}{\partial W} - \frac{X}{N} \frac{\partial^2 r}{\partial X \partial W} > 0 \implies \frac{dX}{dW} > 0.$$

Since X is monotonically increasing in W, we have $X \to \infty$ as $W \to \infty$.

We now establish the properties of the function $X(W; N, \theta, p)$ with respect to θ, p, and N. Proceeding as for W, we have

$$\lambda = \theta \implies c = \frac{\partial R}{\partial \theta} + \frac{X}{N} \frac{\partial^2 R}{\partial X \partial \theta} > 0 \implies \frac{dX}{d\theta} > 0,$$

$$\lambda = p \implies c = \frac{\partial R}{\partial p} + \frac{X}{N} \frac{\partial^2 R}{\partial X \partial p} > 0 \implies \frac{dX}{dp} > 0.$$

To prove that X is increasing in N, refer to Figure 1 and observe that the right-hand side of (15) is increasing in N. Hence, an increase in competition raises the aggregate volume of credit. $\square$

7.3 Proof of Proposition 3

Substituting the expression for $\frac{dX}{d\lambda}$ into (16), we obtain

$$\frac{dx_i}{d\lambda} = -\frac{b_i}{a_i - b_i} \frac{\sum_j \frac{c_j}{a_j - b_j}}{1 + \sum_j \frac{b_j}{a_j - b_j}} - \frac{c_i}{a_i - b_i},$$

27

which yields

$$\frac{dx}{d\lambda} = -\frac{c}{a + N(b-1)}. \tag{19}$$

It is immediate to see that $\frac{dx}{d\lambda} > 0$ for $\lambda = W, \theta, p$. To prove that x is decreasing in N, it is sufficient to rewrite (15) in terms of x,

$$\mu(Nx; \cdot) = -x\mu_X(Nx; \cdot),$$

and observe that the left-hand side is decreasing in N while the right-hand side is increasing in N. The lower panel of Figure 1 illustrates.□

7.4 Proof of Proposition 4

$\pi(0; \cdot) = 0$ follows from $X(0; \cdot) = 0$. We now show that Assumption 1 is necessary for $\pi_W(W; \cdot) > 0$, while Assumption 2 is sufficient. We again follow Dixit [10]. Totally differentiating profit, we obtain

$$d\pi_i = \frac{\partial \pi_i}{\partial x_i} dx_i + \frac{\partial \pi_i}{\partial y_i} dy_i + \frac{\partial \pi_i}{\partial \lambda} d\lambda,$$

where

$$\frac{\partial \pi_i}{\partial \lambda} = x_i \mu_\lambda.$$

At the optimum, $\frac{\partial \pi_i}{\partial x_i} = 0$. Hence,

$$d\pi_i = \frac{\partial \pi_i}{\partial y_i} dy_i + \frac{\partial \pi_i}{\partial \lambda} d\lambda,$$

which says that the change in the bank's profit in response to a generalized improvement in market conditions – either because of an increase in credit demand or because of a reduction in the cost of collecting saving – is the sum of a positive direct effect and a negative indirect effect due to the actions of the other banks. In our oligopoly model this indirect effect has two components: the response of the other $N-1$ banks to the change in market conditions given the level of activity of bank i, and their response to bank i's change in activity given market conditions. Formally,

$$dy_i = \left[\frac{\partial y_i}{\partial \lambda} + \frac{\partial y_i}{\partial x_i} \frac{\partial x_i}{\partial \lambda} \right] d\lambda.$$

Since $-1 < \frac{\partial y_i}{\partial x_i} < 0$, the term in brackets is positive as long as $\frac{\partial y_i}{\partial \lambda} \geq \frac{\partial x_i}{\partial \lambda}$, which is surely true in our symmetric model. Hence, the indirect effect is

negative because the other $N-1$ banks raise their level of activity and this yields an increase in total credit, X, which lowers the interest rate on loans, R, and raises the interest rate on deposits, r. These two effects are captured by the term $\frac{\partial \pi_i}{\partial y_j} < 0$. Exploiting the symmetry of the model, we can write

$$
\begin{aligned}
\frac{d\pi}{d\lambda} &= \frac{\partial \pi}{\partial y}\left[\frac{\partial y}{\partial \lambda} + \frac{\partial y}{\partial x}\frac{\partial x}{\partial \lambda}\right] + \frac{\partial \pi}{\partial \lambda} \\
&= \frac{\partial \pi}{\partial y}\frac{\partial X}{\partial \lambda}\left[1 + \left(\frac{\partial y}{\partial x} - 1\right)\frac{1}{N}\right] + \frac{\partial \pi}{\partial \lambda}.
\end{aligned}
$$

Substituting (17) and (18) and rearranging, we find

$$
\frac{d\pi}{d\lambda} = \frac{X}{N}\left[-\frac{N-1}{N+1+X\frac{\mu_{XX}}{\mu_X}}\frac{N-1+\frac{X}{N}\frac{\mu_{XX}}{\mu_X}(N-2)}{N+\frac{X}{N}\frac{\mu_{XX}}{\mu_X}(N-1)} + \frac{N}{N+X\frac{\mu_{X\lambda}}{\mu_\lambda}}\right].
$$

All the components of the first term inside the bracket are positive because the stability condition $a + b(N-1) < 0$ implies

$$
N - 1 + \frac{X}{N}\frac{\mu_{XX}}{\mu_X}(N-2) < N + \frac{X}{N}\frac{\mu_{XX}}{\mu_X}(N-1) < N + 1 + X\frac{\mu_{XX}}{\mu_X} < 0.
$$

Therefore, Assumption 1 is necessary for $\frac{d\pi}{d\lambda} > 0$ because if it does not hold, the last term inside the bracket is negative and the whole expression is surely negative. Observe now that

$$
\frac{N - 1 + \frac{X}{N}\frac{\mu_{XX}}{\mu_X}(N-2)}{N + \frac{X}{N}\frac{\mu_{XX}}{\mu_X}(N-1)} < 1.
$$

Hence, a sufficient condition for $\frac{d\pi}{d\lambda} > 0$ is

$$
\frac{N + 1 + X\frac{\mu_{XX}}{\mu_X}}{N + X\frac{\mu_{X\lambda}}{\mu_\lambda}} > \frac{N-1}{N}. \tag{20}
$$

With the special functional forms (1) and (4), we can reduce the inequality to

$$
\frac{1}{N(1-\gamma)} > \frac{\left(\frac{1-\alpha}{\alpha}\frac{W}{W-X} + \frac{X}{W-X}\right)\frac{N-1}{N} - \gamma}{\frac{1-\alpha}{\alpha}\frac{W}{W-X}\left[\left(\frac{1-\alpha}{\alpha}\frac{W}{W-X} + \frac{X}{W-X}\right)\gamma + N - 1 + \gamma^2\right]}.
$$

The right-hand side is monotonically increasing in W and converges to zero for $W = 0$ and to a finite upper bound for $W \to \infty$. The inequality is thus satisfied if

$$
\frac{1}{N(1-\gamma)} > \frac{\frac{1-\alpha}{\alpha}\frac{N-1}{N} - \gamma}{\frac{1-\alpha}{\alpha}\left[\frac{1-\alpha}{\alpha}\gamma + N - 1 + \gamma^2\right]}.
$$

This inequality, in turn, is surely satisfied since it reduces to

$$\left(\frac{1-\alpha}{\alpha}\right)^2 \gamma + \left(\frac{1-\alpha}{\alpha}\right)\gamma^2 + \gamma\frac{1-\alpha}{\alpha}\frac{N-1}{N} + N(1-\gamma) > 0.$$

Hence, $\pi(W; N, \theta, p)$ is monotonically increasing in θ, p and W. We illustrate the function $\pi(W; p, \theta, N)$ in Figure 2.$\square$

7.5 Proof of Proposition 5

To better understand the function p, it is useful to review the steps that we followed to construct it. First, we argued that uncertainty about the quality of entrepreneurs, and the informational externality associated with banks' costly production of information, yield that banks screen entrepreneurs with some probability. Formally, we posited a two-stage game whereby in stage one banks determine the probability of screening each entrepreneur and in stage two determine the optimal quantity of credit that they issue. As we emphasized, this structure means that banks split their overall activity between lending to screened entrepreneurs and lending to unscreened entrepreneurs. In other words, the credit market is endogenously segmented in an efficient and an inefficient part. The probability p is given by the mixed strategy equilibrium of a game where the strategy space is the pair $\{YS, NS\}$. An important property of this structure is that the outcome in quantity and profits of the stage-two Cournot game is a linear combination of the two pure strategies $\{YS, NS\}$. We can thus work by backward induction. Given the profit function obtained from the stage two of the game, we set $p = 0$ and $p = 1$ to calculate the profits corresponding to the two pure strategies. We then use these results to characterize the probability p as a function of the wage and structural parameters. The mixed strategy equilibrium is a linear combination of the two extremes where banks always screen or do not screen at all. We use (15) and (14) to solve for X_{YS}, X_{NS}, π_{YS} and π_{NS} by setting, respectively, $p = 1$ and $p = 0$ in (15) and (14). We obtain:

$$\pi_{YS}(W; N) = \frac{X_{YS}(W; N)}{N}\left[f_K(X_{YS}(W; N)) - h(X_{YS}(W; N), W)\right];$$

$$\pi_{NS}(W; N, \theta) = \frac{X_{NS}(W; N, \theta)}{N}\left[\theta f_K(\theta X_{NS}(W; N, \theta)) - h(X_{NS}(W; N, \theta), W)\right].$$

Figure 2 illustrates these expressions, together with the function $\pi(W; N, p, \theta)$ constructed above. Given p, the bank's profit is a weighted average of the

profit that it would make by always playing NS or always playing YS. Let

$$d(W; N, \theta) \equiv \pi_{YS}(W; N) - \pi_{NS}(W; N, \theta)$$

be the profit differential between the two pure strategies. Equation (6) reads

$$p = 1 - \left[\frac{\beta}{\pi_{YS} - \pi_{NS}}\right]^{\frac{1}{N-1}}.$$

As shown in Proposition 1, the mixed strategy equilibrium exists if $\pi_{YS} - \pi_{NS} \geq \beta$. We exploit the properties of the function $d(W; \cdot)$ to characterize the function $p(W; \cdot)$.[23] From inspection of Figure 2, we see that

- $d(0; \cdot) = 0$,

- $d_W(W; \cdot) > 0$,

- $\lim_{W \to \infty} d(W; \cdot) = \infty$,

- $d(W; \cdot)$ decreasing in N and θ.

These properties follow from the properties of the function $\pi(W; \cdot)$ and the construction of π_{YS} and π_{NS}. Figure 3 illustrates the properties listed in the proposition. An intersection exists as long as N is finite and $\theta < 1$. Hence, $p = 0$ for all W *only* when $N \to \infty$ or $\theta = 1$. Since $d(W; \cdot)$ does not depend on β, it is immediate to see that p is decreasing in β.□

7.6 Proof of Proposition 6

To prove continuity, we only need to show that the function is continuous at $K = K_0$, where the regime change occurs. Observe that by construction

$$X_{NS}(W; N, \theta) = X(W; N, p = 0, \theta).$$

---

[23]The reader can now appreciate the importance of the restrictions that ensure monotonicity of the function $\pi(W; N, \theta, p)$ in W. If this function is non-monotonic in W, then the function $d(W; N, \theta)$ is non-monotonic in W, and this implies that there might be more than one solution to the equation $d(W; N, \theta) = \beta$. This in turn means that the set of values of W such that $p > 0$ becomes more complicated than what stated in Proposition 5. Specifically, the properties of the function $d(W; \cdot)$ imply that there is an odd number of solutions. Hence, if there are $s + 1$ solutions $W_1, .., W_s, W_0$, where s is an even integer, then there are $\frac{s}{2}$ finite intervals $(W_1, W_2), ..., (W_{s-1}, W_s)$ followed by the infinite interval (W_0, ∞). Analysis of the model's aggregate dynamics under these conditions is feasible but not particularly insightful.

Since $p(W_0; \cdot) = 0$, at $K = K_0$ we have

$$\theta X_{NS}(W(K_0); \cdot) = X(W(K_0); p(W(K_0); \cdot), \cdot)[p(W(K_0); \cdot)(1 - \theta) + \theta],$$

which yields $\theta X_{NS}(W_0; N, \theta) = \theta X(W_0; N, p = 0, \theta)$. As a consequence,

$$\lim_{K \uparrow K_0} \Phi(K) = \lim_{K \downarrow K_0} \Phi(K)$$

and $\Phi(K)$ is continuous at $K = K_0$. The property $\Phi(0) = 0$ follows from $W(0) = 0$, which implies $X_{NS}(0; \cdot) = 0$. Notice that X_{NS} has the same qualitative properties with respect to W as the function $X(W; \cdot)$ constructed in Proposition 2. $\Phi_K(\cdot) > 0$ then follows from $W_K(K) > 0$, which implies $X_K(K; \cdot) > 0$ since $X_W(W; \cdot) > 0$ (see Proposition 2) and $p_K(K; \cdot) > 0$ since $p_W(W) > 0$ (see proposition 5). $\lim_{K \to 0} \Phi_K(K) = \infty$ and $\lim_{K \to \infty} \Phi_K(K) = 0$ follow from the properties of the production and utility functions (1) and (4) To see this, notice that

$$\lim_{K \to 0} \Phi_K = \lim_{K \to 0} \frac{\partial X_{NS}}{\partial W} \frac{\partial W}{\partial K} = \infty$$

because $\lim_{W \to 0} \frac{\partial X_{NS}}{\partial W} = \infty$ and $\lim_{K \to 0} \frac{\partial W}{\partial K} = \infty$. Similarly,

$$\lim_{K \to \infty} \Phi_K = \lim_{K \to \infty} \left(\frac{\partial X}{\partial W} + \frac{\partial X}{\partial p} \frac{\partial p}{\partial W} \right) \frac{\partial W}{\partial K} = 0$$

because $\lim_{W \to \infty} \frac{\partial X}{\partial W} = 0$, $\lim_{W \to \infty} \frac{\partial X}{\partial p} \frac{\partial p}{\partial W} \in (0, \infty)$, and $\lim_{W \to \infty} \frac{\partial W}{\partial K} = 0$. □

7.7   The case of observable contracts

As discussed in the text, with observable contracts no bank can discriminate between screened and unscreened loans without revealing the result of the screening test. Banks then must set $x_t^S = x_t^U$ and reallocate the credit denied to screened entrepreneurs who failed the test to all the other entrepreneurs. We showed that in this case the relationship between credit and capital is given by (8), which entails a lower efficiency of the credit market. To see this formally, we begin by noticing that

$$p_t + \theta(1 - p_t) > \frac{\theta}{p_t \theta + 1 - p_t} \qquad \forall p_t \in (0, 1). \tag{21}$$

The equilibrium p_t that applies to the two sides of the inequality is different, so we need to be careful in using this result to rank regimes. (In both cases,

however, if banks screen all credit so that $p_t = 1$, then $K_t = X_t$; if they do not screen and $p_t = 0$, then $K_t = \theta X_t$.)

We have shown in Proposition 5 that p_t depends only on the wage, W_t, and the exogenous parameters. This allows us to prove that the regime with unobservable contracts yields faster capital accumulation.

Consider two economies with the same initial level of the capital stock, K_t. One economy has unobservable contracts, the other has observable contracts. Since the capital stock is the same, the two economies have the same wage, W_t, and the same probability of screening, p_t. Hence, the only initial difference in capital accumulation stems from the volume of credit and the fraction of credit that becomes capital. Observe now that given the same p_t, inequality (21) implies that the credit market is more efficient in the economy with observable contracts. Moreover, the demand for credit is higher and, therefore, the equilibrium volume of credit, X_t, is higher. It follows that K_{t+1} is higher because (a) the volume of credit is higher and (b) the fraction of credit that turns into capital is higher. This means that W_{t+1} is higher, which in turn implies that p_{t+1} is higher. Hence, the capital accumulation paths of the two economies diverge: starting from the same initial level of the capital stock, the economy with unobservable contracts follows a higher accumulation trajectory because in all periods the credit market is larger and more efficient.

References

[1] Bencivenga, V.R. and B. Smith, (1991), Financial Intermediation and Endogenous Growth, *Review of Economic Studies*, **LVIII**, 195-209.

[2] Bhattacharya, S. and A.V. Thakor (1993), "Contemporary Banking Theory", *Journal of Financial Intermediation*, 3, 2-50.

[3] Black, S. and P. Strahan (2000), "Entrepreneurship and the Structure of the Banking Industry", Mimeo.

[4] Bonaccorsi Di Patti, E. and G. Dell'Ariccia (2000), "Bank Competition and Firm Creation", Mimeo.

[5] Campbell, T.S. and W.A. Cracaw (1980), "Information Production, Market Signalling, and the Theory of Financial Intermediation", *Journal of Finance*, XXXV, 4, 863-882.

[6] Cao, M. and S.Shi (2000) "Screening, Bidding, and the Loan Market", Mimeo, Queen's University.

[7] Cetorelli, N. (1997), "The Role of Credit Market Competition on Lending Strategies and on Capital Accumulation", Federal Reserve Bank of Chicago, WP-97-14.

[8] Cetorelli, N. and M. Gambera (2001), "Banking Market Structure, Financial Dependence and Growth: International Evidence from Industry Data", *Journal of Finance*, Forthcoming.

[9] Dell'Ariccia, G. (2000), "Learning by Lending, Competition, and Screening Incentives in the Banking Industry", Mimeo, IMF.

[10] Dixit, A. (1986), "Comparative Statics for Oligopoly", *International Economic Review*, 27, 1, 107-122.

[11] Fischer, K.H. (2000), "Acquisition of Information in Loan Markets and Bank Market Power. An Empirical Investigation", Mimeo, Goethe University Frankfurt.

[12] Greenwood, J. and B. Jovanovic, (1990), Financial Development, Growth and the Distribution of Income, *Journal of Political Economics*, **98**, 1076-1107.

[13] Guzman, M. (2000), "Bank Structure, Capital Accumulation, and Growth: A Simple Macroeconomic Model", *Economic Theory*, Forthcoming.

[14] Harris, M. and B. Holmstrom (1982), "A Theory of Wage Dynamics", *Review of Economic Studies*, XLIX, 315-333.

[15] King, R. and R. Levine, (1993), Finance and Growth: Schumpeter Might Be Right, *Quarterly Journal of Economics*, **108**, 717-737.

[16] Leland, H.E. and D.H. Pyle (1977), "Informational Asymmetries, Financial Structure, and Financial Intermediation ", *Journal of Finance*, XXXII, 2, 371-387.

[17] Levine, R. (1997) "Financial Development and Economic Growth: Views and Agenda", *Journal of Economic Literature*, 35, 688-726.

[18] Levine, R. and S. Zervos (1998), "Stock Markets and Economic Growth", *American Economic Review* 88(3), 537–558.

[19] Manove, M., J. Padilla and M. Pagano (2000), "Collateral vs. Project Screening: A Model of Lazy Banks", CEPR Discussion Paper 2439.

[20] Petersen, M.A. and Rajan, R.G. (1995), "The Effect of Credit Market Competition on Lending Relationship", *Quarterly Journal of Economics*, CX, 407–443.

[21] Rajan, R.G., and L.Zingales (1998), "Financial Dependence and Growth", *American Economic Review* 88(3), 559–586.

[22] Rajan, R.G., and L.Zingales (1999), The Politics of Financial Development, Mimeo, University of Chicago.

[23] Ramakrishnan, R.T.S. and A.V. Thakor (1984), "Information Reliability and a Theory of Financial Intermediation", *Review of Economic Studies*, LI, 415-432.

[24] Riordan. M. (1993), "Competition and Bank Performance", in Mayer, C. and X. Vives (eds.) *Capital Markets and Financial Intermediation*, Cambridge University Press.

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

[26] Sharpe, S.A. (1990), "Asymmetric Information, Bank Lending, and Implicit Contracts: A Stylized Model of Customer Relationships", *The Journal of Finance*, XLV, 4, 1069-1087.
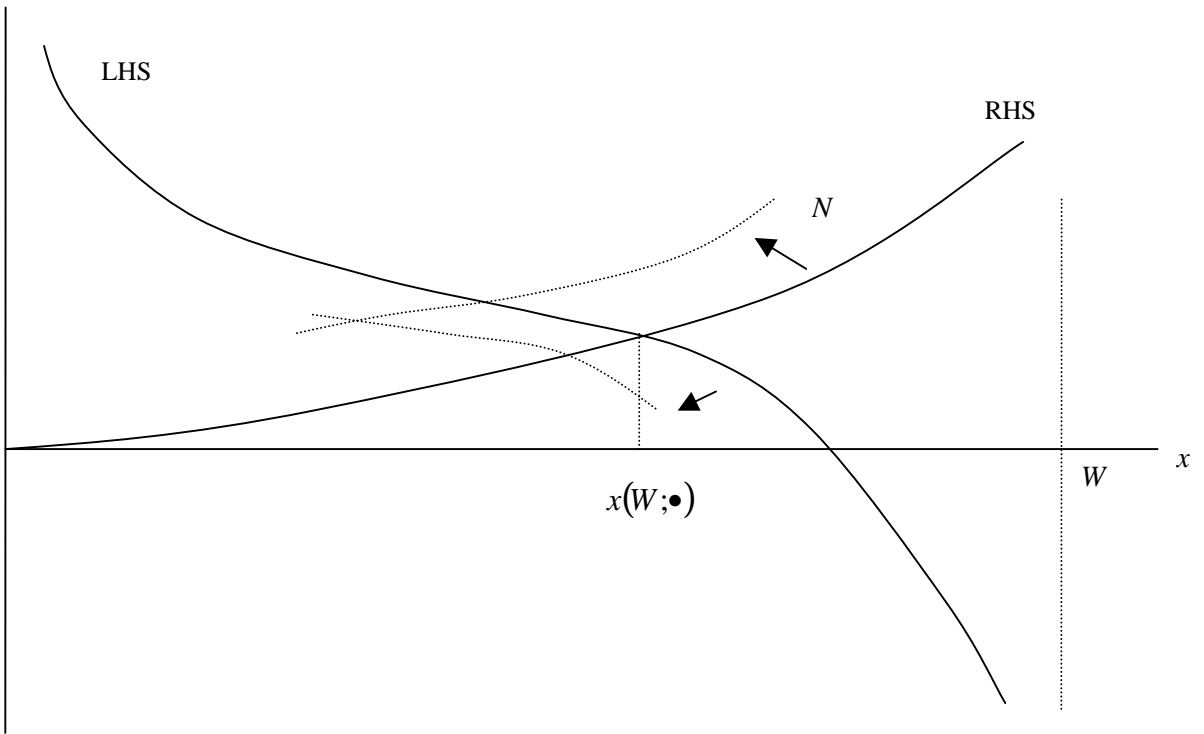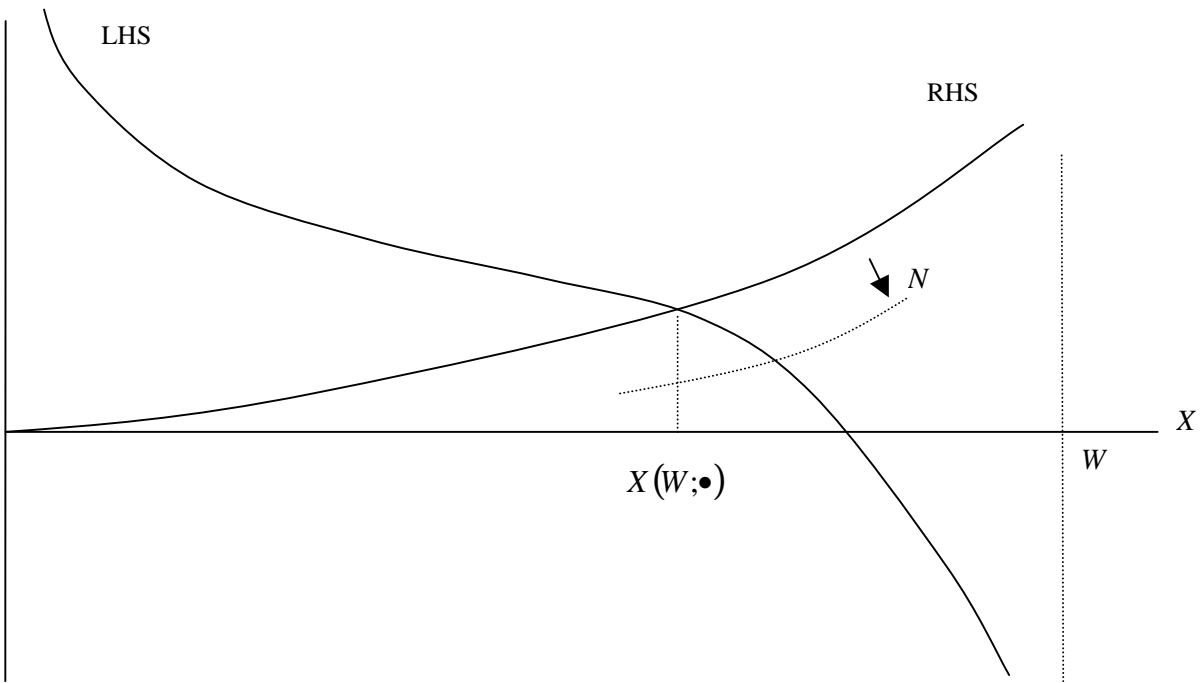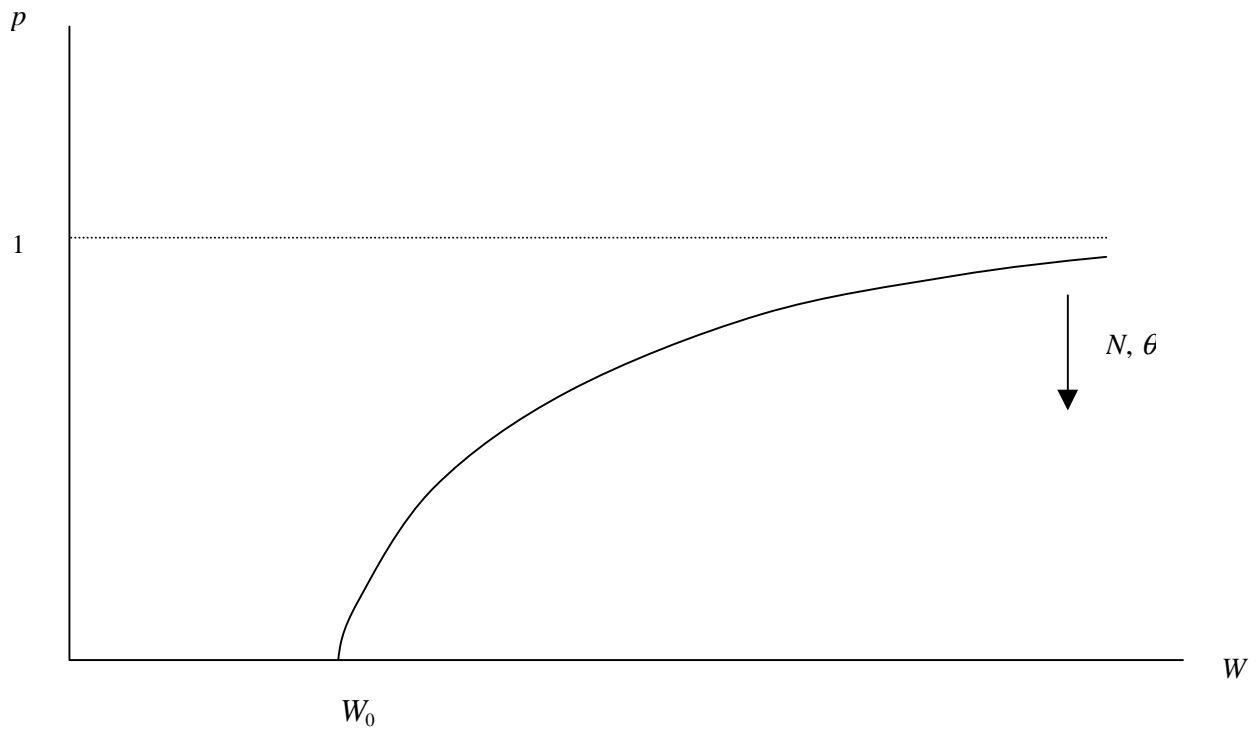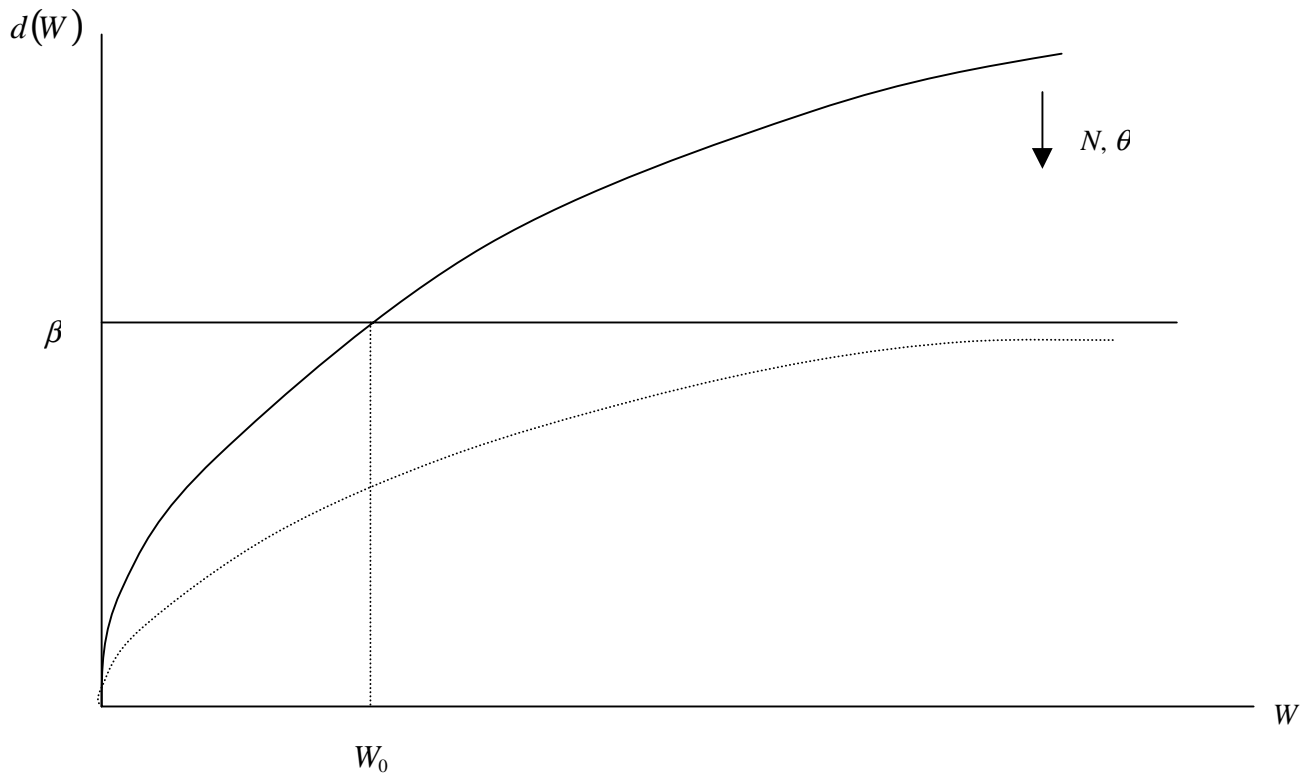
Figure 1
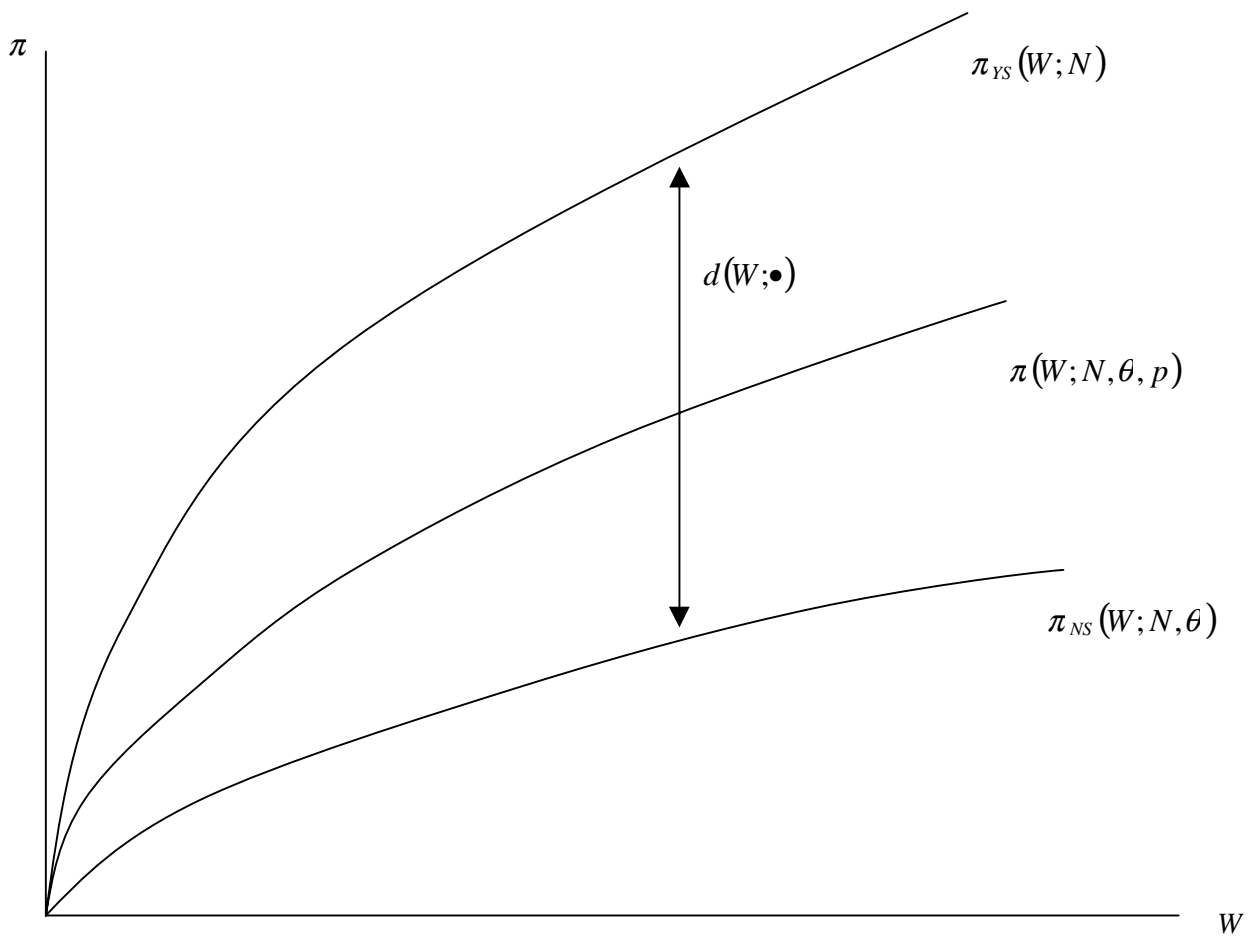
Figure 2

Figure 3

Figure 4

Figure 5